

Tipo de artículo: Artículo original
Temática: Reconocimiento de patrones
Recibido: 13/11/2014 | Aceptado: 03/07/2015

Reglas de Asociación con los datos de una biblioteca universitaria

Association Rules for university library data

María Alejandra Malberti Riveros ^{1*}, Graciela Elida Beguerí ¹

¹ Departamento de Informática. Facultad de Ciencias Exactas, Físicas y Naturales. Universidad Nacional de San Juan. Av. Ignacio de la Roza 590 (O), Complejo Universitario "Islas Malvinas", Rivadavia, San Juan, Argentina. CPA: J5402DCS. [\[amalberti.grabeda\]@gmail.com](mailto:amalberti.grabeda@gmail.com)

* Autor para correspondencia: amalberti@gmail.com

Resumen

Con el objetivo de promover e impulsar el uso de los recursos de la biblioteca y a la vez facilitar la labor por parte de su personal, en este trabajo se describe el uso de Minería de Datos, descubrimiento de Reglas de Asociación, en una biblioteca universitaria. Esta tecnología se aplicó a datos de circulación de libros con el propósito de ayudar a proyectar la conveniencia en la disposición física del material librario. En el análisis se plantearon dos escenarios llamados “Espacios Físicos Cerrados” y “Espacios Físicos Abiertos”. El primer escenario refiere a las estanterías de libros que sólo pueden ser accedidas por el personal de la unidad, mientras que el segundo considera estanterías de libre acceso por parte de las personas que concurren a la biblioteca en búsqueda de libros. Para descubrir reglas de asociación relevantes se consideraron, además de las medidas soporte, confianza y *lift*, las acciones concretas que los encargados del establecimiento realizan habitualmente para satisfacer las demandas de sus socios. Si bien la Minería de Datos ha sido ampliamente utilizada, en el campo de la bibliotecología no se ha explotado al máximo su potencial. Por ello, también se pretende que este estudio ayude a impulsar al personal a incursionar en la aplicación de nuevas tecnologías de manera que puedan lograr un mayor aprovechamiento de los datos que diariamente son recolectados.

Palabras clave: Minería de Datos, Reglas de Asociación, Biblioteca Universitaria, Evaluación

Abstract

This paper aims at promoting and encouraging the use of library resources and, at the same time, facilitating the library staff tasks when using Data Mining and Association Rules at a university library. This technology has been applied in book circulation data in order to decide the best library material arrangement. As far as the analysis is

concerned, two scenarios were presented: “Close Physical Spaces” and “Open Physical Spaces”. The former refers to those bookshelves which can only be accessed by the section staff whereas the latter refers to bookshelves which can be freely accessed by regular book users. To find out the relevant Association Rules, not only support, reliance and lift measures have been considered but also the daily duties that library staff perform regularly to meet their members’ demands. Even though Data Mining has been widely used, its potential has not yet been fully exploited in the library field. Therefore, the present study is also intended to encourage staff to become involved in the application of new technologies so that they can make better use of daily collected data.

Keywords: *Data Mining, Association Rules, University Library, Evaluation*

Introducción

La evaluación y la acreditación, en la educación superior, son procesos que a nivel mundial están contribuyendo al mejoramiento de la educación universitaria.

Argentina cuenta con un organismo descentralizado, CONEAU -Comisión Nacional de Evaluación y Acreditación Universitaria-, que funciona en jurisdicción del Ministerio de Educación de la Nación. Éste es el único organismo público nacional de evaluación y acreditación universitaria y comenzó a funcionar en 1996. Su misión institucional es asegurar y mejorar la calidad de las carreras e instituciones universitarias que operan en el sistema universitario argentino.

De acuerdo con los lineamientos formulados por CONEAU, los servicios de biblioteca cumplen un rol central en el quehacer universitario y lo afirma cuando dice:

“Efectivamente, para poder llevar a cabo las tareas de investigación y docencia es central tener bibliotecas completas y actualizadas, con políticas de adquisición ajustadas a las necesidades de la institución y sus miembros ... Sin acceso a la producción de conocimiento local e internacional no es posible un adecuado uso y distribución del conocimiento. El acceso a la información bibliográfica debe satisfacer, fundamentalmente, las necesidades de las funciones de docencia de grado, posgrado y de investigación de los miembros de la institución. La comunidad universitaria debe tener acceso a servicios y herramientas informáticas para satisfacer las funciones previstas”. (Res. N° 094-CONEAU)

Así, a partir del proceso de evaluación institucional realizado en la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de San Juan, Argentina -FCEF, UNSJ- y enmarcado en los lineamientos mencionados,

se iniciaron acciones tendientes a fortalecer la administración y gestión de la biblioteca perteneciente a la institución referida.

