

Reglas de asociación en el cómputo de utilización de libros en una biblioteca universitaria

(Association rules in computing the use of books in a university library)

María Alejandra Malberti Riveros¹, Raúl Oscar Klenzi¹

Resumen:

Este trabajo recrea una propuesta destinada a evaluar la utilización de diferentes códigos de libros en una biblioteca universitaria. A partir de los datos de circulación de materiales almacenados en el sistema de biblioteca, el modelo aplica un mecanismo de estadísticas de utilización y un mecanismo de descubrimiento de reglas de asociación. Las estadísticas de utilización son computadas en base al grado de importancia, o pertinencia, de cada código respecto a un área de conocimiento, mientras que las reglas de asociación brindan apoyo para determinar la utilización final de los diferentes códigos. En el proceso se tiene en cuenta que los datos almacenados en el sistema corresponden a ejemplares solicitados en calidad de préstamo, renovación o consulta. El estudio introduce una alternativa de descubrimiento de conocimiento en datos, tendiente a enriquecer la gestión de una biblioteca universitaria.

Palabras clave: Minería de Datos; Reglas de Asociación; Estadísticas de uso; Biblioteca Universitaria; Código Dewey

Abstract:

This work recreates a proposal to evaluate the usage of different books categories in a university library. The model employs a mechanism to carry out usage statistics and a mechanism to carry out discovering association rules, from use data stored in the library system. Usage statistics are computed based on the degree of importance, or relevance, with respect to an area of knowledge, and association rules provide support to determine the final use of the various categories. In the process we take into account that stored data correspond to books requested on loan, renewal or consultation. The study presents knowledge discovery in data, aiming to enhance the management of a university library.

Keywords: Data Mining; Association Rules; Usage statistics; University Library; Dewey code

1. Introducción

Una gran cantidad de datos acumulados en las bases de datos de organizaciones, pueden esconder conocimientos valiosos y útiles para la toma de decisiones. La Minería de Datos también conocida como Descubrimiento de Conocimiento en Bases de Datos –Knowledge Discovery in Databases: KDD-, significa un proceso de extracción no trivial, de información implícita, útil, previamente desconocida, desde los datos de una base de datos (Chen, 1996) (Agrawal, 1993). De acuerdo con el Gartner Group (Larose, D, 2005, p.12) "La minería de datos es el proceso de descubrir nuevas correlaciones significativas, patrones y tendencias, tamizando grandes cantidades de datos almacenados en depósitos, usando tecnologías de reconocimiento de

¹ Universidad Nacional de San Juan, San Juan – Argentina, ({amalberti, rklenzi}@unsj-cuim.edu.ar)

patrones, así como técnicas estadísticas y matemáticas." Dentro de ella, 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, 1996, p.7)(Hipp, 2000)(Hahsler, 2006).

La incorporación de esta tecnología en el ámbito de las bibliotecas ha comenzado a adquirir fuerza en los últimos años. Tanto es así que Nicholson define el término Bibliomining (Nicholson, 2006, p. 790) como "la combinación de minería de datos, bibliométricas, herramientas estadísticas, y de reportes, usadas para extraer patrones, basados en comportamientos, de los sistemas de biblioteca". Coincidente con esta idea la asociación ARL -Association of Research Libraries- presenta el trabajo realizado por Barbara Mento (Mento, 2003). Esta publicación resulta de una investigación cuyo objetivo era determinar el uso de la tecnología de minería de datos en las actividades realizadas por distintas bibliotecas. De 124 bibliotecas consultadas, respondieron a la investigación 65. Si bien al momento de la investigación sólo 26 usaban tecnología de minería de datos, las restantes consideraban a esta tecnología muy importante. De las 26 bibliotecas, 18 aplicaban tecnología de minería de datos ya sea para facilitar las tareas administrativas o con propósitos estratégicos. Algunas bibliotecas mencionaron beneficios en la esfera administrativa a partir de la minería de datos, ya que esta tecnología ayudó al desarrollo de la colección, al presupuesto, a las decisiones de desarrollo de la Web, a valorar bases de datos y otros recursos, a determinar las necesidades de los usuarios, a monitorear el desempeño y usabilidad del sistema, a realizar pronósticos y establecer políticas y a mejorar la seguridad de la Web. En las conclusiones, la autora manifiesta que "Las bibliotecas que están usando minería de datos lo están haciendo principalmente para propósitos administrativos tales como facilitar la recolección y el análisis de, por ejemplo, la circulación, la adquisición, el uso de Web, y otros patrones de datos diversos. El objetivo es generalmente reforzar la toma de decisión y las operaciones internas propias de la biblioteca. Mientras que las actividades más importantes de la minería de datos residen en departamentos académicos e involucran investigación académica, hay conciencia y entusiasmo por las posibilidades de la minería de datos como una herramienta para bibliotecas de ARL.". (Mento, 2003,p.13).

