

Ingeniería de Explotación de Información: Una Experiencia de Migración de Resultados de Investigación a Contenidos de Asignatura Electiva

Ramón García-Martínez, Darío Rodríguez

Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de Información
Grupo de Investigación en Sistemas de Información
Universidad Nacional de Lanús
{rgarcia,darodriguez}@unla.edu.ar

Resumen

El proceso de investigación busca generar conocimiento que atienda la resolución de problemas abiertos de una disciplina. Desde hace más de una década, los autores son parte de un grupo de I&D que, inspirados en la Ingeniería de Software, busca desarrollar una Ingeniería de Explotación de Información que contextualice la minería de datos. En este proceso, se han generado resultados conceptualmente estables que pueden ser transmitidos en el marco de una asignatura electiva del ciclo superior de Carreras de Sistemas de Información. En esta comunicación se resume la experiencia de migración de los resultados de la investigación a contenidos de una asignatura electiva.

Palabras claves: migración de resultados de investigación, contenidos de asignatura, ingeniería de explotación de información.

1. Introducción

En [García-Martínez et al., 2011] se señala que la Inteligencia de Negocio propone un abordaje interdisciplinario en el que confluyen, entre otras, la Informática, la Matemática y la Economía, su finalidad se centra en generar un conocimiento que contribuya con la toma de decisiones de gestión y generación de planes estratégicos en las organizaciones [Thomsen, 2003].

La Explotación de Información, por su parte, constituye la subdisciplina de la Informática que aporta a la Inteligencia de Negocio [Negash & Gray, 2008] las herramientas de análisis y síntesis necesarias para extraer

conocimiento no trivial, embebido implícitamente en los datos disponibles en diferentes fuentes de información [Schiefer et al., 2004].

Un Proceso de Explotación de Información se define como un grupo de tareas relacionadas lógicamente [Britos, 2008] que, a partir de una masa de información con un grado de valor para la organización, se ejecutan para lograr piezas de conocimiento sobre el funcionamiento de algún aspecto de esta, con un grado de valor mayor que la información original [Ferreira et al., 2005; Kanungo, 2005]. Identificado el problema de inteligencia de negocio, se debe decidir la secuencia de procesos de explotación que deben ser ejecutados para obtener piezas de conocimiento (patrones) de utilidad para el negocio.

Una de las tareas definidas en estos procesos es la de decidir que técnica de minería de datos debe ser utilizada, y de que manera, para obtener determinados patrones. Las técnicas de minería de datos provienen, en su mayoría, del campo del Aprendizaje Automático [García-Martínez et al., 2003].

El grupo comenzó la línea de investigación en Explotación de Información en el año 2000 [Perichinsky y García-Martínez, 2000] siendo responsables de formación de recursos humanos en el área disciplinar desde entonces en grado [Servente, 2002; Grosser, 2004; Felgaer, 2005; Perversi, I. 2007; Kogan, 2007; Gutierrez-Rüegg, 2008; Reparaz, 2008; Calot, 2009; Baldizzoni, 2012]; y en posgrado [Fernández, 2006; Ferrero, 2006; Merlino, 2006; Ochoa, 2006; Britos, 2008; Flores 2009; Vanrell, 2012; López Nocera, 2012].

En paralelo, se desarrollaron trabajos de campo donde se validaron parcialmente los resultados obtenidos [Perichinsky et al., 2003;

Groser et al., 2005; Cogliati et al., 2006; Felgaer et al., 2006; Ferrero et al., 2006; Pollo Cattaneo et al., 2009; Kuna et al., 2010; Rodríguez et al., 2010].

En el año 2008 uno de los autores dicta por primera vez un curso de posgrado sobre el tema [UNLP, 2008] que se repite en el exterior [UdM, 2011]; y a partir del año 2010 los contenidos de ese curso son adaptados como materias electivas del último año en Carreras de Sistemas [UTN, 2010; UNLa, 2012].

En este contexto, se presenta la motivación del trabajo (sección 2); la propuesta del curso con detalle del proceso de selección de contenidos, propuesta de unidades del programa del curso, metodología de enseñanza elegida y criterios de evaluación (sección 3); se presentan las lecciones aprendidas en un lustro de dictado del curso (sección 4); y se formulan algunas conclusiones preliminares (sección 5).

2. Motivación del Trabajo

La tarea de formación de recursos humanos de grado y posgrado en el área; y la invitación a dictar cursos en el tema, han generado la motivación de los autores a sistematizar el proceso de enseñanza en una propuesta de curso en Ingeniería de Explotación de Información que es la que se presenta en esta comunicación.

