

Estimación Empírica de Carga de Trabajo en Proyectos de Explotación de Información

Rodríguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R.

Grupo Investigación en Sistemas de Información. Departamento Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús.

Grupo de Estudio en Metodologías de Ingeniería de Software. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional.

Grupo de Investigación en Explotación de Información. Sede Andina (El Bolsón). Universidad Nacional de Río Negro.

drodrigu@unla.edu.ar, fpollo@posgrado.frba.utn.edu.ar, paobritos@gmail.com, rgarcia@unla.edu.ar

Resumen. Un proceso de explotación de información puede definirse como un conjunto de tareas relacionadas lógicamente, que se ejecutan para extraer conocimiento no-trivial que reside de manera implícita en los datos disponibles en distintas fuentes de información. Una metodología de explotación de información permite gestionar la complejidad de estos procesos de manera uniforme. Entre estas metodologías, la comunidad científica considera probada a la metodología CRISP-DM. Resultados previos señalan la necesidad de abordar la estimación empírica de proyectos de explotación con base en proyectos realizados para pequeños y medianos emprendimientos. En este artículo, se introduce brevemente la metodología CRISP con énfasis en las tareas asociadas a sus subfases, se presentan resultados experimentales que muestran los porcentuales de tiempo que, sobre el total del proyecto, insumen las fases y las subfases metodológicas.

Palabras Clave. Proceso de explotación de información. Metodología de explotación de información. CRISP-DM. Estimación empírica de proyectos de explotación.

1. Introducción

La Explotación de Información (DM, Data Mining) consiste en la extracción de conocimiento no-trivial que reside de manera implícita en los datos disponibles en distintas fuentes de información [1]. Dicho conocimiento es previamente desconocido y puede resultar útil para algún proceso [2]. Para un experto, o para el responsable de un sistema de información, normalmente no son los datos en sí lo más relevante, sino el conocimiento que se encierra en sus relaciones, fluctuaciones y dependencias. Esta disciplina engloba un conjunto de técnicas encaminadas a la extracción de conocimiento procesable, implícito en el almacén de datos (Data Warehouse, DW) de la organización. Las bases de estas técnicas se encuentran en el análisis estadístico y

en los sistemas inteligentes. Con Explotación de Información se aborda la solución a problemas de predicción, clasificación y segmentación [3].

Un proceso de información o un proceso de transformación de información [4][5], puede definirse como un conjunto de tareas relacionadas lógicamente, que se ejecutan para lograr, a partir de un conjunto de información con un grado de valor para la organización, en otro conjunto de información con un grado de valor mayor que el inicial [6]. Cada proceso de transformación de información define un conjunto de información de entrada, un conjunto de transformaciones y un conjunto de información de salida. Un proceso de transformación de información puede ser parte de un proceso mayor que lo abarque o bien puede incluir otros procesos de transformación de información que deban ser incluidos en él, admitiendo una visión desde varios niveles de granularidad [7]. Identificado el problema de inteligencia de negocio y las técnicas de explotación de información, un proceso de explotación de información describe cuáles son las tareas que hay que desarrollar para que, aplicando las técnicas de explotación a la información que se tenga vinculada al negocio, se obtenga una solución al problema de inteligencia de negocio [8].

Los resultados de investigación obtenidos en distintas áreas de la ingeniería de proyectos de explotación de información tales como: requisitos [9] y metodologías para su educación [10][11][12], procesos de selección de herramientas [13][14], identificación de procesos [15][16], gestión de documentos asociados [17], métodos de transformación de datos orientado al uso de explotación de información [18], para finalmente identificar una serie de problemas abiertos en el área [19][20]; y la experiencia acumulada como consecuencia de trabajar en diversos proyectos de explotación de información en áreas tan diversas como: identificación de patrones humaniformes [21], inteligencia criminal [22] [23] y patrones de comportamiento de población carcelaria [24], descubrimiento patrones en problemas de aprendizaje [25] y deserción universitaria [26], descubrimiento detección de patrones para la prevención de daños y/o averías [27], patrones de métodos en tratamiento de cáncer [28]; señalan la necesidad de abordar la estimación empírica de proyectos de explotación de información con base en proyectos realizados para pequeños y medianos emprendimientos.