Por su parte Galluzzi (Galluzzi, 2014) expresa que los bibliotecarios deben conducir las bibliotecas hacia las direcciones más innovadoras que han surgido en los últimos años y seguir siendo abiertos al cambio y lo más flexible posible.

Sue McKnight describe en (Dale, 2011) la evolución de la biblioteca académica a través del tiempo, lo que demuestra la resiliencia y capacidad de adaptación de los servicios bibliotecarios y de gestión de bibliotecas destacando el importante papel que las bibliotecas académicas tienen en la creación y el apoyo a los entornos digitales de aprendizaje, y su tendencia a participar en ellos hoy y en el futuro.

Si bien existe una amplia variedad de obras que tratan sobre minería de datos, no ocurre lo mismo con bibliomining. Entre los trabajos que aplican minería de datos a los datos de una biblioteca se encuentra el modelo KDBM –Knowledge Discovery Based Model- (Wo, 2004), diseñado para descubrir conocimiento sobre la utilización de las distintas categorías de materiales de una biblioteca universitaria. Este modelo de asignación de presupuesto, se sustenta en la idea de que nuevas asignaciones presupuestarias deben incorporar el conocimiento referido a la utilización realizada anteriormente de los recursos disponibles en la biblioteca.

La Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de San Juan, FCEFN-UNSJ, y en particular su biblioteca, han transitado un proceso de evaluación institucional. La evaluación de las instituciones universitarias del sistema universitario argentino se apoya en los conceptos centrales que surgen del **artículo 44º de la Ley de Educación Superior**. En el documento Lineamientos para la Evaluación Institucional (Resolución CONEAU 094/97) se expone la concepción general de la evaluación institucional que posee la CONEAU-Comisión Nacional de Evaluación y Acreditación Universitaria-, el marco normativo y las características de las evaluaciones externas que se realizan. En el mencionado documento se expresa que “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.” (Lineamientos para la evaluación institucional, p. 20). “Toda evaluación, en cualquiera de sus fases, implica la utilización de medios cuantitativos y cualitativos, que son complementarios y de ninguna manera exclusivos ni aislados.” (Lineamientos para la evaluación institucional, p. 24). Además la Resolución CONEAU 315/00, que fija las pautas generales para la elaboración de los informes de evaluación externa, reconoce explícitamente para la variable Servicios de Biblioteca, de Información e Informáticos, las dimensiones: Presupuesto anual, Servicio e infraestructura, Catalogación y estadísticas, Circulación y préstamos, Informática entre otras.

Por su parte la Norma ISO 11620 -Norma internacional relacionada con la evaluación de bibliotecas de todo tipo- propone un conjunto de indicadores de desempeño entre los que se encuentran los relacionados con el uso, recuperación y préstamos de documentos. Para recabar los datos necesarios para el cómputo de esos indicadores la citada norma propone, explícitamente, la utilización de diferentes métodos entre los que cita el uso de los registros del sistema automatizado.

En conjunto el proceso de evaluación institucional, el marco regulatorio establecido por CONEAU, la Norma ISO-11620, los planes estratégicos, consideraciones y sugerencias realizadas por parte de la IFLA-International Federation of Library Associations and Institutions-, el reconocimiento del profesional bibliotecario como usuario-beneficiario de la tecnología de Minería de Datos en su actuación profesional y el modelo KDBM, han motivado al desarrollo de una propuesta referente

al cómputo de utilización efectiva de los materiales disponibles, con el propósito de asistir al proceso de gestión de los recursos de una biblioteca universitaria, especialmente en instancias en las que se necesitan datos objetivos que apoyen decisiones referentes a la adquisición de nuevos materiales, A la vez, se promueve la incorporación y el reconocimiento de las potencialidades de las nuevas tecnologías, como la Minería de Datos, en entornos no tradicionales tal es el caso de bibliotecas.

En la sección 2 se describe el proceso realizado para el cómputo de utilización de cada categoría de libros, o códigos Dewey², en base a su estadística de utilización y a su utilización conjunta con otras categorías. En la sección 3 se discuten los resultados obtenidos y en la sección 4 se presentan las conclusiones.

2. Materiales y Métodos

Dentro de la minería de datos, la 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, 1996)(Hipp, 2000)(Hahsler, 2006).

La biblioteca de la FCFN-UNSJ registra en el sistema MicroISIS todos los movimientos (préstamos, devoluciones, renovaciones y consultas) de su material librario. En esta biblioteca se aplica el Sistema de Clasificación Decimal Dewey para la catalogación de su colección. A partir de considerar que cada transacción registrada en el sistema, 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 de reglas de asociación al ámbito de la citada biblioteca:

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

$T = \{t_1, t_2, \dots, t_n\}$ 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.

