

Minería de datos en bibliotecas: bibliominería

[[Versió catalana](#)]

JORGE CANDÁS ROMERO

Alumno del Master en Investigación en Documentación
de la Universidad Carlos III de Madrid.

jorgecandas@gmail.com

Opciones

[Imprimir](#) [Recomanar](#) [Citació](#) [Estadístiques](#) [Metadades](#)

Resumen [[Abstract](#)] [[Resum](#)]

Se presenta una introducción teórica a la aplicación de la minería de datos en bibliotecas, denominada bibliominería (propuesta terminológica en español para el inglés *bibliomining*). Asimismo, se presentan algunas de las posibles aplicaciones prácticas y cómo éstas sirven de apoyo a la llamada Biblioteca 2.0 y a la creación y gestión de servicios más y mejor orientados al usuario, basados en nuevas tecnologías. Finalmente se analiza el problema de la privacidad en la aplicación de la bibliominería.

*I never waste memory on things that can easily
be stored and retrieved from elsewhere.*

Albert Einstein (1879-1955)¹

1 Minería de datos: conceptualización y utilidad

La minería de datos es una actividad que está a medio camino entre la Informática, la Estadística y la Documentación, y que se ha estado utilizando en numerosas disciplinas para el análisis de grandes cantidades de datos. En bibliotecas, su práctica está poco extendida a pesar las múltiples aplicaciones que tiene.

El objetivo de este artículo es hacer una introducción teórica a la minería de datos para profundizar después en su aplicación en bibliotecas, presentándose éstas como un campo ideal de trabajo. De esta aplicación, denominada en inglés *bibliomining* (término para cuya traducción al español se propone emplear "bibliominería"), se explican brevemente algunos de los usos prácticos, que van desde la evaluación de bibliotecas, la planificación de adquisiciones, la organización de la colección y los estudios de usuarios. Otra de las aplicaciones de mayor interés es la elaboración de sistemas de recomendaciones tal y como se vienen aplicando en servicios web, y que entra en el campo de la llamada Biblioteca 2.0 (en inglés *Library 2.0*).

La bibliominería, al igual que cualquier actividad de minería de datos, entraña ciertos problemas respecto a la privacidad. Aquí se presentan algunas de las posibles medidas a tomar al respecto.

1.1 Conceptualización

El gran crecimiento de las bases de datos y el aumento de las capacidades de almacenamiento de información, han hecho que todo tipo de organizaciones puedan disponer de una gran cantidad y variedad de datos relativos a su actividad diaria. En muchas de estas organizaciones se han dado cuenta del potencial que tiene esta información para el apoyo a la gestión. Su estudio permite ver la evolución y desarrollo de las organizaciones, y por lo tanto, trazar una línea de tendencia que muestre por dónde pueden moverse en un futuro.

Así, el estudio de los datos y la información almacenados en las bases de datos ofrece una visión perspectiva (qué se está haciendo y cómo se está haciendo) y prospectiva (cómo puede evolucionar la organización en un futuro a corto-medio plazo) de la organización, y es por ello por lo que tiene una función de apoyo a la toma de decisiones.

Este estudio de las bases de datos ha sido, y es, mucho más común en organizaciones económicas y empresariales, desde supermercados hasta grandes multinacionales, pero también organismos científicos que manejan grandes cantidades de información han visto la utilidad de este tipo de estudios.

Debido a que la información en las bases de datos está más desestructurada (siguiendo el modelo DIKW)² que en otras fuentes de información, para su utilización es necesario un proceso de tratamiento y análisis exhaustivo. Estas tareas de análisis de la información de las bases de datos se denominan minería de datos (en inglés *data mining*) o descubrimiento de información en bases de datos (en inglés *knowledge discovery in databases* o KDD)³. Bajo esta perspectiva Hand, Mannila y Smyth definen minería de datos como "la ciencia de extracción de información útil de grandes conjuntos de datos o de bases de datos" (Hand et al., 2001). Para algunos autores minería de datos y KDD se refieren a lo mismo, mientras que para otros KDD se refiere a un proceso de varias fases, entre las que se encuentra la minería de datos (Hernández et al., 2004).

Tal como señalan Fayyad, Piatetsky-Shapiro y Smyth, la gran cantidad de datos que se almacenan en las

organizaciones hace imposible la utilización de métodos manuales para su análisis (Fayyad et al., 1996). Por ello son necesarias técnicas y herramientas informáticas capaces de ayudar al hombre de forma "inteligente" y "automática" en el análisis de grandes cantidades de datos.

Estos mismos autores aportan la definición de minería más citada en la literatura sobre el tema: "minería de datos es un proceso no trivial de identificación de patrones de datos válidos, nuevos, potencialmente usables y comprensibles". Es decir, se trata de un proceso concreto, específico, con un objetivo, que busca identificar repeticiones y/o tendencias en un conjunto de datos que resulten útiles y sean veraces. Para Hernández Orallo, Ramírez Quintana y Ferri Ramírez (Hernández et al., 2004), en esta definición se resumen las propiedades del conocimiento extraído: válido, novedoso, potencialmente útil y comprensible.

Este proceso de minería de datos, Fayyad, Piatetsky-Shapiro y Smyth (Fayyad et al., 1996) lo descomponen en nueve fases lineales e iterativas, del mismo modo, Kudyba y Hoptroff lo hacen desde el punto de vista empresarial (Kudyba; Hoptroff, 2001). Estas fases, aunque con distinta denominación, consisten en tres operaciones: definición de la información que se pretende obtener, toma y procesamiento de datos, análisis de los resultados, obtención y consolidación de nueva información.

En el campo de las bibliotecas, Guenther, basándose en Brethenoux y Strange, simplifica las fases resumiéndolas en tres (Guenther, 2000):

- Selección y adquisición de datos: definir qué datos se quieren recopilar y qué método o procedimiento se va a usar para ello.
- Preparación y proceso de datos.
- Interpretación e integración (de los resultados).

En cualquiera de todos estos pasos, y en cualquiera de las propuestas consultadas, el proceso fundamental es el de la selección de los datos. Si el conjunto de datos (o bases de datos) a utilizar no es el adecuado para el tipo de análisis que se pretende llevar a cabo, el éxito del proceso de minería de datos se verá frustrado.

El proceso de la minería de datos es similar al proceso de elaboración de una estadística, que para Montaña (Montaña Lacambra, 2005) pasa también por tres fases: la recopilación de los datos, su procesamiento y la difusión del resultado final. Si bien el producto final de las estadísticas es una de las materias primas de la minería de datos.

El primer impulso importante para el desarrollo de esta disciplina se dio en *The First International Conference on Knowledge Discovery and Data Mining (KDD-95)* dentro de la *Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, coordinada por Usama Fayyad y Ramasamy Uthurusamy.

Por ese entonces el propio Fayyad, junto con Simoudis, definían minería de datos (aunque predominaba el uso de la expresión *knowledge discovery in databases*) como "un campo emergente que combina técnicas de aprendizaje-máquina, reconocimiento de patrones, estadística, bases de datos y visualización para extraer automáticamente conceptos, conceptos interrelacionados, y patrones de interés desde grandes bases de datos. Su tarea principal es la extracción de conocimiento (o información) de datos de bajo nivel (bases de datos)" (Fayyad; Simoudis, 1995).

Así pues, la minería de datos es un campo inicialmente relacionado con la Inteligencia Artificial y la Estadística, en el que la Documentación y demás Ciencias de la Información han participado posteriormente, desde el punto de vista de la administración y gestión (*management*) de centros de información, y como parte de las tareas de gestión de la información y del conocimiento. De este modo, la minería de datos supone una reivindicación del valor de las fuentes de datos estadísticas para la gestión.

