

TRANSFORMANDO DATOS EN CONOCIMIENTO: NUEVAS PRÁCTICAS

Msc. Raúl Oscar Klenzi

Departamento e Instituto de Informática F.C.E.F. y N. Universidad Nacional de San Juan. Ignacio de la Roza y Meglioli (5400). rauloscarklenzi@gmail.com

Msc. María Alejandra Malberti

Departamento e Instituto de Informática F.C.E.F. y N. Universidad Nacional de San Juan. Ignacio de la Roza y Meglioli (5400). amalberti@gmail.com

Msc. Graciela Elida Beguerí

Departamento e Instituto de Informática F.C.E.F. y N. Universidad Nacional de San Juan. Ignacio de la Roza y Meglioli (5400). grabeda@gmail.com

Lic. Silvina Migani

Departamento e Instituto de Informática F.C.E.F. y N. Universidad Nacional de San Juan. Ignacio de la Roza y Meglioli (5400). silvina.migani@gmail.com

CONTEXTO

En la actualidad, la Facultad de Ciencias Exactas Físicas y Naturales de la Universidad Nacional de San Juan, FCEFNU-SJ, y la Universidad toda, está abocada a la etapa de auto evaluación institucional, destinada a mejorar la calidad de la institución.

Con el propósito de acompañar este proceso, en esta propuesta se realiza la gestión de conocimiento que, entre otros datos, se plasma sobre los pertenecientes a los Sistemas de Biblioteca y de Alumnos de la mencionada facultad. Esto permite, no solo lograr una efectiva transferencia de conocimiento a diferentes cátedras de las carreras Licenciatura en Ciencias de la Computación y Licenciatura en Sistemas de Información pertenecientes al Departamento de Informática, sino también ofrecer a la comunidad distintos cursos de actualización, perfeccionamiento y posgrado.

RESUMEN

La construcción de almacenes de datos – Data Warehouses- DW, la búsqueda de conocimiento inmerso en los datos por medio de OLAP, de estrategias de minería de datos – Data Mining- DM, y su validación mediante herramientas estadísticas, es el área donde el proyecto “Búsqueda estratégica de conocimiento en los datos de Biblioteca y Alumnos de la FCEFNU-SJ”, se desarrolla.

Este proyecto es continuación del denominado "Descubrimiento de conocimiento a través de Data Warehousing y Data Mining, en los datos de la Biblioteca de la FCEFNU-SJ" - 2006/2007- y de anteriores proyectos: "Búsqueda inteligente de Información no relacionada en Grandes Bases de Datos" y "Medición de la biblioteca de la FCEFNU-SJ", los cuales abordaron temáticas relacionadas con la actividad de la biblioteca de la FCEFNU. Actualmente, se avanza sobre los resultados alcanzados, al incorporar los datos correspondientes a los alumnos de la FCEFNU.

Otro tema tratado, es el análisis de encuestas de satisfacción de usuarios de la biblioteca de la FCEFNU, por medio de la aplicación de estrategias de DM como: segmentación, clasificación, y descubrimiento de reglas de asociación, que permiten determinar inductores influyentes en la imagen de los usuarios respecto de su biblioteca. En este proceso, se usan y comparan distintas herramientas de software libre.

Palabras clave: Gestión del Conocimiento - Data Mining - Data Warehouse - Satisfacción de Usuarios.

1. INTRODUCCION

Existen organizaciones que poseen una enorme cantidad de datos generados en el día a día, cuya recolección y organización

insumieron tiempo e importantes sumas de dinero y, sin embargo, están como dormidos en distintos depósitos.

Actualmente la tecnología computacional permite no sólo que se capturen y almacenen esas inmensas cantidades de datos, sino también encontrar patrones, tendencias o anomalías en ellos. Resumirlos a través de modelos cuantitativos simples, no es una tarea trivial, ya que deben convertirse los datos en información, y convertir la información en conocimiento, lo que representa uno de los grandes desafíos de la era de la información (Larose, D. 2006), (Hernández Orallo, J. y otros, 2008).