En este contexto, se introducen un abordaje metodológico para explotación de información aceptado como de uso probado por la comunidad académica (sección 2); se centra el problema de estimación de proyectos de explotación (sección 3); se presenta el experimento realizado (sección 4) y se discuten los resultados (sección 5); finalmente se presentan algunas conclusiones parciales.

2. Metodologías de Explotación de Información: CRISP-DM

Una metodología de Explotación de Información involucra, en general las siguientes fases [29]: comprensión del negocio y del problema que se quiere resolver, determinación, obtención y limpieza de los datos necesarios, creación de modelos matemáticos, ejecución, validación de los algoritmos, comunicación de los resultados obtenidos; e integración de los mismos, si procede, con los resultados en un sistema transaccional o similar. La relación entre todas estas fases tiene una complejidad que

se traduce en una jerarquía de subfases. A través de la experiencia acumulada en proyectos de Explotación de Información se han ido desarrollando metodologías que permiten gestionar esta complejidad de una manera uniforme. Entre estas metodologías, la comunidad científica considera probada a la metodología CRISP-DM.

La metodología CRISP-DM [30] consta de cuatro niveles de abstracción, organizados de forma jerárquica en tareas que van desde el nivel más general hasta los casos más específicos (ver Figura 1). A nivel más general, el proceso está organizado en seis fases (ver Figura 2), estando cada fase a su vez estructurada en varias tareas generales de segundo nivel o subfases. Las tareas generales se proyectan a tareas específicas, donde se describen las acciones que deben ser desarrolladas para situaciones específicas. Así, si en el segundo nivel se tiene la tarea general “limpieza de datos”, en el tercer nivel se dicen las tareas que tienen que desarrollarse para un caso específico, como por ejemplo, “limpieza de datos numéricos”, o “limpieza de datos categóricos”. El cuarto nivel, recoge el conjunto de acciones, decisiones y resultados sobre el proyecto de Explotación de Información específico.

La metodología CRISP-DM proporciona dos documentos distintos como herramienta de ayuda en el desarrollo del proyecto de Explotación de Información: el modelo de referencia y la guía del usuario. El documento del modelo de referencia describe de forma general las fases, tareas generales y salidas de un proyecto de Explotación de Información en general. La guía del usuario proporciona información más detallada sobre la aplicación práctica del modelo de referencia a proyectos de Explotación de Datos específicos, proporcionando consejos y listas de comprobación sobre las tareas correspondientes a cada fase. La metodología CRISP-DM estructura el ciclo de vida de un proyecto de Explotación de Información en seis fases, que interactúan entre ellas de forma iterativa durante el desarrollo del proyecto (Figura 2).

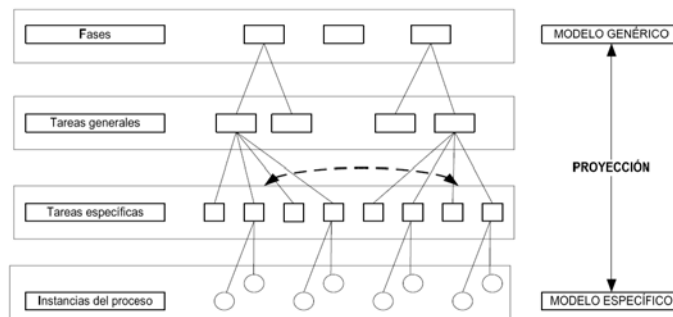


Fig. 1. Esquema de los cuatro niveles de abstracción de la metodología CRISP-DM

Las flechas indican las relaciones más habituales entre las fases, aunque se pueden establecer relaciones entre cualquier fase. El círculo exterior simboliza la naturaleza cíclica del proceso de modelado. En la tabla 1, se detallan las fases que componen la metodología CRISP-DM y se detalla cómo se componen cada una de ellas.

La primera fase de análisis del problema, incluye la comprensión de los objetivos y requerimientos del proyecto desde una perspectiva empresarial, con el fin de convertirlos en objetivos técnicos y en una planificación. La segunda fase de análisis

de datos comprende la recolección inicial de datos, en orden a que sea posible establecer un primer contacto con el problema, identificando la calidad de los datos y estableciendo las relaciones más evidentes que permitan establecer las primeras hipótesis.