1.2 Utilidad de la minería de datos

Actualmente, la minería de datos es una actividad en expansión aplicada cada vez en más disciplinas, que han visto la utilidad del estudio de datos para apoyar la toma de decisiones. Especial énfasis está teniendo en aquellas áreas relacionadas con la Economía, mediante la Econometría, y dentro del campo de inteligencia de los negocios.

El caso más tratado (y tomado como referencia para explicar diferentes tipos de algoritmos de minería de datos) es el de los supermercados (Agrawal; Imielinski; Swami, 1993) (Brin; Motwani; Ullman; Tsur, 1997) (Chi-Wing Wong; Wai-Chee Fu; Wang, 2005) (Hernández et al., 2004), ya que conocer el perfil de compra de los clientes puede ayudar a organizar el stock. De un modo similar, aunque más exhaustiva, es su utilización en comercio electrónico (Kohavi, 2001) (Kohavi; Provost, 2001), ya que en este caso es posible obtener mucha más información del comportamiento global del cliente, y no sólo conocer las compras. A su vez, y en el paso anterior en la cadena de producción, los proveedores y distribuidores pueden planificar la producción conforme a la demanda. Sin embargo, aunque eficiente, este modelo de actuación muestra su debilidad en momentos de excepción, como los casos de supermercados desabastecidos por no disponer de stock durante huelgas, o la incapacidad de hacer frente a grandes demandas en momentos de crisis (equipamiento y material tras el paso del huracán Katrina o el hundimiento del petrolero Prestige).

También relacionados con la Economía, son los estudios para evitar pérdidas de clientes y fraudes (mediante la detección de cambios de comportamiento sospechosos) (Fawcett; Provost, 1997) (Cox; Eick; Wills; Brachman, 1997), así como la concesión de créditos rápidos y de seguros (Hernández et al., 2004). En todos estos casos se pueden realizar previsiones de riesgo mediante la asignación de los individuos a unos perfiles de los que se conoce un patrón de comportamiento, tras analizar unas variables clave.

Asimismo, en la realización de estudios de mercados puede ayudar a mostrar cambios en las tendencias de consumo a gran escala de un país.

En Medicina puede mostrar patrones entre enfermos de la misma dolencia, y ver la evolución de epidemias en grandes poblaciones. También en disciplinas como la Astronomía, la Bio-informática y la Genética ayuda a estudiar las grandes cantidades de datos con que se trabaja.

Otros campos concretos donde la integración de bases de datos y el desarrollo de actividades de minería de datos sería útil, son la posibilidad de compartir datos de investigación científica y de la administración pública para la seguridad ciudadana (Clifton; Doan; Elmagarmid; Kantarcioglu; Schadow; Suci; Vaidya, 2004).

En otro sector totalmente diferente, Ma (Ma; Liu; Wong; Yu; Lee, 2000) presentan una aplicación para el desarrollo de minería de datos en educación con la finalidad de localizar, de manera muy precisa, estudiantes que necesitan clases de apoyo de una determinada materia.

En bibliotecas, la minería de datos puede resultar igualmente muy útil para la toma de decisiones basadas en la evidencia, aún cuando se manejan cantidades de datos mucho menores (Guenther, 2000).

2 Bibliominería: aspectos teóricos

La aplicación de minería de datos en bibliotecas se denomina bibliominería⁴ (*bibliomining*). El término fue acuñado por Nicholson y Stanton (2003)⁵ como una derivación de los términos bibliometría (*bibliometrics*) y minería de datos (*data mining*), con el fin de favorecer la conceptualización y el trabajo de los investigadores del campo. Estos autores definen bibliominería como "la combinación de minería de datos, bibliometría, estadística y herramientas de elaboración de informes y extracción de patrones de comportamiento, basados en sistemas bibliotecarios".

Para Nicholson (2006) la utilización de un nuevo término está justificada porque el término *library* (o biblioteca) asociado a minería de datos se refiere principalmente al conjunto de algoritmos que utiliza el software, por lo que puede dar lugar a errores y dificultades en la descripción y búsqueda de información sobre el tema.

Si bien, aunque la conceptualización es reciente, la bibliominería es una actividad que se viene realizando desde finales de la década de 1990 (Banerjee, 1998). Un ejemplo son los casos (Cullen, 2005) de las bibliotecas de la University of Waterloo (Ontario, Canadá) y la University of Pennsylvania (Filadelfia, EE.UU.). Otro caso reciente (Geyer-Schultz et al., 2003) es el de la biblioteca de la Universität Karlsruhe (TH) (Karlsruhe, Alemania).

Como se ha señalado, la bibliominería suele relacionarse con la bibliometría, pues ambas tareas se ocupan del análisis y cruce de datos mediante técnicas estadísticas para descubrir y establecer patrones y tendencias en los datos como ayuda a la toma de decisiones (obtener el dato por el dato, la mera presentación de resultados es una tarea vacua e inútil). Si bien en bibliominería se trata de datos sobre la actuación previa a la utilización de la información, mientras que la bibliometría trabaja con los datos relativos a la información que ha sido finalmente usada.

Para Nicholson, el proceso de bibliominería está compuesto de seis fases (Nicholson, 2003), que a grandes rasgos coinciden con las propuestas sobre minería de datos indicadas anteriormente:

- Determinación de las áreas de interés.
- Identificación de fuentes de datos internas y externas.
- Recopilar, limpiar y hacer anónimos los datos en el *data warehouse* (almacén de datos).
- Selección de las herramientas de análisis apropiadas.
- Descubrimiento de patrones a través de la minería de datos y creación de informes con herramientas tradicionales de análisis.
- Análisis e implementación de los resultados.

Por fuentes de datos internas se entienden los datos generados por la propia biblioteca en el transcurso de su actividad, lo que se denominan datos observacionales (Hand et al., 2001), es decir, datos que han sido recopilados en la actividad diaria. Por fuentes de datos externas se entienden aquellos tomados de fuentes ajenas a la organización y que sirven para contextualizar los primeros, se trata esencialmente datos demográficos.

En las bibliotecas tradicionales, los datos internos proceden principalmente de dos fuentes: las consultas a los OPAC (*Open Public Access Catalogue*), y los datos de circulación de materiales y préstamo (incluido el préstamo interbibliotecario). Estos datos, como se indica más adelante, muestran información de los materiales que son más usados, el tiempo que son requeridos (renovaciones), materiales relacionados o similares (retirados en un mismo préstamo o por un mismo usuario), y las bibliotecas con que se tiene mayor relación a través del préstamo interbibliotecario.

Sin embargo estos datos no recogen en su totalidad la actuación que el usuario realiza en la biblioteca (en el caso bibliotecas de acceso abierto a la colección), quedándose fuera la consulta de fondos en la estantería, la consulta de fondos que no se prestan (normalmente obras de referencia y publicaciones periódicas en papel) y que a su vez suelen ser los más caros, etc. Por ello, los resultados de bibliominería han de ser considerados como representativos de una gran parte de la utilización de la biblioteca, pero no de su totalidad, por lo que estos datos deben apoyarse en otras informaciones.

En el caso de bibliotecas digitales y fondos en línea, el abanico de datos puede llegar a ser mucho mayor, ya que es posible crear y utilizar *logs* (archivos de registro de actividades) que marquen todo el recorrido y acciones que

realizan los usuarios de la colección de la biblioteca (Nicholson, 2006). La identificación del usuario puede hacerse de dos formas, dependiendo de la política de la biblioteca. Si se trata de un acceso restringido, el usuario se ve obligado a identificarse, y en el caso de bibliotecas digitales de acceso abierto, pueden estudiarse las sesiones que realiza cada usuario mediante la identificación de la dirección IP (*Internet Protocol*) de acceso y el uso de *cookies*.