3. Propuesta del Curso

En esta sección se presenta criterios para la selección de contenidos (sección 3.1), se formula una propuesta de unidades del programa del curso (sección 3.2), se delimita la metodología de enseñanza (sección 3.3), y se señala los métodos de evaluación utilizados (sección 3.4).

3.1. Selección de Contenidos

La primera cuestión a considerar es la de definir cuales de los conocimientos producidos en las investigaciones, cuentan con la

suficiente estabilidad para ser parte de un curso electivo sobre este dominio.

En el campo de Fundamentos se ha argumentado la necesidad de desarrollar una Ingeniería de Proyectos de Explotación de Información para PYMES [García-Martínez et al., 2011]. En el área de Procesos para Proyectos de Explotación de Información (EI), se ha formulado una propuesta de procesos de EI y se han validado sobre un amplio rango de dominios [Britos, 2008]. En el campo de viabilidad se han generado modelos que permiten cuantificarla para proyectos de EI [Pytel et al., 2012a]. En cuanto a la estimación de proyectos de EI se han identificado las tareas críticas en una metodología de desarrollo de proyectos de EI [Pytel et al., 2011], se ha estudiado empíricamente la carga de trabajo en proyectos de este tipo [Rodríguez et al., 2010] y se ha propuesto un modelo analítico de estimación para proyectos desarrollados en PYMES [Pytel et al., 2012b]. En el área de procesos de gestión y desarrollo ha propuesto y validado un modelo de proceso para proyectos de EI [Vanrell et al., 2010; Vanrell, 2012].

Se ha desarrollado un repositorio propio de bases de datos utilizadas para validar resultados parciales de las investigaciones en los dominios de: problemas tuberías de gas [D'Atri et al., 2009], prevención de daños y averías en la industria automotriz [Flores et al., 2009], caracterización de problemas de aprendizaje [Britos et al., 2008], identificación de errores de apropiación de conceptos en el dominio de análisis de sistemas de información [Saavedra-Martínez et al., 2012], prevención de estrés de suelos [Sanson et al., 2009], identificación de causales de abandono de estudios universitarios [Kuna et al., 2010], estimación en gestión de proyectos de software [Bogado, et al., 2011], detección de cambios de consumo de usuarios [Grosser et al., 2005], localización de patrones en eventos meteorológicos [Cogliati et al., 2006], predicción de la salud de una comunidad [Felgaer et al., 2006], detección de daños al corazón [Ferrero et al., 2006], patrones de uso de sitios web [Britos et al., 2007], selección de

protocolos pedagógicos [Britos et al., 2008], y detección de patrones criminales [Valenga et al., 2008].

3.2. Propuesta de Unidades del Programa del Curso

Al ser un curso Ingeniería se centran los contenidos en los aspectos de control y gestión de Proyectos de Explotación de Información (PEI). El curso está estructurado en las siguientes unidades:

- [I] *Procesos de Explotación de Información*: Descubrimiento de reglas de comportamiento. Descubrimiento de grupos. Ponderación de interdependencia de atributos. Descubrimiento de reglas de pertenencia a grupos. Ponderación de reglas de comportamiento o de la pertenencia a grupos.
- [II] *Tecnologías Inteligentes Aplicadas a Procesos de Explotación de Información*: Procesos Unitarios: Algoritmos TDIDT aplicados al descubrimiento de reglas de comportamiento. Mapas auto organizados (SOM) aplicados al descubrimiento de grupos. Redes bayesianas aplicadas a la ponderación de interdependencia entre atributos. Procesos combinados: SOM y TDIDT aplicados al descubrimiento de reglas de pertenencia a grupos. Redes bayesianas aplicadas a la ponderación de reglas de comportamiento o a la pertenencia a grupos.
- [III] *Viabilidad de Proyectos de Explotación de Información*: Evaluación de la viabilidad: plausibilidad, adecuación, éxito. Determinación de las características del proyecto. Conversión de los valores de las características del proyecto a valores en intervalos difusos. Cálculo de la valoración de cada dimensión. Cálculo de la valoración global de la viabilidad del proyecto, interpretación de resultados obtenidos.
- [IV] *Estimación de Proyectos de Explotación de Información*: Factores de costo

relacionados al proyecto: tipo de objetivo de explotación de información, grado de apoyo de los miembros de la organización. Factores de costo relacionados a los datos disponibles: cantidad y tipo de los repositorios de datos disponibles, cantidad de tuplas disponibles en la tabla principal, cantidad de tuplas disponibles en tablas auxiliares, nivel de conocimiento sobre los datos. Factores de costo relacionados a los recursos disponibles: nivel de conocimiento y experiencia del equipo de trabajo, funcionalidad de las herramientas disponibles.