La construcción de DW, la utilización de herramientas OLAP y de DM se presentan como elementos esenciales y complementarios que hacen posible contar por primera vez, con una memoria corporativa que provee información accesible, consistente, adaptable y segura, además de transformarla en conocimiento. Dicha transformación es el cimiento en la toma de decisiones de toda organización. (Kimball y otros, 2004), (Inmon, 2002). Esto provoca considerables beneficios a las organizaciones que la aplican, pues tienen la oportunidad de descubrir patrones rentables, y tendencias novedosas para su negocio.

A partir de ello, la construcción de DW y DM se presenta como una tecnología emergente con varias ventajas:

- Ahorra grandes cantidades de dinero a las empresas y organizaciones en general.
- Abre nuevas oportunidades de negocios.
- Mejora los servicios brindados a los usuarios de cualquier organización ya que, aún en aquellas aplicaciones en las que no se vislumbran importantes beneficios monetarios, existe la posibilidad de incorporar la tecnología de minería de datos.
- Resulta un buen punto de encuentro entre los investigadores y las personas de negocios.

En esta nueva área, relacionada a los Sistemas de Soporte de Decisión, confluyen distintas áreas de conocimiento, tales como: Bases de Datos, Estadística, Inteligencia Artificial, Aprendizaje Automático, Técnicas de Visualización, etc.

Específicamente hablando de DM, el propósito es encontrar modelos matemáticos obtenidos mediante técnicas de aprendizaje, supervisadas o no, que desde los datos rescaten patrones, relaciones, y conocimiento anteriormente no percibido por otras técnicas, y desde allí estimar predicciones sobre las variables analizadas.

DM aplicado a datos inherentes a una biblioteca recibe el nombre Bibliomining (Nicholson, S. y otro, 2003). En el marco de este proyecto el caso de estudio central ha sido la biblioteca de la FCEF-UNSA, desde los datos del movimiento bibliográfico y estructura que la caracterizan, como así también desde la perspectiva que sus usuarios tienen de aquella.

La búsqueda de la satisfacción de usuarios de bibliotecas y la aplicación de técnicas de minería de datos, abarca una primer etapa exploratoria y descriptiva realizada sobre las respuestas obtenidas de una encuesta practicada a los alumnos y usuarios de la biblioteca de la FCEF-UNSA, validada oportunamente en (Beguerí, G., 2006) (Beguerí, G. 2007) y analizadas mediante técnicas de Segmentación, Clasificación, y Reglas de Asociación en (Klenzi, R., 2008) (Malberti, M., 2008). En este proyecto esencialmente se han realizado tareas de descripción, encontrando atributos inductores entre los que conforman la encuesta (Klenzi, R. y otros, 2008) que, conjuntamente con el procesamiento de los datos de circulación de material bibliográfico de diferentes años ha permitido proponer políticas a futuro (redistribuciones presupuestarias, y de material en estanterías, refacciones edilicias, etc.). Similar procedimiento, en cuanto a la satisfacción de usuarios, se está aplicando a encuestas realizadas a usuarios de software libre, intentando encontrar los inductores que permitan inferir ventajas y desventajas de tales softwares según la perspectiva de sus usuarios.

2. LINEAS DE INVESTIGACION y DESARROLLO

En el proceso de desarrollo del proyecto, se realizan tareas de DM, Bibliomining, Text Mining y Web Mining:

- En el DW construido a partir de los datos del departamento alumnos de la FCEFN y de los datos de circulación de material bibliográfico (Migani, y otro, 2006) (Migani, y otros, 2007).
- En encuestas, de satisfacción de usuarios de la Biblioteca de la citada facultad y de usuarios de herramientas código libre.
- En los logs de acceso desde Internet a consulta bibliográfica.

Para ello, se hace uso de las siguientes herramientas:

- R, WEKA-Waikato Environment Knowledge Algorithms y RapidMiner, bajo licencias GNU.
- ORACLE Warehouse Builder y ORACLE Data Miner., del administrador de Base de datos ORACLE, bajo licencia OTN y
- Microsoft SQL Server 2005, en particular SQL Server Business Intelligence Development Studio.

