

PRÁCTICAS Y APLICACIONES DE INGENIERÍA DE REQUISITOS EN PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN

Pollo-Cattaneo, M., Pytel, P., García-Martínez, R., Vegega, C., Amatriain, H., Ramón, H., Mansilla, D., Deroche, A., Cigliuti, P., Saavedra-Martínez, P., Garbarini, R., Rodriguez, D., Britos, P., Tomasello, M.

Grupo de Estudio en Metodologías de Ingeniería de Software

Universidad Tecnológica Nacional.
Facultad Regional Buenos Aires.
Medrano 951 (C1179AAQ) Ciudad Autónoma de Argentina. Buenos Aires Tel +54 11 4867-7511
fpollo@posgrado.frba.utn.edu.ar

Grupo Investigación en Sistemas de Información

Departamento de Desarrollo Productivo y Tecnológico
Universidad Nacional de Lanús
29 de Septiembre 3901 (1826) Remedios de Escalada, Lanús. Argentina.
Tel +54 11 6322-9200 Ext. 194
rgarcia@unla.edu.ar

Grupo de Investigación en Explotación de Información

Universidad Nacional de Río Negro
San Martín esq. Pellegrini (8430) El Bolsón. Río Negro. Argentina.
TE + 54 11 02944 49-8939
paobritos@gmail.com

RESUMEN

La Explotación de Información es la sub-disciplina de la Informática que aporta las herramientas de análisis y síntesis para extraer conocimiento no trivial que se encuentra implícitamente en los datos disponibles en distintas fuentes de información. Aunque la Ingeniería en Software provee muchos métodos, técnicas y herramientas para la construcción de software tradicional, estos en general no son aplicables a proyectos de Explotación de Información por poseer características diferentes. En especial se ha detectado la falta de metodologías que permitan cubrir adecuadamente la gestión de los requisitos de un Proyecto de Explotación de Información, permitiendo tanto la identificación y formalización de las necesidades del cliente como sus expectativas, restricciones y los principales conceptos que son necesarios para realizar el proyecto.

En este contexto, este proyecto se enfocará en sistematizar el cuerpo de conocimiento existente en la Ingeniería en Software y la Ingeniería del Conocimiento para sentar las bases para el desarrollo de una Ingeniería de Requisitos con particular énfasis en Proyectos de Explotación de Información. De esta manera se espera también producir un enriquecimiento en los campos. En relación a la transferencia específica se puede indicar la generación de un Modelo de Proceso para la gestión de Requisitos.

Palabras clave: Proyecto de Explotación de Información, Requerimiento, Ingeniería en Software, Information Mining.

CONTEXTO

Este proyecto de investigación se desarrolla en el marco de la cooperación existente entre el Grupo de Investigación en Sistemas de Información (GISI) del Departamento de Desarrollo Productivo y Tecnológico de la Universidad Nacional de Lanús, el Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS) de la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional, y el Grupo de Investigación en Explotación de Información de la Universidad Nacional de Río Negro (SAEB-UNRN).

INTRODUCCIÓN

La Inteligencia de Negocio propone un abordaje interdisciplinario (dentro del que se encuentra la Informática), que se centra en generar conocimiento que contribuya con la toma de decisiones de gestión y generación de planes estratégicos en las organizaciones [Thomsen, 2003]. La Explotación de Información (EdI) es la sub-disciplina de la Informática que aporta a la Inteligencia de Negocio [Negash & Gray, 2008] las herramientas de análisis y síntesis para extraer conocimiento no trivial que se encuentra (implícitamente) en los datos disponibles de diferentes fuentes de información [Schiefer et al., 2004]. Para un experto, o para el responsable de un Sistema de Información, normalmente no son los datos en sí lo más relevante, sino el conocimiento que se encierra en sus relaciones, fluctuaciones y dependencias. Un Proceso de Explotación de Información puede definirse como un conjunto de tareas relacionadas lógicamente [Curtis et