² La Clasificación Decimal Dewey–DDC, es un sistema desarrollado por Melvil Dewey en 1876, que intenta organizar todo el conocimiento en diez clases principales. Cada clase principal tiene diez divisiones, y cada división tiene diez secciones, por lo que el sistema puede ser resumido en 10 clases principales, 100 divisiones y 1000 secciones. Las obras son clasificadas principalmente por el tema al que pertenecen, produciendo números de clasificación no menores de tres dígitos; mientras más dígitos integran un determinado código, mas especificidad se logra en la catalogación.

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 el marco soporte-confianza (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.

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

En nuestro trabajo, el conjunto está formado por los códigos Dewey que conforman el antecedente y/o el consecuente de la regla de asociación:

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

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 .

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

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 $P(Y/X)$.

El modelo KDBM, que consiste en una propuesta para evaluar la utilización de las diferentes categorías de materiales en una biblioteca universitaria, incorpora un mecanismo de estadísticas de circulación y un mecanismo de descubrimiento de reglas de asociación. Las estadísticas de circulación son computadas en base a la "fuerza semántica", o pertinencia, de las categorías de materiales solicitados por los socios de la biblioteca, respecto a los departamentos involucrados en los préstamos. Las reglas de asociación, que reflejan la correlación entre pares de categorías de materiales, brindan el apoyo adecuado para determinar la importancia de las categorías utilizadas. En este contexto, una regla de asociación toma la forma $A \Rightarrow B(\alpha, \beta)$, donde A y B son

categorías de materiales y α y β son soporte y confianza respectivamente. Por ejemplo, la regla de asociación “Sistemas \Rightarrow Organizaciones y Administración (0.25, 0.33)”, es interpretada como: “ si materiales en la categoría Sistemas fueron pedidos en una transacción, materiales de Organizaciones y Administración fueron también pedidos en la misma transacción con un soporte de 0.25 y una confianza de 0.33” (Wo, 2004, p.714).

La utilización de cada categoría de material es computada por medio de la expresión:

$$MU(A) = n_A + \sum_1^k n_k * (\theta * soporte + \omega * confianza) \quad (1)$$

donde MU es utilización y A una categoría de material representada por su código Dewey

n_A : estadística de utilización de la categoría A

n_k : estadística de utilización de la k-ésima categoría que puede producir A – esto es $k \Rightarrow A$

θ : intensidad de soporte

soporte: valor de soporte de la regla $k \Rightarrow A$

ω : intensidad de confianza

confianza: valor de confianza de la regla $k \Rightarrow A$

A partir de los movimientos registrados en el sistema durante un año, y de (1), se calculó la utilización de cada código Dewey en base a su circulación individual y a su circulación conjunta con otros códigos.

Si bien la FCEF N tiene cuatro departamentos académicos: Geofísica-Astronomía, Biología, Informática, y Geología, en particular se seleccionaron los movimientos realizados por socios de la biblioteca vinculados al Departamento de Informática. Como resultado se obtuvieron 8649 registros concernientes a los préstamos, renovaciones, devoluciones y consultas de libros, realizadas por alumnos y docentes del citado departamento.

2.1. Determinación de pertinencia o fuerza semántica

Se confeccionaron 152 encuestas distintas destinadas a docentes y alumnos del último año de las carreras de cada uno de los 4 departamentos de la FCEF N, las que fueron distribuidas proporcionalmente según el uso que cada departamento realizó de la biblioteca durante un año. Esto llevó a destinar 54 encuestas para Biología, 15 para Geofísica, 40 para Geología y 43 para Informática - Tabla 1.

Tabla 1. Distribución de las encuestas entre referentes de distintos departamentos

	Departamentos Académicos								Total de encuestados
	Geofísica		Geología		Biología		Informática		
	Docente	Alumno	Docente	Alumno	Docente	Alumno	Docente	Alumno	
Cant.	6	9	12	28	16	38	14	29	
Total por Depart.	15		40		54		43		152

Asimismo, se consideraron los 302 códigos Dewey presentes en los movimientos registrados en el Sistema de Biblioteca durante el año tratado. A cada encuestado no se le presentaron códigos, sino diversos títulos de bibliografías catalogados bajo esos códigos. Cada encuesta contiene 20 títulos, los que multiplicados por las 152 encuestas permitieron obtener alrededor de 10 respuestas, provenientes de 10 encuestados distintos, por cada uno de los 302 códigos Dewey. Cabe mencionar que tanto la distribución de los títulos entre las encuestas, como de las encuestas entre los referentes de los departamentos, fue aleatoria. En la Figura 1 se muestra una de las encuestas respondida, en esta oportunidad, por un docente.

Encuesta N° 6

Departamento: de Informática

Autoridad:

Docente: Categoría: Adj. SE.

Alumno: Carrera: Lic. en Cs. de Informática Año que cursa: _____

Marque el grado de pertinencia o relación que para usted existe entre las siguientes bibliografías y el departamento en el que Ud. realiza sus actividades.

