



Bibliomining: minería de datos y descubrimiento de conocimiento en bases de datos aplicados al ámbito bibliotecario

Ricardo Herrera Varela ¹

Resumen

Se estudia las fases del proceso de descubrimiento del conocimiento en bases de datos (KDD) que incluyen datos de tipo complejo y que tiene a la minería de datos como elemento esencial para apoyar este proceso mediante el análisis de datos. La utilización de las herramientas de minería de datos y de visualización de información en las bibliotecas se denomina bibliomining y pretende conocer las actividades del personal o los usuarios de la organización transformándose en una herramienta eficaz para el apoyo de la gestión. Se analizan el concepto y desarrollo de las fases del proceso de bibliomining, como algunas aplicaciones y beneficios que puede significar esta disciplina para el ámbito bibliotecario. Finalmente se intenta esbozar algunas con consideraciones a tener en cuenta antes de desarrollar un proyecto de bibliomining.

Palabras claves: minería de datos / visualización de la información / datos complejos / bibliomining / KDD / bibliotecas

¹Doctorando en Documentación Universidad Carlos III de Madrid 100051023@alumnos.uc3m.es.

1. Introducción

La minería de datos aunque en teoría puede ser aplicada a cualquier tipo de información comúnmente es aplicada a grandes volúmenes de datos de las organizaciones. Las técnicas que aporta esta disciplina se emplean para mejorar el rendimiento de procesos industriales o de negocio en los que se manejan grandes volúmenes de información estructurada y almacenada en bases de datos. Por ejemplo, se usan con éxito en aplicaciones de control de procesos productivos, como herramienta de ayuda en la planificación y la decisión en marketing y finanzas entre otros. Asimismo, la minería de datos es fundamental en la investigación científica y técnica, como herramienta de análisis y descubrimiento de conocimiento a partir de observación de datos o resultados de experimentos.

La búsqueda de patrones en conjuntos de datos tiene una larga tradición en el ámbito académico; en principio en el área estadística y más recientemente en inteligencia artificial, de allí surgen métodos y procesos como KDD². Pero la reciente necesidad de la industria por explotar el potencial de sus enormes acumulaciones de datos en medios informáticos ha impulsado a vendedores de tecnologías y organizaciones de consultoría a crear metodologías o procesos para el uso de las herramientas computacionales disponibles que implantan los algoritmos propios de Minería de Datos.

El uso industrial y científico de la minería de datos, la aplicación en nuevas áreas como es el caso de la bibliotecas y el constante crecimiento del volumen y la tipología de datos, requieren mucho más que la aplicación de sofisticadas técnicas como redes neuronales o árboles de decisión sobre tablas de datos. En el presente documento se presenta a la Minería de Datos como una fase mas dentro del proceso KDD y por otra parte las principales tipologías de datos complejos con el fin ayudar a definir y entender mejor el proceso de extraer los datos y patrones durante un proyecto de descubrimiento del conocimiento.

² Abreviación de las siglas en inglés Knowledge Discovery Databases y traducido al español como Descubrimiento de Conocimiento en Bases de datos.

Las bibliotecas tienen una larga tradición en el uso y análisis de los patrones de conducta especialmente de la colección, con el cual se determina principalmente la utilización que se hace del material o la información por parte de los usuarios. A pesar de este uso, desafortunadamente pocas bibliotecas se han sabido aprovechar estos datos como una manera de mejorar el servicio de cara al cliente, de manejar presupuestos para la adquisición, o que estos datos sirvan para apoyar la toma de decisiones estratégicas sobre la orientación del consumo de información en sus organizaciones.

El objetivo de este trabajo es describir los pasos del proceso de extracción de conocimiento en bases de datos y como se pueden utilizar la explotación minera de los datos en bibliotecas para entender, en primer lugar, los patrones del comportamiento entre usuarios y el personal de la biblioteca, y en una segunda instancia comprender los patrones de uso de los recursos de información en la institución. La investigación intentará examinar además fuentes de datos, los posibles beneficios del uso de las técnicas de minería de datos en bibliotecas y algunas consideraciones éticas y legales para implementar un proyecto de bibliomining.

