

Bibliomining para descubrir reglas de asociación en el Centro de Documentación Regional “Juan Bautista Vázquez”

Valeria Haro, Wilson Pérez, Víctor Saquicela

Departamento de Ciencias de la Computación, Universidad de Cuenca, Av. 12 de Abril, Cuenca, Ecuador, 010150.

Autores para correspondencia: valeria.harov@ucuenca.ec,
{wilson.perez,victor.saquicela}@ucuenca.edu.ec

Fecha de recepción: 19 de junio del 2016 - Fecha de aceptación: 24 de julio del 2016

ABSTRACT

The technological advances in hardware and software in the last years allows storing large volumes of information, which has complicated its management and exploitation. Data mining is a process to discover hidden knowledge in databases and to support decision-making, taking advantage of historic data of a business or institution. This article describes the implementation of data mining techniques in the Regional Documentation Centre “Juan Bautista Vazquez” of the University of Cuenca. The study follows the methodology proposed by Nicholson and uses the algorithms of Weka based on the loan transactions data of Data Warehouse, implemented previously in the documentation center. Finally, the association rules between the bibliographic resources are validated to determine the users’ preference.

Keywords: Data mining, association rules, bibliomining, library.

RESUMEN

Los avances tecnológicos en hardware y software de los últimos años, han permitido almacenar un gran volumen de información, lo cual ha complicado los procesos de gestión y explotación de los mismos. La minería de datos es un proceso para descubrir información oculta en las bases de datos, y permite dar soporte a la toma de decisiones, aprovechando los datos históricos de una empresa o institución. Este artículo describe la implementación de las técnicas de minería de datos en el Centro de Documentación Regional “Juan Bautista Vázquez” de la Universidad de Cuenca. Para ello, este estudio sigue la metodología propuesta por Nicholson y usa los algoritmos de Weka, basándose en los datos de préstamos contenidos en un Data Warehouse implementado previamente en este centro de documentación. Finalmente, se encuentra y valida las reglas de asociación entre los recursos bibliográficos lo cual determina las preferencias de los usuarios.

Palabras clave: Minería de datos, reglas de asociación, bibliomining, biblioteca.

1. INTRODUCCIÓN

A lo largo de los años, las empresas/instituciones acumulan grandes volúmenes de datos generados día a día por los sistemas transaccionales, esto debido al avance tecnológico tanto en hardware como software que se ha desarrollado en los últimos años. Esta gran cantidad de datos ha generado complicaciones en los procesos de gestión y explotación. Si dichos datos, son administrados de forma óptima pueden dar soporte a las decisiones financieras, administrativas y económicas en una empresa/institución, caso contrario únicamente se convierten en “datos basura”.

Dentro de la academia, las bibliotecas han recopilado datos sobre sus colecciones y usuarios por varios años, pero no han sido utilizados de manera eficiente para la toma de decisiones. Una de las

opciones para descubrir conocimiento valioso que se encuentre oculto es el Data mining (Han *et al.*, 2011). La combinación de técnicas de Data mining y la bibliometría¹ es conocida como Bibliomining (Nicholson & Stanton, 2003), la misma que permite a encargados e investigadores de las bibliotecas tener una idea global de los recursos en sus organizaciones.

Tradicionalmente las decisiones que se toman diariamente se basan en la experiencia de las personas que administran las bibliotecas, sin embargo, el análisis de los datos con las técnicas y metodologías actuales pueden ser el apoyo eficaz para los directivos y de esta manera la biblioteca puede ofrecer un mejor servicio a los usuarios finales donde, se reflejen las soluciones a las necesidades reales como la compra de nuevo material bibliográfico, la ubicación del material en estantería, la sugerencia en los préstamos, etc. En este artículo se presenta la implementación de técnicas de Bibliomining, sobre la información que posee el Centro de Documentación Regional “Juan Bautista Vázquez” (CDRJBV) de la Universidad de Cuenca, con la finalidad de descubrir patrones de asociación de los recursos que ofrece. Durante el desarrollo del proyecto, se usa las diferentes fases de descubrimiento de información descrita por Nicholson (2003).