3 Utilidad de la bibliominería

De modo general, la utilización de los datos almacenados (generados y recopilados) por la biblioteca puede realizarse de tres maneras: en primer lugar mediante la elaboración de informes periódicos de variables determinadas (lo que se denomina OLAP, *Online Analytical Processing*, o minería de datos dirigida); en segundo lugar, mediante preguntas concretas a la base de datos; y en tercero, mediante exploración aleatoria de variables (lo que se denomina minería de datos no dirigida).

Con ello se pueden realizar tres tareas básicas:

- Asociaciones: ver qué elementos están relacionados ya sea por derivación, causa/efecto o por similitud.
- Agrupaciones (*clustering*): crear grupos de datos con características similares.
- Resumen: presentar de modo abreviado los datos sobre la actividad diaria para una mejor comprensión de los mismos.

Una cuestión de suma importancia en el estudio, interpretación y utilización de los resultados de bibliominería, es que, al igual que con los datos obtenidos mediante estudios bibliométricos, éstos no deben ser tomados como algo definitivo y autoexplicativo, sino que han de ser tenidos en cuenta en su contexto y siendo comparados, contrastados y estudiados en función de otras variables y otros datos.

Kohavi y Provost presentan cinco desideratas para el éxito de la minería de datos en el comercio electrónico, desideratas que pueden hacerse extensibles a cualquier otro campo y que en el caso de las bibliotecas (por tanto, bibliominería) se cumplen a la perfección (Kohavi; Provost, 2001):

- Datos con descripciones ricas. En las bibliotecas, los ítems cuentan (o deben contar) con una completa, rica y adecuada descripción, que además se ha elaborado de acuerdo a unas normas, lo cual facilita enormemente su uso.
- Una gran cantidad de datos. Los autores indican que a mayor cantidad de datos se pueden extraer modelos más fidedignos.
- Colección de datos controlada y correcta. Como se ha indicado, en las bibliotecas la normalización de la información es una asignatura cubierta.
- La capacidad de evaluar resultados. Éste es el fin último del proceso, estudiar los resultados y tener en cuenta las conclusiones para futuras actuaciones. En caso contrario, es una tarea inútil.
- Facilidad de integración con los procesos existentes. La automatización de las bibliotecas a través de las bases de datos es una realidad desde hace años, y éstas forman parte ineludible de la actividad diaria, por lo que llevar a cabo tareas de bibliominería sólo supondría añadir un paso más a la cadena de la base de datos.

En cuanto a aplicaciones concretas en bibliotecas, las posibilidades son múltiples y han sido tratadas por diferentes autores. Un caso es Papatheodorou, quien, centrándose en bibliotecas digitales, indica que la bibliominería puede ayudar a las bibliotecas de diferentes maneras (Papatheodorou; Kapidakis; Sfakakis; Vassilou, 2003):

- Optimización de servicios: ayuda a los administradores a reorganizar el contenido de la biblioteca, autoridades e interfaces.
- Apoyo a la toma de decisiones.
- Personalización: ayuda a los usuarios a identificar información de interés para ellos por recomendación de materias similares. Lo que supondría una mejora de los tan útiles sistemas de DSI (Difusión Selectiva de Información).

En cuanto a aplicaciones más concretas, Nicholson y Stanton señalan las siguientes (Nicholson; Stanton, 2003):

- Predicción de necesidades de los usuarios: ver la evolución de las temáticas consultadas puede ayudar a predecir cuáles serán consultadas posteriormente.
- Identificación de materiales no consultados: por un mal proceso de selección o por una catalogación o clasificación incorrecta.
- Justificación del mantenimiento o supresión de acuerdos de préstamo interbibliotecario: gasto del préstamo frente a gasto de adquisición.

En los apartados siguientes se explican con mayor detalle éstas y otras de las posibles aplicaciones de la bibliominería.

3.1 Evaluación bibliotecaria

En primer lugar, la bibliominería puede ofrecer una nueva perspectiva en la evaluación de las bibliotecas. Por una parte porque proporciona datos directos y objetivos de los usuarios y del uso que hacen de la biblioteca, frente a otros métodos con mayor o menor grado de parcialidad como las encuestas, las entrevistas, el método Delphi y el incidente crítico (Sanz Casado, 1994). Y por otra parte porque permite definir y calcular una serie de indicadores de comparación (*benchmarking*), tal como recomienda la IFLA (*International Federation of Library Associations and*

Instituciones) en sus directrices para bibliotecas públicas, como son préstamos, uso de fondos, usuarios reales y potenciales, etc.

El análisis estadístico de datos cuantitativos es un criterio tradicionalmente muy usado en la evaluación de bibliotecas (como en el caso de la Red de Bibliotecas Universitarias Españolas <<http://www.crue.org/rebiun/>> , la Agència per a la Qualitat del Sistema Univeristari de Catalunya <<http://www.aqucatalunya.org>>, y en menor medida la Agencia Nacional de Evaluación de la Calidad y Acreditación <<http://www.aneca.es>> entre otras). Pero la bibliominería permite ir más allá, pues estos indicadores simples serán puestos en un contexto más amplio y relacionados entre sí para la obtención de nuevos indicadores. Es decir, no se limita a un recuento, sino que permite estudiar la relación entre indicadores para ver patrones de repetición en la actividad en tiempo real.

Para Lancaster, ya en 1988, el estudio de patrones de uso de la colección y la identificación de ítems poco usados es una de las formas de evaluación de la colección (Lancaster, 1993). Con este enfoque, se podría seguir de primera mano, al menos en lo que atañe al uso de la colección, una "calidad basada en el cliente/usuario",⁶ es decir, la idea de que los usuarios tienen diferentes necesidades y aquellos productos que satisfacen mejor dichas necesidades presentan mayor calidad. Cuánto más y mejor se conozcan las necesidades y opiniones de los usuarios, mejor se podrán adaptar los servicios a estos criterios, y por tanto, se obtendrá una mayor calidad, centrándose en lo que también se ha llamado "calidad percibida por el cliente" (Domínguez Sanjurjo, 1996).

3.2 Gestión económica: selección y adquisición de fondos

En la gestión de las bibliotecas, como en cualquier organización, la toma de decisiones desde la dirección es una actividad fundamental. Especialmente delicada es en la asignación y utilización de los recursos económicos, y más teniendo en cuenta que la mayor parte de las bibliotecas son financiadas con fondos públicos. En este sentido se pronuncian la IFLA y la UNESCO en las directrices para bibliotecas públicas, que establecen que tanto la planificación como la administración financieras resultan fundamentales para que la biblioteca funcione con eficiencia (obteniendo resultados óptimos), de modo económico (al menor coste) y con eficacia (dando máximos beneficios); es por ello por lo que la utilización de estos recursos económicos debe realizarse con la mayor transparencia y el mejor criterio posible.

Así, la utilización de datos sobre el comportamiento de los usuarios en el uso de la colección resulta de gran ayuda e importancia para la toma de decisiones respecto a la selección y adquisición de fondos (en 2000 las bibliotecas públicas destinaron el 14'5% de su presupuesto a adquisiciones) (Hernández, 2003). Es decir, conocer qué tipo de material se usa más y cuáles de los ítems más solicitados cuentan con pocos ejemplares, resulta crucial y muy valioso para orientar futuras adquisiciones.

Un estudio de la utilización de la bibliominería en la asignación de presupuestos para adquisiciones es el modelo presentado por Kao, Chang y Lin para bibliotecas universitarias. Dicho modelo, denominado ABAMDM (*Acquisition Budget Allocation Model via Data Mining*), está basado principalmente en los datos de circulación de materiales y en el "peso" que tienen los diferentes departamentos universitarios en la asignación del presupuesto de adquisición de materiales (Kao; Chang; Lin, 2003).

