

MODELOS DE PROCESO PARA INGENIERÍA DE EXPLOTACIÓN DE INFORMACIÓN PARA PYMES: ABORDAJE ÁGIL Y ABORDAJE ROBUSTO

Ramón García-Martínez, Eduardo Diez, Roberto García, Sebastian Martins, Ezequiel Baldizzoni

Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de Información
Grupo Investigación en Sistemas de Información
Departamento de Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús
29 de Septiembre 3901 (1826) Remedios de Escalada, Lanús. Argentina. Tel +54 11 6322-9200 Ext. 194
rgm1960@yahoo.com, smartins089@gmail.com

RESUMEN

Las metodologías de desarrollo de proyectos de explotación de información existentes, se centran en el proceso de descubrimiento de patrones de conocimiento en la masa de datos. Sin embargo no existe una visión de proceso que identifique fases, tareas, técnicas de representación y procedimientos de ejecución de la tarea; ni que permitan soportar de forma completa e integrada las actividades de administración y desarrollo de proyectos. Resultados de investigación recientes en el área de métricas de proyecto han posibilitado clasificar un determinado proyecto en microproyecto o proyecto regular en función de los valores asociados a dimensión y carga de trabajo. Por analogía con la Ingeniería de Software, que propone metodologías ágiles para cierto tipo de proyectos y metodologías robustas para otros, esta investigación propone cubrir la vacancia de Modelos de Proceso para Proyectos de Ingeniería de Explotación de Información que integre en fases las técnicas desarrolladas para la administración y desarrollo de proyectos de explotación de información en dos versiones, una orientada a microproyectos y otra orientada a proyectos regulares.

Palabras clave: Modelos de Proceso, Ingeniería de Explotación de Información, Proceso Ágil, Proceso Robusto.

CONTEXTO

El proyecto que se presenta se encuadra a nivel nacional en “PLAN ARGENTINA INNOVADORA 2020 > Temas estratégicos para convocatorias 2014 > SECTOR DESARROLLO Y TECNOLOGÍA SOCIAL > Grandes Datos (Big Data) e Informática > Búsqueda y minería en grandes datos” del Ministerio de Ciencia, Tecnología e Innovación, [MINCyT, 2014]; y en los lineamientos estratégicos que para la CADENA DE VALOR DEL SOFTWARE Y SERVICIOS INFORMÁTICOS

que establece el Plan Industrial 2020 del Ministerio Industria [MI, 2014].

El proyecto que se presenta continua con las investigaciones iniciadas por el Proyecto 33A167 “Instrumentos para la Gestión de Proyectos de Explotación de Información” (período 2013-2014) y por el Proyecto 33A105 “Ingeniería de Proyectos de Explotación de Información” (período 2011-2012).

El proyecto prevé formular aportaciones concretas al núcleo de herramientas informáticas del área de Ingeniería de Explotación de Información con énfasis en Minería en Grandes Datos para realizar inteligencia de negocio para la Industria.

Contar con herramientas de este tipo permite a las PYMES del sector, explorar una nueva línea de servicios informáticos. Disponer de modelos de proceso para este tipo de proyectos, habilita a las PYMES disponer de una guía que ordene el trabajo interno y permita establecer y ofrecer a los requerentes hitos para los entregables del proyecto.

INTRODUCCIÓN

Hace ya una década, un estudio de la Universidad de California en Berkeley [Lyman y Varian, 2003] señaló que la información disponible en Internet crecía a razón de 92 petabytes [1015 bytes] por año. En [Maimon y Rokach, 2005] se ha señalado que esta información esta disponible para procesos de descubrimiento de conocimiento con independencia que se encuentre en fuentes estructuradas [Rudin y Cressy, 2003; Moss, 2003] ó desestructuradas [Vuori, 2006].

La inteligencia de negocio propone un abordaje interdisciplinario (dentro del que se encuentra la Informática), que tomando todos los recursos de información disponibles y el uso de herramientas analíticas y de síntesis con capacidad de transformar la información en conocimiento, se centra en generar a partir de éstos, conocimiento que contribuya con la toma de decisiones de gestión y generación de planes estratégicos en las

organizaciones [Thomsen, 2003; Negash y Gray, 2008].