Una biblioteca es una *“Organización o parte de una organización, cuyos principales objetivos son crear y mantener una colección y facilitar el uso de estos recursos de información y de las instalaciones cuando se requieran para satisfacer las necesidades de información, de investigación, educativas, culturales o de ocio de sus usuarios”* (Norma ISO-11620 Información y Documentación - Indicadores de Desempeño para Bibliotecas). El concepto de biblioteca ha evolucionado desde su consideración como “guarda y custodia de libros” a la idea actual de "servicio" (IFLA/UNESCO). En particular, las bibliotecas universitarias constituyen un servicio clave de apoyo a las dos funciones que constituyen la razón de ser de una institución universitaria: la investigación o creación de conocimiento y la enseñanza o comunicación de dicho conocimiento. Estas unidades, como cualquier otro servicio, deben analizar y evaluar su rendimiento. En tal sentido los Estándares del Sistema de Bibliotecas de la Universidad de Buenos Aires (SISBI, 2013) y Ana Pérez López (Pérez López, 2002), entre otros, reconocen dos enfoques de evaluación, independientemente de que las técnicas empleadas sean cuantitativas o cualitativas: el enfoque centrado en los materiales y el enfoque centrado en los usuarios, entendiendo como usuario a aquella persona que utiliza un libro, una publicación periódica o un segmento de la colección.

En lo que refiere específicamente a la evaluación de uso de una colección existe como presunción que el uso en el pasado es un buen predictor del uso presente o futuro, ya que los intereses de los usuarios pueden cambiar, pero esto ocurre muy lentamente. Al respecto, la mencionada autora afirma que “Los estudios de uso hacen posible que los bibliotecarios observen qué métodos de distribución y promoción son mejores. Promocionar los títulos mejorando su visibilidad y accesibilidad afecta a su uso. En una serie de estudios, Goldhor (1973) y Taylor (1981) demostraron que los libros expuestos en las estanterías se prestaban más que los libros no expuestos en las estanterías. Profundizando en esta investigación, Baker (1986) descubrió que la exposición de los libros en estanterías puede incrementar el uso, independientemente del tipo de libro de que se trate, pero sólo si están localizados en áreas fácilmente accesibles y muy visibles.” (Pérez López, 2002, p. 352)

Si bien bibliotecas y servicios de información pertenecientes a distintas organizaciones, capturan la historia de circulación de sus materiales e información sobre sus usuarios y sobre sus recursos en la colección, desafortunadamente pocas de ellas han aprovechado estos datos para mejorar el servicio al cliente, dirigir presupuestos de adquisición de recursos, o influir en la toma de decisiones estratégicas sobre los usos de la información en sus organizaciones (Nicholson y otros, 2003). Ricardo Herrera Varela (Herrera Varela, 2006)

concluye que mediante la minería de datos, la visualización de los mismos, y la estadística, estas organizaciones de información pueden conseguir una visión más clara de las necesidades de entrega y de gestión de la información.

Con la gran cantidad de datos recolectados durante años mediante el sistema MicroISIS, sistema realizado por la UNESCO y disponible en la Biblioteca de la FCEFN, se decidió buscar en ellos conocimiento por medio de la aplicación de Minería de Datos.

La Minería de Datos refiere al proceso de extracción no trivial de información implícita, útil y previamente desconocida, desde los datos de una base de datos (Chen y otros, 1996) (Agrawal, 1993). Dentro de ella, la minería de datos, en su estrategia de descubrimiento de reglas de asociación, propone encontrar conjuntos de elementos que co-ocurren juntos frecuentemente en una base de datos (Agrawal, 1993) (Chen y otros, 1996, p.7) (Hipp y otros 2000) (Hahsler y otros, 2006). La aplicación de esta tecnología, también denominada Minería de Reglas de Asociación (MRA), puede generar una gran cantidad de reglas, por lo que surge la necesidad de seleccionar aquellas reglas de asociación relevantes desde la perspectiva de los usuarios de la tecnología. Este aspecto es destacable pues se pretende que bibliotecarios, usuarios en este caso, tengan en cuenta y apliquen la minería de reglas de asociación en varias instancias de gestión de la biblioteca.

En particular este trabajo propone utilizar Minería de Reglas de Asociación, como una tecnología que respalda la toma de decisiones relacionadas con la disposición física del material librario disponible en la biblioteca, con el propósito de favorecer el servicio de préstamo de su colección. Siguiendo con esta idea, se desarrollaron las actividades expuestas en el presente trabajo.

Materiales y métodos

Dentro de la MD se destaca el problema de descubrir asociaciones a partir de los datos; en otras palabras, identificar grupos de variables que se correlacionan fuertemente entre sí. Se cuenta con un conjunto de ítems y un gran conjunto de transacciones que son subconjuntos de esos ítems. La MRA procura encontrar relaciones entre ítems, llamadas Reglas de Asociación, a partir de la presencia frecuente de varios ítems dentro de las transacciones.

Formalmente, se consideran los siguientes modelos matemáticos para dirigir el problema de búsqueda de reglas de asociación:

$D = \{d_1, d_2, \dots, d_m\}$ conjunto de ítems.

$T = \{t_1, t_2, \dots, t_n\}$ conjunto de transacciones, donde cada transacción t_i es un conjunto de ítems tal que $t_i \subseteq D$ $1 \leq i \leq n$.