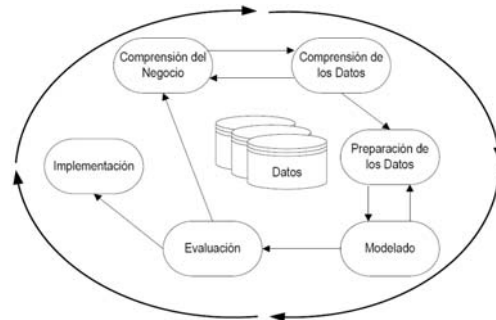


Fig. 2. Fases del proceso de modelado metodología CRISP-DM.

FASE	TAREAS COMPONENTES	• ACTIVIDADES ASOCIADAS
Comprensión del negocio	Determinar los objetivos de negocio	<ul style="list-style-type: none"> • Background • Objetivos del negocio • Criterios de éxito del negocio
	Evaluar de la situación	<ul style="list-style-type: none"> • Inventarios de recursos • Requisitos, supuestos y requerimientos • Riesgos y contingencias • Terminología • Costos y beneficios
	Determinar objetivos de explotación de información	<ul style="list-style-type: none"> • Las metas del Proyecto de Explotación de Información • Criterios de éxito del Proyecto de Explotación de Información
	Producir el plan del proyecto	<ul style="list-style-type: none"> • Plan de proyecto • Valoración inicial de herramientas
Entendimiento de datos	Recolección inicial de los datos	• Reporte de recolección de datos iniciales
	Descripción de los datos	• Reporte de descripción de los datos
	Exploración de los datos	• Reporte de exploración de datos
	Verificación de calidad de los datos	• Reporte de calidad de datos
Preparación de los datos	Tareas preparatorias	<ul style="list-style-type: none"> • Conjunto de Datos • Descripción del Conjunto de Datos
	Seleccionar los datos	• Inclusión / exclusión de datos
	Limpieza de datos	• Reporte de calidad de datos limpios
	Construcción de datos	<ul style="list-style-type: none"> • Derivación de atributos • Generación de registros
	Integración de los datos	• Unificación de datos
	Formato de datos	• Reporte de calidad de los datos
Modelado	Selección de la técnica de modelado	<ul style="list-style-type: none"> • La técnica modelada • Supuestos del modelo
	Generación del diseño del ensayo	• Plan de pruebas de ensayo
	Construcción del modelo	<ul style="list-style-type: none"> • Configuración de parámetros • Modelo • Descripción del modelo
	Evaluación del modelo	<ul style="list-style-type: none"> • Evaluar el modelo • Revisión de la configuración de parámetros
Evaluación	Evaluar los resultados	<ul style="list-style-type: none"> • Valoración de resultados mineros con respecto al éxito del negocio • Modelos aprobados
	Proceso de revisión	• Revisión del proceso
	Determinación de los próximos pasos	• Listar posibles acciones
Implantación	Plan de implantación	• Plan de implantación
	Plan de vigilancia y mantenimiento	• Plan de vigilancia y mantenimiento
	Producción final	<ul style="list-style-type: none"> • Informe final • Presentación Final
	Revisión del proyecto	• Documentación de la experiencia

Tabla 1. Tareas de cada fase de la metodología CRISP-DM

Una vez realizado el análisis de datos, la metodología establece que se proceda a la preparación de los datos, de tal forma que puedan ser tratados por las técnicas de modelado. La preparación de datos incluye las tareas generales de selección de datos a los que se va a aplicar la técnica de modelado (variables y muestras), limpieza de los datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato. La fase de preparación de los datos, se encuentra muy relacionada con la fase de modelado, puesto que en función de la técnica de modelado que vaya a ser utilizada los datos necesitan ser procesados en diferentes formas. Por lo tanto las fases de preparación y modelado interactúan de forma sistemática. En la fase de modelado se seleccionan las técnicas de modelado más apropiadas para el proyecto de Explotación de Información específico. Las técnicas a utilizar en esta fase se seleccionan en función de los siguientes criterios: ser apropiada al problema, disponer de datos adecuados, cumplir los requerimientos del problema, tiempo necesario para obtener un modelo y conocimiento de la técnica. Antes de proceder al modelado de los datos se debe de establecer un diseño del método de evaluación de los modelos, que permita establecer el grado de bondad de los modelos. Una vez realizadas estas tareas genéricas se procede a la generación y evaluación del modelo. Los parámetros utilizados en la generación del modelo dependen de las características de los datos. En la fase de evaluación, se evalúa el modelo, no desde el punto de vista de los datos, sino desde el cumplimiento de los criterios de éxito del problema. Se debe revisar el proceso seguido, teniendo en cuenta los resultados obtenidos, para poder repetir algún paso en el que, a la vista del desarrollo posterior del proceso, se hayan podido cometer errores. Si el modelo generado es válido en función de los criterios de éxito establecidos en la primera fase, se procede a la explotación del modelo.