La Explotación de Información es la sub-disciplina de los Sistemas de Información que aporta a la Inteligencia de Negocio [Langseth y Vivatrat, 2003] las herramientas para la transformación de información en conocimiento [Srivastava et al., 2000]. Ha sido definida como la búsqueda de patrones interesantes y de regularidades importantes en grandes masas de información [Abraham, 2003; Coley, 2003].

Un Proceso de Explotación de Información se define, como un grupo de tareas relacionadas lógicamente [Curtis et al., 1992] que, a partir de un conjunto de información con un cierto grado de valor para la organización, se ejecuta para lograr otro, con un grado de valor mayor que el inicial [Ferreira et al., 2005; Kanungo, 2005]. Adicionalmente, existe una variedad de técnicas de minería de datos, en su mayoría provenientes del campo del Aprendizaje Automático [García-Martínez, 1997; García-Martínez et al., 2003], susceptibles de ser utilizadas en cada uno de estos procesos.

La ingeniería de explotación de información entiende en los procesos y las metodologías utilizadas para: ordenar, controlar y gestionar la tarea de encontrar patrones de conocimiento en masas de información [García-Martínez et al. 2011]. Los proyectos de ingeniería de explotación de información poseen características muy distintas a los proyectos de desarrollo de software tradicional [Vanrell et al., 2010], sobre todo en la parte operativa del proyecto. La diferencia se presenta en los procesos de desarrollo y mantenimiento en los cuales el ciclo de fases de un proyecto de software tradicional: inicio, requisitos, análisis y diseño, construcción, integración y pruebas no resultan naturales en un proyecto de explotación de información [Vanrell et al., 2012]. Por otra parte, al evaluar las principales metodologías existentes para los proyectos de explotación de información [Chapman et al., 2000; SAS, 2014; Pyle, 2003], se observa la falta de herramientas que permitan soportar de forma completa la fase de administración de proyectos.

Una de las lecciones aprendidas sobre desarrollo de software en Informática derivada de los estadios tempranos de la disciplina, es que, la ausencia de una ingeniería de software conllevaba a un desarrollo artesanal de los artefactos software [Ochoa et al., 2008]. El desarrollo artesanal implicaba la imposibilidad de poder establecer dentro de valores racionales, parámetros tales como: [a] cantidad y calificación de los recursos humanos

a emplear en el proyecto, [b] tiempos de desarrollo, [c] modelos de proceso que guiarán el desarrollo y permitirán establecer hitos de entrega, [d] formalismos de documentación que dieran cuenta de lo hecho en el proyecto de desarrollo del artefacto software y de las decisiones de diseño asumidas, constituyendo el punto de partida para futuras ampliaciones de funcionalidades, [e] modelos de costo de proyecto [Böehm, 1981], entre otros. De hecho la estimación de estos parámetros se hacía en base a la experiencia de individuos sin ninguna base ingenieril; y lo que para un grupo de desarrollo podía hacerse en meses, para otro podía hacerse en años. Actualmente, los abordajes de la Ingeniería de Software se mueven entre dos extremos, el llamado abordaje robusto representado por metodologías como MÉTRICA v.3 [Pae, 2014] y el llamado abordaje ágil como el propuesto por el núcleo SEMAT [Jacobson, 2013].

La ingeniería de explotación de información esta en sus primeros estadios y, al igual que lo que ocurría con el desarrollo de artefactos software, adolece de un marco de trabajo que integre instrumentos para un adecuado control, gestión y desarrollo de proyectos de este tipo. En particular, en el caso de las PyMEs del sector, reviste importancia por cuanto contar con un marco de trabajo integrado para proyectos de explotación de información, les permite explorar una nueva línea de servicios informáticos.

Los modelos de proceso para este tipo de proyectos, habilita que la PyME disponga de una guía que ordene el trabajo interno y permita establecer y ofrecer al requirente, hitos para los entregables del proyecto. Disponer de modelos que permitan cuantificar: la viabilidad, los recursos y los tiempos a emplear en el proyecto de explotación de información, hace predecibles (y acotables) los costos al momento que una PyME ofrezca este tipo de servicios en el marco de aplicaciones gubernamentales o comerciales. Por otra parte disponer de formalismos de documentación facilita a la PyME repetir procesos con el consecuente ahorro de recursos.