La implicación $X \Rightarrow Y$ es una **Regla de Asociación** donde $X \subset D$, $Y \subset D$, $X \cap Y = \phi$ y $X \cup Y \subseteq t_i$, los conjuntos X e Y son mutuamente excluyentes, t_i es el conjunto de ítems formado por aquellos que corresponden al antecedente o al consecuente de la regla de Asociación. El conjunto $X \cup Y$ debe estar contenido o ser igual a alguna de las transacciones pertenecientes a T .

Las medidas más populares en MRA son soporte, confianza y *lift* (Brown, 2014) (Lucas, 2010). En el marco soporte-confianza, originariamente (Agrawal, 1993) (Brin, 1997) (Silverstein, 1998), la búsqueda de reglas de asociación adopta los factores soporte y confianza para evaluar las reglas descubiertas.

La regla $X \Rightarrow Y$ tiene **soporte** s en el conjunto de transacciones T , $0 \leq s \leq 1$, si $s\%$ de las transacciones de T contienen tanto a X como a Y . El soporte puede ser considerado como la probabilidad de que las transacciones contengan un conjunto de ítems.

$$\text{soporte}(X) = P(X) \quad (1)$$

Para el caso de las reglas de asociación, el conjunto está formado por los ítems que conforman el antecedente y el consecuente de la regla de asociación:

$$\text{soporte}(X \Rightarrow Y) = \text{soporte}(X \cup Y) = P(X \cup Y) \quad (2)$$

La regla $X \Rightarrow Y$ se mantiene en el conjunto de transacciones T , con factor de **confianza** c , $0 \leq c \leq 1$, si $c\%$ de las transacciones de T que satisfacen X también satisfacen Y ; esto es, el porcentaje de transacciones que contienen a X e Y - $X \cup Y$ - respecto al número total de transacciones que contienen X .

$$\text{confianza}(X \Rightarrow Y) = \frac{\text{soporte}(X \cup Y)}{\text{soporte}(X)} = \frac{P(X \cup Y)}{P(X)} \quad (3)$$

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 . La confianza

puede ser considerada, entonces, como un estimador de la probabilidad condicional $P\left(\frac{Y}{X}\right)$. La confianza está dirigida y computa valores diferentes para las reglas $X \Rightarrow Y$ e $Y \Rightarrow X$.

Según Romero Morales (Romero Morales, 2003) la confianza no es capaz de detectar la independencia estadística. Este mismo autor expresa que normalmente se ha creído que mientras mayor es el soporte, mejor es el conjunto de elementos, pero esto puede no ser siempre cierto ya que un conjunto de elementos con alto soporte puede ser fuente de engaños, debido a que aparecen en la mayoría de las transacciones.

Las restricciones que pueden presentar los factores de soporte y confianza, unidos a la necesidad de rescatar las reglas interesantes de un conjunto posiblemente numeroso de reglas generadas, dificulta la actividad de los expertos en el campo en el que se aplica MD. A raíz de ello, distintos autores han desarrollado otras medidas para evaluar la importancia de las reglas generadas. Entre estas medidas se rescata la denominada *lift* o medida de independencia (Brin, 1997; Romero Morales, 2003 y Silverstein, 1998). El factor *lift* representa una prueba para medir la dependencia estadística y se define como:

$$\text{lift}(X \Rightarrow Y) = \text{lift}(Y \Rightarrow X) = \frac{P(X \cup Y)}{P(X)P(Y)} \quad (4)$$

Este factor establece 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. El *lift* es simétrico, es decir que $\text{lift}(X \Rightarrow Y) = \text{lift}(Y \Rightarrow X)$, de forma que este valor sólo mide el grado de dependencia y no la implicación en ambas direcciones. Según Hahsler (Hashler, 2006) *las reglas recuperadas por medio de soporte y confianza, deberían ser filtradas usando sus valores de lift*, pues valores de *lift* mayores a 1 indican asociación entre ítems, mientras que valores menores a 1 pueden indicar su independencia. Reglas de asociación con valores de *lift* menores a 1 no deberían ser tenidas en cuenta para la toma de decisiones.

Lenca (Lenca, 2007) provee una descripción comparativa entre un conjunto de medidas, a partir de la identificación de una serie de propiedades *relevantes desde la perspectiva de los usuarios*. Las medidas consideradas son objetivas, esto es dirigidas por los datos, pues en su cómputo sólo se toman en cuenta la cardinalidad de los mismos; en otras palabras, las medidas empleadas para extraer reglas de asociación interesantes, son definidas por medio del uso de contadores de ocurrencias o de frecuencias relativas. La noción de medidas interesantes –*interestingness measures*–, es tomada generalmente como una medida del valor del patrón o modelo descubierto, que combina validez, novedad,

utilidad y sencillez. Un patrón provee conocimiento si excede algún umbral de interés. Veinte medidas interesantes, propuestas por Zhao (Zhao, 2004) y Vaillant (Vaillant, 2004) son retomadas por Lenca, y evaluadas según propiedades que cubren un amplio espectro de potenciales preferencias de los usuarios. Entre esas propiedades están: **Procesamiento asimétrico** de X e Y , **Disminución con respecto al número de registros que verifican Y** - n_Y , **Independencia** entre X e Y , **Sensibilidad al número total de registros**- N , **Facilidad para fijar un umbral** y Habilidad de la medida de expresar una idea exhaustiva del interés de una regla- **Inteligibilidad**. En la Tabla 1 se presenta parte de la matriz de evaluación propuesta por Lenca (Lenca, 2007).