- [V] *Modelo de Proceso de Proyectos de Explotación de Información*: Proceso de administración de proyectos: subproceso de planificación / entendimiento del negocio, subproceso de realización, subproceso de evaluación y control, subproceso de cierre / entrega. Proceso de Desarrollo del Proyecto: subproceso de entendimiento del negocio, subproceso de entendimiento de los datos, subproceso de preparación de los datos, subproceso de modelado, subproceso de evaluación, subproceso de entrega.

3.3. Metodología de Enseñanza

Se propone centrar el proceso de enseñanza en la utilización de guías de estudio. El uso de este método se fundamenta en la concepción que el alumno se apropia de conceptos con alto grado de abstracción a través de un proceso con tres instancias diferenciadas:

- [a] Lectura dirigida del material provisto por la cátedra mediante la utilización de guías de estudio. En esta instancia el alumno tiene un primer acercamiento a los conceptos y da comienzo la apropiación de los mismos. En la figura 1 se ilustra una guía de estudio.
- [b] Discusión con pares. En esta instancia, el trabajo de resolución de las guías de estudio en el ámbito de grupos de trabajo permite que el alumno discuta con pares la

validez de sus apropiaciones conceptuales, y genere las primeras ratificaciones o rectificaciones de estas apropiaciones.



UNIVERSIDAD NACIONAL DE LANUS
LICENCIATURA EN SISTEMAS
Tecnologías de Explotación de Información
Prof. Tit.: Dr. Ramón García-Martínez
JTP: Lic. Darío Rodríguez

PROPUESTA DE PROCESOS DE EXPLOTACIÓN DE INFORMACIÓN

Material: 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.

GUIA DE PREGUNTAS

1. Defina los objetivos de la inteligencia de negocio y el aporte de la explotación de información.
2. Defina Explotación de Información (EI) e EI basada en Sistemas Inteligentes (EIBSI).
3. De una lista de Procesos de Explotación de Información.
4. Defina cuando aplica el proceso de descubrimiento de reglas de comportamiento. Ejemplifique.
5. Enuncie la técnica utilizada para el descubrimiento de reglas de comportamiento, describa el proceso y de un esquema.
6. Defina cuando aplica el proceso de descubrimiento de grupos. Ejemplifique.
7. Enuncie la técnica utilizada para el descubrimiento de grupos, describa el proceso y de un esquema.
8. Defina cuando aplica el proceso de ponderación de interdependencia de atributos. Ejemplifique.
9. Enuncie la técnica utilizada para ponderar la interdependencia de atributos, describa el proceso y de un esquema.
10. Defina cuando aplica el proceso de descubrimiento de reglas de pertenencia a grupos. Ejemplifique.
11. Enuncie las técnicas utilizadas para el descubrimiento de reglas de pertenencia a grupos, describa el proceso y de un esquema.
12. Defina cuando aplica el proceso de ponderación de reglas de comportamiento o de pertenencia a grupos. Ejemplifique.
13. Enuncie las técnicas utilizadas para la ponderación de reglas de comportamiento o de pertenencia a grupos, describa el proceso y de un esquema.

Figura 1. Ejemplo de Guía de Estudios

Este proceso de ajuste (ratificación / rectificación) permite al alumno identificar conceptos cuyo intento de apropiación basada en los saberes logrados hasta ese momento, esta fuera de su alcance. Esta situación motiva el proceso de consulta al docente.

[c] Discusión de la resolución de guía de estudio. En esta instancia, el docente revisa con el alumno los aprendizajes realizados y mediante discusión mayéutica induce los ajustes en las apropiaciones del alumno.

Las guías de estudio se correlacionan con guías de trabajos prácticos de resolución escrita (figura 2) y de resolución en herramientas software de explotación de información (figura 3).



UNIVERSIDAD NACIONAL DE LANUS
LICENCIATURA EN SISTEMAS
Tecnologías de Explotación de Información
Prof. Tit.: Dr. Ramón García-Martínez
JTP: Lic. Darío Rodríguez

PROCESOS DE EXPLOTACIÓN DE INFORMACIÓN

GUIA DE TRABAJOS PRACTICOS

Ejercicio 1: Genere con su grupo de trabajo un problema de inteligencia de negocio que describa los aspectos de negocio que se estimen pertinentes para la mejor comprensión de dicho problema y que requieran la utilización del proceso de descubrimiento de reglas de comportamiento.