2. La minería de datos y el proceso de descubrimiento de conocimiento en bases de datos

2.1. Concepto de minería de datos

La minería de datos es definida por Kopanmakis & Theodoluidis (2003), como el proceso de descubrimiento de conocimiento sobre repositorios de datos complejos mediante la extracción oculta y potencialmente útil en forma de patrones globales y relaciones estructurales implícitas entre datos. Otros como Written & Frank (2003) apuntan que la minería de datos como aquel proceso en que se extrae conocimiento útil y comprensible, previamente desconocido, y a partir de grandes conjuntos de datos almacenados en distintos formatos.

Hernández et al. (2005) aclara que existen muchos términos que se relacionan o utilizan como sinónimos de la minería de datos, una de ellas es el KDD el cual Fayyad et al. (2002) define como el proceso no trivial de identificar patrones validos, novedosos, potencialmente útiles y en última instancia comprensibles a partir de los datos. A diferencia de la minería de

datos es un proceso mas complejo que lleva no solo a obtención de modelos o patrones, que es el objetivo de la minería de datos, sino que incluye además una evaluación y una posible interpretación de los mismos. Fig 1



Figura 1. Proceso de Extracción del conocimiento según Hernández et al. (2005)

2.2. Tipo de datos

En muchas organizaciones se han preocupado de recopilar gran cantidad de información de todo tipo. Es fácil digitalizar información, ya no es excesivamente caro almacenarla y en principio, los datos recogidos creemos que pueden llegar a sernos útiles. Es por eso que aunque la minería de datos puede ser aplicada a cualquier tipo de información variando solo las técnicas de minería de datos a utilizar en cada tipo estructura de datos analizados, principalmente la extracción de datos procede de bases de datos relacionales, bases de datos espaciales, bases de datos temporales, bases de datos documentales y bases de datos multimedia, existe además desde el advenimiento de Internet una fuerte tendencia de extraer información especialmente desde la World Wide Web³.

2.3. Tipo de modelos

La minería de datos tiene como función principal encontrar en los datos patrones para extraer información oculta, con el fin de descubrir conocimiento. Para lograr este objetivo se deben concretar relaciones entre los datos que constituyen el modelo, para ello existen formas diferentes de representar estos últimos y con ello también diferentes técnicas que se

³ A pesar de lo que se cree comúnmente Internet no es lo mismo que www, ya que esta última al igual que otros servicios, utiliza esta red de redes llamada Internet que se compone de millones de ordenadores interconectados.

utilizan para deducirlos. Así los modelos pueden ser de carácter descriptivo o predictivo, las predicciones nos sirven para prever el comportamiento futuro de algún tipo de entidad, mientras que una descripción nos puede ayudar a su comprensión. Los modelos predictivos además pueden ser descriptivos (hasta donde sean comprensibles por personas) y a su vez los modelos descriptivos pueden emplearse para realizar predicciones.

3. Fases de un proyecto de minería de datos

La reciente necesidad sobre todo de la industria por explotar el potencial de sus enormes y crecientes acumulaciones de datos en medios informáticos ha impulsado a vendedores de tecnologías y organizaciones de consultoría a crear metodologías o procesos para el uso de las herramientas computacionales disponibles que implantan los algoritmos propios de minería de datos, de allí que se hable comúnmente del proceso de Minería de datos e incluso se denomine de igual forma como si fuesen la misma cosa. Al respecto Rhodes (2002) hace la distinción entre KDD como el proceso de descubrir conocimiento mientras que la explotación minera de los datos es el método de extraer patrones de los datos.

Algunos de los estándares como el CRIPS-DM (Cross Industry Standard Process for Data Mining)⁴ es utilizado en el ámbito industrial por mas de 160 empresas e instituciones de todo el mundo y su génesis se debe a la necesidad de concretar una estandarización, el CRIPS-DM propone un modelo neutral para la industria y herramientas, como también un modelo general de procesos para proyectos de minería de datos.

El otro de los modelos es el desarrollo de un proyecto de minería de datos o más específicamente el descubrimiento de conocimiento en bases de datos, comúnmente conocido como KDD , independiente de la técnica específica de extracción del conocimiento siempre son las mismas, consta de una secuencia en la que intervienen distintas etapas, el número de estas fases son cinco (Fayyad et al., 1996; Gamberger, 2001; Hernández et al, 2005) tal como lo demuestra la Figura 2, sin embargo, existe una fase previa que (Edelstein, 1999) hace referencia y que incluye el modelo presentado por Two Crows Corporation, también llamado por ellos Minería de Datos para el Descubrimiento de Conocimiento, este

⁴ Para ver más sobre este estándar se puede visitar la guía completa para la minería de datos en : <http://www.crisp-dm.org/CRISPWP-0800.pdf>

modelo toma muchas cosas de su propia experiencia, de los procesos de KDD y CRISP-DM.