3. El Problema

La gestión de un proyecto de software comienza con un conjunto de actividades que se denominan planificación del proyecto. Antes de que el proyecto comience, se debe realizar una estimación: del trabajo a ejecutar, de los recursos necesarios y del tiempo que transcurrirá desde el comienzo hasta el final de su realización [31]. Los proyectos de explotación de información también requieren de un proceso de planificación que permita estimar sus tiempos, sin embargo, dada las diferencias que existente entre un proyecto convencional de construcción de software y un proyecto de explotación de información, los métodos usuales de estimación no son aplicables. En este contexto, este trabajo buscó obtener empíricamente la distribución porcentual de la carga de trabajo en proyectos de explotación de información.

4. El Experimento

El objetivo del experimento es obtener una estimación empírica del porcentaje del tiempo del proyecto de explotación de información que insume la ejecución de cada una de las tareas asociadas a las subfases de la Metodología CRISP-DM con foco en proyectos para pequeños y medianos emprendimientos.

Los pasos seguidos se identifican en la tabla 2.

PASOS EXPERIMENTALES	
Paso 1:	Entrenar a estudiantes avanzados (5º año de Ingeniería en Sistemas de Información) en la identificación de objetivos de negocio, la identificación de los procesos de explotación de información asociados y la documentación que se debe desarrollar durante la aplicación de la metodología CRISP-DM. Dividir en grupos de trabajo con capacidades homogéneas.
Paso 2:	Identificar N proyectos de explotación de información para pequeños y medianos emprendimientos.
Paso 3:	Asignar cada uno de los proyectos de explotación de información a un grupo de los formados en el paso 1.
Paso 4:	Desarrollo por cada grupo del proyecto de explotación de información con registro del tiempo utilizado para desarrollar cada tarea de cada subfase de la metodología CRISP-DM.
Paso 5:	Integración en los resultados del experimento de los tiempos obtenidos por cada grupo .

Tabla 2. Pasos experimentales seguidos para obtener los tiempos experimentales

Los proyectos de explotación de información para pequeños y medianos emprendimientos se muestran en la tabla 3.