La resolución deberá contener:

- [a] Descripción del negocio y de los aspectos del negocio pertinentes al problema de inteligencia de negocio.
- [b] Descripción del problema de inteligencia de negocio.
- [c] Descripción de los datos del negocio que se estiman de utilidad para aplicar el proceso de explotación de información de referencia.

Ejercicio 2: Ídem Ejercicio 1 para el proceso de descubrimiento de grupos.

Ejercicio 3: Ídem Ejercicio 1 para el proceso de ponderación de interdependencia de atributos.

Ejercicio 4: Ídem Ejercicio 1 para el proceso de descubrimiento de reglas de pertenencia a grupos.

Ejercicio 5: Ídem Ejercicio 1 para el proceso de ponderación de reglas de comportamiento o de pertenencia a grupos.

Figura 2. Ejemplo de Guía de Trabajos Practicos de resolucion escrita



UNIVERSIDAD NACIONAL DE LANUS
LICENCIATURA EN SISTEMAS
Tecnologías de Explotación de Información
Prof. Tit.: Dr. Ramón García-Martínez
JTP: Lic. Darío Rodríguez

HERRAMIENTA DE EXPLOTACIÓN DE INFORMACIÓN

Material: Herramienta TANAGRA

GUIA DE EJERCICIOS

Ejercicio 1: Una bodega ha recolectado la información relativa a la composición química de sus vinos producidos durante los últimos años (archivo: "TP1-Calidad-Vinos.xls"). Con fines de confeccionar futuras líneas de vinos que tengan diferentes calidades, se desea realizar un análisis de los vinos producidos para conocer los tipos de calidad disponibles con sus características más relevantes.

- [a] Determine el o los procesos de explotación de información a utilizar.
- [b] Utilice la herramienta Tanagra para desarrollar los procesos de explotación de información identificados.
- [c] Discuta los resultados y de una interpretación.

Ejercicio 2: Una compañía de telefonía celular ha formulado una encuesta a un grupo de usuarios para obtener información (archivo: "IP1-fidelidadtelefonía.xls") sobre la visión de dicho grupo sobre la calidad del servicio prestado. Con carácter posterior a dicho trabajo de investigación de mercado surge el interés de caracterizar el comportamiento de los que continúan comprando los servicios y los que no.

- [a] Determine el o los procesos de explotación de información a utilizar.
- [b] Utilice la herramienta Tanagra para desarrollar los procesos de explotación de información identificados.
- [c] Discuta los resultados y de una interpretación.

Ejercicio 3: Una compañía importadora de autos de alta gama, ha decidido revisar su política de venta dirigida a las concesionarias minoristas, en la que las mismas pueden ofrecer la promoción de reparaciones sin cargo para el cliente (pero con cargo a la compañía importadora) durante el primer año, a partir de la información que tiene de la red de talleres que le prestan dicho servicio (archivo: "TP1-Reparacion-autos.xls"). Esta revisión parte de la sospecha de maniobras dolosas surgidas de la posible connivencia entre algunas concesionarias y algunos talleres respecto de algún tipo de reparación no verificable.

- [a] Determine el o los procesos de explotación de información a utilizar.
- [b] Utilice la herramienta Tanagra para desarrollar los procesos de explotación de información identificados.
- [c] Discuta los resultados y de una interpretación.

Ejercicio 4: Con base en el análisis de su cartera histórica de clientes hipotecarios (archivo: "TP1-Otorgamiento-Creditos.xls") un banco desea relanzar la línea de créditos en el área para lo cual ha establecido las siguientes necesidades: [I] sistematizar los criterios de otorgamiento de créditos, [II] identificar y caracterizar grupos de clientes en orden a estudiar líneas de crédito diferenciales por grupo.

- [a] Determine el o los procesos de explotación de información a utilizar.
- [b] Utilice la herramienta Tanagra para desarrollar los procesos de explotación de información identificados.
- [c] Discuta los resultados y de una interpretación.

Figura 3. Ejemplo de Guía de Trabajos Prácticos de resolución en máquina

3.4. Métodos de Evaluación

Durante el desarrollo del curso, la valoración de manejo de conceptos, aplicación de conocimientos y dominio de técnicas se realiza mediante evaluaciones parciales.

En el diseño de las evaluaciones parciales se utilizan las siguientes técnicas: preguntas multiple-choice, sentencias con determinación verdadero/falso y ejercicios de asociación de conceptos.