En línea con área disciplinar del proyecto, el grupo de docentes investigadores ha trabajado en los siguientes temas de Ingeniería de Explotación de Información:

- [a] La educación de requerimientos de proyectos [Britos et al., 2008] y el correspondiente modelado [Vegega et al., 2012a; 2012b].
- [b] Se ha propuesto un modelo de proceso para elicitación de requerimientos [Mansilla et al., 2012; 2013].

- [c] Se ha trabajado en la determinación de la viabilidad de proyectos [Pytel et al., 2012a; 2012c; 2012d; 2013a] y la estimación de recursos necesarios [Pytel et al., 2012b; 2013b; 2013c, 2015].
- [d] Se han diseñado procesos de explotación de información basados en sistemas inteligentes [Britos, 2008; García-Martínez et al., 2013; Kuna et al., 2014a], y se ha estudiado el comportamiento resultante de la integración de varios procesos variando la complejidad del dominio [Rancan et al., 2010; López Nocera y García-Martínez, 2012; López Nocera et al., 2011; 2012].
- [e] Se han formulado estimaciones empíricas de carga de trabajo [Rodríguez et al., 2010] identificándose las tareas críticas en una metodología de desarrollo de proyectos de explotación de información [Pytel et al., 2011].
- [f] Se ha definido un proceso de derivación del proceso de explotación de información desde el modelado del negocio [Martins, 2014; Martins et al., 2014].
- [g] Se han propuesto métricas para la gestión de desarrollo de proyectos de explotación de información [Basso et al., 2013], que permiten, en función de los valores asociados a dimensión y carga de trabajo, clasificar un determinado proyecto en microproyecto o proyecto regular [Basso, 2014].

CONVENCIONES, PREGUNTAS PROBLEMA, HIPÓTESIS Y OBJETIVOS GENERALES Y ESPECÍFICOS

Convenciones:

Término: Explotación de Información:

El término minería de datos esta fuertemente ligado al concepto de base de datos y se remonta a la definición de algoritmos de búsqueda de patrones en grandes bases de datos [Fayyad et al., 1996]. Sin embargo, hoy existen líneas de investigación en campos tales como: minería de textos [Tan, 1999], minería de imágenes [Hsu et al., 2002], minería de patrones en flujos de información [Gaber et al., 2010], minería en la web [Kosala y Blockeel, 2000], entre otras. En este contexto, se conviene utilizar el término “explotación de información” como referencia genérica a cualquiera de los tipos de minería precitados.

Término: Ingeniería de Explotación de Información:

Con base en que la Ingeniería de Software ha sido definida en el SWEBOOK [Abran et al., 2004] como:

“la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo, operación y mantenimiento de software, y el estudio de estos enfoques, es decir, la aplicación de la ingeniería al software”; se conviene en definir a la Ingeniería de Explotación de Información como la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo de proyectos de explotación de información, y el estudio de este enfoque, es decir, la aplicación de la ingeniería a la explotación de información.

Hipótesis:

Hipótesis I:

La actividad de descubrimiento de patrones en masas de información utilizando algoritmos de descubrimiento de conocimiento (knowledge discovery) conocidos hoy como algoritmos de minería de datos (data mining), acredita un desarrollo consolidado a lo largo de las últimas tres décadas. La sistematización de la aplicación de estos algoritmos está sustentada en metodologías orientadas a transformar datos en conocimiento. Sin embargo, no existe una visión generalizada como la propuesta en [Britos, 2008] que permita bajo un paradigma equivalente al de procesos de transformación, agrupar en familias a algoritmos que busquen el mismo tipo de patrón de conocimiento.

Hipótesis II:

Las metodologías de desarrollo de proyectos de explotación de información existentes, se centran en el proceso de descubrimiento de patrones de conocimiento en la masa de datos. Sin embargo no existe una visión de proceso que identifique fases, tareas, técnicas de representación y procedimientos de ejecución de la tarea; que permita sistematizar la concreción del proyecto.