Dentro de los elementos que incluye el KDD se destaca el determinar y analizar las necesidades de la organización y la definición del problema, lo que Hernández et al. (2005) a pesar que la menciona, no la considera como una fase propia del proceso. El desarrollo del descubrimiento del conocimiento es además iterativo e interactivo, por lo que las fases del proceso pueden ser en cualquier momento interrumpidas para volver a comenzar en alguno de los pasos anteriores, siendo este proceso de iteración muchas veces necesario para poder lograr un descubrimiento de conocimiento de alta calidad.

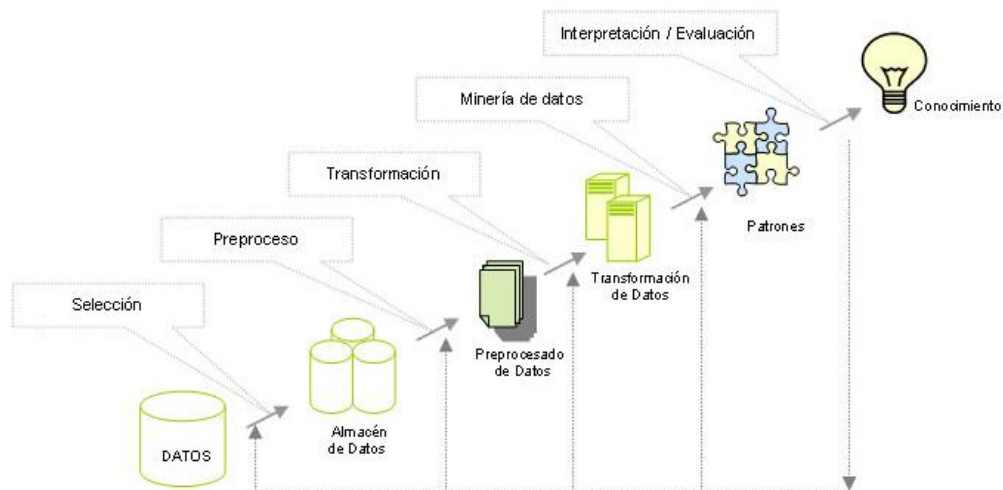


Figura. 2 Fases del Proceso de Extracción del Conocimiento (KDD) según Fayyad.

3.1. Integración y Recopilación

En esta primera fase, intervienen los datos o fuentes de información en bruto que se cree pueden ser útiles para la investigación, además aquí se define como y de donde extraemos esa información. Quizás este es uno de los procesos más lentos y el cual requiere mayor atención ya que comúnmente debemos integrar gran cantidad de datos en un almacén

de datos también llamado Datawarehouse⁵. Estos datos tienen la particularidad de proceder en la mayoría de los casos de distintas bases de datos. Este almacén de datos que consta de información similar, relevante y generalmente única, sirve para poder detectar inconsistencia o redundancia de datos y como un primer acercamiento con los datos a estudiar o que potencialmente interesan ser estudiados (Inmon, 1996^a). Según Hernández et al. (2005) la creación de este almacén puede ser obviado si la cantidad de datos es relativamente pequeña, para estos casos este almacén de datos puede ser sustituido por datos heterogéneos en una hoja de cálculo, o de texto. Una vez terminado el proceso de integración y recopilación de información es necesario limpiar, identificar errores, o datos faltantes siendo esta la segunda fase en un proyecto de minería de datos.

3.2. Filtrado datos; limpieza, selección y transformación

Después de recoger y almacenar los datos que potencialmente son factibles de minar se necesita eliminar la cantidad de aquellos que sean redundantes y dejar o filtrar aquellos que sean de mejor calidad para el proceso de minado. Se destaca en esta etapa tener en cuenta aquellos datos que faltan o se obviaron por error, ya que la no identificación de estos en esta fase puede determinar la calidad de los resultados finales. Al igual que los datos faltantes la identificación o selección de aquellos datos que son relevantes es de suma importancia para poder lograr una buena tarea de minería de datos (Hernández et al., 2005 p. 23). En lo que respecta a la transformación en la etapa de filtrado de datos, se puede transformar los datos para facilitar y/o mejorar la calidad de los resultados, por ejemplo se pueden asignar números a aquellos datos que no sean susceptibles en enumerar, lo que conlleva a utilizar un mejor espacio y aplicar técnicas numéricas. El caso inverso, transformar datos numéricos en datos discretos o nominales se denomina discretización.