3.3 Organización de la colección

Aunque quizás de una menor riqueza, el estudio del propio catálogo bibliográfico mediante técnicas de minería de datos puede dar resultados interesantes. Así Lavoie, Dempsey y Silipigni indican que mediante estos análisis se pueden descubrir nuevos modos de presentación y organización de la colección que la hagan más interesante a los usuarios suponiendo, por tanto, un sistema más eficaz de acceso a la colección (Lavoie et al., 2006). Asimismo indican que la bibliominería combinada con el empleo del modelo *Functional Requirements for Bibliographic Records* (FRBR) también es de utilidad a la hora de identificar las múltiples manifestaciones e ítems de que se dispone en el centro. Un ejemplo sería relacionar las diferentes ediciones de *La Tragedia de Hamlet, Príncipe de Dinamarca* con las críticas, con las versiones cinematográficas, con obras que hagan referencia, o con las adaptaciones de todo tipo de que se disponga en el centro, y así recomendar todo este conjunto de ítems a aquellos usuarios que estén interesados en Shakespeare.

3.4 Lenguajes de clasificación

También a través de la bibliominería y analizando el uso del OPAC (más adelante se profundizará en la bibliominería aplicada a los OPAC), puede estudiarse la manera en que los usuarios llegan a los registros. De especial importancia sería ver qué términos se utilizan, tanto si éstos pertenecen al lenguaje documental usado por la biblioteca como si no, para así enriquecer la descripción temática y acercarla a la realidad del usuario. En el bien entendido que no se propone la adición de los términos empleados por los usuarios al lenguaje usado por la biblioteca, sino su utilización como sinónimos y para enriquecer un sistema de recomendaciones como se explica en el apartado siguiente.

Puesto que el empleo de cualquier sistema de clasificación responde a un modo de percepción de la realidad, el lenguaje documental será más eficaz y comprensible cuanto más se adapte a la visión que tiene el usuario de las cosas, por tanto será más útil, y por tanto mejor. Esto es, buscar la adopción de la "sabiduría de las masas" (Surowiecki, 2004) y de las folksonomías.

Aunque ciertos estudios (Ríos García, 1991) muestran que las necesidades normales de los usuarios quedan satisfechas con un registro que contenga sólo un mínimo de descripción bibliográfica -lo que Bereijo denomina planteamientos reduccionistas (Bereijo Martínez, 1998)- cuanto más rica sea la descripción de los fondos, más fácilmente llegarán los usuarios a ellos y, por tanto, mayor uso harán de los mismos. Y es mediante bibliominería

como más fácilmente se pueden estudiar las tendencias de uso de términos de indización para una mayor adecuación de los mismos al usuario.

Así, con todo ello se avanzaría en la solución de los dos principales problemas a los que se enfrentan los usuarios de los OPAC según Borgman (1986), citada por Fernández Molina y Moya Anegón, (1998): problemas mecánicos (errores tipográficos, no recuperaciones por errores en los nombres de los autores, errores en el sistema) y sobre todo los problemas conceptuales (no entendimiento del proceso de búsqueda, uso de términos incorrectos).

3.5 OPAC: presentación de resultados y sistema de recomendaciones

En cuanto al catálogo, con el empleo de la bibliominería pueden estudiarse el comportamiento y los tipos de consulta de los usuarios en el OPAC, para la presentación de resultados. Así, conocer qué ítems fueron considerados relevantes anteriormente en búsquedas iguales o similares, permitiría establecer resultados recomendados para las búsquedas, y también implementar sistemas de ordenación de resultados de búsqueda por relevancia basados en cuestiones de adecuación temática, y no únicamente en la concurrencia del término de búsqueda en el registro. Igualmente se podrían recomendar términos similares a los empleados en la búsqueda, que servirían a los usuarios para enriquecer y reorientar su estrategia de búsqueda tal como se ha explicado anteriormente. Además se podrían detectar errores ortográficos comunes en términos de búsqueda y, bien corregirlos de forma automática o bien proponer correcciones (al estilo de buscadores web como Google (<http://www.google.com>)), ya que siempre es mejor que tras una búsqueda se ofrezca un resultado erróneo (aún corrigiendo automáticamente al usuario) que no ofrecer ninguno, o un simple mensaje de error.

Con todo esto, lo que se pretende es desarrollar para los OPAC un sistema similar al utilizado en portales comerciales como Amazon.com (<http://www.amazon.com>) (en la utilización de la información sobre el comportamiento de los clientes, Amazon.com es un claro ejemplo a seguir por las bibliotecas), basados en lo que se denomina un análisis asociativo, del tipo *Quienes compraron este álbum también compraron...*, *Búsquedas relacionadas*, *Te gustaría...* o *También podría gustarte*, del propio Amazon.com. La obtención de relaciones asociativas a través de minería de datos ha sido y sigue siendo uno de los aspectos más estudiados dentro de esta disciplina.

Si bien, lo más correcto y completo sería que, junto a estas recomendaciones (de resultados y términos de búsqueda) obtenidas por minería de datos, se incluyeran otras hechas por los profesionales de las bibliotecas mediante el empleo de técnicas como el *clustering*⁷ y el propio conocimiento de la colección, y además permitir a los usuarios definir recomendaciones propias. Así se convertiría en un trabajo de colaboración, que además permitiría que la biblioteca se adaptase mejor a la comunidad en la que se ubica.

En resumen, el OPAC, como medio principal de acceso a la colección, es la herramienta fundamental de la biblioteca. Siguiendo a Lubeztky (Bereijo Martínez, 1998) el catálogo puede funcionar como simple herramienta de recuperación (*finding list*), con lo que bastaría con una relación de las entradas del material disponible en el centro, o también puede funcionar como un instrumento de referencia (*reference tool*), explotando así todos los recursos documentales e informativos del centro. Con una correcta utilización de bibliominería es esta segunda utilización del OPAC la que primaria, aumentando así su utilidad.

3.6 Estudios de usuarios: socialización y grupos de usuarios

En bibliotecas especializadas y universitarias, la bibliominería debe ser un método complementario a la realización de estudios bibliométricos sobre su comunidad. Como ya se ha señalado, ambos estudios son complementarios, pues la bibliominería ofrecerá la perspectiva de utilización de la colección como usuarios, mientras que los estudios bibliométricos ofrecerán la perspectiva de estos usuarios en su papel de autores, marcando el nivel de utilidad efectiva de la colección y mostrando también los elementos que se utilizan y que no forman parte de la colección (lo que orientará futuras adquisiciones).

Otra aplicación similar, también para bibliotecas de investigación y universitarias, es el descubrimiento de tendencias de investigación de sus usuarios (Bollen et al., 2003), que ayudaría también a orientar la política de selección y adquisición de fondos. Esto es lo que se denomina modelos de comportamiento: el estudio y definición de perfiles de los usuarios a través del comportamiento que tienen respecto a la colección del centro.

La bibliominería ofrece otra posibilidad. Como consecuencia del estudio de datos sobre búsquedas (y préstamos) y los estudios de usuarios, pueden configurarse grupos de usuarios con perfiles y gustos similares. Si en bibliotecas académicas esto puede ayudar a establecer y afianzar grupos de investigación, en bibliotecas públicas puede servir de elemento socializador entre los usuarios. Esta función podría desarrollarse incluso mediante la creación de perfiles públicos por parte de los usuarios, que les permitiera intercambiar opiniones y sugerencias. Se crearía de este modo una red social con la biblioteca como punto de unión, algo similar a lo que ofrecen servicios web como Last.fm - The Social Music Revolution (<http://www.last.fm>) en relación con la música.