Hipótesis III:

Trabajos de investigación recientes han consolidado resultados en el área de gestión de proyectos de explotación de información tales como: modelado de requisitos, métodos para la determinación de la viabilidad de proyecto y la estimación de recursos necesarios, métricas del proyecto, entre otros. Sin embargo, las metodologías usuales de explotación de información no contemplan el nivel de administración del control y la gestión de proyecto que los nuevos modelos de proceso de software prevén.

Hipótesis IV:

La Ingeniería de Software propone metodológicamente abordajes ágiles como el descrito por el núcleo SEMAT. Sin embargo, las metodologías

usuales de explotación de información no contemplan este tipo de abordajes orientados específicamente a la gestión y el desarrollo de microproyectos de explotación de información.

Pregunta Problema:

¿Se puede cubrir la vacancia de Modelos de Proceso para Proyectos de Ingeniería de Explotación de Información que integre en fases las técnicas desarrolladas para el control, gestión y desarrollo de proyectos (microproyectos y regulares) de explotación de información y en cuyo marco la elección de un algoritmo de minería de datos sea una actividad prevista en una de sus fases?

Objetivo:

El objetivo de este proyecto es sistematizar el conocimiento existente sobre ingeniería de explotación de información y formular dos abordajes de modelo de proceso de gestión y desarrollo de proyecto; un proceso ágil orientado a microproyectos y un proceso robusto orientado a proyectos regulares que integren métodos, técnicas y herramientas de desarrollo de proyectos en esta sub-disciplina informática.

Objetivos Específicos:

Objetivo específico vinculado a Hipótesis I (OE1):

Desarrollar una taxonomía de algoritmos de minería de datos basada en la familia de procesos propuesta por Britos [2008].

Objetivo específico vinculado a Hipótesis II (OE2):

Desarrollar un modelo de proceso de desarrollo de proyectos de ingeniería de explotación de información que identifique fases, tareas, técnicas de representación y procedimientos de ejecución de la tarea que permita sistematizar el desarrollo de proyectos en el área.

Objetivo específico vinculado a Hipótesis III (OE3):

Desarrollar un modelo de proceso de control y gestión de proyectos de ingeniería de explotación de información que identifique fases, tareas, técnicas de representación y procedimientos de ejecución de la tarea; que permita sistematizar el control y la gestión de proyectos en el área.

Objetivo específico vinculado a Hipótesis II y III (OE4):

Desarrollar un modelo de proceso integrado de ingeniería de explotación de información orientado a proyectos regulares que articule los modelos de proceso de control y gestión y de desarrollo de proyectos de ingeniería de explotación de información.

Objetivo específico vinculado a Hipótesis IV (OE5):

Desarrollar un modelo de proceso integrado de ingeniería de explotación de información orientado a microproyectos equivalente al propuesto por el Núcleo SEMAT.

METODOLOGÍA DE TRABAJO

Para construir el conocimiento asociado al presente proyecto de investigación, se seguirá un enfoque de investigación clásico [Riveros y Rosas, 1985; Creswell, 2002] con énfasis en la producción de tecnologías [Sábato y Mackenzie, 1982]; identificando métodos, materiales y abordaje metodológico necesarios para desarrollar el proyecto.

Métodos:

Revisiones Sistemáticas:

Las revisiones sistemáticas [Argimón, 2004] de artículos científicos siguen un método explícito para resumir la información sobre determinado tema o problema. Se diferencia de las revisiones narrativas en que provienen de una pregunta estructurada y de un protocolo previamente realizado.

Prototipado Evolutivo Experimental (Método de la Ingeniería):

El prototipado evolutivo experimental [Basili, 1993] consiste en desarrollar una solución inicial para un determinado problema, generando su refinamiento de manera evolutiva por prueba de aplicación de dicha solución a casos de estudio (problemáticas) de complejidad creciente. El proceso de refinamiento concluye al estabilizarse el prototipo en evolución.

Materiales:

Para el desarrollo de los formalismos y procesos propuestos se utilizarán:

- Formalismos de modelado conceptual usuales en la Ingeniería de Software [Rumbaugh et al., 1999] y en la Ingeniería del Conocimiento [Gómez et al., 1997; García-Martínez y Britos, 2004].
- Modelos de Proceso usuales en Ingeniería de Software [IEEE, 1997; ANSI/IEEE, 2007; Oktaba et al., 2007].