PROYECTO	OBJETIVOS DE NEGOCIO	OBJETIVOS DE EXPLOTACIÓN DE INFORMACIÓN
1	Determinar cuáles son las necesidades que se buscan satisfacer en un auto la clase media.	<ul style="list-style-type: none"> ▪ Descubrir las reglas de negocio que permiten determinar a qué grupo social pertenece un auto. ▪ Determinar cuáles son las características que busca la clase media en un auto. ▪ Determinar cuál es la característica que más influye a la hora de comprar un auto para dicha clase.
2	La Secretaría Académica de una Facultad desea diseñar nuevos cursos extracurriculares de especialización (arancelados), de acuerdo al interés de su población universitaria.	<ul style="list-style-type: none"> ▪ Descubrir grupos de asignaturas electivas por los que tomen opción los estudiantes de los últimos años de las distintas carreras que se desarrollan en la Facultad ▪ Descubrir las reglas de negocio que permitan caracterizar la pertenencia a los grupos descubiertos.
3	Establecer la mejor forma de armar packs de servicios secundarios en forma de promoción para conseguir la mayor cantidad de clientes posibles, de acuerdo a sus necesidades, dependiendo estos del plan primario al que pertenezcan y el sector en que se encuentren.	<ul style="list-style-type: none"> ▪ Determinar la mejor manera de agrupar los productos ofrecidos por la empresa. ▪ Descubrir grupos de clientes que utilizan grupos semejantes de servicios ▪ Identificar el servicio característico de cada pack descubierto
4	El objetivo del Gobierno Provincial es proveer asistencia vial para ordenar el tránsito en las situaciones en que el clima puede provocar dificultades, como visibilidad reducida o calzada resbaladiza.	<ul style="list-style-type: none"> ▪ Descubrir reglas que permitirán determinar, en base a los registros meteorológicos de un día en particular, condiciones climatológicas (lluvia, tormenta o niebla)
5	Mejorar las metodologías utilizadas por el área de producción de vinos a partir de las formulas obtenidas en la implementación del proyecto y obtener diferentes tipos de productos de acuerdo a su calidad (buena, muy buena, excelente).	<ul style="list-style-type: none"> ▪ Determinación de fórmulas de vinos "predeterminadas" ▪ Predecir la calidad de un vino en relación a la variación de los distintos parámetros comprendiéndolos y de esta manera poder innovar en la producción de forma más orientada.
6	Preservar y fidelizar a los clientes, aumentar la cobertura de los servicios y mejorar la tecnología.	<ul style="list-style-type: none"> ▪ Buscar patrones o grupos de zonas específicas con el fin de hacer un plan de migración tecnológica. ▪ Encontrar los atributos relevantes y la interdependencia de ellos.
7	Una empresa mayorista de blanquecía busca alternativas para generar promociones que puedan generar más ventas.	<ul style="list-style-type: none"> ▪ Descubrir comportamiento de compra de los diferentes.

Tabla 3. Proyectos de explotación de información identificados para pequeños y medianos emprendimientos

5. Discusión de los Resultados

Los resultados destacan aquellas fases y subfases que insumen una cantidad significativa de tiempo (más del 50%).

Las fases de "Comprensión del Negocio" y "Modelado" insumen más del 50% del tiempo del proyecto, utilizando "Modelado" el 34,41% y "Comprensión del Negocio" el 20,70% (ver tabla 4).

FASE	% de TEMPO
Fase 1 COMPRENSIÓN DEL NEGOCIO	20,70
Fase 2 ENTENDIMIENTO DE LOS DATOS	10,90
Fase 3 PREPARACION DE DATOS	15,61
Fase 4 MODELADO	34,41
Fase 5 EVALUACIÓN	7,45
Fase 6 IMPLANTACION	10,93

Tabla 4. Carga de trabajo (en % de tiempo) de cada fase de la Metodología CRISP-DM

En la fase "Comprensión del Negocio" las subfases "Determinar los Objetivos de Negocio" y "Evaluar la Situación" utilizan más del 70% del tiempo de la fase, utilizando "Determinar los Objetivos de Negocio" el 35,59% y "Evaluar la Situación" el 32,24% (ver tabla 5).

En la fase "Entendimiento de los Datos" las subfases "Recolección Inicial de Datos" y "Descripción de los Datos" emplean más del 60% del tiempo de la fase, utilizando "Recolección Inicial de Datos" el 36,72% y "Descripción de los Datos" el 25,70% (ver tabla 6).

En la fase "Preparación de Datos" las subfases "Tareas Preparatorias" y "Formato de Datos" ocupan más del 60% del tiempo de la fase, utilizando "Formato de Datos" el 34,62% y "Tareas Preparatorias" el 28,03% (ver tabla 7).

En la fase "Modelado" la subfase "Construcción del Modelo" requiere el 62,97% del tiempo de la fase (ver tabla 8).

En la fase "Evaluación" la subfase "Evaluar Resultados" lleva el 50,36% del tiempo de la fase (ver tabla 9).

En la fase "Implantación" la subfase "Producción Final" lleva el 56,29% del tiempo de la fase (ver tabla 10).

Fase 1 COMPRENSIÓN DEL NEGOCIO	
SUBFASE	% de TEMPO
Subfase 1.1 Determinar los objetivos de negocio	35,59
Subfase 1.2 Evaluar la situación	32,24
Subfase 1.3 Determinar objetivos explotación de información	12,83
Subfase 1.4 Producir el plan del proyecto	19,34

Tabla 5. Carga de trabajo (en % de tiempo) de la fase "Comprensión del Negocio".