El alumno debe desarrollar un proyecto de explotación de información como actividad integradora de los saberes apropiados de cada una de las unidades del curso. A este efecto se dispone de la documentación descriptiva correspondiente a varios casos, donde se describe un negocio, se plantea el problema de inteligencia de negocio y se pone a disposición las bases de datos asociadas al negocio. Para desarrollar el proyecto de explotación de información el alumno debe:

- [i] Determinar la viabilidad que el problema de inteligencia de negocio sea resuelto con un Proyecto de Explotación de Información (Unidad III).
- [ii] En la situación que esto sea posible, formular la estimación del proyecto (Unidad IV).
- [iii] Desarrollar y documentar el proyecto conforme al modelo de proceso propuesto (Unidad V).
- [iv] Utilizar en el subproceso de desarrollo alguno de los procesos de explotación de información presentados en el curso (Unidad I).
- [v] Utilizar en la instrumentación del procesos de explotación de información seleccionado alguna de las tecnologías inteligentes presentadas en el curso (Unidad II).

Para la evaluación final el alumno debe realizar y defender un trabajo monográfico sobre algún tema de la asignatura siguiendo el siguiente procedimiento:

- [i] Elegir uno de los artículos propuestos en la página de trabajos finales e informarlo

al profesor por e-mail quien confirmará la elección o podrá indicar otro artículo.

- [ii] Realizar una investigación documental (búsqueda de otros artículos) con relación al artículo elegido o propuesto.
- [iii] Redacción de una monografía que sumalice la investigación documental realizada.
- [iv] Enviar el trabajo monográfico por e-mail al profesor con una anticipación no menor a quince días de la fecha prevista para la evaluación final.
- [v] Presentarse a la evaluación final con una copia impresa de la monografía.
- [vi] Defender en un coloquio con el docente el trabajo realizado y su relación con las distintas unidades del programa del curso.

4. Lecciones Aprendidas en el Proceso de Migración de Resultados a Contenidos

En el proceso de tratar de migrar de resultados de las investigación a contenidos de asignatura los autores han aprendido algunas lecciones:

Lección 1: *No Todo Resultado de Investigación Se Puede Enseñar.*

Durante el proceso de defensa de una tesis en el área de procesos de explotación de información, el tesista señaló a los Mapas Auto Organizados (SOM por Self Organized Maps) como algoritmo posible para el proceso de descubrimiento de grupos. En el entendimiento que la tesis que se defendía era sobre ingeniería de procesos software (y no sobre algoritmia), para sorpresa del tesista (y del director de tesis), uno de los jurados cuestionó al algoritmo puesto como ejemplo, indicando que había otro del que se contaba experimentos que acreditaban mejores resultados que SOM para la tarea.

Con independencia del uso del algoritmo A o el algoritmo B, no se ponía en duda la necesidad de un proceso de explotación de información orientado al descubrimiento de grupos. La situación generó la inquietud del

director de tesis quien motivó a un tesista de maestría para que investigara la cuestión. Esa investigación generó como resultado la conclusión, empíricamente validada, que los resultados que proporcionan los algoritmos de inducción y los de agrupamiento, son fuertemente dependientes de la complejidad del dominio de conocimiento al cual son aplicados [Lopez-Nocera, 2012].

Si bien el resultado sobre la relación entre algoritmos de inducción/agrupamiento y complejidad de dominio resultó de interés para la explotación de información (y 5 años más tarde salvo el honor del tesista y del director); es un resultado solo de utilidad teórica, ya que la complejidad de un dominio no se puede determinar sin conocerlo en su totalidad y en consecuencia, no se puede decidir que algoritmo tendrá el mejor comportamiento para un determinado proceso de explotación de información.

Lección 2: El Resultado Obtenido Debe Ser Aplicable para Ser Enseñable

Si bien esta lección parecería desprenderse de la Lección 1, surgió en un contexto diferente. Durante los primeros seis años de la investigación (2000-2006) el interés del grupo estuvo orientado a identificar aplicaciones de sistemas inteligentes en minería de datos. Estos resultados se resumen en [Britos et al., 2005]. Cuando surgió la necesidad de transmitir a nuevos tesisistas lo investigado con el propósito de explorar otras aplicaciones, se observó que distintos algoritmos de minería de datos podían ser utilizados para realizar el mismo proceso de descubrimiento de patrones (aunque con resultados diferentes). Es entonces donde emerge el concepto “proceso de explotación de información” y se concluye que es un concepto más general que el de “algoritmo de minería de datos”. A partir de esta conclusión, solo medio un paso a la formulación de la pregunta de investigación: “¿Cuántos y cuales son los procesos canónicos de explotación de información?”. Este resultado se presenta en la tesis de Britos [2008] y es el primer resultado de carácter

general que es transmisible y en el que se puede entrenar a alguien para ser aplicado.