El artículo está organizado de la siguiente manera: en la sección 2 se realiza una revisión bibliográfica de trabajos relacionados, describiendo los conceptos de Data mining, Bibliomining y la metodología aplicada para descubrir conocimiento. Luego, en la sección 3 se implementa dicha metodología; posteriormente, en la sección 4 se puede observar los resultados obtenidos con sus respectivas evaluaciones. Finalmente, en la sección 5 se sintetiza las conclusiones y trabajos futuros.

2. ANTECEDENTES Y TRABAJOS RELACIONADOS

En esta sección, se describen conceptos previos requeridos para entender el alcance de la propuesta. Concretamente se introduce a los conceptos de Data mining, Bibliomining y el proceso para el descubrimiento de conocimiento. También se describe los conceptos de las reglas de asociación y una referencia de los trabajos relacionados.

2.1. Data mining

Data mining (DM) es el proceso de descubrir conocimiento que se encuentra oculto sobre grandes volúmenes de datos (Han *et al.*, 2011), generados por las empresas a lo largo del tiempo para convertir esa información en conocimiento potencialmente útil (Witten *et al.*, 2011). La utilización de técnicas de limpieza e integración de datos, hace posible generar conocimiento aplicando estadística y técnicas de inteligencia artificial (Nicholson & Stanton 2005), para determinar patrones de comportamiento de los procesos empresariales, generando nuevas oportunidades de negocio.

Las técnicas de DM son conjuntos de cálculos y reglas que permite crear un modelo de DM a partir de los datos (Liao *et al.*, 2012). Los algoritmos analizan los datos de entrada, en busca de patrones, encontrando todas las conexiones posibles que pueda haber en toda la información. Las técnicas de DM pueden ser categorizadas en asociación, clasificación, agrupación y regresión (Siguenza-Guzman *et al.*, 2015).

2.2. Bibliomining

Bibliomining es una combinación de la bibliometría y DM (Nicholson & Stanton, 2003 & 2005), que lo definen como la aplicación de técnicas de DM en el ámbito bibliotecario para el análisis de la información generada por las bibliotecas y extraer conocimiento que ayude a tomar decisiones. Además, está relacionado con la biblioteconomía y documentación, debido a que pretende que los resultados obtenidos de la aplicación de las técnicas de DM sean útiles para entender las comunidades de usuarios.

En el trabajo presentado por Nicholson (2003) se define el proceso de Bibliomining, que consta

¹ El estudio de las regularidades en las citas, la autoría y temas utilizando técnicas cuantitativas y de visualización (Nicholson & Stanton, 2005).

de seis fases descritas a continuación.

- *Determinar el área de interés* - Esta fase se enfoca en identificar la problemática a solucionar, sea esta un problema específico en la biblioteca o la exploración de un área para tomar una decisión.
- *Identificar las fuentes de datos internas y externas* - Partiendo del área de interés, con el objetivo de ver que datos son útiles para la problemática a solucionar, se analiza cada una de las fuentes internas (fuentes propias de información que se pueden obtener y explorar sin necesidad de acudir a terceros) y externas (fuentes provenientes de terceros a la biblioteca).
- *Creación de un Data Warehouse* - La primera parte de esta fase es la recopilación, limpieza, anonimizado y la integración de los datos. Posteriormente los datos resultantes son almacenados en un Data Warehouse (DW). Un DW es definido como “una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis”, determinadas a partir de los requerimientos del negocio considerando el tiempo (Kimball & Ross, 2002), integrando todas las fuentes de datos involucradas estratégicamente en la organización con el objetivo de ayudar a una adecuada toma de decisiones.
- *Seleccionar herramientas de análisis apropiadas* - Una vez que se tiene la información lista para el análisis, se debe seleccionar la herramienta que más facilidades ofrezca para la manipulación de la información.
- *Descubrir patrones utilizando DM y creación de reportes* - En esta fase se aplica las diferentes técnicas de DM sobre los datos con el objetivo de obtener patrones de comportamiento.
- *Analizar e implementar los resultados* - Una vez descubiertos los patrones de comportamiento, el siguiente paso es validar que dichos patrones sean correctos; para ello se testea un determinado porcentaje de datos, con el fin de probar que los resultados obtenidos sean los apropiados.

2.3. Algoritmos de asociación

Las técnicas de asociación permiten descubrir reglas que ocurren en común entre elementos que pertenecen a un conjunto de datos, para lo cual se considera todas las posibles combinaciones de atributo-valor de todos los datos almacenados en un conjunto (Chen *et al.*, 1996). Donde, dado un conjunto de registros se encontrará reglas que predicen la ocurrencia de un ítem, basándose en las ocurrencias de otros ítems en el registro.