Tabla 1. Propiedades de las medidas soporte, confianza y *lift*.

Prop. Medida	Procesamiento asimétrico de X e Y	Disminuye con n_Y	Independencia	Sensibilidad a N	Facilidad para fijar un umbral	Inteligibilidad
confianza	Asym	no-dec	var	Desc	easy	a
<i>lift</i>	Sym	dec	cst	Desc	easy	a
soporte	Sym	no-dec	var	Desc	easy	a

Nota: sym/asym: simétrica/asimétrica; dec/no-dec: decreciente/no decreciente; var/cst: variable/constante; Desc: invariable; easy:fácil; a: semántica fácilmente comprensible.

Aplicación de Minería de Reglas de Asociación

La Biblioteca de la FCEFNU-SJ, dispone del Sistema MicroISIS, en el cual se registra diariamente la circulación de sus colecciones. Los bibliotecarios que trabajan en esta unidad, además de facilitar cada recurso librario y registrar su movimiento en el sistema de biblioteca –préstamo, renovación, consulta o devolución- ante requerimientos concretos por parte de los socios, colaboran en las tareas vinculadas con el desarrollo y acceso a la colección, y participan en las actividades relacionadas con la asignación y organización del material librario en las estanterías. Esta distribución física es realizada según código Dewey, asignado a cada libro por un bibliotecario experto dedicado al proceso de catalogación del material cuando éste es adquirido. Además de ser usado para establecer su ubicación física en las estanterías, cada código Dewey describe el área temática específica a la que cada libro pertenece.

Descubrimiento de Conocimiento

A partir del proceso de MD propuesto por Mehmed Kantardzic (Kantardzic, 2003) en este trabajo se siguieron los siguientes pasos:

1. Establecer el problema.
2. Recolectar datos.
3. Realizar preprocesamiento.
4. Aplicar Minería de Reglas de Asociación.
5. Interpretar el modelo y obtener conclusiones

1-Establecer el Problema

Se procura descubrir conocimiento, plasmado en reglas de asociación, para ser aplicado en la toma de decisiones relativas a la disposición física del material librario en cada uno de los siguientes escenarios: **Espacios Físicos Cerrados**, aquellos espacios sólo accesibles por el personal de la biblioteca, y **Espacios Físicos Abiertos**, aquellos lugares con estanterías de libre acceso para las personas que requieren libros de la biblioteca.

2-Recolectar Datos

Los datos de todos los recursos y movimientos de la biblioteca están en tablas del sistema MicroISIS, de las que se usaron particularmente las relativas a Préstamos y Catálogo.

La tabla Préstamos contiene aproximadamente 80000, 74000 y 75000 registros correspondientes a tres años. Cada registro tiene los datos surgidos del movimiento de cada ejemplar -libro- ante el requerimiento concreto del mismo por parte de un usuario socio de la biblioteca. Estos datos son Número de Socio, Número de Inventario del libro, Tipo de Movimiento -Préstamo, Devolución, Renovación, Consulta-, Fecha de Préstamo - Año/Mes/Día, entre otros.

La tabla Catálogo tiene entre los datos que describen a cada libro: Número de Inventario del libro, Título, ISBN / ISSN y Código Dewey.

3-Realizar Preprocesamiento

A partir de considerar que cada transacción consiste de los códigos Dewey correspondientes a los libros (ejemplares) –más de uno-, gestionados por un socio- Número de Socio- en una misma visita –Fecha- a la biblioteca, se adaptaron las definiciones del modelo matemático al contexto de biblioteca:

$D = \{d_1, d_2, \dots, d_m\}$ conjunto de los códigos Dewey correspondientes a los libros de la biblioteca de la FCEF, N,

$T = \{t_1, t_2, \dots, t_n\}$ conjunto de transacciones, en el que cada transacción $t_i \subseteq D$ $1 \leq i \leq n$, consiste de los códigos Dewey correspondientes a los libros requeridos por un socio en una sola visita a la biblioteca.

La implicación $X \Rightarrow Y$ es una **Regla de Asociación** donde $X \subset D$, $Y \subset D$, $X \cap Y = \phi$ y $X \cup Y \subseteq t_i$, esto es, el conjunto de códigos Dewey 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 .

En este paso se realizaron las siguientes tareas:

a) Selección de transacciones: Las transacciones ocurridas en los distintos años tratados están inmersas en la tabla Préstamos. Cada transacción consta de **todos** los registros, más de uno, correspondientes a los **distintos movimientos**: pedido, devolución, consulta, renovación de libros; realizados por **un socio** (Número de socio), en **una visita** (Fecha), a la biblioteca.

b) Construcción de pares de códigos Dewey, potenciales reglas de asociación: En una transacción específica, cada uno de los libros -que tiene un código Dewey particular-, puede estar involucrado en un préstamo, devolución, renovación o consulta. De ello se desprende que **no solo debe tenerse en cuenta el código Dewey del libro, sino el tipo de movimiento en el que ese libro está involucrado**. De este análisis surgieron los siguientes interrogantes:

¿Qué combinaciones de movimientos son significativas al momento de seleccionar los pares de códigos Dewey que integren reglas de asociación aceptables?