Lección 3: Utilizar La Visión Ordenadora de la Ingeniería de Procesos.

Los resultados del 2008 y la posibilidad concreta de utilizarlos en nuevos proyectos de explotación de información, permitieron tomar conciencia sobre la utilidad de pensar en definir los aspectos ingenieriles de la explotación de información. Pensar la línea de investigación desde la ingeniería de procesos, asimiló la Ingeniería de Explotación de Información a la Ingeniería de Software [Garcia-Martinez et al., 2011].

Es en este punto, donde el grupo reorienta la línea de investigación ya que los proyectos de explotación de información requieren:

- [a] Una ingeniería de requisitos específica [Pollo-Cattaneo et al., 2009; Vegega et al., 2012].
- [b] Modelos de proceso para este tipo de proyectos [Vanrell et al., 2010, 2012; Vanrell, 2012].
- [c] Técnicas de determinación de viabilidad [Pytel et al., 2012a; 2012b] y de estimación [Rodríguez et al., 2010; Pytel, 2011; Pytel et al., 2011; 2012c] para este tipo de proyectos.

Lección 4: Involucrar a los Alumnos en el Proceso de Investigación de Campo.

Los procesos, modelos y técnicas desarrolladas, requieren su testeo en pruebas de campo para que ganen en verosimilitud y validez. Desde el año 2010, con respuesta favorable, se ha invitado a los alumnos de los cursos de Ingeniería de Explotación de Información a cargo de docentes-investigadores del grupo a participar en distintos experimentos de validación, contrastación de resultados o construcción de muestras. El resultado de esta participación es reconocido y agradecido en las correspondientes publicaciones [Rodríguez et al., 2010; Pytel et al., 2012c].

5. Conclusiones

Se ha reseñado en esta comunicación el proceso de migración de resultados de investigación a contenidos de asignatura electiva en el área de ingeniería de explotación de información, desarrollado por los docentes-investigadores del grupo de trabajo.

Se han propuesto los contenidos del curso, se ha fundamentado la derivación de los mismos desde resultados de las investigaciones en curso y se ha señalado como la necesidad de transmitir a los nuevos tesisistas los resultados de la investigación actuó como disparador estructurante de la organización del curso.

Se enuncian las lecciones aprendidas, que sin pretensión de generalidad, formulan algunos principios subyacentes al proceso de migración de resultados de investigación obtenidos a contenidos enseñables.

Como futura línea de trabajo se está considerando migrar a contenidos del curso, resultados recientes vinculados a limpieza de datos, ciclo de vida y mapa de actividades para proyectos de explotación de información.

6. Financiamiento

Las investigaciones que se reportan en este artículo han sido financiadas parcialmente por el Proyecto de Investigación 33A167 de la Universidad Nacional de Lanús; y por la Gerencia de Investigación y Desarrollo de Staffing IT Software & Services.

7. Referencias

Baldizzoni, E. 2012. *Un Proceso de Transformación de Datos para Proyectos de Explotación de Información*. Trabajo Final de Licenciatura en Sistemas. Universidad Nacional de Lanús.

Bogado, V., Dapozo, G., García Martínez, R. 2011. *Estimación en Gestión de Proyectos de Software Basada en Explotación de Información*. Proceedings XIII Workshop de Investigadores en Ciencias de la Computación. Artículo 3894.

Britos, P. 2008. *Procesos de Explotación de Información Basados en Sistemas Inteligentes*. Tesis Doctoral. Facultad de Informática. Universidad Nacional de La Plata.

Britos, P., Cataldi, Z., Sierra, E., García-Martínez, R. 2008. *Pedagogical Protocols Selection Automatic Assistance*. Notes in Artificial Intelligence 5027: 331-336.

Britos, P., Hossian, A., García-Martínez, R., Sierra, E. 2005. *Minería de Datos Basada en Sistemas Inteligentes*. Editorial Nueva Librería. ISBN 987-1104-30-8.

Britos, P., Martinelli, D., Merlino, H., García-Martínez, R. 2007. *Web Usage Mining Using Self Organized Maps*. International Journal of Computer Science and Network Security, 76:45-50. ISSN : 1738-7906.

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.

Calot, E. 2009. *Reconocimiento de Patrones en Imágenes Médicas Basado en Sistemas Inteligentes*. Tesis de Grado en Ingeniería Informática. Facultad de Ingeniería. Universidad de Buenos Aires..