al., 1992], que se ejecutan para lograr, a partir de un conjunto de información con un grado de valor para la organización, otro conjunto de información con un grado de valor mayor que el inicial [Ferreira et al., 2005; Kanungo, 2005]. Identificado el problema de Inteligencia de Negocio, un Ingeniero de Explotación de Información decide la secuencia de Procesos de Explotación de Información que deben ser ejecutados para obtener una solución al problema de Inteligencia de Negocio. Cada Proceso de Explotación de Información tiene varias técnicas de Minería de Datos que pueden ser elegidas para realizarlas [García-Martínez et al., 2011a]. Muchas de esas técnicas vienen del campo del Aprendizaje Automático [García-Martínez et al., 2003].

En [García-Martínez et al., 2011b] se ha observado el uso indiscriminado de los términos “Minería de Datos” (o “Data Mining” en inglés) y “Explotación de Información” (o “Information Mining” en inglés) para referirse al mismo cuerpo de conocimientos. Esto es un tipo de confusión similar a utilizar como sinónimos “Ciencias de la Computación” y “Sistemas de Información”. La Minería de Datos está relacionada a la tecnología (algoritmos) necesaria para la obtención de conocimiento y la Explotación de Información está relacionada con los procesos y las metodologías necesarias para realizar el proyecto con éxito. De esta forma, la Minería de Datos se encuentra más cerca de las tareas de programación mientras que la Explotación de Información está cercana a la Ingeniería en Software.

Aunque la Ingeniería en Software y la Ingeniería del Conocimiento [García-Martínez & Britos, 2004] provee muchos métodos, técnicas y herramientas, estos no son útiles ya que no se ocupan de los aspectos específicos que poseen los Proyectos de Explotación de Información (PEI). Mientras, que las herramientas tradicionales de elicitación de la Ingeniería en Software se enfocan en la descripción de los diferentes tipos de requerimientos haciendo hincapié en las características que debe cumplir el producto software final [Wieggers, 2003], un Proyecto Explotación de Información no busca la

construcción del sistema software sino la aplicación de un proceso que convierta los datos disponibles en conocimiento. Por lo tanto, se considera necesario el desarrollo y la validación de métodos, técnicas y herramientas que puedan asistir a los practicantes del área de ingeniería de requisitos y proveer la necesaria objetividad, racionalidad, generalización y confiabilidad a los Proyectos de Explotación de Información.

En este contexto, se ha identificado como un problema abierto la necesidad de organizar un nuevo cuerpo de conocimiento relacionado a la identificación, formalización y administración de los deseos y necesidades de los interesados en un Proyecto de Explotación de Información. Una de las razones para desarrollar una Ingeniería de Requisitos de Explotación de Información ha sido el descubrimiento de una falta de actividades y técnicas asociadas a la ejecución de las fases correspondientes en las metodologías, normalmente aplicadas para el desarrollo de Proyectos de Explotación de Información. Si bien se destacan las metodologías SEMMA [SAS, 2008], P3TQ [Pyle, 2003] y CRISP-DM [Chapman et al., 2000] las cuales se consideran probadas y tienen un buen nivel de madurez, éstas dejan de lado aspectos a nivel operativo sobre las tareas necesarias para elicitar, formalizar y administrar los requisitos. Por otro lado, existen aproximaciones que tratan de integrar el conocimiento, como es el ciclo de vida propuesto por [Kimball et al., 2011] que es utilizado en iniciativas de Data Warehouse & Business Intelligence. Sin embargo, es necesario adaptar este ciclo de vida para considerar las particularidades de los Proyectos de Explotación de Información.

En este sentido, se han detectado la falta de procesos que permitan cubrir adecuadamente los siguientes aspectos de la Ingeniería de Requisitos en Proyectos de Explotación de Información:

- identificación de las principales necesidades del cliente, sus expectativas, restricciones y los principales conceptos que son necesarios para realizar el proyecto [Pollo-Cattaneo et al., 2009].
- el proceso de elicitación de requisitos mediante la transformación de la

experiencia en el uso de las herramientas disponibles para la elicitación en el dominio de los sistemas tradicionales [Mansilla et al., 2012].