¿Son las dieciséis combinaciones posibles de movimientos –Tabla 2–, convenientes para los escenarios propuestos?

Estos interrogantes condujeron al **análisis de las características funcionales de cada uno de los escenarios considerados**. Para descubrir reglas de asociación relevantes no solo fue importante tener en cuenta los códigos Dewey de los libros, sino también el tipo de movimiento en el que dichos libros estaban involucrados.

- Escenario 1. Espacios Físicos Cerrados- Estanterías de acceso restringido

En este escenario, el socio que concurre a la biblioteca en búsqueda de libros, tiene generalmente una idea precisa sobre el material que desea solicitar. El bibliotecario es quien accede a las estanterías para recuperar cada ejemplar en caso de préstamo o consulta, pues la renovación no involucra acceso a la estantería, tampoco para restituirlo a su lugar de origen si se trata de una devolución. En este último tipo de movimiento, los bibliotecarios de la FCEFNB suelen depositar los libros devueltos junto al mostrador de atención al público, para luego acomodarlos en sus lugares en los momentos en los que no hay socios esperando atención o al final de la jornada.

- Escenario 2. Espacios Físicos Abiertos - Estanterías de libre acceso

En este caso, es el usuario quien extrae el libro de la estantería al acudir a la biblioteca con la intención de solicitar un libro en préstamo o consulta. En una devolución de un libro por parte de un socio, es el bibliotecario quien retorna el libro al lugar que le corresponde en la estantería. Si el socio requiere una renovación de un libro, este movimiento es resuelto en el mostrador y no se genera acceso a la estantería.

El análisis realizado promovió la selección, para cada escenario, de un subconjunto de pares de movimientos respecto a las dieciséis alternativas posibles, los que son presentados en la Tabla 2 por medio del símbolo X. Para el caso de Espacios Físicos Cerrados, se consideraron las combinaciones de movimientos causantes que el bibliotecario realice más de una visita a las estanterías, para obtener libros demandados por los socios. Es por ello que los pares de movimientos seleccionados son Préstamo-Consulta, Consulta-Préstamo, Préstamo-Préstamo, Consulta-Consulta.

Tabla 2. Combinaciones de movimientos para Espacios Físicos Cerrados y Espacios Físicos Abiertos

Escenarios Movimientos	Espacios Físicos	
	Cerrados	Abiertos
Devolución, Préstamo		X
Devolución, Renovación		
Devolución, Consulta		X
Préstamo, Devolución		X
Préstamo, Renovación		X
Préstamo, Consulta	X	X
Renovación, Devolución		
Renovación, Préstamo		X
Renovación, Consulta		X
Consulta, Devolución		X
Consulta, Préstamo	X	X
Consulta, Renovación		X
Devolución, Devolución		
Préstamo, Préstamo	X	X
Renovación, Renovación		
Consulta, Consulta	X	X

Respecto al escenario de Espacios Físicos Abiertos se consideró que, si bien el socio suele concurrir a la biblioteca con una idea sobre el material que necesita, la ubicación física de los libros puede promover en él la revisión de materiales que están en una ubicación próxima a los libros que en un comienzo solicitaría. Esta proximidad física entre los libros requeridos frecuentemente, podría evitarle al socio tener que realizar una nueva visita a la biblioteca al asistirlo, de manera silenciosa, en sus posibles necesidades de información. Con esta finalidad se tuvieron en cuenta

también combinaciones de movimientos en las que intervienen renovaciones y devoluciones, además de préstamos y consultas.

En particular, en el trabajo se presenta lo desarrollado para Espacios Físicos Cerrados. Del preprocesamiento surgieron 192605 registros de movimientos realizados en la biblioteca en el período de tres años. A partir de ellos y al combinar entre sí, de a pares, todos los códigos Dewey involucrados en cada transacción conjuntamente con sus tipos de movimientos-Tabla 2, Espacios Físicos Cerrados-, por medio de consultas SQL, se generaron 345296 registros.

4-Aplicar Minería de Reglas de Asociación

Para descubrir los pares de códigos Dewey que constituyen reglas de asociación que provean conocimiento para la toma de decisiones, se realizaron dos etapas: Descubrimiento de reglas de asociación y Evaluación de las reglas de asociación generadas.

a) Descubrimiento de reglas de asociación: Entre los pares de códigos Dewey construidos en la etapa de preprocesamiento, habían pares de códigos coincidentes, y/o pares que ocurrieron desde una sola vez en todas las transacciones. Estas dos situaciones llevaron a buscar respuestas al siguiente interrogante:

¿Pueden ser considerados reglas de asociación que apoyen la toma de decisión, todos los pares de códigos Dewey generados?

