

Líneas de I+D+I del Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de Información (LIDIEI GISI UNLa)

R. García-Martínez, D. Rodríguez, E. Baldizzoni, S. Martins

Grupo Investigación en Sistemas de Información
Licenciatura en Sistemas. Universidad Nacional de Lanús
<http://www.unla.edu.ar/sistemas/gisi/LIDIEI.htm>
{rgarcia,darodriguez,ebaldizzoni,smartins}@unla.edu.ar

Resumen

Los Proyectos de Ingeniería de Explotación de Información (PIEI) difieren sustancialmente de los de Ingeniería de Software tradicional. Las fases clásicas de procesos de desarrollo de artefactos, y las técnicas/herramientas involucradas en las actividades de cada fase son distintas aunque la visión ingenieril del proceso constructivo es común. Así como la codificación es solo una actividad en una fase de un proceso de Ingeniería de Software, la elección de un algoritmo de minería de datos es una actividad en una fase de un Proceso de Ingeniería de Explotación de Información. En este contexto, el LIDIEI busca, a partir de las líneas de investigación que se presentan en esta comunicación, formular contribuciones al cuerpo de conocimientos de esta nueva subdisciplina Informática, con atención a las necesidades de su aplicación industrial.

Palabras Clave: inteligencia de negocio, ingeniería de proyectos de explotación, gestión, procesos, especificación de requisitos, viabilidad y estimación, modelo de procesos, métricas.

Contexto

El Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de

Información (LIDIEI) integra el Grupo de Investigación en Sistemas de Información (GISI) de la Licenciatura en Sistemas de la Universidad Nacional de Lanús. Articula líneas de trabajo de proyectos de investigación acreditados institucionalmente por la Secretaria de Ciencia y Técnica de la UNLa; y tiene radicados planes de trabajo correspondientes a tesis doctorales, tesis de maestría y trabajos finales de licenciatura propios y de otras instituciones universitarias.

Introducción

El LIDIEI organiza sus actividades de investigación, desarrollo e innovación a través de diez líneas de trabajo: Sistemas de Información para Inteligencia de Negocio; Ingeniería de Proyectos de Explotación de Información para PYMES; Instrumentos para la Gestión de PIEI; Procesos de Explotación de Información Georeferenciada; Metodología para la Especificación de Requisitos en PIEI (en colaboración con UTN-FRBA); Viabilidad y Estimación de PIEI; Procedimientos de Explotación de Información para la Identificación de Datos Faltantes, con Ruido e Inconsistentes (en colaboración con UNaM); Modelo de Procesos para PIEI; Descubrimiento de Conocimiento Mediante la Integración de

Algoritmos de Explotación de la Información; y Métricas para PIEI.

Líneas de Investigación, Desarrollo e Innovación

Sistemas de Información para Inteligencia de Negocio

Esta línea busca: [a] caracterizar las aplicaciones de los procesos de explotación de información asociados a los problemas de inteligencia de negocio; y [b] identificar las tecnologías de acceso libre para una PyME que pueden utilizarse para los procesos caracterizados, validando estos procesos a través de casos aceptados por la comunidad internacional.

Ingeniería de Proyectos de Explotación de Información para PYMEs

Esta línea intenta, utilizando las metodologías de investigación documental exploratoria, prototipado evolutivo y casos de estudio, el desarrollo de los siguientes artefactos de Ingeniería de Proyectos de Explotación de Información: [a] una batería de técnicas de educación y formalismos de documentación de requerimientos; [b] un modelo de procesos y las métricas asociadas; y [c] un modelo de ciclo de vida.

Instrumentos para la Gestión de Proyectos de Explotación de Información

En esta línea se propone desarrollar, mediante la metodología de prototipado evolutivo, los siguientes instrumentos de control y la gestión de PIEI: [a] test de viabilidad, [b] procedimiento de derivación del problema de negocio en problema de explotación de información, y [c] técnicas y métodos de aseguramiento de la calidad del proceso y del producto.