2.4. Trabajos relacionados

En el trabajo presentado en Siguenza-Guzman *et al.* (2014), los autores utilizan una metodología holística, propuesto por Siguenza-Guzman *et al.* (2015) para la evaluación integral de bibliotecas. Dicha metodología evalúa la colección (fuentes externas e internas) y los servicios para la gestión de bibliotecas, tales como: el desempeño de los servicios, el control de calidad, el uso de la colección y la interacción con el usuario. En base a dicho análisis holístico, Haro *et al.* (2014), diseñan e implementan un sistema de soporte de decisiones basado en técnicas de DW para la biblioteca de la Universidad de Cuenca y propone una arquitectura de DW que integra, procesa y almacena los datos.

En el trabajo presentado por Riveros *et al.* (2015), los autores desarrollan un modelo de DM, sobre el cual emplean un mecanismo de descubrimiento de reglas de asociación para determinar el uso de los diferentes recursos bibliotecarios por parte de los usuarios finales. Adicionalmente, en Decker & Höppner (2006) los autores aplican reglas de asociación para determinar la usabilidad de los libros en una biblioteca; para la aplicación de dichas reglas, los autores parten de un sistema informático de administración de biblioteca basado en un DW.

3. APLICACIÓN DE LAS REGLAS DE ASOCIACIÓN

En esta sección se presenta un análisis de la información disponible y posteriormente se describe el procedimiento para la aplicación de técnicas de Bibliomining definidas en la sección 2.2. La aplicación de las fases se realizó en el CDRJBV perteneciente a la Universidad de Cuenca, Ecuador.

3.1. Análisis de datos

El CDRJBV está conformado por las Bibliotecas de los Campus Central, El Paraíso (Áreas de la Salud) y Yanuncay (Áreas Agropecuarias y de Artes). Su conjunto documental está constituido por publicaciones convencionales en todas sus formas, así como por diversos soportes digitalizados, audiovisuales y bases de datos en línea. El CDRJBV ofrece servicios de reservas de libros, préstamos a usuarios, préstamos interbibliotecarios, evaluación de bibliotecarios y encuestas de satisfacción. Los usuarios finales del CDRJBV son estudiantes, docentes y personal administrativo de la Universidad de Cuenca, a quienes se ofrece el servicio de préstamos y reservas de libros (Haro *et al.*, 2014).

Tabla 1. Las 5 Facultades con mayor préstamo de material bibliográfico.

Facultad	Porcentaje (%)
Ciencias Médicas	21.30
Ciencias Económicas y Administrativas	17.42
Filosofía, Letras y Ciencias de la	12.47
Facultad de Ingeniería	10.39
Arquitectura y Urbanismo	9.57

Tabla 2. Libros más populares en transacciones que solicitan más de un recurso bibliográfico

Identificador	Título
80772	Contabilidad general 4
62184	Cuaderno de prácticas contables: contabilidad general 4
7975	Algebra superior
49959	Ejercicios de Algebra Superior: soluciones de los ejercicios propuestos en la obra Algebra Superior
69329	Matemáticas para administración y economía
109649	Anuario de la escuela de medicina tomo 1
114800	Anuario de la escuela de medicina tomo 2
48977	Acuarela
60030	Introducción a la acuarela
113863	Pedrosa diagnóstico por imagen. Tratado de radiología clínica
13848	Diagnóstico por imagen – Cerebro
73542	Auditoría informática: guía para su aplicación
84325	Auditoría informática. Un enfoque práctico
53888	Contabilidad de costos
153067	Contabilidad de costos enfoque gerencial de gestión
44185	Geometría analítica
44937	Teoría y problemas de Geometría analítica
53409	El modulator
76163	El Modulator 2: Ensayo sobre una medida armónica a la escala humana aplicable universalmente a la Arquitectura y a la mecánica
69329	Código de procedimiento penal, legislación conexas, concordancias