Conjunto a la indagación teórica inherente a cada una de las estrategias de DM abordadas, se realizan análisis de los algoritmos que las integran, y se promueve el análisis comparativo entre las distintas herramientas usadas, tanto en lo referente a la capacidad de las mismas en la búsqueda de conocimiento, como de los algoritmos que proveen. En este aspecto se ha realizado un sucinto análisis comparativo entre herramientas de software libre, aplicables en el área de la minería de datos (Gonzalez, y otros, 2008). Las herramientas analizadas han sido **WEKA - Waikato Environment for Knowledge Analysis**- versión 3.5.8, Universidad de Waikato en Nueva Zelanda con licencia GNU General Public License (GPL) (Witten I. H. & Frank E. 2005), **RAPIDMINER** (Versión Original YALE) 4.2.000 con licencia AGPL versión 3. **R Project** que es un lenguaje y entorno de programación para análisis estadístico y gráfico (R Development Core

Team. 2006). R se distribuye bajo la licencia GNU GPL. Las tres herramientas están disponibles para los sistemas Operativos Windows, Macintosh, Unix, y GNU/Linux.

3. RESULTADOS OBTENIDOS/ESPERADOS

Los resultados alcanzados hasta este momento, han sido publicados y divulgados en diferentes eventos, tal es el caso de:

- **Divulgación De Las Actividades Del Proyecto 21/E824.** Beguerí; Malberti; Klenzi,. Seminario de Bibliotecas Universitarias, Los Cambios de Paradigmas en su Gestión. Universidad Nacional de San Juan- Junta de Bibliotecarios. 29 y 30 de Agosto de 2008.
- **Problemas de Asignación de Aulas, Docentes y Materias en Instituciones Educativas.** Aballay; Klenzi. CACIC 2008.
- **Análisis comparativo entre estrategias de clasificación y reglas de asociación, en la determinación de inductores, en encuesta de satisfacción de usuarios de una biblioteca universitaria.** Klenzi, Malberti, Beguerí. Jornadas Chilenas de Computación 2008, XX Encuentro Chileno de Computación 2008.
- **Minería de datos con herramientas de software libre.** González; Aciar; Martínez. 1° Congreso Nacional de Software Libre-San Juan
- **JOOMLA en la implementación de un prototipo de autoevaluación, en una cátedra universitaria.** Echeagaray, Klenzi, Malberti. 5° Jornada de Informática y Educación. U. N. V. M. Argentina.

En particular, en el ámbito de la FCEFN, conjuntamente con otro proyecto de la facultad, se organizó un Workshop de actualización en Telecomunicaciones y Gestión del Conocimiento en Bibliotecas, en el marco del cual se presentaron los siguientes trabajos:

- *Datamarts y Olap en Oracle 11g*. Migani, Lépez, González.
- *Minería de texto. Conceptos y Herramientas*. Gutiérrez, Klenzi
- *Computación Paralela y sus Aplicaciones*. Dra. Piccoli (Invitada)
- *IsisMarc Cliente-Servidor. I+D aplicado a la reingeniería de un producto*. Lic. Marmonti. (Invitado).
- *Catalogo de Bibliotecas en el marco de la web 2.0 y las redes sociales*. Lic. Marmonti.

Actualmente, desde un constructo generado y validado destinado a usuarios de software libre, se realiza el análisis y procesamiento con el propósito de determinar las necesidades implícitas y explícitas de aquellos usuarios.

Asimismo, se espera ampliar el ámbito de aplicación de la tecnología de minería de datos, al análisis de datos sísmicos perteneciente al área de ciencias de la tierra, temática totalmente pertinente a nuestra facultad.

4. FORMACION DE RECURSOS HUMANOS

En lo referente a la formación de recursos humanos, en el marco del proyecto se han desarrollado diferentes tesis de posgrado, de grado y dirección de becarios.

Se realizó la defensa de dos Tesis de Maestría, correspondientes a la Maestría en Informática dependiente de la Universidad Nacional de La Matanza. Ellas son:

- *Aplicación de minería de reglas de asociación en una biblioteca universitaria: una tecnología apropiada*. Lic. A. Malberti
- *Aplicación de minería de datos a la gestión bibliotecaria*. Ing. R. Klenzi

Se dirigió una Tesis de Maestría, correspondiente a la Maestría en Logística dependiente de la Universidad Nacional de Cuyo:

- *Propuesta de Autoevaluación de una Biblioteca Universitaria en un marco de Gestión Logística*. Prof. V. Márquez.