3.7 Biblioteca 2.0

Toda esta actitud de colaboración de los usuarios en la construcción y mejora de servicios descrita en los apartados anteriores, es la base de la llamada Web 2.0. En el campo de las bibliotecas se habla de Biblioteca 2.0 o B2 (*Library 2.0* o *L2*) término acuñado por Casey (2005) y cuya definición más aceptada es la dada por Houghton quien dice que "Biblioteca 2.0 simplemente significa hacer tu espacio bibliotecario más interactivo, con mayor colaboración, y dirigido hacia las necesidades de la comunidad [...]" (Houghton, 2005). Es decir, básicamente lo que cualquier biblioteca debe hacer, pero apoyada por las tecnologías y con una mayor cantidad

de información.

Para Crawford (quien además recopila 62 visiones y siete definiciones de B2), "Biblioteca 2.0 abarca un abanico de nuevas, y no tan nuevas, metodologías software (software social, interactividad, API,⁸ software modular...) que puede ser y será útil para que muchas bibliotecas ofrezcan nuevos servicios y hagan accesibles los ya existentes de una manera nueva e interesante". Si bien, como también dice este autor, no todo es nuevo en la llamada B2, así los conceptos son los mismos y las tecnologías son nuevas, con lo cual se aparecen nuevos horizontes, y muchos de los existentes se ven ampliados (Crawford, 2006).

Que las bibliotecas están (o deben estar) abiertas a los cambios y necesidades de los usuarios es lógico, pero la relación con el usuario es siempre difícil, pues ¿cuántos usuarios que no encuentran el libro que buscan realizan una desiderata? Sin embargo, si a través de las estadísticas del OPAC, mediante bibliominería, se ven reiteradas búsquedas de un libro del que no se dispone en la biblioteca, ya se sabrá de esa demanda y se conocerá el tipo de usuario que buscaba ese libro pudiéndosele avisar una vez esté disponible. Es decir, la B2 no es "sólo" tecnología, pero "también" es tecnología.

Así, desde mi punto de vista, los OPAC, y más concretamente los interfaces con el usuario, serán el aspecto que más puede beneficiarse y mejorar si la B2 se consolida. Especialmente en el fomento de la personalización y adecuación de la información a las necesidades y características del usuario en tiempo real y de manera inteligente.

A la B2 la bibliominería le ofrece la base de estudio de datos necesaria para llevar a cabo los procesos de colaboración y personalización que promueve mediante los explicados estudios de los usuarios y sus grupos, de la colección, de las consultas a OPAC, de la utilización de lenguajes de clasificación, y de otros. Así, las propuestas y mejoras de la B2 suponen el elemento cristalizador y aunador de la útil información que se puede obtener mediante bibliominería.

4 Bibliominería y privacidad

Uno de los elementos problemáticos de la minería de datos, y por tanto de la bibliominería, está en cómo se respeta la privacidad de los usuarios. Esta situación hace que muchos usuarios puedan recelar del hecho de ceder sus datos (propios o generados) para su explotación, más cuando la biblioteca es considerada como un lugar libre y seguro donde informarse, y donde se respetan (o deben respetarse) todos los puntos de vista y opiniones de modo imparcial y objetivo. Para romper con esta reticencia se puede intentar enfatizar un sentimiento de empatía hacia el conjunto de usuarios: si tú ayudas y aportas algo a la comunidad, la comunidad te aportará algo a ti. Nuevamente la filosofía Web 2.0 o Biblioteca 2.0 y su énfasis en la colaboración será un factor que ayude a rebasar esta situación.

La privacidad en bibliotecas es un asunto de gran amplitud y de preocupación por lo que reaparece periódicamente en los debates profesionales y es un aspecto que debe ser regulado no sólo por la legislación vigente, como se indica más adelante, sino también en los códigos éticos de la profesión (Gómez-Pantoja Fernández-Salguero; Pérez Pulido, 1998).

Para Zwass (1996) la privacidad es el aspecto ético más importante de los sistemas de información. E indica que esta cuestión ha estado presente desde antes de la aparición de las nuevas tecnologías en las bibliotecas, pero son éstas las que han aumentado el riesgo de un posible mal uso. Ramos Simón (2004) también muestra preocupación por el uso de la información generada por la actividad de los usuarios, en su caso en los sistemas de DRM (*Digital Right Management* o gestión de derechos digitales).

Domingo-Ferrer y Torra (quienes además presentan una breve recopilación de artículos sobre el tema) señalan que la cuestión de la privacidad en la minería de datos (*Privacy-Preserving Data Mining* o PPDM) ha sido estudiada desde diferentes disciplinas (Domingo-Ferrer; Torra, 2005):

- Estadística: la mayoría de las leyes nacionales de estadística incluyen referencias a la salvaguarda de la privacidad.
- Filosofía: la ética en la sociedad de la información recoge la privacidad como uno de los valores a defender.
- Informática: la privacidad se incluye como parte de la seguridad de los datos.

Como indican Geyer-Schultz, Neumann y Thede, por parte de las bibliotecas esta cuestión ha sido una de las que ha impedido la aplicación de bibliominería, como también lo han sido, también según estos autores, las limitaciones presupuestarias y algunos aspectos técnicos (Geyer-Schultz et al., 2003). Aunque no es objetivo de este documento analizar los elementos tecnológicos de la bibliominería, los problemas relativos a la tecnología son cada vez menores debido a la bajada de precios y el aumento de la potencia del hardware, así como el desarrollo de software específico y fácil de usar, para estas tareas, siendo muchas veces utilizable el propio sistema de gestión bibliotecaria (Michail, 2000).

Como se ha señalado, en ocasiones se hace necesario cruzar varias fuentes y bases de datos para obtener resultados más precisos. En esta línea, para Clifton actualmente resulta crítico desarrollar técnicas que permitan integrar y compartir datos sin violar la privacidad. Para ello, estos autores diseñan un esquema de organización de los campos de las bases de datos que garantiza el mantenimiento de la privacidad aún utilizando datos personales (Clifton et al., 2004). Por su parte Yang, Zhong y Wright proponen un procedimiento de encriptado de los datos, que impida que el sistema sea capaz de establecer correspondencias entre los datos y su origen, protegiéndose así también la privacidad de los usuarios (Yang et al., 2005).

Si bien, para el caso que aquí se trata, realmente los datos necesarios para la bibliominería no tienen por qué

contener información personal ni permitir una identificación del usuario. Esto se llevaría a cabo utilizando un código aleatorio para la identificación de cada usuario, o bien utilizando lo que Nicholson (2006)⁹ denomina "substitutos demográficos" (*demographic surrogates*), es decir, almacenando simplemente unos datos del perfil de usuario que permitan clasificarlo sin identificarlo. Nicholson y Smith describen una metodología para este proceso; hablan de la "des-identificación" (*deidentification*) del usuario (Nicholson; Smith, 2005), lo que también en Estadística se denomina "datos anonimizados" (Montaña Lacabra, 2005). Si bien, ambas técnicas son perfectamente compatibles.

Así, en el caso de datos sobre los préstamos en una biblioteca universitaria, de cara a la bibliominería no se necesita saber que en el curso 2004-2005 María López, alumna de tercer curso de la Licenciatura en Derecho, se ha llevado prestado cuatro veces el libro *Derecho del trabajo* de Ignacio Albiol Montesinos, sino que este año ese libro ha sido prestado cuatro veces al usuario 1010011010, estudiante de la Licenciatura en Derecho.

Otra cuestión que merece reflexión es si resulta interesante para la gestión de la biblioteca el mantener una relación de los préstamos que ha realizado cada usuario. Estos datos pueden ser útiles incluso para el propio usuario, y protegidos correctamente no tienen por qué dar problemas de respeto a su privacidad. Es aquí donde intervienen cuestiones técnicas de protección de información y también cuestiones éticas para los profesionales de las bibliotecas en la utilización de los datos de que disponen.