Bibliografía	Valoración		Pertinencia				
	Total	Alta	Media	Baja	Nula		
FILOSOFIA PRACTICA Y TEORIA DE LA HISTORIA						X	
ELECTRICIDAD Y MAGNETISMO						X	
GEOLOGIA DE AMERICA DEL SUR						X	
MANUAL DE PALEONTOLOGIA						X	
TEORIA Y PROBLEMAS DE MATRICES		X					
SEDIMENTOLOGIA						X	
ECOLOGIA						X	
CURSO DE PROSPECCION GRAVIMETRICA						X	
ACTAS DE LAS SEGUNDAS JORNADAS SOBRE GEOLOGIA DE LA PRECORDILLERA 1997						X	
MANUAL DE LA FLORA DE LOS ALREDEDORES DE BUENOS AIRES						X	
LISP			X				
PROGRAMACION DE ACTIVE SERVER PAGES		X					
TEORIA DE FUNCIONES DE VARIABLE COMPLEJA				X			
LOS ASPECTOS TEORICOS DEL ENLACE QUIMICO						X	
PALEOAMBIENTES SEDIMENTARIOS						X	
ROCAS SEDIMENTARIAS						X	
LA GEOLOGIA EN EL DESARROLLO DE LOS PUEBLOS						X	
GEOLOGIA DE SAN JUAN						X	
LOS GENEROS DE GRAMINEAS DE AMERICA AUSTRAL						X	
ECKERT FISILOGIA ANIMAL						X	

Para cada categoría coloque una X en el valor que considere adecuado

Figura 1. Encuesta respondida por un docente del Departamento de Informática

El procesamiento de las encuestas se realizó por medio de una planilla de cálculo. Las X se reemplazaron por los siguientes valores: columna Pertinencia Total por valor 0,4; columna Pertinencia Alta por valor 0,3; columna Pertinencia Media por valor 0,2; columna Pertinencia Baja por valor 0,1 y columna Pertinencia Nula por valor 0. Dado que diferentes títulos bibliográficos corresponden a un mismo código Dewey, el valor representativo de pertinencia de ese código para un área de conocimiento particular, se obtuvo mediante el cálculo de la mediana desde las respuestas obtenidas de referentes de ese área. Se computa la mediana, por representar los valores numéricos procesados valores semánticos inherentes a las pertinencias brindadas por cada encuestado.

Para el cómputo de la mediana se ordenaron los valores de pertinencia asociados a un código Dewey de mayor a menor, luego se podaron los extremos hasta alcanzar un único resultado representativo de ese código Dewey.

En la Tabla 2 se pueden apreciar las respuesta brindadas, para el Dewey 005.71, por encuestados pertenecientes a diversas áreas de conocimiento respecto de la pertinencia entre el título del libro y las carreras en las que el encuestado trabaja. La columna 2 tiene el número de encuesta en la que se ha obtenido la respuesta considerada, siendo la número 6 la presentada en la Figura 1.

En el caso del Dewey considerado, 005.71, el valor de pertinencia obtenido desde el cálculo de la mediana de las respuestas de referentes del área Informática es 0,3 (Alta).

2.2. Generación de estadísticas de utilización

El valor de estadística de utilización de cada código Dewey surgió de multiplicar la cantidad de veces que dicho código ocurrió en préstamos, renovaciones y consultas, por su fuerza semántica. Se computó n_A de (1), pero considerando los códigos Dewey distintos ocurridos en los préstamos, renovaciones y consultas durante el año tratado.

De los códigos analizados, 55 tenían valor cero para estadísticas de utilización. Asimismo cada uno de ellos había ocurrido entre 1 y 887 veces, pero en promedio la presencia de cada código en las solicitudes de libros realizadas en el año fue de 47 veces.

2.3. Generación de asociación de utilización

La asociación de utilización de un código Dewey toma en cuenta las reglas de asociación en las que dicho código participa como consecuente.

A partir de (1), la circulación conjunta se evaluó en base a las estadísticas de utilización de los códigos -k-, que pueden producir el código en cuestión -A, esto es reglas de asociación de la forma $k \Rightarrow A$. La circulación conjunta es la suma de las estadísticas de utilización de todos los

códigos k, ponderadas éstas por su soporte -que expresa la probabilidad de ocurrencia de la regla-, y confianza -que refiere a la probabilidad de ocurrencia del código A, habiendo ocurrido k. Si bien en (1) aparecen los factores θ - intensidad de soporte, y ω - intensidad de confianza, éstos refieren a valores establecidos por el usuario de la tecnología en base a su experiencia, pero en este caso ambos fueron omitidos pues no se disponía de información sobre los mismos.