3.3. Minería de datos

La palabra descubrimiento en el ámbito de la minería de datos está relacionada con el hecho de que mucha de la información valiosa es desconocida con anterioridad, por lo que

⁵ Un almacén de los datos o Data Warehouse es un expediente de una organización más allá de la información transaccional y operacional, almacenada en una base de datos diseñada para favorecer análisis y la divulgación eficientes de datos.

esta fase dentro de un proceso de KDD es el momento en que gracias a las técnicas existentes para ello se puede confirmar cualquier sospecha sobre el comportamiento del sistema en un particular contexto. Por lo tanto, el objetivo del proceso de minería de datos es poder lograr crear conocimiento a partir de los datos recogidos con el fin de que este conocimiento sea utilizado por el usuario. Es partir de esta fase en donde se debe crear un modelo que se debe basar en los datos recopilados en las fases anteriores.

Las nuevas y abundantes técnicas disponibles y de los algoritmos involucrados en la fase de modelación hacen de esta la fase más interesante del proceso de Minería de Datos. Los pasos importantes en la fase de modelación son:

- Selección de la técnica de modelación: Al principio del proceso se establece el problema a resolver y la meta de Minería de Datos implicada, siguiendo luego el momento de seleccionar una técnica de Minería de Datos en concreto. Cuando se escoge una técnica apropiada entre numerosas técnicas de modelación disponibles en la minería de datos se debe tener en cuenta el objetivo principal del proyecto. La primera división de las técnicas de modelación de Minería de Datos está hecha en base al tipo de tarea de descubrimiento de conocimiento que se desea conseguir, estas tareas pueden ser la Predicción o Descripción. La primera intenta en base a los resultados del análisis adelantarse a ciertos eventos que tengan relación con el objeto de estudio, la tarea de descripción por su parte pretende crear un panorama sobre el estado actual del fenómeno estudiado. La figura 3 muestra algunas clases de tareas de modelación y las técnicas de Minería de Datos adecuadas.

TAREA DE MODELACION	TECNICAS DE MINERIA DE DATOS
Clasificación	Métodos de inducción de reglas, Árboles de Decisión, K vecinos más cercanos, razonamiento basado en casos.
Predicción	Análisis de regresión, Árboles de regresión, redes neuronales, K vecinos más cercanos.
Análisis de Dependencia	Análisis de Correlación, Análisis de regresión, Reglas de Asociación, Redes Bayesianas, programación con lógica inductiva.
Segmentación o Agrupación	Técnicas de Agrupación, redes neuronales, técnicas de visualización.

Figura 3. Tareas de modelación y las técnicas de minería de datos adecuadas

- **Pruebas para el Modelo:** Luego de construido un modelo, se debe generar un procedimiento o mecanismo para probar la calidad y validez del mismo.
- **Construcción del Modelo:** Una vez que la técnica de modelación ha sido evaluada y seleccionada, se procede a ejecutarla sobre los datos previamente preparados para generar un modelo. Todas las técnicas de modelación tienen un conjunto de parámetros que determinan las características del modelo a generar. La selección de los parámetros óptimos para la técnica de modelación es un proceso iterativo y se basa exclusivamente en los resultados generados, estos por su parte deben ser interpretados y su rendimiento justificado.
- **Calificar el Modelo:** Una vez que los modelos son generados, estos son interpretados de acuerdo al conocimiento preexistente del dominio y los criterios de éxito preestablecidos.

Algunos autores como (Hernández et al., 2005) ponen énfasis en el carácter dinámico del proceso de extracción del conocimiento ya que menciona reiteradamente la posibilidad de que en la fase del modelado quizás se tenga que volver a comenzar desde otras fases anteriores debido a que se necesitaría recurrir a nuevos datos, agrega que incluso esta iteración podría llegar a modificar la definición del problema.

3.4. Interpretación y evaluación

Cada vez que se llega a esta fase los patrones deben ser evaluados y revisados, incluso en el caso de que sea necesario se debe volver a repetir. Por lo tanto, después de construir un modelo se debe evaluar sus resultados e interpretar sus significados. Debe tenerse presente que la confiabilidad calculada para el modelo sólo es aplicable para los datos sobre los que se realizó el análisis.