Fase 2 ENTENDIMIENTO DE LOS DATOS	
SUBFASE	% de TEMPO
Subfase 2.1 Recolección inicial de datos	36,72
Subfase 2.2 Descripción de los datos	25,70
Subfase 2.3 Exploración de los datos	22,89
Subfase 2.4 Verificación de calidad de los datos	14,96

Tabla 6. Carga de trabajo (en % de tiempo) de la fase "Entendimiento de los Datos".

Fase 3 PREPARACION DE DATOS	
SUBFASE	% de TEMPO
Subfase 3.0 Tareas preparatorias	28,03
Subfase 3.1 Selección de datos	13,85
Subfase 3.2 Limpieza de datos	10,26
Subfase 3.3 Construcción de datos	8,89
Subfase 3.4 Integración de los datos	4,36
Subfase 3.5 Formato de datos	34,62

Tabla 7. Carga de trabajo (en % de tiempo) de la fase "Preparación de los Datos".

Fase 5 EVALUACIÓN	
SUBFASE	% de TEMPO
Subfase 5.1 Evaluar los resultados	50,36
Subfase 5.2 Proceso de revisión	21,51
5.3 Determinación de los próximos pasos	28,14

Tabla 9. Carga de trabajo (en % de tiempo) de la fase "Evaluación".

Fase 4 MODELADO	
SUBFASE	% de TEMPO
Subfase 4.1 Selección de la técnica de modelado	15,70
Subfase 4.2 Generación del diseño del ensayo	7,99
Subfase 4.3 Construcción del modelo	62,97
Subfase 4.4 Evaluación del modelo	13,34

Tabla 8. Carga de trabajo (en % de tiempo) de la fase "Modelado".

Fase 6 IMPLANTACION	
SUBFASE	% de TEMPO
Subfase 6.1 Plan de implantación	17,70
Subfase 6.2 Plan de vigilancia y mantenimiento	12,82
Subfase 6.3 Producción final	56,29
Subfase 6.4 Revisión del proyecto	13,19

Tabla 10. Carga de trabajo (en % de tiempo) de la fase "Implantación".

6. Conclusiones

La construcción de métodos de estimación de proyectos de software que logren resultados predictivos sobre los recursos a emplear que se ajusten de la mejor manera posible a la realidad obtenible, es un problema abierto en el campo de los sistemas de información. Los proyectos de explotación de información no escapan a esta necesidad y la historia de la Ingeniería en general y la Informática en particular, registran que los primeros abordajes son siempre de naturaleza empírica.

En este contexto, los resultados experimentales que se describen en este trabajo, permiten tener una estimación sobre cuál es la carga relativa de trabajo asociada a una subfase determinada de la metodología CRISP-DM. Esto posibilita, conociendo el tiempo utilizado en alguna de las subfases, tener una aproximación a los tiempos de las otras subfases y a la estimación global del proyecto.

7. Agradecimientos

Deseamos agradecer a los señores Pablo Amura, Marcelo Barusso, Rubén Berman, Bruno Bonanno, Nicolás Calligaro, Hernán Castelo, David Cohan, Hernán De Cabo Pérez, Adrián Díaz, Javier Gardella, Mauro González, Jose Lugo, Pablo Morales, José Murillo, Juan Lucas Pierdomenico, Hugo Ramón, Maria Sol Rodrigo, Ariel Sánchez y Pablo Sualdo que en calidad de noveles ingenieros de explotación de información; reportaron los tiempos de sus respectivos proyectos.

8. Financiamiento

Las investigaciones que se reportan en este artículo han sido parcialmente financiadas por el Proyecto de Investigación 33A081 del Departamento de Desarrollo Productivo y Tecnológico de la Universidad Nacional de Lanús.