Dado que $X \cap Y = \phi$, sólo deberían tenerse en cuenta pares de códigos Dewey no coincidentes. Asimismo, y dado que se detectaron pares de códigos que habían ocurrido en un número muy reducido de transacciones, se decidió solo evaluar aquellas reglas de asociación surgidas de pares de códigos Dewey que ocurrieron *una cantidad significativa de veces*. Para establecer esta cantidad, se computó el promedio de transacciones diarias realizadas en la biblioteca durante los 3 años tratados, el cual fue 101. A partir de ello, se consideraron sólo los pares de códigos Dewey presentes al menos en 101 transacciones. Esto podría interpretarse como que el bibliotecario accede a las estanterías por pares de libros identificados por los mismos códigos Dewey 101 veces como mínimo en un mismo día (fecha de evaluación parcial o final de una asignatura, en la que los libros solicitados integrarían su listado bibliográfico). La cercanía física de los libros solicitados podría ayudar a los bibliotecarios a optimizar el servicio que brindan.

b) Evaluación de reglas de asociación: Se adoptó un marco de evaluación apoyado en las medidas soporte, confianza y *lift*. Para la estimación de las probabilidades en las expresiones (1), (2), (3) y (4), siendo T el conjunto de transacciones, se consideró:

$n(X)$: Número de transacciones de T que contienen al conjunto X

$n(Y)$: Número de transacciones de T que contienen al conjunto Y

$n(X \cup Y)$: Número de transacciones de T que contienen al conjunto $X \cup Y$

N : Número total de transacciones, esto es cardinal del conjunto T : $|T|$

$P(X)$: Frecuencia relativa asociada a X , obtenida como $P(X) = \frac{n(X)}{N}$

$P(Y)$: Frecuencia relativa asociada a Y , obtenida como $P(Y) = \frac{n(Y)}{N}$

$P(X \cup Y) = \frac{n(X \cup Y)}{N}$

$$\text{soporte}(X \Rightarrow Y) = P(X \cup Y) = \frac{n(X \cup Y)}{N} \quad (5)$$

$$\text{confianza}(X \Rightarrow Y) = \frac{P(X \cup Y)}{P(X)} = \frac{n(X \cup Y)}{n(X)} \quad (6)$$

$$\text{lift}(X \Rightarrow Y) = \text{lift}(Y \Rightarrow X) = \frac{n(X \cup Y) N}{n(X) n(Y)} \quad (7)$$

Las expresiones (5), (6) y (7) se aplicaron para evaluar las reglas descubiertas.

Resultados y discusión

Los modelos obtenidos a partir de la aplicación de MRA, deberían asistir en el proceso de toma de decisiones. Los usuarios de esta tecnología, bibliotecarios en este caso particular, no pueden resumir, interpretar y usar para una exitosa toma de decisiones, cientos de datos numéricos –valores de soporte, confianza y *lift* correspondientes a las reglas generadas. Presentar esa información abstracta por medio de esquemas 2D, incrementa la posibilidad de asimilar las distintas dimensiones de los datos en una forma amplia e inmediatamente comprensible, promoviendo que estos usuarios de la tecnología se transformen efectivamente en beneficiarios.

En la Figura 1, y con asistencia de la Tabla 3, puede observarse que los pares de códigos Dewey 574-551 o 551-574, 540-530.02 o 530.02-540 y 551-540 540-551, correspondientes a, “Biología- Geología, Hidrología, meteorología”, “Química y ciencias afines.- Física Manuales y Geología”, “Hidrología, meteorología - Química y ciencias afines”, son todos estadísticamente independientes entre sí dado que el valor *lift* de las reglas en las que están involucrados es menor que 1. El resto de los pares presentan dependencia.

El valor de confianza para la regla $540.02 \Rightarrow 540$ es mayor que el de la regla $540 \Rightarrow 540.02$. Como el código 540.02 denota "Manuales de Química" y el 540 a "Química y Ciencias Afines", la regla $540.02 \Rightarrow 540$ puede leerse de la siguiente manera: si los "Manuales de Química" fueron pedidos en una transacción, los libros de "Química y Ciencias afines" también fueron pedidos en la misma transacción con un soporte de 0,022, una confianza de 0,242 y *lift* 2,395. El valor de soporte expresa que en el 2% de las transacciones estuvieron involucrados libros con los códigos 540.02 y 540. Según el valor de confianza, aproximadamente el 24% de las transacciones que contienen el código 540.02, también contienen el código 540. Además entre estos códigos existe co-ocurrencia, *lift* es mayor que 1, con una diferencia entre los valores de confianza computados para las reglas que los involucra de alrededor del 2%.