La Universidad de Cuenca cuenta con un total de 12 Facultades, distribuidas en los diferentes campus de estudio. La Tabla 1 presenta las cinco facultades que más préstamos realizan con su respectivo porcentaje. En el CDRJBV la mayoría de préstamos registra un solo libro solicitado por el usuario, representando el 71.83% del total de préstamos realizados y el restante 28.17% de los usuarios solicitan más de un libro a la vez. La muestra de datos para la aplicación de algoritmos de asociación forma parte del 21.78% de préstamos, considerando los libros más populares de la

colección que dispone el CDRJBV dentro de ésta muestra que se presentan en la Tabla 2. Es necesario restringir la cantidad de libros a analizar dentro de la muestra de datos ya que cada libro representa una variable a ser analizada con los algoritmos de asociación

3.2. Proceso para la aplicación de reglas de asociación

Una vez realizada la segmentación de datos de los préstamos realizados en el CDRJBV, se procede a la implementación de las fases de descubrimiento del conocimiento en Bibliomining, presentada en la sección 2.2 de este artículo.

- *Determinar el área de interés* - Los principales usuarios del CDRJBV son los estudiantes de la Universidad de Cuenca, quienes constantemente realizan investigaciones en diferentes áreas. Se considera necesario aportar al estudiante, sugerencias de libros sean estas de la misma o diferentes temáticas cuando solicite un préstamo, considerando las preferencias de otros usuarios, siendo este nuestra área de interés para el análisis de las reglas de asociación.
- *Identificar las fuentes de datos internas y externas* - En el trabajo presentado en Haro *et al.* (2014), se describe que el CDRJBV posee fuentes propias de información que se pueden obtener y explorar sin necesidad de acudir a terceros, los datos a los que se tiene acceso internamente son los relacionados a los procesos de: Catalogación, Préstamos, Análisis de costos de procesos, Repositorio digital, Acceso remoto (EZProxy), Estadísticas de acceso a bases de datos digitales y Evaluación de servicios. Además de las fuentes internas, el CDRJBV dispone de acceso a otras fuentes externas provenientes del Departamento de Tecnologías de la Información y Comunicación de la Universidad de Cuenca (DTIC), relacionadas con ciertos procesos académicos, socioeconómicos y administrativos. Es así que, para el descubrimiento de las reglas de asociación, este artículo se centra en las fuentes relacionadas al proceso de préstamos, donde los atributos de fecha de préstamo, fecha de devolución, campus, categoría, autores, título y bibliotecario son recuperados desde las fuentes propias del centro de documentación; y los atributos de usuario, facultad y carrera son obtenidos de las fuentes del DTIC (ver Figura 1).

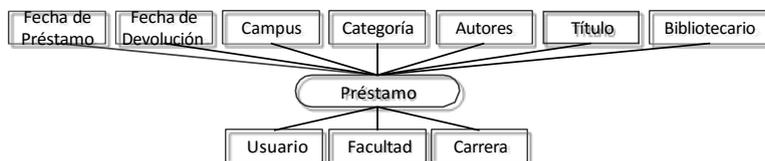


Figura 1. Modelo conceptual del proceso de préstamos.

- *Creación de un Data Warehouse* - Como se presentó en la sección 2.4 de este artículo, el CDRJBV dispone de un DW diseñado bajo las necesidades del centro de documentación (Haro *et al.*, 2014). En dicho trabajo, los autores realizan los diferentes procesos de limpieza, la respectiva integración de las diferentes fuentes y posteriormente la información es almacenada en un DW.
- *Seleccionar herramientas de análisis apropiadas* - Siguiendo los lineamientos establecidos en Haro *et al.* (2014), la herramienta seleccionada para la implementación de DM es Weka², debido a que ofrece un completo banco de herramientas y algoritmos de asociación.
- *Descubrir patrones utilizando DM y creación de reportes* - Las técnicas de asociación permiten descubrir reglas que ocurren en común entre elementos que pertenecen a un conjunto de datos (Chen *et al.*, 1996). Donde, dado un conjunto de registros se encontrará reglas que predicen la ocurrencia de un ítem, basándose en las ocurrencias de otros ítems en el registro. Considerando la regla de asociación presentada en Chen *et al.* (1996) y Riveros *et al.* (2015), se define:

² Waikato Environment for Knowledge Analysis es un software para el análisis del conocimiento, desarrollado por la Universidad de Waikato que presenta una colección de algoritmos de aprendizaje automático para tareas de Data mining (<http://www.cs.waikato.ac.nz/ml/weka/>).