Cogliati, M., Britos, P., García-Martínez, R. 2006. *Patterns in Temporal Series of Meteorological Variables Using SOM & TDIDT*. Volume 217, En Artificial Intelligence in Theory and Practice, ed. M. Bramer, IFIP Boston: Springer., 217: 305, 314.

D'Atri, M., Rodríguez, D., García-Martínez, R. 2009. *Improving Pipeline Risk Models by Using Data Mining Techniques*. 24th World Gas Conference Proceedings CD. Paper 663.

Felgaer, P. 2005. *Optimización de Redes Bayesianas Basada en Técnicas de Aprendizaje por Inducción*. Tesis de Grado en Ingeniería Informática. Facultad de Ingeniería. Universidad de Buenos Aires.

Felgaer, P., Britos, P. and García-Martínez, R. 2006. *Prediction in Health Domain Using Bayesian Network Optimization Based on*

- Induction Learning Techniques*. International Journal of Modern Physics C 173: 447-455.
- Felgaer, P., Britos, P., García-Martínez, R. 2006. *Prediction in Health Domain Using Bayesian Network Optimization Based on Induction Learning Techniques*. International Journal of Modern Physics C 173.: 447, 455.
- Fernández, E. 2006. *Asistente para la gestión de documentos de proyectos de explotación de datos*. Tesis de Magíster en Ingeniería del Software. Universidad Politécnica de Madrid.
- 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. 471, 474.
- Ferrero, G. 2006. *Detección de patrones en imágenes médicas*. Tesis de Magíster en Ingeniería del Software. Universidad Politécnica de Madrid.
- Ferrero, G., Britos, P., García-Martínez, R., 2006. *Detection of Breast Lesions in Medical Digital Imaging Using Neural Networks*, IFIP, Volume 218, Professional Practice in Artificial Intelligence, eds. J. Debenham, Boston: Springer.. 1, 10.
- Ferrero, G., Britos, P., García-Martínez, R., 2006. *Detection of Breast Lesions in Medical Digital Imaging Using Neural Networks*. In IFIP International Federation for Information Processing, Volume 218, Professional Practice in Artificial Intelligence, eds. J. Debenham, Boston: Springer, Pág. 1-10.
- 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 Congreso Argentino de Ciencias de la Computación. Workshop de Base de Datos y Minería de Datos. Págs. 1021-1030. 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., Servente, M., Pasquini, D. 2003. *Sistemas Inteligentes*. Editorial Nueva Librería.
- Grosser, H. 2004. *Detección de Fraude en Telefonía Celular Usando Redes Neuronales*. Tesis de Grado en Ingeniería Informática. Facultad de Ingeniería. Universidad de Buenos Aires.
- Grosser, H., Britos, P. y García-Martínez, R. 2005. *Detecting Fraud in Mobile Telephony Using Neural Networks*. Lecture Notes in Artificial Intelligence 3533: 613-615.
- Grosser, H., Britos, P., García-Martínez, R. 2005. *Detecting Fraud in Mobile Telephony Using Neural Networks*. Lecture Notes in Artificial Intelligence 3533:613, 615
- Gutierrez-Rüegg, P. 2008. *Caracterización de la Población Carcelaria en Argentina Mediante la Aplicación de Minería de Datos para la Prevención de Hechos Delictivos*. Tesis de Grado en Ingeniería Industrial. Instituto Tecnológico de Buenos Aires.
- 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. 231, 240.
- Kogan, A. 2007. *Integración de Algoritmos de Inducción y Agrupamiento. Estudio del Comportamiento*. Tesis de Grado en Ingeniería Informática. Facultad de Ingeniería. Universidad de Buenos Aires.
- 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, 22.: 275, 286.
- Kuna, H., García Martínez, R. Villatoro, F. 2010c. *Pattern Discovery in University Students Desertion Based on Data Mining*. Advances and Applications in Statistical