Tabla 2. Respuestas obtenidas para Títulos asociados al código Dewey 005.71

Dewey	Nº Encuesta (área de conocimiento del encuestado)	Total	Alta	Media	Baja	Nula	Título
005.71	4 (Biología)						HTML DINAMICO, ASP Y JAVASCRIPT A TRAVES DE EJEMPLOS
005.71	6 (Informática)		X				PROGRAMACION DE ACTIVE SERVER PAGES
005.71	8 (Informática)		X				HTML DINAMICO, ASP Y JAVASCRIPT A TRAVES DE EJEMPLOS
005.71	10 (Geología)						PROGRAMACION DE ACTIVE SERVER PAGES
005.71	12 (Geofísica)				X		HTML DINAMICO, ASP Y JAVASCRIPT A TRAVES DE EJEMPLOS
005.71	14 (Geología)						PROGRAMACION DE ACTIVE SERVER PAGES
005.71	16 (Informática)	X					HTML DINAMICO, ASP Y JAVASCRIPT A TRAVES DE EJEMPLOS
005.71	18 (Geofísica)			X			PROGRAMACION DE ACTIVE SERVER PAGES
005.71	20 (Geología)						HTML DINAMICO, ASP Y JAVASCRIPT A TRAVES DE EJEMPLOS
005.71	22 (Informática)		X				PROGRAMACION DE ACTIVE SERVER PAGES

Cada movimiento registrado en el sistema corresponde a un préstamo, renovación, devolución o consulta de un libro por parte de un socio de la biblioteca. Ahora bien, si cada regla de asociación tiene dos códigos Dewey, correspondientes al antecedente y al consecuente, surge entonces el siguiente interrogante ¿Son las dieciséis combinaciones de movimientos de ejemplares, convenientes para el análisis propuesto? , en otras palabras ¿Qué combinaciones de movimientos son significativas al momento de seleccionar pares de códigos Dewey que representen posibles

reglas de asociación relevantes? De las dieciséis alternativas posibles de pares, las combinaciones de movimientos relevantes aceptadas en este trabajo para el cómputo de la asociación de utilización son las marcadas con X en la Tabla 3. Esta elección se basó en la observación directa de las acciones concretas que el personal de la biblioteca realiza al atender las solicitudes conjuntas de libros efectuadas por cada socio al momento de asistir a la biblioteca.

Tabla 3. Combinaciones de movimientos reconocidas en las reglas de asociación

Combinaciones de movimientos posibles	Combinaciones de movimientos aceptadas
Devolución, Préstamo	
Devolución, Renovación	
Devolución, Consulta	
Préstamo, Devolución	
Préstamo, Renovación	X
Préstamo, Consulta	X
Renovación, Devolución	
Renovación, Préstamo	X
Renovación, Consulta	X
Consulta, Devolución	
Consulta, Préstamo	X
Consulta, Renovación	X
Devolución, Devolución	
Préstamo, Préstamo	X
Renovación, Renovación	X
Consulta, Consulta	X

La tarea encarada requería no solo reconocer reglas de asociación conformadas por códigos Dewey sino que, además, tanto el antecedente como el consecuente de cada una de las reglas debían estar involucrados en los pares de movimientos de materiales previamente aceptados - Tabla 3-.

Si bien Weka, R y Rapidminer proveen algoritmos de descubrimiento de reglas de asociación, ninguno de ellos contempla esta situación. Es así que, por medio de SQL los 8649 registros almacenados en el sistema se agruparon en 1974 transacciones, a partir de las que se construyeron 5492 reglas de asociación.

3. Resultados y discusión

Si bien en el sistema de biblioteca estaba registrado el uso de 302 códigos Dewey, solo 164 de ellos correspondían a movimientos realizados por socios pertenecientes al Departamento de Informática. En esta instancia se computó la utilización de esos 164 códigos Dewey, a partir de su

estadística de utilización y de su asociación de utilización. La relación entre los valores computados de utilización, estadística de utilización y asociación de utilización, para 53 de los 164 códigos es mostrado en la Figura 2.