Desde otro lado, un caso de utilización de la minería de datos como medio de control de las actividades bibliotecarias de los ciudadanos (como los registros de circulación, los perfiles del usuario, las peticiones de préstamo, y archivos registro de uso de Internet) es el USA Patriot Act (Taylor; Black, 2004). Caso al que la ALA (*American Library Association*) se ha opuesto rotundamente ya que choca con su código ético.

Actualmente no hay una política generalizada sobre el tratamiento de los datos bibliotecarios, y la problemática legal que implica su manejo difiere entre los estados. Por ello no se puede generalizar una medida de actuación sobre el tema y habría de estudiarse detenidamente cada caso y cada situación legal.

No obstante, como bien sostienen Million y Fisher, aunque no exista una ley que regule la confidencialidad de los registros bibliotecarios, las bibliotecas deben tener una política sobre el tema, "apropiada, aprobada, bien comunicada al personal e incluida en el manual de procedimiento" (Million; Fisher, 1998).

En caso de existir una legislación al respecto, la política del centro deberá ser congruente con ésta. En el caso español, ya la Constitución de 1978 en su artículo 18¹⁰ recoge la protección de la información y datos personales al garantizar el derecho a la intimidad. Pero es la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal (normalmente abreviada en LOPD) la que regula esta actividad, y es la Agencia Española de Protección de Datos la autoridad administrativa independiente encargada de velar por la aplicación de la legislación al respecto. Además la LOPD "regula el deber de guardar el secreto profesional a quienes intervienen en cualquier proceso automático de recogida de datos de carácter personal, e impone las oportunas sanciones en el caso de manipular ilegalmente datos sensibles estrechamente vinculados al concepto de intimidad" (Gómez-Pantoja Fernández-Salguero; Pérez Pulido, 1998)

Conforme al artículo 2 de la LOPD,¹¹ que fija que su ámbito de aplicación son todos los datos de carácter personal registrados en soporte físico, y que se aplica a toda la modalidad de uso posterior de esos datos por parte del sector público y privado, los ficheros de lectores de bibliotecas están regulados por esta ley. Para su utilización de cara a bibliominería, el artículo 4¹² de dicha ley señala que el uso de los datos con fines estadísticos no es incompatible con la actividad con la que fueron recogidos, por lo que éste puede ser considerado un uso legítimo.

También es una cuestión que aparece protegida en diferentes instrumentos del derecho internacional y de protección de los derechos humanos como son la Declaración Universal de Derechos Humanos (art. 12), el Convenio Internacional sobre Derechos Civiles y Políticos (art. 17), la Convención Europea sobre Derechos Humanos (art. 8), y la Convención Internacional de Telecomunicaciones (art. 22) (Gómez-Pantoja Fernández-Salguero; Pérez Pulido, 1998).

Como indican Agrawal y Srinikant (Agrawal; Srinikant, 2000), son muchos los aspectos que plantea la cuestión de privacidad y que hay que estudiar, algunos de los cuales son señalados por Nicholson y Smith (2005):

- ¿Cuánta información se pierde en la "des-identificación"? ¿Hay algoritmos que puedan suplir o simular los datos perdidos?
- ¿Qué estándares se podrían crear para motivar a las bibliotecas a mantener el mismo tipo de datos des-identificados?¹³ ¿Estos estándares permitirían desarrollar *data warehouses* (repositorios de datos) en redes o consorcios bibliotecarios?
- ¿Qué opinión tienen los usuarios de sus datos "des-identificados"?

5 Conclusiones

Como se ha venido explicando, la Biblioteconomía y la minería de datos no son dos actividades incompatibles. Es más, se podría decir que las bibliotecas son un espacio ideal para desarrollar tareas de minería de datos por tratarse de un espacio controlado, con una descripción correcta (eso es lo esperable) de los elementos con que trabaja, acostumbrada a recopilar datos (préstamos, usuarios, fondos, etc.) y con profesionales (eso también es lo esperable) que saben manejar información y sacar partido de ella. Se tienen los datos, las herramientas y los profesionales, la utilidad de la bibliominería está clara y es alta, ¿a qué se espera para su aplicación? Ya se han enriquecido los OPAC en contenido, permitiendo acceso a múltiples fondos y bases de datos, ahora deben y pueden enriquecerse en funcionalidades.

Con el aumento de información disponible y la mejora de servicios que se pueden obtener de la bibliominería, en primer lugar se debe romper con la idea de que el usuario vaya a la biblioteca a buscar un libro concreto para después irse. Hay que potenciar la biblioteca como un lugar rico, abierto y completo donde descubrir y ampliar sus ideas y gustos.

Como se han señalado, son muchos los aspectos a seguir estudiando y muchas las vías de investigación existentes en la aplicación de la bibliominería. En el presente artículo simplemente se han presentado algunas de las posibles utilidades, pero las posibilidades son muchas y las vías de investigación muy amplias.

La bibliominería es un campo donde la innovación y la amplitud de miras a la hora de estudiar y cruzar datos, es importante. Cada centro, conociendo su forma de trabajo, sus usuarios y sus características propias, deberá identificar unos parámetros de interés y estudiará unos patrones concretos. Ideas más trabajo siempre equivalen a éxito. Si bien, puede que de la bibliominería o de la llamada Biblioteca 2.0 falle la aplicación real de todas estas ideas. ¿Faltan sólo recursos o es también un problema de falta de motivación profesional?

Bibliografía consultada

Ackoff, Russell L. (1989). "From data to wisdom". *Journal of applied systems analysis*, no. 16, p. 3-9.

Agrawal, Rakesh; Imielinski, Tomasz; Swami, Arun (1993). "Mining association rules between sets of items in large databases". *Proceedings of the 1993 ACM SIGMOD Conference*. <<http://rakesh.agrawal-family.com/papers/sigmod93assoc.pdf>>. [Consulta: 1-05-2006]

Agrawal, Rakesh; Srinikant, Ramakrishnan (2000). "Privacy-preserving data mining". *Proceedings of the 2000 ACM SIGMOD conference on management of data*. p. 439-450. <<http://doi.acm.org/10.1145/342009.335438>> . [Consulta: 1-05-2006].

American Library Association. *The USA patriot act in the library*. <<http://www.ala.org/ala/oif/ifissues/usapatriotactlibrary.htm>>. [Consulta: 30-04-2006].

Banerjee, K. (1998). "Is data mining right for your library?" *Computers in libraries*, vol. 18, no. 10, p. 28-31.

Bellinger, Gene; Castro, Durval; Mills, Anthony (1994). *Data, information, knowledge, and wisdom*. <<http://www.systems-thinking.org/dikw/dikw.htm>>. [Consulta: 1-05-2006].

Bereijo Martínez, Antonio (1998). "Caracterización del concepto de 'calidad' en la catalogación descriptiva: factores que atañen al diseño de objetivos". *Boletín Millares Carlo*, núm. 17, p. 319-355. <http://dialnet.unirioja.es/servlet/fichero_articulo?articulo=1700760&orden=37277>. [Consulta: 12/04/2006].

Bollen, Johan; Luce, Rick; Vemulapalli, Soma Sekhara; Weining, Xu (2003). "Usage analysis for the identification of research trends in digital libraries". *D-lib magazine*, May, vol. 9, no. 5, <<http://www.dlib.org/dlib/may03/bollen/05bollen.html>>. [Consulta: 31-03-2006].

Borgman, Christine L. (1986). "Why are online catalogs hard to use? Lessons learned from information-retrieval studies" *Journal of the American society for information science*, vol. 37, no. 6, p. 387-400. <<http://www3.interscience.wiley.com/cgi-bin/abstract/57783/>>. [Consulta: 7-05-2006].

Brin, Sergey; Motwani, Rajeev; Ullman, Jeffrey D.; Tsur, Shalom (1997). "Dynamic itemset counting and implication rules for market basket data". *Proceedings of the 1997 ACM SIGMOD conference*. p. 255-264 <<http://doi.acm.org/10.1145/253260.253325>>. [Consulta: 1-05-2006].