- el proceso de formalización de requisitos a través de la utilización de técnicas de representación de conocimiento provistas por la Ingeniería del Conocimiento [Vegega et al., 2012].

OBJETIVOS E HIPOTESIS DE INVESTIGACION

Este proyecto se inscribe en una línea de investigación que busca adaptar buenas prácticas ingenieriles (es decir procesos, métodos, técnicas y herramientas) existentes en la Ingeniería en Software y la Ingeniería del Conocimiento como también desarrollar nuevas para ser utilizadas en la Gestión de Requisitos de los Proyectos de Explotación de Información teniendo en cuenta las particularidades de las Pequeñas y Medianas Empresas.

Entre los supuestos (o hipótesis) que guían el proyecto se encuentran los siguientes:

- *Hipótesis I:* Los proyectos de Explotación de Información poseen características muy distintas a las de los proyectos de desarrollo de software tradicional y no tradicional. En consecuencia, muchas de los procesos, métodos, técnicas y herramientas no pueden ser aplicables a proyectos de Explotación de Información.
- *Hipótesis II:* A pesar de existir metodologías para el desarrollo de proyectos de Explotación de Información, muchas de estas metodologías dejan de lado aspectos a nivel operativo por no indicar el método y las técnicas que deberían ser utilizadas para la gestión de requisitos.
- *Hipótesis III:* Los riesgos y problemas encontrados durante el desarrollo de un proyecto de Explotación de Información pueden ser reducidos mediante la aplicación de buenas prácticas provenientes de la Ingeniería en Software y la Ingeniería del Conocimiento.

Teniendo en cuenta dichas hipótesis se buscará:

- *Objetivo General:* En el marco de proyectos de Explotación de Información, este proyecto se enfocará en sistematizar el

cuerpo de conocimientos existente en la Ingeniería en Software y la Ingeniería del Conocimiento para sentar las bases para el desarrollo de prácticas ingenieriles para la Gestión de Requisitos en Proyectos de Explotación de Información.

Asociados a este objetivo general se definen los siguientes objetivos específicos:

- *Objetivo Específico 1:* Identificar las principales actividades, técnicas y herramientas para ser utilizadas en cada fase del proceso de elicitación de requisitos de un proyecto de Explotación de Información.
- *Objetivo Específico 2:* Identificar las principales actividades, técnicas y herramientas para ser utilizadas en cada fase del proceso de formalización de requisitos de un proyecto de Explotación de Información.
- *Objetivo Específico 3:* Proponer un modelo que permita la gestión de requisitos en proyectos de Explotación de Información dentro del marco de un ciclo de vida definido ad-hoc.

FUNDAMENTACION E IMPORTANCIA

La necesidad de desarrollar una Ingeniería de Requisitos para Proyectos de Explotación de Información surge del relevamiento efectuado en el campo metodológico, en el que se identifica la carencia de técnicas asociadas a la ejecución de las primeras fases planteadas en las metodologías identificadas. En este contexto, este proyecto promueve el desarrollo y la validación de métodos, técnicas y herramientas, conllevando a una mejora en el campo de la Ingeniería de Requisitos. Los métodos con abordaje ingenieril permiten dotar al proceso de desarrollo de: objetividad, sistematicidad, racionalidad, generalidad y fiabilidad, contribuyendo al avance del conocimiento científico mediante el uso de técnicas consistentes. De esta manera, los resultados de este proyecto pueden ser transferidos, en sus distintos niveles de gestión, para formular políticas en varias dimensiones. El crecimiento de los proyectos informáticos con aplicación de técnicas inteligentes en la actualidad delimita la importancia de utilizar

eficientemente los recursos disponibles para el desarrollo de las actividades correspondientes.