RESULTADOS OBTENIDOS/ESPERADOS

Como resultado de este proyecto, se esperan contar al fin de su desarrollo con un proceso robusto para ingeniería de proyectos de explotación de información de gran tamaño y una versión ágil de proceso para proyectos de pequeño porte.

Se estima también poder contribuir a otras líneas de investigación del grupo en las áreas de formalismos de modelado de requisitos de proyectos explotación

de información y métricas de desarrollo para los mismos.

FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo se encuentra formado por dos investigadores formados y tres investigadores en formación. En su marco se desarrolla una Tesis de Doctorado en Ciencias Informáticas y 2 Tesis de Maestría en Tecnología Informática.

FINANCIAMIENTO

Las investigaciones que se proponen en esta comunicación cuentan con financiamiento como Proyecto de Investigación 33A205 de la Secretaría de Ciencia y Técnica de la Universidad Nacional de Lanús (Argentina).

REFERENCIAS

Abraham, A. 2003. Business Intelligence from Web Usage Mining. *Journal of Information & Knowledge Management*, 24: 375-390.

Abrañ, A., Moore, J. W., Bourque, P., Dupuis, R., Tripp, L. 2004. *Guide to the Software Engineering Body of Knowledge (2004 version)*. IEEE Computer Society Press. ISBN 0-7695-2330-7.

ANSI/IEEE. 2007. *Draft IEEE Standard for software and system test documentation*. ANSI/IEEE Std P829-2007.

Argimón, J. 2004. *Métodos de Investigación Clínica y Epidemiológica*. Elsevier España. 84-8174-709-2.

Basilii 1993. *The Experimental Paradigm in Software Engineering*. En *Experimental Software Engineering Issues: Critical Assessment and Future Directions* (Ed. Rombach, H., Basili, V., Selby, R.). *Lecture Notes in Computer Science*, Vol. 706. ISBN 978-3-540-57092-9.

Basso, D., Rodríguez, D., García-Martínez, R. 2013. *Propuesta de Métricas para Proyectos de Explotación de Información*. *Workshop de Bases de Datos y Minería de Datos*. Proceedings XIX Congreso Argentino de Ciencias de la Computación. Pág. 983-992. ISBN 978-987-23963-1-2.

Basso, D. 2014. *Propuesta de Métricas para Proyectos de Explotación de Información*. *Revista Latinoamericana de Ingeniería de Software*, 2(4): 157-218, ISSN 2314-2642.

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.

Britos, P., Dieste, O., García-Martínez, R. 2008. *Requirements Elicitation in Data Mining for Business Intelligence Projects*. In *Advances in Information Systems Research, Education and Practice*. David Avison, George M. Kasper, Barbara Pernici, Isabel Ramos, Dewald Roode Eds. (Boston: Springer), IFIP International Federation for Information, 274: 139-150.

Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. 2000. *CRISP-DM 1.0 Step by step BI guide*. Edited by SPSS.

Cooley, R. 2003. *The Use of Web Structure and Content to Identify Subjectively Interesting Web Usage Patterns*. *ACM Transactions on Internet Technology*, 32: 93-116.

Creswell, J. 2002. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Prentice Hall. ISBN 10: 01-3613-550-1.

Curtis, B., Kellner, M., Over, J. 1992. *Process Modelling*. *Communications of the ACM*, 35(9): 75-90.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. 1996. *From data mining to knowledge discovery in databases*. *AI Magazine*, 17(3): 37-54.

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.

Gaber, M., Zaslavsky, A., Krishnaswamy, S. 2010. *Data stream mining*. En Maimon, O. and Rokach, L., eds. *Data mining and knowledge discovery handbook*. Springer, Pág. 759-787. ISBN 978-0-387-09823-4.

García Martínez, R. 1997. *Sistemas Autónomos. Aprendizaje Automático*. Editorial Nueva Librería. ISBN 950-9088-84-6. García Martínez, R. y Britos, P. 2004. *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería. ISBN 987-1104-15-4.