Procesos de Explotación de Información Georeferenciada

Este proyecto tiene como objetivo caracterizar los procesos de explotación de

información georeferenciada asociados a detección de patrones y regularidades importantes en artefactos de información geográfica e identificar las tecnologías de Sistemas Inteligentes que se pueden utilizar para los procesos caracterizados. Se proponen los siguientes objetivos específicos: [a] caracterizar y desarrollar el proceso de descubrimiento de reglas de asociación en espacios georeferenciados; [b] caracterizar y desarrollar el proceso de descubrimientos de grupos en espacios georeferenciados; y [c] caracterizar y desarrollar el proceso de descubrimiento de interdependencias en espacios georeferenciados.

Metodología para la Especificación de Requisitos en Proyectos de Explotación de Información (en colaboración con UTN-FRBA)

Esta línea tiene como objetivo definir una metodología de educación de requisitos para PIEI en el marco del ciclo de vida utilizado en este tipo de proyectos. Se enfoca a construir una metodología que permita la educación de requisitos teniendo presente los distintos procesos que involucra el ciclo de vida usado en este tipo de proyectos. Implica investigación básica en la definición del proceso de educación de requisitos en PIEI.

Viabilidad y Estimación de Proyectos de Explotación de Información

Muchos de los métodos, técnicas y herramientas que provee la Ingeniería en Software no son aplicables a los proyectos de explotación de información dada sus características diferenciales. Si bien existen metodologías que acompañan el desarrollo de proyectos de explotación de información entre las cuales se destacan CRISP, P3TQ y SEMMA, que se consideran probadas y tienen un buen nivel de madurez en cuanto al desarrollo del proyecto, estas metodologías dejan de lado aspectos a nivel gestión de los proyectos y de empresa.

En este contexto, esta línea busca desarrollar procesos ad-hoc que permitan en el marco de los PIEI: [a] evaluar la viabilidad a partir de la información disponible al comienzo del proyecto, y [b] estimar los recursos necesarios con énfasis en su utilización en PyMEs.

Procedimientos de la Explotación de Información para la Identificación de Datos Faltantes, con Ruido e Inconsistentes (en colaboración con la Universidad Nacional de Misiones)

Existen antecedentes de procedimientos para la implementación de procesos de explotación de información en auditoría, pero no procedimientos específicos de explotación de información para datos faltantes, con ruido e inconsistentes, en ese contexto en esta línea se propone establecer procedimientos que identifiquen este tipo de datos. Se busca establecer una taxonomía relacionada con la calidad de los datos, analizando las técnicas de minería que mejor aplican, se explorarán esas técnicas analizando las ventajas y desventajas de cada una de ellas.

Modelo de Procesos para Proyectos de Explotación de Información

El objetivo de esta línea es desarrollar un modelo de procesos para PIEI, atendiendo los siguientes objetivos específicos: [a] establecer diferencias y relaciones entre los PIEI y los proyectos clásicos, [b] determinar límites, alcance y componentes de modelos existentes aplicables a PIEI, [c] identificar los procesos específicos de PIEI a través de las distintas metodologías existentes que se aplican a dichos proyectos y que no estén cubiertos por los modelos de procesos de software considerados, [d] seleccionar los procesos a incluir en el nuevo modelo, [e] desarrollar el nuevo modelo basándose en modelos existentes a partir de la documentación de los procesos identificados y utilizando el patrón usual de procesos, y [f] validar el modelo mediante

su uso en proyectos reales o a través de juicio experto.

Descubrimiento de Conocimiento Mediante la Integración de Algoritmos de Explotación de Información

En Minería de Datos se parte del supuesto que los dominios a los cuales aplican los algoritmos son equivalentes o directamente indistinguibles entre sí a los efectos experimentales. En esta línea se postula que esto no es así, y que la caracterización de los dominios en uso durante la fase experimental tiene amplia injerencia en los resultados observados y finalmente obtenidos. Se ha propuesto un experimento, diseñado para estudiar la combinación de algoritmos de agrupamiento con algoritmos de inducción y su aplicación al descubrimiento de conocimiento en ejemplos de dominios con distinto grado de complejidad. Se espera que de los resultados experimentales surja una caracterización del comportamiento de cada combinación de algoritmos en función de la complejidad de cada dominio.