METODOLOGÍA DE TRABAJO

Las tareas a realizar para desarrollar los procesos serán las siguientes:

- a) Se realizará una investigación documental exploratoria tratando de identificar los conceptos de interés para caracterizar un proyecto de Explotación de Información.
- b) Se buscará identificar relaciones existentes entre los diversos conceptos identificados, sus posibles fuentes y su posible aplicación para realizar la gestión de requisitos de un proyecto de Explotación de Información.
- c) Para el proceso de elicitación de requisitos se realizarán las siguientes actividades:
 1. Se identificarán y analizarán las técnicas y herramientas existentes en la Ingeniería en Software aplicables para proyectos de Explotación de Información.
 2. Se determinará el orden de las técnicas y herramientas para realizar la elicitación de requisitos en proyectos de explotación de información indicando las entradas y salidas necesarias.
 3. Se realizarán pruebas de concepto para validar el proceso de elicitación desarrollado.
- d) Para el proceso de formalización de requisitos se realizarán las siguientes actividades:
 1. Se identificarán y analizarán las técnicas y herramientas existentes en la Ingeniería en Software e Ingeniería del Conocimiento aplicables para proyectos de Explotación de Información.
 2. Se determinará el orden de las técnicas y herramientas para realizar la formalización de requisitos en proyectos de Explotación de Información indicando las entradas y salidas necesarias.
 3. Se realizarán pruebas de concepto para validar el proceso de formalización desarrollado.
- e) A partir de los resultados de los proceso de elicitación y formalización de requisitos, se propondrá un modelo que permita la gestión de requisitos en proyectos de Explotación

de Información dentro del marco de un ciclo de vida definido ad-hoc.

- f) Se realizarán pruebas de concepto para validar el modelo de proceso desarrollado.

Para el desarrollo de las tareas propuestas se utilizarán los siguientes materiales:

- Metodologías CRISP-DM [Chapman, et al., 2000], P3TQ [Pyle, 2003] y SEMMA [SAS, 2008].
- Ciclo de Vida para iniciativas de Data Warehouse & Business Intelligence [Kimball et al., 2011].
- Trabajos previos de los integrantes del proyecto relacionados con la Explotación de Información.
- Métodos, técnicas y herramientas existentes en Ingeniería en Software.
- Métodos, técnicas y herramientas existentes en Ingeniería del Conocimiento.
- Casos de estudio tomados tanto de proyectos reales desarrollados por los integrantes del proyecto como de trabajos de tesis de maestría y/o especialización disponibles (como por ejemplo son [Cigliuti et al., 2012] y [Saavedra-Martínez et al., 2012]).
- Reportes sobre proyectos de Explotación de Información de alumnos de la carrera Ingeniería en Sistemas de Información de la UTN-FRBA y UTN-FRCU.
- Hemerotecas y Repositorios Científicos accesibles por Internet.
- Acceso a bibliotecas digitales de IEEE, ACM y SCOPUS.

RESULTADOS OBTENIDOS/ESPERADOS

Como resultado de este proyecto, se esperan los siguientes aportes:

- Generar un modelo de proceso para la gestión de requisitos dentro de proyectos de Explotación de Información con la especificación de un ciclo de vida ad-hoc, la cual será transferible a una Software Factory.
- Mejorar la comprensión de las potenciales aplicaciones de la disciplina por parte de los docentes integrantes del proyecto.

FORMACIÓN DE RECURSOS HUMANOS

Este proyecto busca tanto la obtención de nuevos conocimientos como la motivación de los implicados para que asciendan dentro del escalafón de la carrera de investigadores.

Los recursos humanos deben poseer una firme vocación de trabajar en el área de la ingeniería en software, los sistemas inteligentes con su vinculación a la Explotación de Información, de manera de canalizar y proveer una base sustentable de aporte al proyecto. De esta manera, se logra un doble beneficio, el proyecto obtiene e incorpora el conocimiento tecnológico de los recursos humanos en el área de la especialidad, a la vez que plantea un esquema de formación de especialistas de punta en el proceso de gestión.

Además, en el marco de este proyecto de investigación se prevé la radicación de una Tesis de Doctorado en Ciencias Informáticas y cinco tesis de magister en distintas áreas de la informática en codirección con otras instituciones. También se espera que los alumnos becarios que participan del proyecto tengan la posibilidad de articular su Trabajo Final de Carrera de grado y posgrado con temas relacionados a este proyecto.