García Martínez, R. y Britos, P. 2004. *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería. ISBN 987-1104-15-4.

García Martínez, R., Servente, M. y Pasquini, D. 2003. *Sistemas Inteligentes*. Editorial Nueva Librería. Buenos Aires. ISBN 987-1104-05-7.

García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell, J. 2011. *Towards an Information Mining Engineering*. En *Software Engineering, Methods, Modeling and Teaching*. Sello Editorial Universidad de Medellín. ISBN 978-958-8692-32-6. Páginas 83-99.

García-Martínez, R., Britos, P., Rodríguez, D. 2013. *Information Mining Processes Based on Intelligent Systems*. *Lecture Notes on Artificial Intelligence*, 7906: 402-410. ISBN 978-3-642-38576-6.

Hsu, W., Lee, M., Zhang, J. 2002. *Image mining: Trends and developments*. *Journal of Intelligent Information Systems*, 19(1): 7-23.

IEEE. 1997. *IEEE Standard for Developing Software Life Cycle Processes*. IEEE Std 1074-1997 (Revision of IEEE Std 1074-1995; Replaces IEEE Std 1074-1-1995)

Jacobson, I., Ng, P., McMahon, P., Spence, I., Lidman, S. 2013. *The Essence of Software Engineering: Applying the SEMAT Kernel*. Addison Wesley. ISBN 9780321885951.

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.

Kosala, R., Blockeel, H. 2000. *Web mining research: A survey*. *ACM SIGKDD Explorations Newsletter*, 2(1): 1-15.

Kuna, H., García-Martínez, R., Villatoro, F. 2014. *Outliers Detection in Audit Logs for Application Systems*. *Information Systems Journal*, 44: 22-33 Elsevier. ISSN 0306-4379.

Langseth, J., Viatrat, N. 2003. *Why Proactive Business Intelligence is a Hallmark of the Real-Time Enterprise: Outward Bound*. *Intelligent Enterprise* 518: 34-41.

Lopez-Nocera, M., García-Martínez, R. 2012. *Variabilidad del Comportamiento de Agrupamiento e Inducción Basado en las Características del Dominio*. *Proceedings Latin American Congress on Requirements Engineering and Software Testing*. Pág. 23-28. ISBN 978-958-46-0577-1.

Lopez-Nocera, M., Britos, P., Rodríguez, D., García-Martínez, R. 2012. *Impacto de la Complejidad del Dominio en las Variaciones del Comportamiento de Procesos de Explotación de Información*. *Proceedings IX Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento*. Pág. 55-62. Sello Editorial de la Pontificia Universidad Católica del Perú. ISBN 978-612-4057-85-4.

Lopez-Nocera, M., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2011. *Un Protocolo de Caracterización Empírica de Dominios para Uso en Explotación de Información*. *Proceedings XVII Congreso Argentino de Ciencias de la Computación*. Pág. 1047-1055. ISBN 978-950-34-0756-1.

Lyman, P., Varain, H. 2003. *How Much Information?*. School of Information Management & Systems. University of California Berkeley. <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>. Ultimo acceso 23/03/2014.

Maimon, O. y Rokach, L. (Eds.). 2005. *Data mining and knowledge discovery handbook*. Springer.

Mansilla, D., Pollo, F., Britos, P., García-Martínez, R. 2012. *Modelo de Proceso para Elicación de Requerimientos en Proyectos de Explotación de Información*. *Proceedings Latin American Congress on Requirements Engineering and Software Testing*. Pág. 38-45. ISBN 978-958-46-0577-1.

Mansilla, D., Pollo, F., Britos, P., García-Martínez, R. 2013. *A Proposal of a Process Model for Requirements Elicitation in Information Mining Projects*. *Lecture Notes in Business Information Processing*, 139: 165-173. ISBN 978-3-642-36610-9.

Martins, S. 2014. *Derivación del Proceso de Explotación de Información Desde el Modelado del Negocio*. *Revista Latinoamericana de Ingeniería de Software*, 2(1): 53-76, ISSN 2314-2642.