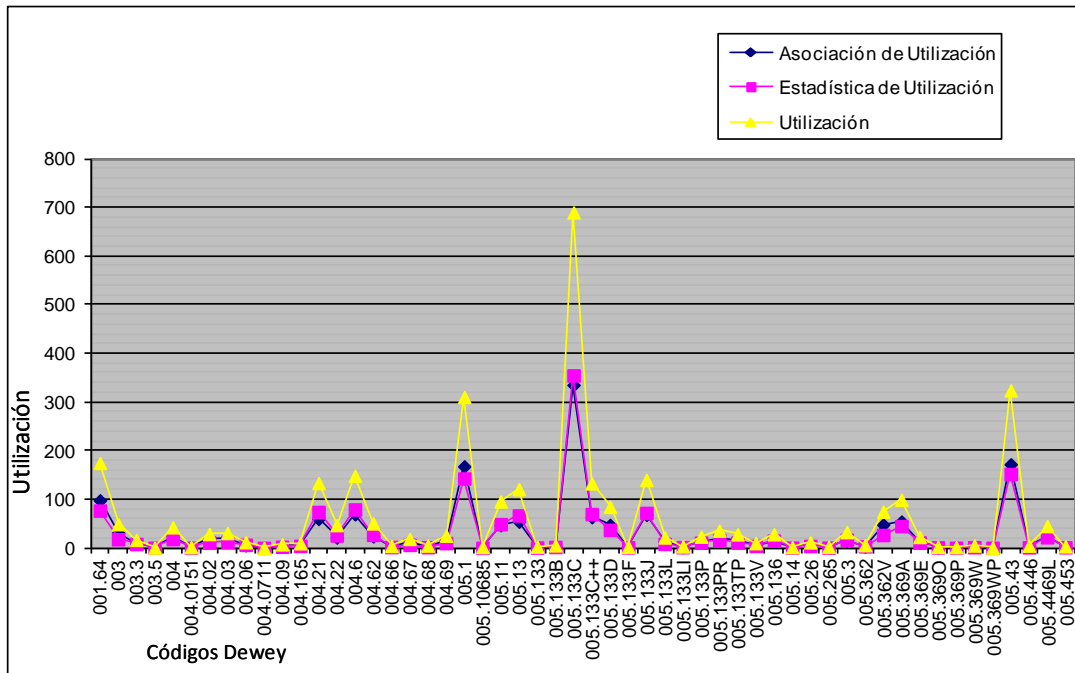


Figura 2. Utilización - Estadística de Utilización - Asociación de Utilización

Por ejemplo, la estadística de utilización, n_A de (1), de los códigos 005.1, 005.133C y 005.43 es 142.8, 354.79 y 152; mientras que su asociación de utilización, n_k de (1), es 167.71, 335.22 y 171.93 respectivamente. La descripción de cada código Dewey contenido en la Figura 2 es presentada en la Tabla 4, conjuntamente con su valor de utilización - MU de (1)-, y el porcentaje de asociación de utilización computado.

Tabla 4. Descripción-Utilización -% Asociación de Utilización de Figura 2

Dewey	Descripción	Utilización	% Asociación de Utilización respecto a Utilización
001.64	Conocimiento. Procesamiento digital	174,63	56,02%
003	Sistemas. Investigación operativa	48,96	63,44%
003.3	Modelos y simulación por computador	16,72	51,57%
003.5	Teoría de la comunicación y control	0,80	50,20%
004	Ciencias de la Computación. Procesamiento de Datos.	42,54	56,98%
004.0151	Procesamiento electrónico de datos	1,57	80,85%
004.02	Informática. Manuales	28,63	61,23%
004.03	Informática.Diccionarios	30,71	59,62%
004.06	Análisis y diseño de sistemas, arquitectura del computador, evaluación de desempeño	11,77	45,63%

004.0711	<i>Informática.Enseñanza. Universidades.Planes de estudio</i>	0,20	0,00%
004.09	<i>Computadoras. Historia</i>	7,48	67,89%
Dewey	Descripción	Utilización	% Asociación de Utilización respecto a Utilización
004.165	<i>Microcomputadores digitales específicos</i>	9,71	56,76%
004.21	<i>Análisis y diseño de sistemas</i>	133,50	44,57%
004.22	<i>Arquitectura del computador</i>	47,07	45,62%
004.6	<i>Interfases y comunicaciones</i>	148,29	46,59%
004.62	<i>Protocolos para interfases y comunicaciones</i>	50,54	46,97%
004.66	<i>Modalidades de transmisión de datos y métodos de transmisión de datos</i>	4,28	57,92%
004.67	<i>Redes extensas -Wan</i>	18,56	66,06%
004.68	<i>Redes locales- Lan</i>	4,91	67,40%
004.69	<i>Clases específicas de comunicaciones por computador</i>	25,25	60,39%
005.1	Programación	310,51	54,01%
005.10685	<i>Ingeniería de software</i>	1,92	89,56%
005.11	<i>Técnicas especiales de programación</i>	95,50	48,90%
005.13	<i>Lenguajes de programación</i>	120,81	44,71%
005.133	<i>Lenguajes específicos de programación</i>	2,76	70,99%
005.133B	<i>Lenguajes específicos de programación-Basic</i>	5,33	62,51%
005.133C	Lenguajes específicos de programación-C	690,01	48,58%
005.133C++	<i>Lenguajes específicos de programación-C++</i>	132,87	47,62%
005.133D	<i>Lenguajes específicos de programación-Delphi</i>	84,52	55,99%
005.133F	<i>Lenguajes específicos de programación-Fortran</i>	2,40	50,08%
005.133J	<i>Lenguajes específicos de programación-Java</i>	139,92	48,54%
005.133L	<i>Lenguajes específicos de programación-Lisp</i>	21,60	61,11%
005.133LI	<i>Lenguajes específicos de programación-Lisp</i>	3,25	50,77%
005.133P	<i>Lenguajes específicos de programación-Pascal</i>	23,23	53,51%
005.133PR	<i>Lenguajes específicos de programación-Prolog</i>	35,61	55,35%
005.133TP	<i>Lenguajes específicos de programación-Turbo Pascal</i>	28,02	60,03%
005.133V	<i>Lenguajes específicos de programación-Visual Basic</i>	9,95	57,80%
005.136	<i>Lenguajes de máquina y lenguajes ensambladores</i>	28,15	46,71%
005.14	<i>Verificación, prueba, medición, depuración</i>	1,66	63,75%
005.26	<i>Programación para microcomputadores digitales</i>	11,55	65,36%
005.265	<i>Programación para computadores específicos</i>	2,17	72,38%
005.3	<i>Software y programas de aplicaciones</i>	32,55	48,39%