- Sciences Journal, 22: 275-286. ISSN 0974-6811.
- Lopez Nocera, M. 2012. *Descubrimiento de Conocimiento Mediante la Integración de Algoritmos de Explotación de la Información*. Tesis de Magister en Ingeniería de Sistemas de Información. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional.
- Merlino, H. 2006. *Un ambiente abstracto de minería de datos*. Tesis de Magister en Ingeniería del Software. Universidad Politécnica de Madrid.
- Negash, S., Gray, P. 2008. *Business Intelligence*. En Handbook on Decision Support Systems, 2 ed., Eds. F. Burstein y C. Holsapple (Heidelberg, Springer), 175, 193.
- Ochoa, A. 2006. *Uso de técnicas de educación para la comprensión del negocio*. Tesis de Magister en Ingeniería del Software. Universidad Politécnica de Madrid.
- Perichinsky, G. y Garcia Martinez, R. 2000. *A Data Mining Approach to Computational Taxonomy*. Proceedings del Workshop de Investigadores en Ciencias de la Computación. Páginas 107-110. Editado por Departamento de Publicaciones de la Facultad de Informática. Universidad Nacional de La Plata. Mayo.
- Perichinsky, G., Servetto, A., García-Martínez, R., Orellana, R., Plastino, A. 2003. *Taxomic Evidence Applying Algorithms of Intelligent Data Mining Asteroid Families*. Proceedings International Conference on Computer Science, Software Engineering, Information Technology, e-Business & Applications. Río de Janeiro Brasil. 308, 315.
- Perversi, I. 2007. *Aplicación de Minería de Datos para la Exploración y Detección de Patrones Delictivos en Argentina*. Tesis de Grado en Ingeniería Industrial. Instituto Tecnológico de Buenos Aires.
- 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. 467, 469.
- Pytel, P. 2011. *Método de Estimación de Esfuerzo para Proyectos de Explotación de Información. Herramienta Para Su Validación*. Tesis de Magister en Ingeniería del Software. Convenio Universidad Politécnica de Madrid e Instituto Tecnológico Buenos Aires.
- Pytel, P., Amatriain, H., Britos, P., Garcia-Martinez, 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. *Viabilidad y Estimación de Proyectos de Explotación de Información*. Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación. Pág. 217-221. ISBN 978-950-766-082-5.
- Pytel, P., Britos, P., García-Martínez, R. 2012c. *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., 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.
- Pytel, P., Tomasello, M., Rodríguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2011. *Estudio del Modelo Paramétrico DMCoMo de Estimación de Proyectos de Explotación de Información*. Proceedings XVII Congreso Argentino de Ciencias de la Computación. Pág. 979-988. ISBN 978-950-34-0756-1.
- Reparaz, R. 2008. *Aplicación de Minería de Datos para Determinar la Eficacia de la*

- Braquiterapia en el Tratamiento de Cáncer de Prostata*. Tesis de Grado en Ingeniería Industrial. Instituto Tecnológico de Buenos Aires. 25/06/08.
- 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., Rodríguez, D., Britos, P., García-Martínez, R. 2012. *Proceso de Identificación de Errores de Apropiación de Conceptos Basado en Explotación de Información*. Proceedings VII Congreso de Tecnología en Educación y Educación en Tecnología. ISBN 978-987-28186-0-9. Universidad Nacional del Noroeste de Buenos Aires. Pergamino. Buenos Aires. Argentina.
- Sanson, E., Britos, P., Rodríguez, D., García-Martínez, R. 2009. *Clasificación Automática para la Prevención del Estrés de los Suelos y la Fatiga de Soja en el Noroeste Argentino*. Proceedings XI Workshop de Investigadores en Ciencias de la Computación. Pág. 333-335. ISBN 978-950-605-570-7.
- 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. 162, 169.
- Servente, M. 2002. *Algoritmos TDIDT aplicados a la Minería de Datos Inteligente*. Tesis de Grado en Ingeniería Informática. Facultad de Ingeniería. Universidad de Buenos Aires.
- Thomsen, E. 2003. *BI's Promised Land*. Intelligent Enterprise, 6(4): 21, 25.
- UdM, 2011. *Seminario Ingeniería de Explotación de Información*. Programa de Maestría en Ingeniería del Software de la Universidad de Medellín.
- UNLa, 2012. *Asignatura Tecnologías de Explotación de Información*. Materia Electiva de 5to Año de la Carrera de Licenciatura en Sistemas de la Universidad Nacional de Lanús.
- UNLP, 2008. *Curso de Doctorado Tecnologías Inteligentes de Explotación de Información*. Facultad de Informática de la Universidad Nacional de La Plata.
- UTN, 2010. *Asignatura Tecnologías de Explotación de Información*. Materia Optativa de 5to Año de la Carrera de Ingeniería en Sistemas de Información de la Facultad Regional de Buenos Aires de la Universidad Tecnológica Nacional.
- 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.
- Vanrell, J. 2012. *Un Modelo de Procesos para Proyectos de Explotación de Información*. Tesis de Magister en Ingeniería de Sistemas de Información. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional.
- 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., 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 de la Pontificia Universidad Católica del Perú. ISBN 978-612-4057-85-4.