- $D = \{d_1, d_2, d_3, \dots, d_m\}$, como el conjunto de los códigos de libros disponibles en el CDRJBV.
- $T = \{t_1, t_2, t_3, \dots, t_n\}$, como el conjunto de transacciones, donde cada transacción $t_i \subseteq D$ $1 \leq i \leq n$ consiste de los códigos de libros que son requeridos por un usuario en una misma transacción.

Una regla se define como una implicación de la forma $X \Rightarrow Y$, donde $X \subset D, Y \subset D, X \cap Y = \emptyset$ y $X \cup Y \subseteq t_i$. El conjunto de códigos de libros formado por aquellos que corresponden al antecedente o al consecuente, no contiene códigos repetidos, y debe estar contenido o ser igual a alguna de las transacciones pertenecientes a T .

- *Analizar e implementar los resultados* - Después de aplicar los algoritmos de asociación sobre la muestra de datos seleccionada, se encontró 12 reglas que se visualizan en la Tabla 3, los datos presentados y la validación se describen en la sección 4. Los resultados se obtuvieron utilizando la herramienta Weka haciendo uso del algoritmo *Apriori*. Dicho algoritmo utiliza el conocimiento previo de las propiedades de conjunto de elementos frecuentes, reduciendo el espacio de búsqueda y aumentando la eficiencia (Han *et al.*, 2011).

Tabla 3. Reglas de asociación encontradas en el CDRJBV.

Número de regla	Libro 1	Libro 2	Soporte	Confianza	Lift
R1	80772	62184	0,00048	0,3098	282,05
R2	62184	80772	0,00048	0,44	282,05
R3	7975	49959	0,00092	0,2210	89,03
R4	49959	7975	0,00092	0,3716	89,03
R5	109649	114800	0,00015	0,7	3185,98
R6	44185	44937	0,00015	0,0507	37,23
R7	53409	76163	0,00018	0,6666	3034,26
R8	44937	49959	0,00002	0,0178	90,305
R9	48977	60030	0,00018	0,1951	740,06
R10	113863	13848	0,00018	0,1212	106,09
R11	73542	84325	0,00022	0,5	650,2
R12	53888	153067	0,00015	0,0823	70,72

4. RESULTADOS Y VALIDACIÓN

A continuación, se procede con la evaluación de las reglas de asociación generadas en este modelo presentado en la sección anterior. Las medidas más comunes para evaluar las reglas de asociación son el soporte, la confianza y el *lift* (Brown, 2014); donde, la regla $X \Rightarrow Y$ tiene soporte s en el conjunto de transacciones T , $0 \leq s \leq 1$, si el porcentaje de s de las transacciones de T contienen tanto a Y como a X . El soporte puede ser considerado como la probabilidad de que las transacciones contengan un conjunto de ítems *soporte* (X) = $P(X)$. Para las reglas de asociación, el conjunto está formado por los ítems que conforman el antecedente y el consecuente de la regla de asociación, *soporte* ($X \Rightarrow Y$) = $N(X \cup Y)/N(T)$, donde $N(T)$ es el número total de coincidencias del conjunto T .

La regla $X \Rightarrow Y$ se mantiene en el conjunto de transacciones T , con factor de confianza c , $0 \leq c \leq 1$, si el porcentaje de c de las transacciones de T que satisfacen X también satisfacen Y ; esto es, el porcentaje de transacciones que contienen a X e Y respecto al número total de transacciones que contienen X . La confianza se define como la probabilidad de que las transacciones que contienen el antecedente de la regla, también contengan el consecuente; esto es, la probabilidad de que ocurra Y dado que ya ocurrió X , $P(Y/X)$.

$$confianza (X \Rightarrow Y) = \frac{N(X \cup Y)}{N(X)}$$

La medida de independencia estadística *lift*, es definida como una relación entre la ocurrencia simultánea de *X* e *Y*, cuando los conjuntos de ítems que conforman el antecedente y el consecuente de la regla sean estadísticamente independientes. En el trabajo presentado por (Hahsler, Hornik, & Reutterer, 2006) se concluye que las reglas de conocimiento descubiertas por medio de soporte y confianza, deberían ser filtradas usando sus valores de *lift*, dado que los valores de *lift* mayores a 1 indican asociación entre ítems; y los valores menores a 1 indican su independencia entre ítems y no deben ser consideradas para la toma de decisiones.