005.362	Programas en lenguajes específicos de programación	5,81	43,19%
Dewey	Descripción	Utilización	% Asociación de Utilización respecto a Utilización
005.362V	Programas en lenguajes específicos de programación-Visual Basic	74,50	64,29%
005.369A	Programas específicos-Access	99,24	54,65%
005.369E	Programas específicos-Excel	22,64	50,96%
005.369O	Programas específicos-Office	2,47	51,41%
005.369P	Programas específicos-Powerpoint	1,58	81,01%
005.369W	Programas específicos-Word	5,03	70,15%
005.369WP	Programas específicos-Word Perfect	0,68	70,43%
005.43	Programas de Sistemas. Sistemas operativos	323,93	53,08%
005.446	Sistemas operativos para microcomputadores digitales	4,80	75,00%
005.4469L	Sistemas operativos específicos-Linux	45,07	51,19%

Los tres códigos con mayor valor de utilización son 005.133C, 005.43 y 005.1, de los cuales los correspondientes a "Programas de sistemas. Sistemas operativos" y "Programación", tienen un porcentaje de asociación de utilización que supera el 50%. Asimismo, de los 53 códigos Dewey presentados, 39 tienen un porcentaje de asociación de utilización de más de 50%.

Tabla 5. Códigos Dewey con más de 1% de Utilización respecto al cómputo total

Dewey	Descripción	% Utilización respecto Utilización Total	% Asociación de Utilización
005.133C	Lenguajes específicos de programación-C	6,29%	48,58%
005.74	Archivos de datos y Base de datos	3,13%	50,26%
005.43	Programas de Sistemas. Sistemas operativos	2,95%	53,08%
005.1	Programación	2,83%	50,77%
005.73	Estructuras de datos	2,79%	44,42%
515.15	Análisis. Cálculo y geometría analítica	2,24%	44,76%
423.6	Diccionarios Inglés-Español	2,24%	35,78%
001.64	Conocimiento. Procesamiento digital	1,59%	56,02%
519.2	Probabilidades	1,42%	39,33%
004.6	Interfases y comunicaciones	1,35%	46,59%
005.133J	Lenguajes específicos de programación-Java	1,27%	48,54%
004.21	Análisis y diseño de sistemas	1,22%	44,57%
005.133C++	Lenguajes específicos de programación-C++	1,21%	47,62%
005.13	Lenguajes de programación	1,10%	44,71%

El cómputo acumulado de utilización de los 164 códigos arrojó el valor 10978. En la Tabla 5 se muestran los 14 códigos Dewey que tienen un porcentaje de utilización que supera el 1% de la

utilización total computada. De ellos, 12 presentan un porcentaje de asociación de utilización respecto a su utilización que supera el 40%.

Además se detectó que para 120 de los 164 códigos Dewey ocurridos efectivamente en los movimientos registrados durante el año tratado, la asociación de utilización representó más del 50% de su utilización.

Se evidencia entonces que existen materiales bibliográficos que individualmente no son muy requeridos, sin embargo su uso aumenta al ser solicitados conjuntamente con otros.

4. Conclusiones

Se han empleado esquemas 2D para facilitar al bibliotecario, considerado en esta oportunidad como usuario-beneficiario de la tecnología de descubrimiento de conocimiento a partir de los datos, la detección inmediata de las relaciones existentes entre utilización, estadísticas de utilización y asociación de utilización para las diferentes categorías de libros. Los códigos Dewey que computan mayor valor de utilización presentan altos valores de asociación de utilización, lo que significa que dichos códigos estuvieron asociados con otros en la mayoría de los movimientos de libros registrados en el sistema. Ello implica que la utilización de un código Dewey particular no es totalmente dependiente de su estadística de utilización, sino que también depende de su co-ocurrencia con otros códigos Dewey. Se propone incorporar este conocimiento en instancias de gestión de la biblioteca, en especial al momento de establecer criterios de asignación de presupuesto destinado a la adquisición de nuevos ejemplares. En particular los valores obtenidos de utilización, estadísticas de utilización y asociación de utilización permiten objetivamente distribuir el presupuesto disponible para compras de libros entre distintos códigos Dewey. Esto representa un importante aporte a esta tarea realizada habitualmente por el bibliotecario responsable de la biblioteca de la FCEF, para la cual frecuentemente recurre a valoraciones subjetivas respecto al uso de los libros por parte de sus socios.