Métricas para Proyectos de Explotación de Información

El objetivo de la línea es definir una propuesta de métricas aplicables a proyectos de Explotación de Información para PyMEs (Pequeñas y Medianas Empresas), que permitan medir y controlar el avance del desarrollo de un proyecto de este tipo. Constituyen objetivos específicos: [a] establecer las categorías aplicables al proceso de desarrollo de proyectos de Explotación de Información, como forma de clasificación de las métricas a proponer; [b] proponer un conjunto de métricas aplicables para proyectos de explotación de información siguiendo los lineamientos, subprocesos y tareas enunciadas en el proceso de desarrollo definido en [Vanrell, 2012]. Las métricas propuestas, constituirán el principal aporte por tratarse de métricas novedosas, que no se encuentran definidas

en ningún modelo de procesos de explotación de información.

Resultados obtenidos/esperados

Se ha identificado la necesidad de desarrollar la Ingeniería de Explotación de Información como una subdisciplina de la Informática [García-Martínez et al., 2011]. Se ha trabajado, en colaboración con docentes investigadores de la UTN-FRBA, en la definición de un proceso de deducción de requerimientos para PIEI [Mansilla et al., 2012; 2013; Pollo-Cattaneo et al., 2013]. Se ha dado la definición inicial de un modelo de ciclo de vida [Arboleya, 2013] y de un modelo de proceso para PIEI [Vanrell et al., 2012]. Se han propuesto cinco procesos de explotación de Información [Britos y García-Martínez, 2009; García-Martínez et al., 2013]. Se han presentado propuestas de modelos de viabilidad y estimación para PIEI [Rodríguez et al., 2010; Pytel et al., 2012, 2013a; 2013b; 2013c]. Se han propuesto formalismos de conceptualización del dominios de negocio basado en Técnicas de Ingeniería del Conocimiento [Vegega et al., 2012] y procedimientos de derivación de procesos de explotación de información a partir de este tipo de conceptualizaciones [Martins et al., 2014; Martins, 2014]. Se ha introducido un proceso de transformación de datos para PIEI [Baldizzoni, 2013]; y se ha comenzado la descripción y prueba de un conjunto de métricas para PIEI [Basso et al., 2013]. Se ha estudiado el comportamiento de la integración de diversos algoritmos de agrupamiento e inducción en el marco del proceso de descubrimiento de reglas de pertenencia a grupos [Rancan et al., 2010; García-Martínez y López Nocera, 2012; Lopez Nocera y Garcia-Martinez, 2012, López Nocera et al., 2012]. Se han explorado, en colaboración con docentes investigadores de la Universidad Nacional

de Misiones, procesos de explotación de información aplicados a la detección de datos anómalos [Kuna et al., 2010a; 2012], y la identificación de causales de abandono de estudios universitarios [Kuna et al., 2009; 2010b]. También se a utilizado explotación de información a la identificación de fallas de apropiación de conceptos en cursos iniciales de carreras de Sistemas [Saavedra-Martínez et al., 2012].

Formación de Recursos Humanos

El grupo de trabajo se encuentra integrado por: dos investigadores formados, dos investigadores en formación. Colabora como investigadora formada externa la Dra. Paola Britos del Laboratorio de Informática Aplicada de la Universidad Nacional de Río Negro. A la fecha de esta comunicación en el LIDIEI tienen radicadas las investigaciones: tres candidatos a doctor, dos candidatos a magister y dos alumnos avanzados que están desarrollando su trabajo final de Licenciatura en Sistemas. Han concluido su desarrollo: tres tesis de maestría y tres trabajos finales de licenciatura.