Casey, Michael (2005). "Working towards a definition of Library 2.0". *Library Crunch: bringing you a library 2.0 perspective*, October 21, 2005. <http://www.librarycrunch.com/2005/10/working_towards_a_definition_o.htm>. [Consulta: 16-05-2006].

Chi-Wing Wong, Raymond; Wai-Chee Fu, Ada; Wang, Ke (2005). "Data mining for inventory item selection with cross-selling considerations". *Data mining and knowledge discovery*, July 2005, vol. 11, no. 1, p. 81-112.

Cleveland, Harland (1982). "Information as resource". *The futurist*, December, p. 34-39.

Clifton, Chris; Doan, Anhai; Elmagarmid, Ahmed; Kantarcioglu, Murat; Schadow, Gunther; Suciu, Dan; Vaidya, Jaideep (2004). "Privacy preserving, data integration and sharing". *Data mining and knowledge discovery archive. Proceedings of the 9th ACM SIGMOD workshop on research issues in data mining and knowledge discovery*. p. 19-26. <<http://doi.acm.org/10.1145/1008694.1008698>>. [Consulta: 21-04-2006].

Cox, Kenneth C.; Eick, Stephen G.; Wills, Graham J.; Brachman, Ronald J. (1997). "Brief application description; visual data mining: recognizing telephone calling fraud". *Data mining and knowledge discovery*, June, vol. 1, no. 2, p. 225-231.

Crawford, Walt (2006). "Library 2.0 and 'Library 2.0'". *Cites & insights*, vol. 6, no. 2, p. 1-32. <<http://cites.boisestate.edu/civ6i2.pdf>> [Consulta: 16-05-2006].

Cullen, Kelvin (2005). "Delving into data". *Library journal* August, vol. 130, no. 13, p. 30-32. <<http://www.libraryjournal.com/article/CA633325.html>>. [Consulta: 26-04-2006].

Domingo-Ferrer, Josep; Torra, Vicenç (2005). "Privacy in data mining". *Data mining and knowledge discovery*, September, vol. 11, no. 2, p. 117-119.

Domínguez Sanjurjo, M^a Ramona (1996). *Nuevas formas de organización y servicios en la biblioteca pública*. Gijón: Trea

España. "Ley orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal". *Boletín Oficial del Estado*, 14 de diciembre de 1999, núm. 298, p. 43088-43099.

Fawcett, Tom; Provost, Foster (1997). "Adaptive fraud detection". *Data mining and knowledge discovery*, September, vol. 1, no. 3, p. 291-316.

Fayyad, Usama M.; Pietetsky-Shapiro, Gregory; Smyth, Padhraic (1996). "From data mining to knowledge discovery: an overview". En: Fayyad, Usama M.; Pietetsky-Shapiro, Gregory; Smyth, Padhraic (ed.). *Advances in knowledge discovery and data mining*. California: AAAI Press, The MIT Press, p. 1-36.

Fayyad, Usama; Simoudis, Evangelos (1995). *Knowledge discovery and data mining*. <<http://www-aig.jpl.nasa.gov/public/kdd95/tutorials/IJCAI95-tutorial.html>> . [Consulta: 1-05-2006].

Fayyad, Usama; Uthurusamy, Ramasamy (1996). "Data mining and knowledge discovery in databases". *Communications of the ACM*, November, vol. 39, no. 11, p. 24-26. <<http://doi.acm.org/10.1145/240455.240463>> . [Consulta: 1-05-2006].

Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas; Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (2001). *Directrices de la IFLA/UNESCO para el desarrollo del servicio de bibliotecas públicas*. 94 p. <<http://www.ifla.org/VII/s8/news/pg01-s.pdf>> . [Consulta: 4-04-2006].

Fernández Molina, Juan Carlos; Moya Anegón, Félix de (1998). *Los catálogos de acceso público en línea: el futuro de la recuperación de información bibliográfica*. Málaga: Asociación Andaluza de Bibliotecarios, 197 p.

Geyer-Schulz, Andreas; Neumann, Andreas; Thede, Anke (2003). "An architecture for behavior-based library recommender systems". *Information technology and libraries*, 2003, vol. 22, no. 4, p. 165-174. <<http://www.ala.org/ala/lita/litapublications/ital/2204geyer.htm> . [Consulta: 23-08-2006]

Gómez-Pantoja Fernández-Salguero, Aurora; Pérez Pulido, Margarita (1998). "El concepto de privacidad en servicios bibliotecarios actuales". *FESABID 98. VI Jornadas Españolas de Documentación: los sistemas de información al servicio de la sociedad*. <http://fesabid98.florida-uni.es/Comunicaciones/a_gomez.htm> . [Consulta: 15-06-2006].

Guenther, Kim (2000). "Applying data mining principles to a library data collection". *Computers in libraries*, vol. 20, no. 4, p. 60-63.

Hand, David; Mannila, Heikki; Smyth, Padric (2001). *Principles of data mining*. Cambridge: Massachusetts Institute of Technology.

Hernández, Hilario (dir.) (2003). *Las colecciones de las bibliotecas públicas en España: informe de situación*. Salamanca: Fundación Germán Sánchez Ruipérez.

Hernández Orallo, José; Ramírez Quintana, M^a José; Ferri Ramírez, César. (2004). *Introducción a la minería de datos*. Madrid: Pearson - Prentice Hall.

Houghton, Sarah (2005). "Library 2.0 Discussion: Michael Squared". *LibraryInBlack.net: resources and discussions for the 'tech-librarians-by-default' among us...* December 19. <http://librarianinblack.typepad.com/librarianinblack/2005/12/library_20_disc.html> . [Consulta: 16-05-2006].

Kao, S.-C.; Chang, H.-C.; Lin, C.-H. (2003). "Decision support for the academic library acquisition budget allocation via circulation database mining". *Information processing and management*, no. 39, p. 133-147. *Knowledge discovery and data mining conference (KDD-95)*. <<http://www-aig.jpl.nasa.gov/public/kdd95/>> . [Consulta: 1-05-2006].

Kohavi, Ron (2001). "Mining e-commerce data: the good, the bad, and the ugly". *Conference on Knowledge Discovery in Data Archive: proceedings of the seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. p. 8-13. <<http://doi.acm.org/10.1145/502512.502518>> . [Consulta: 1-05-2006].

Kohavi, Ron; Provost, Foster (2001). "Applications of data mining to electronic commerce". *Data mining and knowledge discovery*, January, vol. 5, no. 1-2, p. 5-10.

Kudyba, Stephan; Hoptroff, Richard (2001). *Data mining and business intelligence: a guide to productivity*. Hershey: Idea

Lancaster, Frederick Wilfrid (1993). *If you want to evaluate your library...* London: Library Association Publishing.

Lavoie, Brian; Dempsey, Lorcan; Silipigni Connaway, Lynn (2006). "Making data work harder". *Library journal*, January, vol. 131. <<http://www.libraryjournal.com/article/CA6298444.html>> [Consulta: 26-04-2006].

Ma, Yiming; Liu, Bing; Wong, Ching Kian; Yu, Philip S.; Lee, Shuik Ming (2000). "Targeting the right students

using data mining". *Conference on Knowledge Discovery in Data Archive: proceedings of the seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. p. 457-464. <<http://doi.acm.org/10.1145/347090.347184>>. [Consulta: 1-05-2006].

Michail, Amir (2000). "Data mining library reuse patterns using generalized association rules". *International Conference on Software Engineering Archive: proceedings of the 22nd International Conference on Software Engineering*. p. 167-176. <<http://doi.acm.org/10.1145/337180.337200>>. [Consulta: 2-05-2006].