Martins, S., Rodríguez, D., García-Martínez, R. 2014a. *Derivación del Proceso de Explotación de Información desde el Dominio de Negocio*. *Capítulo X en "Ingeniería del Software e Ingeniería del Conocimiento: Dos Disciplinas Interrelacionadas"*. Pág. 159-177. Sello Editorial de la Universidad de Medellín. ISBN 978-958-8815-31-2.

Martins, S., Rodríguez, D., García-Martínez, R. 2014. *Deriving Processes of Information Mining Based on Semantic Nets and Frames*. *Lecture Notes on Artificial Intelligence*, 8482: 150-159. ISBN 978-3-319-07466-5.

MI. 2014. *Capítulo XII: Cadena de valor del software y servicios informáticos*. Plan Estratégico Industrial 2020. Ministerio de Industria. Presidencia de la Nación. Republica Argentina. <http://www.industria.gov.ar/software-2/>. Pagina vigente al 14/08/2014

MINCYT. 2014. *Argentina Innovadora 2020*. Plan Nacional de Ciencia, Tecnología e Innovación. *Lineamientos Estratégicos 2012-2015*. Ministerio Nacional de Ciencia y Tecnología. Presidencia de la Nación. Republica Argentina. <http://www.mincyt.gov.ar/adjuntos/archivos/000/022/0000022576.pdf>. Pagina vigente al 14/08/2014.

Moss, L. 2003. *Non-technical Infrastructure of BI Applications*. *DM Review* 131: 42-45.

Negash, S., Gray, P. 2008. *Business Intelligence*. En *Handbook on Decision Support Systems 2*, ed. F. Burstein y C. Holsapple Heidelberg. Springer, Pág. 175-193.

Ochoa, A., Fernández, E., Britos, P., García-Martínez, R. 2008. *Metodologías de Ingeniería Informática*. Editorial Nueva Librería. ISBN 978-987-1104-54-3.

Okta, H., García, F., Plattini, M., Ruiz, F., Pino, F., Alcúrciga, C. 2007. *Software Process Improvement: The Competisofit Project*. *IEEE Computer*, 40(10): 21-28. ISSN 0018-9162.

PAE. 2014. *Métrica v.3*. Portal de Administración Electrónica. Gobierno de España. http://administracionelectronica.gob.es/pae_home/pae_Documentacion/pae_Metodolog/pae_Metrica_v3.html#U_DLqYpP-TQ. Pagina vigente al 15/08/2014.

Pyle, D. 2003. *Business Modeling and Business Intelligence*. Morgan Kaufmann Publishers.

Pytel, P., Britos, P., García-Martínez, R. 2013a. *Proposal and Validation of a Feasibility Model for Information Mining Projects*. *Proceedings 25th International Conference on Software Engineering and Knowledge Engineering*. Pág. 83-88. ISBN 978-1-891706-33-2.

Pytel, P., Amatrián, H., Britos, P., García-Martínez, R. 2012a. *Estudio del Modelo para Evaluar la Viabilidad de Proyectos de Explotación de Información*. *Proceedings IX Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento*. Pág. 63-70. Sello Editorial de la Pontificia Universidad Católica del Perú. ISBN 978-612-4057-85-4.

Pytel, P., Britos, P., García-Martínez, R. 2012b. *Comparación de Métricas de Estimación para Proyectos de Explotación de Información*. *Proceedings of Latin American Congress on Requirements Engineering and Software Testing*. Pág. 29-37. ISBN 978-958-46-0577-1.

Pytel, P., Britos, P., García-Martínez, R. 2012c. *Propuesta de un Modelo para Evaluar la Viabilidad de Proyectos de Explotación de Información*. *Proceedings del XVIII Congreso Argentino de Ciencias de la Computación*. Pág. 1039-1048. N 978-987-1648-34-4.

Pytel, P., Britos, P., García-Martínez, R. 2012d. *Initial Activities Oriented to Reduce Failure in Information Mining Projects*. *Capítulo 2 en Software Engineering: Methods, Modeling, and Teaching*. Volume 2. Pág. 11-20. Sello Editorial de la Pontificia Universidad Católica del Perú. ISBN 978-612-4057-84-7.

Pytel, P., Britos, P., García-Martínez, R. 2013b. *A Proposal of Effort Estimation Method for Information Mining Projects Oriented to SMEs*. *Lecture Notes in Business Information Processing*, 139: 58-74. ISBN 978-3-642-36610-9.