$$lift(X \Rightarrow Y) = lift(Y \Rightarrow X) = \frac{N(X \cup Y) * N(T)}{N(X) * N(Y)}$$

Las 12 reglas identificadas en la sección 3.2 aportan al conocimiento generado, debido a que su valor de *lift* es mayor que 1 lo que significa que sus asociaciones son estadísticamente dependientes. La Tabla 3 se lee de la siguiente manera: Considerando la regla R10, si el libro “*Pedrosa diagnóstico por imagen. Tratado de radiología clínica*” con código 113863 fue pedido, por lo tanto el libro “*Diagnóstico por imagen - Cerebro*” con código 13848 también fue pedido en una misma transacción con un soporte del 0,00018, una confianza de 0,1212 y un *lift* de 106,9. De modo similar para la regla R6, si el libro “*Geometría analítica*” con código 44185 fue pedido, por lo tanto el libro “*Teoría y problemas de Geometría analítica*” con código 44937 también fue pedido en una misma transacción con un soporte del 0,00015, una confianza de 0,0507 y un *lift* de 37,23; y el mismo libro en la regla R8 posee asociación con el libro “*Ejercicios de Algebra Superior: soluciones de los ejercicios propuestos en la obra Algebra Superior*” de código 49959 con un soporte del 0,00002, una confianza de 0,0178 y un *lift* de 90,305. Por el valor de confianza se puede afirmar de la regla R10 que el 12,12% de transacciones que involucra el libro con código 113863, también contiene el libro 13848; la regla R6, que el 5,07% de transacciones que involucra el libro con código 44185, también contiene el libro 44937; y en la regla R8 el 1,78% de transacciones que involucra el libro con código 44937, también contiene el libro 49959.

En la Figura 2 se visualiza los valores de confianza y soporte para las 12 reglas descubiertas. Las reglas R5, R7 y R11 presentan una confianza superior al 50%, significando que en la mayoría de las transacciones que involucra un libro X también se presta el libro Y.

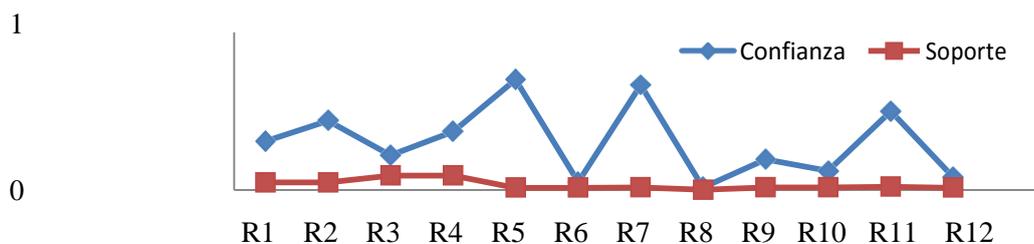


Figura 2. Reglas de Asociación con valores de soporte y confianza

5. CONCLUSIONES Y TRABAJOS FUTUROS

La necesidad de tener un conocimiento profundo de los datos y descubrir reglas de negocio que ayuden a una adecuada toma de decisiones, ha generado que se inicie la investigación y aplicación de Bibliomining en el CDRJBV. En este artículo se presentó la implementación de técnicas de Bibliomining, sobre la información que posee el centro de documentación de la Universidad de Cuenca con la finalidad de descubrir patrones de asociación de los recursos bibliográficos. Dicho proceso fue guiado por las fases de Bibliomining presentadas por Nicholson (2003) y posteriormente, se presentó la evaluación de resultados considerando el nivel de soporte, confianza y *lift* que poseen, para que finalmente puedan ser consideradas como reglas válidas. De esta manera, las reglas de

asociación en el CDRJBV ayudan a determinar preferencias de los usuarios al solicitar recursos bibliográficos y permite ofrecer un servicio más personalizado en la biblioteca, al sugerir bibliografía de acuerdo al comportamiento de los otros usuarios.

El éxito de Data mining depende del valor de la información que se disponga, razón por la cual, la etapa de limpieza de datos y la elección de la muestra de datos debe realizarse minuciosamente. Cuando el Data mining es aplicado sobre un Data Warehouse se puede obviar la fase de limpieza, ya que posee información de mejor calidad, que la encontrada en las bases de datos de los sistemas transaccionales.