Existen múltiples técnicas de evaluación, la más básica es la de validación simple. Esta técnica de validación reserva un porcentaje de la base de datos como prueba y se aísla después de trabajar con la base de datos “real”, el segundo grupo de datos es utilizado para construir un segundo modelo pero este último será utilizado para predecir los resultados del grupo de datos originales. En segundo lugar se pueden emplear múltiples medios para interpretación de resultados, uno de los más utilizadas son las matrices de confusión (Edelstein, 1999) siendo muy empleadas en problemas de clasificación. Consiste en una tabla que indica cuantas clasificaciones se han hecho para cada tipo, en donde la diagonal de la tabla representa las clasificaciones correctas.

Predicción	Clase A	Clase B	Clase C
Clase A	45	2	3
Clase B	10	38	2
Clase C	4	6	40

Figura 4. Matriz de Confusión

3.5. Despliegue de los resultados: difusión, uso y monitorización

Una vez que el modelo ha sido construido y validado puede ser empleado en una o dos formas. La primera de ellas consiste en que un analista familiarizado con la organización recomiende acciones basadas simplemente en la observación del modelo y sus resultados. La segunda consiste en aplicar el modelo a diferentes conjuntos de datos, por ejemplo, para marcar ciertos registros según su clasificación o asignarles puntuaciones tales como la probabilidad de acción, un ejemplo de ello es la probabilidad de respuesta favorable a una campaña de publicidad telefónica.

El Monitoreo y Mantenimiento es importante para ver como evoluciona el modelo implantado. Si los modelos resultantes del proceso de Minería de Datos son desplegados en el dominio del problema como parte de la rutina diaria, se aconseja preparar estrategias de monitoreo y mantenimiento para ser construidas sobre los

modelos. La retroalimentación generada por el monitoreo y mantenimiento pueden indicar si el modelo está siendo utilizado apropiadamente. El reporte final es la conclusión del proyecto de Minería de Datos. Resume los puntos importantes del proyecto, la experiencia ganada y explica los resultados obtenidos.

4. La minería de datos complejos

Cuando hablamos de proyectos de minería de datos en los que se analizan datos extraídos de repositorios podemos distinguir aquellos en los que se extraen información simple y otros en los que los datos son más complejos y/o heterogéneos. Dentro de este segundo grupo destacan la minería de datos basada en base de datos espaciales, temporales o con elementos multimedia. Dado el creciente aumento de Internet cada vez se hace necesario el crear y utilizar técnicas de minería de datos para el desarrollo de proyectos en los que los datos son extraídos desde documentos o bases de datos web, algunas de estas técnicas son la minería de datos de textos, minería de documentos en la web y de documentos xml.

4.1. Minería datos heterogéneos en bases de datos

La minería de datos heterogéneos, también llamada minería de datos complejos es una de las áreas de la minería de datos en la que la información extraída de los repositorios es especialmente difícil de analizar debido a su naturaleza. Este tipo de información esta patente en casi todas las bases de datos, un ejemplo de ello son las transacciones que se producen a diario en un sistema de préstamo automatizado de una biblioteca donde cada día esta información es parte de una secuencia de eventos, así la cantidad de usuarios que realizan las transacciones van concadenando una evolución determinada por dos factores principales; durante un período de tiempo y por un grupo de personas determinado.

El factor espacio es determinante para concluir el análisis de bases de datos en donde la información extraída es principalmente de tipo geográfica. Siguiendo el ejemplo de la biblioteca quizás sería importantísimo en un estudio en bases de datos de bibliotecas públicas poder conseguir un censo de todas las localidades en la que los usuarios menores de 18 años sobrepasan el umbral de préstamos diarios con respecto a los mayores de 18 años. En este ejemplo de análisis el aspecto espacial de la información que sería vital para el análisis serían

los datos correspondientes a las localidades en que están insertas el grupo de bibliotecas a estudiar.

La información multimedia es otra tipología de datos heterogéneos, se entiende el significado de esta como el utilizar conjunta y simultáneamente diversos medios, como imágenes, sonidos y texto, en la transmisión de una información. Muchas bases de datos hoy en día se componen de esta clase de información y en variadas ocasiones de forma conjunta ya que por su naturaleza la información multimedia se compone de audio, video, texto e imagen integrados.