Se dirigieron dos trabajos finales, de la carrera Licenciatura en Ciencias de la Información:

- *Implementación de Autoevaluaciones en entorno WEB*. F. Echegaray.
- *Minería de Datos Aplicada a un DW. Utilización de la Herramienta ORACLE para el desarrollo del DW y la extracción de información oculta*. H. Lépez, A. González.
- *Búsqueda de satisfacción en usuarios de software libre*. Martínez, Aciar (en ejecución)

En la actualidad se dirigen dos Becas de Investigación para alumnos avanzados, de la UNSJ:

- *Proceso de Extracción, Transformación, Carga y Refresco de un DW con datos de provenientes del sistema de alumnos y de biblioteca*. Becario H. Lépez.
- *Aplicación de Minería de Datos a Encuesta de Satisfacción de Usuarios de una Biblioteca Universitaria*. Becario A. González.

En el bienio 2006-2007, se dirigió una beca de innovación tecnológica en las ingenierías, "Digitalización de Archivos históricos" que fue financiada con \$10000 por la fundación Banco San Juan y estuvo a cargo del becario Ing. Cristian de los Ríos.

5. BIBLIOGRAFIA

Beguerí, G., Olguín, L. (2006). *Estudio sobre la Percepción del Usuario en una Biblioteca Universitaria. Normas ISO 11620, IRAM – ISO 11620*. Publicado en: <http://www.uniram.com.ar/jornadas/XXV/TC-14.pdf>

Beguerí, G. (2007) *Logística como garantía de satisfacción del usuario*. Tesis de Maestría-Universidad Nacional de Cuyo.

Gonzalez A., Aciar J., Martinez S. (2008) *Minería de Datos con Herramientas de Software Libre (Análisis Comparativo de*

Herramientas de Software Libre Aplicadas al Área de la Minería De Datos) 1° Congreso Nacional de Software libre. San Juan 2008

Hernández Orallo J., Ramirez Quintana, J, Ferri Ramirez, C. (2008) *Introducción a la Minería de Datos*. Pearson-Prentice Hall.

Inmon (2002). *Building the Data Warehouse Third Edition*. Wiley Computer Publishing

Kimball R., Reeves L., Ross M., Thornthwhite W. *The Data Warehouse Life Cycle Toolkit Expert Methods for Designing, Developing and Deploying Data Warehouses*. John Wiley & Sons.405 pag.

Klenzi, Raúl O. (2008). *Aplicación de minería de datos a la gestión bibliotecaria*. Tesis de Maestría. Maestría en Informática. Universidad Nacional de la Matanza.

Klenzi, R.; Malberti, M. A.; Beguerí, G. (2008) *Análisis comparativo entre estrategias de clasificación y reglas de asociación, en la determinación de inductores, en encuesta de satisfacción de usuarios de una biblioteca universitaria*. Jornadas Chilenas de Computación 2008 (JCC '08) en el marco del XX Encuentro Chileno de Computación 2008 (ECC '08). Punta Arenas- Chile

Larose, D. (2006). *Data Mining Methods and Model*,. Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Malberti, María A. (2008). *Aplicación de Minería de Reglas de Asociación en una Biblioteca Universitaria*. Tesis de Maestría. Maestría en Informática. Universidad Nacional de la Matanza.

Migani, S.; Chavez, S. (2006) *Diseño Conceptual de un Data Warehouse*. Revista Ciencias. N°1 2006 Pag. 83 a 90.

Migani, S.; Chavez, S. Klenzi, R; (2007) *Construcción de un Data Warehouse-Parte I*. Revista Ciencias. N°1 2007 Pag. 45 a 56.

Nicholson, S. and Stanton, J. (2003). *Gaining Strategic Advantage through Bibliomining: Data Mining for Management Decisions in Corporate, Special, Digital, and Traditional Libraries*. Organizational Data Mining: Leveraging Enterprise Data Resources for Optimal Performance, H. Nemati and C. Barko, eds. (Hershey, Penn.: Idea Group Pub.).

R Development Core Team. (2006). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.

Witten I. H. & Frank E. (2005). *Data Mining, practical Machine Learning Tools and techniques*. Second Edition. Morgan Kaufmann Publisher.