La distribución del material bibliográfico en definidos espacios físicos está expresada también por la codificación Dewey. Lo experimentado en la presente aplicación permitió, en determinados casos, proponer reubicación de material bibliográfico según el perfil propio y característico de la biblioteca de la FCEF sobre la que se realizó el estudio.

Con la intención de profundizar las tareas de bibliomining y en camino de futuras aplicaciones, en el ámbito de la biblioteca de la FCEF se está trabajando sobre una propuesta de automatización de obtención de código Dewey referidos a publicaciones en español, generadas en el ámbito de la propia Facultad o de la UNSJ en general. Hasta el momento la tarea de asignar su código Dewey a material bibliográfico que no lo posee, e inclusive escrito en diferentes idiomas, está a cargo de un integrante de la biblioteca de la FCEF quien desde la lectura del libro o publicación, define su

área temática de pertenencia y desde allí su correspondiente numeración. Similar objetivo se pretende, desde la aplicación automática de métricas de minería de texto, determinar la pertenencia y pertinencia de una titulación bibliográfica e índice asociado con los contenidos mínimos abarcados por las diferentes carreras que se dictan en el ámbito de la FCFN.

Un incipiente trabajo se está llevando adelante en lo referente a la definición y tratamiento de expurgo (no prestable) de material bibliográfico. Mediante la aplicación de algoritmos de segmentación se pretende llegar a determinar qué material bibliográfico está en condiciones de pasar a estado de expurgo y por ende la consideración de establecer por parte de la autoridad bibliotecaria la necesidad o no, de un nuevo lugar de almacenamiento. En el tratamiento de esta problemática también es tenido en cuenta el desarrollo realizado y los resultados obtenidos en el presente trabajo,

Bibliografía

Agrawal, R., Imielinski, T., Swami, A. (1993). Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD International Conference on Management of Data. Washington D.C., p. 207-216.

Association of Research Libraries (ARL)- <http://www.arl.org/>

Brin, S., Motwani, R., Ullman, J. D., Tsur, S. (1997). Dynamic itemset counting and implication rules for market basket data. Proceedings ACM SIGMOD International Conference on Management of Data. Tucson, Arizona, USA. p. 255-264.

Chen, M., Han, J. and Yu P. (1996). Data Mining: An Overview from a Database Perspective. IEEE Transactions on Knowledge and Data Engineering, V.8 N.6, p.866-883.

CONEAU- <http://www.coneau.gov.ar/>

Dale, P., Beard, J., & Holland, M. (2011). University Libraries and Digital Learning Environments.

Galluzzi, A. (2014). Libraries and Public Perception: A Comparative Analysis of the European Press. Elsevier.

Hahsler, Michael; Hornik, Kurt; Reutterer, Thomas. (2006). Implications of probabilistic data modeling for mining association rules. En From Data and Information Analysis to Knowledge Engineering. Springer Berlin Heidelberg, p. 598-605.

Hipp, Jochen; güntzer, Ulrich; Nakhaeizadeh, Gholamreza. (2000). Algorithms for association rule mining—a general survey and comparison. ACM sigkdd explorations newsletter, vol. 2, no 1, p. 58-64.

- IFLA -The International Federation of Library Associations and Institutions-
<http://www.ifla.org/III/index.htm>
- Larose, D. (2005). *Discovering Knowledge In Data. An Introduction To Data Mining*. Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
- Lineamientos para la Evaluación Institucional- Aprobado por Resolución N° 094-CONEAU-1997.
<http://www.coneau.gov.ar/archivos/482.pdf>
- Liu, S. T., & Chuang, M. (2009). Fuzzy efficiency measures in fuzzy DEA/AR with application to university libraries. *Expert Systems with Applications*, 36(2), 1105-1113.
- Mento, Bárbara. (2003). *Data Mining and Data Warehousing*. SPEC KIT 274. ARL. 77 pp. [Consulta: 01/02/2013]. <http://old.arl.org/bm~doc/spec274webbook.pdf>
- Nicholson, S. (2006). 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, Pages 785–804 [Consulta: 01/03/2013].
<http://www.sciencedirect.com/science/article/pii/S0306457305000658>
- Norma ISO 11620- <http://redc.revistas.csic.es/index.php/redc/article/viewFile/503/557>
- Resolución CONEAU 315/00- <http://www.coneau.edu.ar/archivos/485.pdf>
- Silverstein, Craig; Brin, Sergey; Motwani, Rajeev. (1998), Beyond market baskets: Generalizing association rules to dependence rules. *Data mining and knowledge discovery*, vol. 2, no 1, p. 39-68.
- Wu, C.; Lee, T. and Kao, S. (2004). *Knowledge discovery applied to material acquisitions for libraries*. *Information Processing and Management: an International Journal*, Vol. 40 Issue 4, Pergamon Press, Inc.