Debido a las características propias de los datos heterogéneos, los algoritmos tradicionales de extracción de información no permiten que sean tratados directamente (Hernández et al. p. 526). Esto provoca muchas veces distorsión del análisis y del conocimiento propiamente tal extraído como resultado de la aplicación de las técnicas de minería de datos tradicionales. Para ellos según el propio Hernández actualmente se están creando nuevas técnicas aplicables a bases de datos que posean este tipo de información en particular.

4.2. Minería Web

Según Hernández et al. (2005) Etziony (1996) definió la minería web como el uso de técnicas de minería de datos para descubrir y extraer información automáticamente desde la Word Wide Web utilizando para ello herramientas de minería web que analizan y procesan los datos de la web con el fin de producir información significativa. Debido a que los contenidos de Internet se componen de varios tipos de datos, como el texto, imagen, vídeo, metadatos o los enlaces, investigaciones recientes usan el término Multimedia Data Mining (Minería de Datos Multimedia) como una forma de describir la minería web para tratar ese tipo de información (Zaiane, 1998).

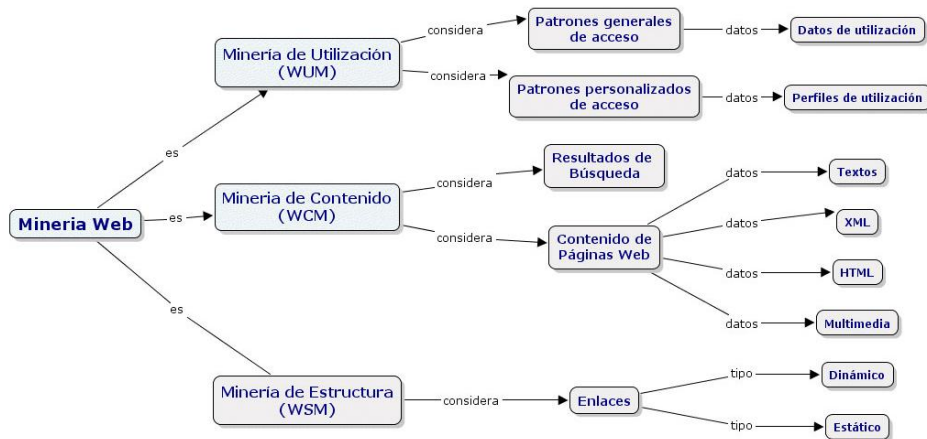


Fig 5. Clasificación minería web según Juan Carlos Dürsteler

Los accesos totales por dominio, horarios de accesos más frecuentes y visitas por día, entre otros datos, son registrados por herramientas estadísticas que complementan todo el proceso de análisis de la minería de datos. Para lograrlo la minería web se descompone en las siguientes tareas:

- a) Descubrimiento de recursos: localización de documentos relevantes o no de la red. Ésta es la función de los índices buscadores (que extraen contenido basándose en palabras, zonas del documento, idioma) y de los índices temáticos (que clasifican los documentos).
- b) Extracción de información: extracción de determinada información a partir de un documento, independientemente de su formato (HTML, XML, texto, etc.).
- c) Generalización: descubrimiento de patrones generales a partir de sitios web individuales (agrupamiento de documentos o clustering, asociaciones entre documentos).
- d) Análisis, validación e interpretación de los patrones.

Por lo tanto la minería web utiliza las técnicas de la minería de datos para descubrir automáticamente los documentos y servicios de la web y extraer información de ellos, información que implica distintos tipos de datos: texto, semi-estructurado, imágenes, audio, entre otros. Así el ámbito de acción intenta involucra problemas que permitan entender mejor la Web, el como aprovechar la información que contiene y facilitar el acceso a ella.

4.2.1. Minería de uso web

La minería de uso web o también llamada minería de utilización y en inglés Web User Mining (WUM), intenta descubrir patrones de uso de los objetos digitales en un ambiente web, para ello se divide en dos áreas principales de trabajo. Una de ellas es descubrir patrones de generales de acceso en los que los datos de utilización son la materia prima para lograr el objetivo. La otra área se encarga de descubrir patrones personalizados o individuales de acceso y para ello utiliza los datos que hacen referencia al perfil de los usuarios de los documentos digitales presentes en la web.