Referencias

- Arboleya, H.. 2013. *Propuesta de Ciclo de Vida y Mapa de Actividades para Proyectos de Explotación de Información*. Revista Latinoamericana de Ingeniería de Software, 1(3): 107-124, ISSN 2314-2642
- Arboleya, H.. 2013. *Propuesta de Ciclo de Vida y Mapa de Actividades para Proyectos de Explotación de Información*. Revista Latinoamericana de Ingeniería de Software, 1(3): 107-124, ISSN 2314-2642
- Baldizzoni, E. 2013. *Un Proceso de Transformación de Datos Para Proyectos de Explotación de Información*. Revista Latinoamericana de Ingeniería de Software, 1(2): 57-70, ISSN 2314-2642.
- Baldizzoni, E. 2013. *Un Proceso de Transformación de Datos Para Proyectos de Explotación de Información*. Revista Latinoamericana de Ingeniería de Software, 1(2): 57-70, ISSN 2314-2642.
- 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. Pag. 983-992. ISBN 978-987-23963-1-2.
- 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.
- Britos, P., García-Martínez, R. (2009). *Propuesta de Procesos de Explotación de Información*. Proceedings XV Congreso Argentino de Ciencias de la Computación. Workshop de Base de Datos y Minería de Datos. Págs. 1041-1050. ISBN 978-897-24068-4-1.
- 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.
- García-Martínez, R., Lopez Nocera, M. 2012. *Behavioral Variability of Datamining Algorithms Based on Domain Complexity*. Proceedings of VI Meeting on Dynamics of Social and Economic Systems. Pág. 37-38. Ushuaia, Octubre 1-4, Argentina.
- Kuna, H., Caballero, S., Rambo, A., Meinl, E., Steinhilber, A., Pautch, G., Rodríguez, D., García-Martínez, R., Villatoro, F. 2010a. *Identification of Noisy Data in Databases by Means of a Clustering Process*. En Ingeniería de Software e Ingeniería del Conocimiento: Tendencias de Investigación e Innovación Tecnológica en Iberoamérica (Editores: R. Aguilar, J. Diaz, G. Gómez, E- León). Pág. 264-273. Alfaomega Grupo Editor. ISBN 978-607-707-096-2.
- Kuna, H., García Martínez, R. Villatoro, F. 2010b. *Pattern Discovery in University Students Desertion Based on Data Mining*. Advances and Applications in Statistical Sciences Journal, 2(2): 275-286. ISSN 0974-6811.
- Kuna, H., García Martínez, R., Villatoro, F. 2009. *Identificación de Causales de Abandono de Estudios Universitarios. Uso de Procesos de Explotación de Información*. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología 5: 39-44.
- Kuna, H., García-Martínez, R., Villatoro, F. 2012. *Automatic Outliers Fields Detection in Databases*. Journal of Modelling and Simulation of Systems, 3(1): 14-20. ISSN 1737-9377.
- 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.
- 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.
- Mansilla, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2012. *Modelo de Proceso para Elicitación 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.
- 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. 2014. *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.
- Pollo-Cattaneo, M., Mansilla, D., Vegega, C., Pytel, P., Pesado, P., García-Martínez, R., Britos, P. 2013. *Propuesta Integral de Manejo de Requerimientos en Proyectos de Explotación de Información*. En "Reflexiones sobre Ingeniería de Requisitos y Pruebas de Software" (Ed. Jaime Echeverri). Pág. 26-44. Editorial de la Corporación Universitaria Remington y Organización LACREST. ISBN 978-958-58070-3-7.
- 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., Britos, P., García-Martínez, R. 2012. *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.
- 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.
- 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.
- Saavedra-Martínez, P., Pollo-Cattaneo, F., Britos, P., Rodríguez, D., García-Martínez, R. 2012. *Explotación de Información Aplicada a Identificación de Fallas de Apropiación de Conceptos*. Proceedings IX Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento. Pág. 103-110. Sello Editorial de la Pontificia Universidad Católica del Perú. ISBN 978-612-4057-85-4.
- 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., Amatriain, H., Pytel, P., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2012. *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 Pontificia Universidad Católica del Perú. ISBN 978-612-4057-85-4.