9. Referencias

- [1] Schiefer, J., Jeng, J., Kapoor, S., Chowdhary, P. (2004). *Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence*. Proceedings 2004 IEEE International Conference on E-Commerce Technology. Pág. 162-169.
- [2] Stefanovic, N., Majstorovic, V., Stefanovic, D. (2006). Supply Chain Business Intelligence Model. Proceedings 13th International Conference on Life Cycle Engineering. Pág. 613-618.
- [3] Umapathy, K. (2007). *Towards Co-Design of Business Processes and Information Systems Using Web Services*. Proceedings 40th Annual Hawaii International Conference on System Sciences. Pág. 172-181.
- [4] Curtis, B., Kellner, M., Over, J. (1992). *Process Modelling*. Communications of the ACM, 35(9): 75-90.
- [5] Britos, P., García-Martínez, R. 2009. *Propuesta de Procesos de Explotación de Información*. Proceedings XV CACIC. Workshop de Base de Datos y Minería de Datos. Págs. 1041-1050. ISBN 978-897-24068-4-1.
- [6] Ferreira, J., Takai, O., Pu, C. (2005). *Integration of Business Processes with Autonomous Information Systems: A Case Study in Government Services*. Proceedings Seventh IEEE International Conference on E-Commerce Technology. Pág. 471-474.
- [7] Kanungo, S. (2005). *Using Process Theory to Analyze Direct and Indirect Value-Drivers of Information Systems*. Proceedings of the 38th Annual Hawaii International Conference on System Sciences. Pág. 231-240.
- [8] Musen, M., Ferguson, R., Grosso, W., Noy, N., Crubezy, M., Gennari, J. (2000). Component-Based Support for Building Knowledge-Acquisition Systems. Reporte SMI- 2000-0838. Stanford Medical Informatics. Universidad de Stanford.
- [9] Ochoa, M., Britos, P. y García-Martínez, R. 2006. *Una Prototipo de Entendimiento del Negocio para Metodologías de Desarrollo de Sistemas*. XII CACIC. San Luis. Argentina.
- [10] Britos, P., Enrique Fernandez, E., García-Martínez, R. 2007. *Propuesta metodológica para la educación de requisitos en proyectos de explotación de información*. Proceedings IX Workshop de Investigadores en Ciencias de la Computación. Pág. 497-501. ISBN 978-950-763-075-0.
- [11] Britos, P., Dieste, O., García-Martínez, R. 2008. *Requirements Elicitation in Data Mining for Business Intelligence Projects*. En Advances in Information Systems Research, Education and Practice. David Avison, George M. Kasper, Barbara Pernici, Isabel Ramos, Dewald Roode Eds. (Boston: Springer), IFIP Series, 274: 139-150.
- [12] Pollo-Cattaneo, F., Britos, P. Pesado, P., García-Martínez, R. 2009. *Metodología para Especificación de Requisitos en Proyectos de Explotación de Información*. Proceedings XI Workshop de Investigadores en Ciencias de la Computación. Pág. 467-469. ISBN 978-950-605-570-7.
- [13] Britos, P., Fernández, E., Ochoa, M., Merlino, H., Diez, E. y García Martínez, R. 2005. *Metodología de Selección de Herramientas de Explotación de Datos*. Proceedings X CACIC. Pág. 113-123.
- [14] Britos, P., Merlino, H., Fernández, E., Ochoa, M., Diez, E. y García Martínez, R. 2006. *Tool Selection Methodology in Data Mining*. Proceedings V Ibero-American Symposium on Software Engineering. Pág. 85-90.
- [15] Britos, P. 2008. *Procesos de Explotación de Información Basados en Sistemas Inteligentes*. Tesis de Doctorado en Ciencias Informáticas. Facultad de Informática. Universidad Nacional de La Plata.