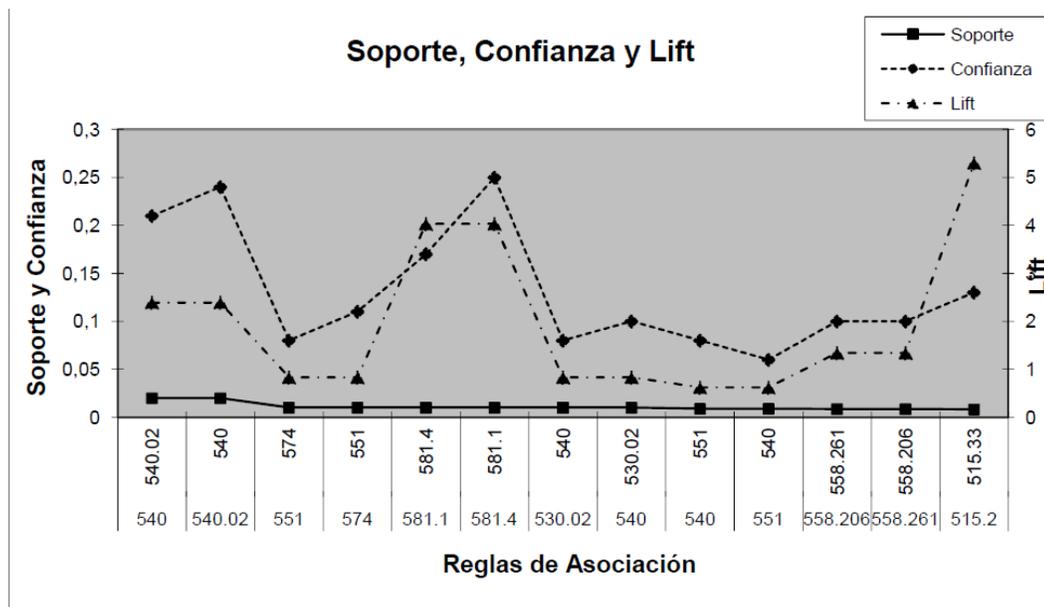


Figura 1. Reglas de Asociación para Espacios Físicos Cerrados con sus valores de soporte, confianza y *lift*.

Tabla 3. Descripción de Reglas de Asociación presentadas en Figura 1.

Antec.	Consec.	Descip. Antec.	Descip. Consec.
540	540.02	Química y ciencias afines.	Química- Manuales
540.02	540	Química- Manuales	Química y ciencias afines.
551	574	Geología, hidrología, meteorología	Biología.
574	551	Biología.	Geología, hidrología, meteorología.
581.1	581.4	Fisiología vegetal	Botánica. Anatomía y Morfología
581.4	581.1	Botánica. Anatomía y Morfología	Fisiología vegetal
530.02	540	Física-Manuales	Química y ciencias afines.

540	530.02	Química y ciencias afines.	Física-Manuales
540	551	Química y ciencias afines.	Geología, hidrología, meteorología.
551	540	Geología, hidrología, meteorología.	Química y ciencias afines.
558.206	558.261	Geología Argentina-Congresos	Geología Argentina-San Juan
558.261	558.206	Geología Argentina-San Juan	Geología Argentina-Congresos
515.2	515.33	Análisis. Aspectos generales	Análisis. Cálculo diferencial

Conclusiones

La adecuación a la problemática abordada, del marco teórico adoptado para la MRA – concepto de transacción, marco de evaluación para las reglas generadas entre otros-, es el primer paso que debe realizarse cuando se decide encarar la búsqueda de conocimiento a partir de los datos. De los cinco pasos del proceso de MRA aplicado, especial atención se les dedicó a los tres últimos:

-Realizar preprocesamiento: De la observación directa de las acciones realizadas por los bibliotecarios de la Biblioteca de la FCFN para satisfacer los requerimientos de los socios, y del planteo y análisis del funcionamiento de los 2 escenarios tratados, destinados a contener el material librario, se reconoció que los tipos de movimientos de libros y combinaciones de los mismos también debían ser considerados.

-Aplicar Minería de Reglas de Asociación: Si bien existen entornos de MD que proveen algoritmos para descubrir reglas de asociación, tal es el caso de Weka, RapidMiner y R, estos no brindan la posibilidad de tratar situaciones particulares, como en el caso en el que tanto los antecedentes como los consecuentes de las reglas de asociación candidatas a integrar el modelo están constituidos por categorías de materiales ponderados por el tipo de movimiento. A la vez, los entornos mencionados son restringidos en cuanto a permitir aplicar diferentes factores simultáneamente para evaluar las reglas generadas.

Con el propósito de que los bibliotecarios se vean favorecidos con el uso de esta tecnología, y a la luz de las características particulares de una amplia gama de factores destinados a valorar las reglas de asociación, en este trabajo la evaluación de las reglas se realizó en el marco soporte, confianza y *lift*. Asimismo, y dado que no existe un rango de valores prefijado de *lift*, las reglas resultantes podrían también clasificarse según valores decrecientes de esta medida.

-Interpretar el modelo y obtener conclusiones: Las reglas de asociación obtenidas son presentadas en esquemas 2D, para facilitarle al bibliotecario una selección visual e inmediata de las reglas más relevantes.

A partir del análisis de uso del material librario disponible, los resultados alcanzados asisten al personal de la biblioteca al momento de planificar y decidir sobre la distribución física de los libros en las estanterías.