4.2.2. Minería de contenido web

La minería de contenido web se relaciona tal como lo indica su nombre a los contenidos inmersos en las parcelas web analizadas en el proceso de minería de datos. Los diversos contenidos pueden ser divididos en dos áreas principales de acción. El contenido de los resultados de búsqueda en la web y los contenidos incluidos en las paginas web, estos contenidos que insertos pueden contener información o datos de diversa índole, en los que se encuentran los datos de las páginas html, los datos multimedia, datos xml y de textos.

4.2.3. Minería de estructura.

Este tipo de minería web analiza los patrones susceptibles de encontrar en los enlaces dispuestos en documentos hipertextuales en Internet. Es por eso que a diferencia de las dos técnicas anteriores su materia prima no se basa en tipos de datos sino más bien en la clase de enlaces analizados ya que consiste en estudiar la estructura de los enlaces entre documentos e intradocumentos. Las técnicas se inspiran en el estudio de las redes sociales y el análisis de citas (Chakrabarti, 2000), así una página (persona o artículo) está reforzada por la cantidad de referencias (amistades o citas) que tiene. El análisis de la estructura es muy útil a la hora de descubrir:

- Autoridades, que proporcionan la mejor fuente sobre un tema determinado.
- "Hubs", que proporcionan colecciones de enlaces a autoridades.

5. Bibliomining y la minería de datos aplicada a los entornos documentales

Una de las posibles aplicaciones de la minería de datos y del KDD es el llamado bibliomining, esta variante de aplicación esta relacionada con la biblioteconomía y documentación ya que se pretende que los resultados obtenidos de la aplicación de las técnicas de minería de datos sean útiles para entender las comunidades de usuarios o bien para aplicarlos en cualquier entorno relacionado con el ámbito bibliotecario. El nombre es citado por primera vez por (Nicholson & Stanton, 2003) y nace por la necesidad de diferenciar o unificar los términos librería y minería de datos, ya que estos en su conjunto hacen referencia al un grupo de herramientas de software para la minería de datos. Por otro lado el término se asemeja bastante a otra disciplina muy parecida llamada bibliometría, la cual tiene como fin, descubrir patrones de comunicación en la literatura científica (Pritchard, 1969).

El uso de la minería de datos en el ámbito bibliotecario viene de la mano con el advenimiento de las nuevas tecnologías en las bibliotecas, con la adopción de catálogos automatizados paralelamente se mejoraron las técnicas y los métodos estadísticos de la bibliometría y de la visualización para localizar patrones no comunes inmersos en grandes cantidades de datos. Por consiguiente el bibliomining se refiere al uso de estas técnicas que permiten sondear las enormes cantidades de datos generados por las bibliotecas automatizadas (Nicholson, 2003).

Autores pioneros en el campo de las bibliotecas comenzaron a explorar aplicaciones sofisticadas a los datos extraídos de este tipo de organizaciones algunos años antes de que el concepto de la minería de datos se popularizara. Uno de ellos es Nutter (1987) el cual exploró los datos de la biblioteca con el fin de tener herramientas de apoyo a la toma de decisiones. Weibel y Oskins desarrollaron un sistema para la catalogación automatizada basada en los títulos de los libro. También Mancini (1996) habla sobre el como de extraer de los datos información útil para la toma de decisiones a nivel de dirección de una biblioteca, por su parte Peters (1996) habla sobre el como examinar los datos extraídos de la colección con el fin de contar con información fidedigna para apoyar la administración de los fondos de la biblioteca.

Nicholson (2005) menciona que más recientemente, Banerjee (1998) se centró en describir cómo es el proceso de minería de datos y las maneras de usarlo para proporcionar un acceso mejor a la colección. Comenta además como Doszkocs (2001) discutió el potencial para aplicar redes neuronales a los datos de la biblioteca para descubrir asociaciones posibles entre los documentos. Liddy (2000) combinó el proceso del lenguaje natural con la minería de datos textual para descubrir la información oculta en colecciones de las bibliotecas digitales. El mismo Nicholson aclara que estos proyectos compartieron una visión en común y este es el de mejorar y automatizar dos de las funciones de una biblioteca; las adquisiciones y administración de la colección. A pesar de lo anterior, se desprende que en ninguno de estos proyectos se planteó el uso de los datos extraídos de la biblioteca para ser utilizados como herramienta para apoyar las decisiones estratégicas a nivel de gerencia global para las bibliotecas. Algunos autores han comenzado a tratar este campo centrándose en los usuarios de la biblioteca por ejemplo Schulman (1998) utilizó la minería de datos para examinar las tendencias que cambiaban en comportamiento del usuario de la biblioteca, por su parte Chau (2000) planteó el uso de la explotación de la minería de datos Web para personalizar los servicios de referencia electrónica.