- [16] Britos, P., García-Martínez, R. (2009). *Propuesta de Procesos de Explotación de Información*. Proceedings XV CACIC. Workshop de Base de Datos y Minería de Datos. Págs. 1041-1050. ISBN 978-897-24068-4-1.
- [17] Fernández, E., Merlino, H., Ochoa, M., Diez, E., Britos, P. y García-Martínez, R. 2005. *Gestión Asistida de Documentos. en una Metodología de Explotación de Información*. Proceedings del II Workshop de Ingeniería de Software y Bases de Datos del X CACIC. Pág. 504-514.
- [18] Merlino, H., Britos, P., Ierache, J., Diez, E. y García-Martínez, R. 2005. *Un Método de Transformación de Datos Orientado al Uso de Explotación de Información*. Proceedings del II Workshop de Ingeniería de Software y Bases de Datos del X CACIC. Pág. 22-32.
- [19] García-Martínez, R., Britos, P., Diez, E., Ochoa, M., Merlino, H., Fernández, E., Rancán, C. 2007. *Minería de Datos en Inteligencia de Negocios*. Proceedings IX Workshop de Investigadores en Ciencias de la Computación. Pag. 130-133. ISBN 978-950-763-075-0.
- [20] Pollo-Cataneo, F., Amatriain, H., Rodríguez, D., Pytel, P., Ciccolella, E., Vegega, C., Dearriba, M., Rodríguez Aubert, M., Bose, F., Giordano, L., Britos, P., García-Martínez, R. (2010). *Ingeniería de Proyectos de Explotación de Información*. Proceedings XII Workshop de Investigadores en Ciencias de la Computación. Pág. 172-176. ISBN 978-950-34-0652-6.
- [21] Britos, P., Abasolo, M., García-Martínez, R. y Perales, F. 2005. *Identification of MPEG-4 Patterns in Human Faces Using Data Mining Techniques*. Proceedings 13 th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2005. Pág. 9-10.
- [22] Britos, P., Fernández, E., Merlino, H., Pollo-Cataneo, F., Rodríguez, D., Procopio, C., Rancan, C., García-Martínez, R. 2008. *Explotación de Información Aplicada a Inteligencia Criminal en Argentina*. Proceedings del XIV CACIC, Workshop de Ingeniería de Software y Bases de Datos, Artículo 1866. ISBN 978-987-24611-0-2.
- [23] Valenga, F., Fernández, E., Merlino, H., Rodríguez, D., Procopio, C., Britos, P., García-Martínez, R. 2008. *Minería de Datos Aplicada a la Detección de Patrones Delictivos en Argentina*. Proceedings VII Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento. Pág. 31-39. ISSN 1390-292X.
- [24] Gutiérrez Rüegg, P., Merlino, H., Rancan, C., Procopio, C., Rodríguez, D., Britos, P., García-Martínez, R. 2008. *Identificación de Patrones Característicos de la Población Carcelaria Mediante Minería de Datos*. Proceedings X Workshop de Investigadores en Ciencias de la Computación. Pág. 461-465. 978-950-863-101-5.
- [25] Britos, P., Jiménez Rey, E., García-Martínez, E. 2008. *Work in Progress: Programming Misunderstandings Discovering Process Based On Intelligent Data Mining Tools*. Proceedings 38th ASEE/IEEE Frontiers in Education Conference. Session F4H: Assessing and Understanding Student Learning. ISBN 978-1-4244-1970-8.
- [26] Kuna, H., García Martínez, R. Villatoro, F. 2010. *Pattern Discovery in University Students Desertion Based on Data Mining*. Advances and Applications in Statistical Sciences Journal, 2(1): Aceptado. ISSN 0974-6811.
- [27] Reparaz, D., Merlino, H., Rancan, C., Rodríguez, D., Britos, P., García-Martínez, R. 2008. *Determinación de la Eficacia de la Braquiterapia en Tratamiento de Cáncer Basada en Minería de Datos*. Proceedings X Workshop de Investigadores en Ciencias de la Computación. Pág. 456-460. 978-950-863-101-5.
- [28] Flores, D., García-Martínez, R. Fernandez, E., Merlino, H., Rodríguez, D., Britos, P. 2009. *Detección de Patrones para la Prevención de Daños y/o Averías en la Industria Automotriz*. Proceedings XV CACIC. Workshop de Base de Datos y Minería de Datos. Págs. 1021-1030. ISBN 978-897-24068-4-1.
- [29] Maimon, O., Rokach, L. 2005. *The Data Mining and Knowledge Discovery Handbook*. Springer Science + Business Media Publishers.
- [30] Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. 2000. *CRISP-DM 1.0 Step by step BGuide. Edited by SPSS*. <http://www.crisp-dm.org/> CRISPWP-0800.pdf. Ultimo acceso 01/06/08.
- [31] Pressman, R. 2004. *Software Engineering: A Practitioner's Approach*. Editorial Mc Graw Hill.