Pytel, P., Britos, P., García-Martínez, R. 2013c. *Modelos para Asistir la Gestión de Proyectos de Explotación de Información*. *Revista Latinoamericana de Ingeniería de Software*, 1(1): 8-17, ISSN 2314-2642.

Pytel, P., Hossain, A., Britos, P., García-Martínez, R. 2015. *Feasibility and Effort Estimation Models for Medium and Small Size Information Mining Projects*. *Information Systems Journal*, 47: 01-14. Elsevier. ISSN 0306-4379.

Pytel, P., Pollo-Cattaneo, F., Rodríguez, D., Britos, P., García-Martínez, R. 2011. *Identificación de Tareas Críticas en una Metodología de Desarrollo de Proyectos de Explotación*. *Proceedings XVII Congreso Argentino de Ciencias de la Computación*. Pág. 989-998. ISBN 978-950-34-0756-1.

Rancan, C., Pesado, P., García-Martínez, R. 2010. *Issues in Rule Based Knowledge Discovering Process*. *Advances and Applications in Statistical Sciences Journal* [ISSN 0974-6811], 2(2): 303-314. ISSN 0974-6811.

Riveros, H. y Rosas, L. 1985. *El Método Científico Aplicado a las Ciencias Experimentales*. Editorial Triliter. México. ISBN 96-8243-893-4.

Rodríguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2010. *Estimación Empírica de Carga de Trabajo en Proyectos de Explotación de Información*. *Anales del XVI Congreso Argentino de Ciencias de la Computación*. Pág. 664-673. ISBN 978-950-9474-49-9.

Rudin, K., Cressy, D. 2003. *Will the Real Analytic Application Please Stand Up?* *DM Review* 133: 30-34.

Rumbaugh, J., Jacobson, I., Booch, G. 1999. *The Unified Modeling Language, Reference Manual*. Addison Wesley, ISBN-10: 02-0130-998-X.

Sábato, J. y Mackenzie, M. 1982. *La Producción de Tecnología*. Editorial Nueva Imagen. México. ISBN 968-429-348-8.

SAS. 2008. *SAS Enterprise Miner: SEMMA*. <http://www.sas.com/technologies/analytics/datamining/miner/semma.html>. Ultimo acceso 23/03/2014.

Srivastava, J., Cooley, R., Deshpande, M., Tan, P. 2000. *Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data*. *SIGKDD Explorations*, 12: 12-23.

Tan, A. 1999. *Text mining: The state of the art and the challenges*. In *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*, pp. 65-70.

Thomsen, E. 2003. *BI's Promised Land*. *Intelligent Enterprise*, 64: 21-25.

Vanrell, J., Bertone, R., García-Martínez, R. 2010. *Modelo de Proceso de Operación para Proyectos de Explotación de Información*. *Anales del XVI Congreso Argentino de Ciencias de la Computación*. Pág. 674-682. ISBN 978-950-9474-49-9.

Vanrell, J., Bertone, R., García-Martínez, R. 2012. *Un Modelo de Procesos para Proyectos de Explotación de Información*. *Proceedings Latin American Congress on Requirements Engineering and Software Testing*. Pág. 46-52. ISBN 978-958-46-0577-1.

Vegega, C., Amatrián, H., Pytel, P., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2012a. *Formalización de Dominios de Negocio basada en Técnicas de Ingeniería del Conocimiento para Proyectos de Explotación de Información*. *Proceedings IX Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento*. Pág. 79-86. Sello Editorial de la Pontificia Universidad Católica del Perú. ISBN 978-612-4057-85-4.

Vegega, C., Pytel, P., Ramón, H., Rodríguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2012b. *Formalización de Dominios de Negocio para Proyectos de Explotación de Información basada en Técnicas de Ingeniería del Conocimiento*. *Proceedings del XVIII Congreso Argentino de Ciencias de la Computación*. Pág. 1049-1058. ISBN 978-987-1648-34-4.

Vuori, V. 2006. *The Employees as a Source of External Business Information*. *Proceedings European Productivity Conference EPC'06*. Pág. 29-36.