A futuro se pretende extender este trabajo considerando la información académica y financiera de los usuarios, para aprovechar la gran cantidad de datos, y así descubrir nuevo conocimiento. Además, la aplicación de técnicas de predicción, clasificación y agrupación, permitirá al CDRJBV proporcionar programas más apropiados, resolver las necesidades de información del usuario y observar problemas en su colección bibliográfica.

AGRADECIMIENTOS

Se agradece al Departamento de Ciencias de la Computación por el apoyo y asesoramiento brindado, así como también al personal del Centro de Documentación Regional “Juan Bautista Vázquez” que siempre están prestos a colaborar en los proyectos planteados.

REFERENCIAS

- Brown, M. S. (2014). *Data mining for dummies*. Hoboken, NJ: John Wiley & Sons.
- Chen, M.-S., J. Han, P. Yu, 1996). *Data mining: An overview from database perspective*. IEEE Transactions on Knowledge and Data Engineering, 866-883.
- Decker, R., M. Höppner, 2006. *Strategic planning and customer intelligence in academic libraries*. Library Hi Tech, 504-514.
- Hahsler, M., K. Hornik, T. Reutterer, 2006. *Implications of probabilistic data modeling for mining association rules*. Chapter in: From Data and Information Analysis to Knowledge Engineering, Part of the series Studies in Classification, Data Analysis, and Knowledge Organization, pp 598-605.
- Han, J., M. Kamber, J. Pei, 2011. *Data mining: Concepts and techniques* (3rd ed.). Elsevier, 704 pp.
- Haro, V., W. Pérez, L. Siguenza-Guzman, D. Cattrysse, V. Saquicela, 2014. *Diseño e implementación de un sistema de soporte de decisiones para el Centro de Documentación Regional “Juan Bautista Vázquez”*. MASKANA Número especial: ACTAS: Congreso TIC.EC (2014), 245-256.
- Kimball, R., M. Ross, 2002. *The data warehouse toolkit: The complete guide to dimensional modeling*. Hoboken, NJ, John Wiley & Sons.
- Liao, S.-H., P.-H. Chu, P.-Y. Hsiao, 2012. Data mining techniques and applications - A decade review from 2000 to 2011. *Expert Systems with Applications*, 39(12), 11303-11311.
- Nicholson, S., 2003. The bibliomining process: Data warehousing and data mining for library decision-making. *Transinformação*, 16(3), 146-151.
- Nicholson, S., J. Stanton, 2003. *Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries*. In: Organizational Data Mining: Leveraging Enterprise Data Resources, Chapter 17, pp. 247-262.
- Nicholson, S., J. Stanton, 2005. *Bibliomining for library decision-making*. Disponible en [http://ebooks.narotama.ac.id/files/Encyclopedia%20of%20Information%20Science%20and%20Technology%20\(2nd%20Edition\)/Bibliomining%20for%20Library%20Decision-Making.pdf](http://ebooks.narotama.ac.id/files/Encyclopedia%20of%20Information%20Science%20and%20Technology%20(2nd%20Edition)/Bibliomining%20for%20Library%20Decision-Making.pdf), pp. 341-345.

- Riveros Malberti, M.A., R.O. Klenzi, 2015). Reglas de asociación en el cómputo de utilización de libros en una biblioteca universitaria. *Enfoque UTE*, 6(2), 86-101.
- Siguenza-Guzman, L., V. Saquicela, D. Cattrysse, 2014. *Design of an integrated decision support system for library holistic evaluation*. International Association of University Libraries Conference. Disponible en <https://lirias.kuleuven.be/bitstream/123456789/455176/1/Conference+Paper.pdf>, 12 pp.
- Siguenza-Guzman, L., V. Saquicela, E. Avila-Ordóñez, J. Vandewalle, D. Cattrysse, 2015. Literature review of data mining applications in academic libraries. *The Journal of Academic Librarianship*, 41(4), 499-510.
- Siguenza-Guzman, L., A. Van Den Abbeele, J. Vandewalle, H. Verhaaren, D. Cattrysse, 2015. A holistic approach to supporting academic libraries in resource allocation processes. *The Library Quarterly: Information, Community, Policy*, 85(3), 295-318.
- Witten, I.H., E. Frank, M.A. Hall, 2011. *Data mining: Practical machine learning tools and techniques*. Burlington, MA: Morgan Kaufmann.