Million, Angela C.; Fisher, Kim N. (1998). "Library records: a review of confidentiality laws and policies". *The journal of academic librarianship*, vol. 11, no. 6, p. 346-349.

Montaña Lacambra, Rosa (2005). "Fuentes estadísticas y bibliotecas: estado de la cuestión". *BiD: Textos universitarios de biblioteconomía i documentació*, diciembre, núm. 15. <http://www2.ub.edu/bid/consulta_articulos.php?fichero=15monta2.htm> [Consulta: 23-08-2006].

Nicholson, Scott (2003). "The bibliomining process: data warehousing and data mining for library decision-making". *Information technology and libraries*, December, vol. 22, no. 4, p. 146-151. <<http://www.ala.org/ala/lita/litapublications/ital/2204nicholson.htm>>. [Consulta: 23-08-2006]

Nicholson, Scott (2006). "The basis for bibliomining: frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services". *Information processing & management*, vol. 42. iss. 3, May, p. 785-804. <<http://dx.doi.org/10.1016/j.ipm.2005.05.008>>. [Consulta: 22-08-2006]. (ver nota 9)

Nicholson, Scott; Smith, Catherine Arnott (2005). "Using lessons from health care to protect the privacy of library users: guidelines for the de-identification of library data based on HIPAA". En: Andrew Grove (ed.). *Proceedings 68th annual meeting of the American Society for Information Science and Technology (ASIST)* vol. 42. <http://eprints.rclis.org/archive/00005255/01/Nicholson_Using.pdf>. [Consulta: 29-04-2006].

Nicholson, Scott; & Stanton, Jeffrey (2003). "Gaining strategic advantage through bibliomining: data mining for management decisions in corporate, special, digital, and traditional libraries". En: Nemati, Hamid R.; Barko, Christopher D. (eds.). *Organizational data mining: leveraging enterprise data resources for optimal performance*. Hershey, PA: Idea Group Publishing, p. 247-262. (ver nota 5)

Papatheodorou, Christos; Kapidakis, Sarantos; Sfakakis, Michalis; Vassilou, Alexandra (2003). "Mining user communities in digital libraries". *Information technology and libraries*, December, vol. 22, no. 4, p. 152-157.

Ramos Simón, Luis Fernando (2004). "DRM: protección versus accesibilidad de la información digital". *Hipertext.net*, mayo, núm. 2. <<http://www.hipertext.net/web/pag208.htm>>. [Consulta: 23-08-2006].

Ríos García, Yolanda (1991). "Catálogos en línea de acceso público". *Revista española de documentación científica*, vol. 14, núm. 2, p. 121-141.

Sanz Casado, Elías (1994). *Manual de estudios de usuarios*. Madrid: Fundación Germán Sánchez Ruipérez.

Sharma, Nikhil (2005). *The origin of the 'data information knowledge wisdom' hierarchy*. <http://www-personal.si.umich.edu/~nsharma/dikw_origin.htm>. [Consulta: 1-05-2006].

Surowiecki, James (2004). *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. New York: Doubleday.

Taylor, Mayo; Black, William (2004). "In search of reason: libraries and the USA patriot act". *Journal of librarianship and information science*, vol. 36, no. 2, p. 51-54.

Yang, Zhiqiang; Zhong, Sheng; Wright, Rebecca N. (2005). "Anonymity-preserving data collection". *Conference on Knowledge Discovery in Data Archive: proceedings of the seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. p. 334-343. <<http://doi.acm.org/10.1145/1081870.1081909>>. [Consulta: 21-04-2006].

Zeleny, Milan (1987). "Management support systems: towards integrated knowledge management". *Human systems management*, vol. 7, no. 1, p. 59-70.

Zwass, Vladimir (1996). "Ethical issues in information systems". En: Kent, Allen; Lancour, Harold. *Encyclopedia of library and information science*. New York: Marcel Dekker, p. 175-195.

Fecha de recepción: 28/07/2006. Fecha de aceptación: 10/09/2006.

Notas

¹ Cita tomada de Fayyad (Fayyad; Uthurusamy, 1996).

² El modelo DIKW (*data, information, knowledge, wisdom*: datos, información, conocimiento, sabiduría) es uno de los

más conocidos y aceptados para representar la estructura jerárquica conceptual de la información. La teoría más ampliamente aceptada (Sharma, 2005; Bellinger, Castro, Mills, 1994) sostiene que este esquema está basado en un poema de T.S. Eliot titulado *The Rock* (1990), siendo desarrollado teóricamente por Ackoff (1989), Cleveland (1982) y Zeleny (1987), entre otros.

³ En la literatura consultada, el uso de la frase en español "minería de datos" para la inglesa *data mining* es más común que "descubrimiento de información en bases de datos" para *knowledge discovery in databases*, por lo que se optado por utilizar de aquí en adelante la primera en español y la segunda en inglés.

⁴ Aunque se podría utilizar la expresión "minería de datos en bibliotecas", por las mismas razones que utiliza Nicholson (2006) para el uso del término *bibliomining*, se ha optado por la adaptación directa al español, como "bibliominería".

⁵ También disponible en línea la edición preliminar actualizada en 2004: Nicholson, Scott; Stanton, Jeffrey (2003). "Gaining strategic advantage through bibliomining: data mining for management decisions in corporate, special, digital, and traditional libraries." <<http://www.bibliomining.com/nicholson/odmcom.html>>. [Consulta: 30-03-2006].

⁶ Smart, Seawright y Detiene (Bereijo Martínez, 1998) definen un esquema estructural para estudiar el concepto de calidad, basado en el esquema bidimensional de Seawright y Young, sobre el que aplican las seis categorías de Garvin: calidad trascendente, calidad basada en el proceso de manufactura, calidad basada en el producto, calidad basada en el cliente, calidad basada en el valor, y calidad estratégica.

⁷ Para Fernández Molina y Moya Anegón, las técnicas de *clustering*, o análisis cluster, "consisten en clasificar de manera automática el contenido de las bases de datos. [...] En términos generales, estas [técnicas] se ocupan de la forma en que se agrupan los términos de indización asignados a los documentos, con objeto de poner de manifiesto la relación entre los documentos [...] de materias similares" (Fernández Molina; Moya Anegón, 1998).

⁸ *Application Programming Interface*, Interfaz de Programación de Aplicaciones. Puede definirse, en términos muy básicos, como el conjunto de operaciones que indica cómo han de interactuar los módulos de software entre sí.

⁹ También disponible en línea la edición preliminar en: Nicholson, Scott (2005?). "The basis for bibliomining: frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services." <<http://bibliomining.com/nicholson/nicholsonbibliointro.html>>. [Consulta: 30-03-2006].

¹⁰ Artículo 18. 1. Se garantiza el derecho al honor, a la intimidad personal y familiar y a la propia imagen. [...]

¹¹ Artículo 2. 1. La presente Ley Orgánica será de aplicación a los datos de carácter personal registrados en soporte físico, que los haga susceptibles de tratamiento, y a toda modalidad de uso posterior de estos datos por los sectores público y privado. [...]

¹² Artículo 4. 1. Los datos de carácter personal sólo se podrán recoger para su tratamiento, así como someterlos a dicho tratamiento, cuando sean adecuados, pertinentes y no excesivos en relación con el ámbito y las finalidades determinadas, explícitas y legítimas para las que se hayan obtenido.

Artículo 4. 2. Los datos de carácter personal objeto de tratamiento no podrán usarse para finalidades incompatibles con aquellas para las que los datos hubieran sido recogidos. No se considerará incompatible el tratamiento posterior de éstos con fines históricos, estadísticos o científicos. [...]

¹³ Actualmente el establecimiento de estándares para compartir la información obtenida por minería de datos es una de las principales tareas del área.