En el caso de las estanterías de acceso restringido, esto favorece a una mejora en los tiempos de respuesta de los bibliotecarios ante solicitudes de libros. Además de lo expresado, en un escenario con estanterías de libre acceso, la ubicación física de los libros promueve la difusión y revisión de materiales que están en una ubicación próxima a los libros que en un comienzo un socio solicitaría, pudiendo de esta forma evitarle una nueva visita a la biblioteca al apoyarlo, de manera silenciosa, en sus posibles requerimientos de información.

Agradecimientos

Agradecemos la colaboración prestada por María José Marcovecchio, docente e investigadora de la Universidad Nacional de San Juan, en la revisión y confección del abstract.

Referencias

- AGRAWAL, R., IMIELINSKI, T., SWAMI, A. Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD International Conference on Management of Data. 1993. Washington D.C., p. 207-216.
- BROWN, Meta S. *Data mining for dummies*. John Wiley & Sons, 2014.
- BRIN, S., MOTWANI, R., ULLMAN, J. D., TSUR, S. Dynamic itemset counting and implication rules for market basket data. Proceedings ACM SIGMOD International Conference on Management of Data. 1997. Tucson, Arizona, USA. p. 255-264
- CHEN, M., HAN, J. AND YU P. Data Mining: An Overview from a Database Perspective. IEEE Transactions on Knowledge and Data Engineering, 1996. V.8 N.6, p.866-883.
- CONEAU- <http://www.coneau.gov.ar/>
- HAHSLER, Michael; HORNIK, Kurt; REUTTERER, Thomas. Implications of probabilistic data modeling for mining association rules. En From Data and Information Analysis to Knowledge Engineering. Springer Berlin Heidelberg, 2006. p. 598-605.
- HERRERA VARELA, Ricardo. Bibliomining: minería de datos y descubrimiento de conocimiento en bases de datos aplicados al ámbito bibliotecario [en línea]. "Forinf@", vol. 33, 2006. [Consulta: 01/09/2014]. <<http://lemi.uc3m.es/est/forinf@/index.php/Forinfa/rt/printerFriendly/122/127>>

- HIPP, Jochen; GÜNTZER, Ulrich; NAKHAEIZADEH, Gholamreza. Algorithms for association rule mining—a general survey and comparison. *ACM sigkdd explorations newsletter*, 2000, vol. 2, no 1, p. 58-64.
- KANTARDZIC, M. *Data Mining: Concepts, Models, Methods, and Algorithm*. John Wiley & Sons .2003.343 pages.
- LENCA, P., MEYER, P., VAILLANT, B., LALLICH, S. On selecting interestingness measures for association rules: user oriented description and multiple criteria decision aid. [en línea]. *European Journal of Operational Research*. Volume 184, Issue 2, 16 January 2008, Pages 610–626. [Consulta: 01/03/2008]. <http://www.sciencedirect.com/science/article/pii/S0377221706011465>
- NICHOLSON, S. The basis for bibliomining: Frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services. [en línea]. *Information Processing & Management*. Volume 42, Issue 3, May 2006, Pages 785–804 [Consulta: 01/02/2008]. <http://www.sciencedirect.com/science/article/pii/S0306457305000658>
- LÓPEZ, Ana Pérez. La evaluación de colecciones: métodos y modelos. *Documentación de las Ciencias de la Información*, 2002, vol. 25, p. 321-360.
- LUCAS, Joel Pinho. *Métodos de clasificación basados en asociación aplicados a sistemas de recomendación*. 2010. Tesis Doctoral. Universidad de Salamanca.
- ROMERO, C. Aplicación de técnicas de adquisición de conocimiento para la mejora de cursos hipermedia adaptativos basados en Web. Tesis Doctoral.. Universidad de Granada. E.T.S. Ingeniería Informática. 2003.
- SAHU, Hemlata; SHRMA, Shalini; GONDHALAKAR, Seema. A Brief Overview on Data Mining Survey. *International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume*, 2011, vol.1.
- SILVERSTEIN, Craig; BRIN, Sergey; MOTWANI, Rajeev. Beyond market baskets: Generalizing association rules to dependence rules. *Data mining and knowledge discovery*, 1998, vol. 2, no 1, p. 39-68.
- UNIVERSIDAD DE BUENOS AIRES. SISTEMA DE BIBLIOTECAS Y DE INFORMACIÓN. COMISIÓN TÉCNICA DE ESTÁNDARES. *Estándares del Sistema de Bibliotecas de la Universidad de Buenos Aires*. SISBI, Universidad de Buenos Aires, Secretaría de Ciencia y Técnica, Sistema de Bibliotecas y de Información, 2013.
- VAILLANT, Benoît; LENCA, Philippe; LALLICH, Stéphane. A clustering of interestingness measures. *En Discovery Science*. Springer Berlin Heidelberg, 2004. p. 290-297.
- ZHAO, Yanchang; ZHANG, Chengqi; ZHANG, Shichao. Discovering interesting association rules by clustering. *En AI 2004: Advances in Artificial Intelligence*. Springer Berlin Heidelberg, 2005. p. 1055-1061.