REFERENCIAS

- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step Data Mining Guide. <http://tinyurl.com/crispdm> Último acceso Enero de 2013.
- Cigliuti, P., Pollo-Cattaneo, F., García-Martínez, R. (2012). Procesos de Identificación de Comportamiento de Comunidades Educativas Centradas en EVEAs. Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación. Pág. 954-956. ISBN 978-950-766-082-5.
- Curtis, B., Kellner, M., Over, J. (1992). Process Modelling. Communications of the ACM, 35(9): 75-90.
- Ferreira, J., Takai, O. & Pu, C. (2005). Integration of Business Processes with Autonomous Information Systems: A Case Study in Government Services. Proceedings Seventh IEEE International Conference on E-Commerce Technology. Pág. 471-474.
- García Martínez, R. y Britos, P. (2004). Ingeniería de Sistemas Expertos. Editorial Nueva Librería.
- García-Martínez, R., Servente, M. y Pasquini, D. (2003). Sistemas Inteligentes. Editorial Nueva Librería. ISBN 987-1104-05-7
- García-Martínez, R., Britos, P., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P. (2011a). Information Mining Processes Based on Intelligent Systems. Proceedings of II International Congress on Computer Science and Informatics (INFONOR-Chile 2011). Pág. 87-94. ISBN 978-956-7701-03-2.
- García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell, J. (2011b). Towards an Information Mining Engineering. En Software Engineering, Methods, Modeling and Teaching. Sello Editorial Universidad de Medellín. ISBN 978-958-8692-32-6. Pág. 83-99.
- Kanungo, S. (2005). Using Process Theory to Analyze Direct and Indirect Value-Drivers of Information Systems. Proceedings of the 38th Annual Hawaii International Conference on System Sciences. Pág. 231-240.
- Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., & Becker, B. (2011). The data warehouse lifecycle toolkit. Wiley & Sons.
- Mansilla, D., Pollo-Cattaneo, M., Pytel, P., García-Martínez, R. (2012). Modelo de Proceso para Elicitación de Requerimientos en Proyectos de Explotación de Información. Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación. Pág. 222-225. ISBN 978-950-766-082-5
- Negash, S. & Gray, P. (2008). Business Intelligence. En In Handbook on Decision Support Systems 2, ed.eds. F. Burstein y C. Holsapple (Heidelberg, Springer), Pág.Pp. 175-193.
- Pollo-Cattaneo, F., Britos, P. Pesado, P., García-Martínez, R. (2009). Metodología para Especificación de Requisitos en Proyectos de Explotación de Información. Proceedings XI Workshop de Investigadores en Ciencias de la Computación. Pág. 333-335. ISBN 978-950-605-570-7.
- Pyle, D. (2003). Business Modeling and Business intelligence. Morgan Kaufmann Publishers.
- Saavedra-Martínez, P., Pollo-Cattaneo, M., Pytel, P., Rodríguez, D., García-Martínez, R. (2012). Detección de Problemas de Aprendizaje Basado en Explotación de Información. Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación. Pág. 962-964. ISBN 978-950-766-082-5.
- SAS (2008). SAS Enterprise Miner: SEMMA. <http://tinyurl.com/semmaSAS> Último acceso Enero de 2013.
- Schiefer, J., Jeng, J., Kapoor, S. & Chowdhary, P. (2004). Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence. Proceedings 2004 IEEE International Conference on E-Commerce Technology. Pág.Pp. 162-169.
- Thomsen, E. (2003). BI's Promised Land. Intelligent Enterprise, 6(4): 21-25.
- Vegega, C., Mansilla, D., Pollo-Cattaneo, M. F., Pytel, P., Rodríguez, D., Diez, E., García-Martínez, R. (2012). Documentación de Requisitos en Proyectos de Explotación de Información. Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación. Pág. 226-229. ISBN 978-950-766-082-5.
- Wieggers, K. (2003). Software Requirements. Microsoft Press.