En definitiva el bibliomining a pesar que en sus comienzos su foco de atención era la colección de la biblioteca, hoy en día tiene como meta encontrar patrones de comportamiento de grandes grupos de usuarios, del propio personal de las bibliotecas o de ambos, con el fin de ayudar a los encargados de biblioteca a tomar mejores decisiones. La necesidad de encontrar patrones entre los datos de grupos de usuarios asociados a las bibliotecas puede ser que sea de mucha utilidad al ser explorados en amplios grupos demográficos, sin embargo, cabe destacar que el utilizar algunos de estos datos sin el previo permiso de cada individuo en algunos casos puede ser poco ético y en otros ilegal.

5.1. Las fases del proceso de bibliomining

El bibliomining al igual que un proceso de extracción de conocimiento en cualquier base de datos debe cumplir con un proceso sistemático que pueda desencadenar en el descubrimiento de conocimiento. A pesar de que el proceso de bibliomining se basa en planteamientos básicos para la extracción del conocimiento, Nicholson (2003) pone de manifiesto que el almacenamiento y el aislamiento de los patrones es casi la tarea más importante dentro del proceso. Propone y destaca además que en él se debe procurar aislar los datos propios de la

base de datos con los datos de los usuarios. Así una vez este creado el almacén de datos se procede a utilizar herramientas de exploración y visualización para lograr un conocimiento mas acabado de las comunidades de usuarios y la utilización de los recursos de la biblioteca. El mismo autor propone una división de las fases del proceso de bibliomining en las siguientes etapas:

5.1.1. Determinación del área de análisis

Como todo proyecto, la determinación de empezar un proyecto de bibliomining es desencadenado por una necesidad o problema específico que se tiene en la biblioteca. Así el problema a resolver puede ser específico de la biblioteca, alguna área parcial o bien que abarque un problema global que necesite de extracción de información para la toma de decisiones. Según Nicholson (2003) existen dos formas de lograr determinar el área de análisis; el primero es el enfocar el problema encontrado a una exploración de datos, esto es ; hay un problema específico que conduce la exploración de los datos. La segunda de ellas es utilizar la minería de datos por parte del bibliotecario para lograr una solución o idea mejor de un área concreta de la organización.

5.1.2. Identificar fuente de datos internas y externas de la organización

Como todo trabajo de extracción de conocimiento el identificar y descubrir los datos realmente útiles para nuestro propósito es el paso a seguir después de que hemos determinado el área de foco o problema que queremos resolver con nuestra investigación. Esto se hace aun más difícil en el ámbito bibliotecario cuando se desea saber el comportamiento de las transacciones de los usuarios de nuestro centro de información, ya que se deben aislar y proteger los datos pertenecientes a nuestros “clientes” por que en la mayoría de los países la protección de los datos de los ciudadanos es primordial. Se debe tener en cuenta además que los datos a identificar deben ser internos y externos, el primero de ellos son los datos extraídos dentro del propio entorno del sistema bibliotecario. Los datos externos por su parte son aquellos datos que sin ser propios del sistema bibliotecario son importantes debido que a proceden de los entornos en el cual esta inmerso el sistema bibliotecario, un

ejemplo de esto puede ser el uso de los datos de nivel socio-económico de los estudiantes que visitan una biblioteca determinada dentro de un campus universitario cualquiera.

5.1.3. Crear almacén de datos

Como muchos otros autores, Nicholson (2003) propone que para poder cruzar o combinar diversos tipos de datos la manera mas eficaz y segura es crear un almacén de datos, este almacén es una base de datos paralela a los sistemas operacionales y con la garantía de estar “limpios” y adaptados en su formato para facilitar su análisis. Inmon (1996b) ha propuesto una serie de factores a tener en cuenta en la construcción de un almacén de datos y que consiste en un conjunto de ideas que se deben tener en cuenta a la hora de considerar qué es un almacén de datos y cuando es necesario proceder a su construcción.

Por lo tanto la tarea del almacén de datos es crear una fuente de datos que contenga información filtrada, y en la que a su vez la información personal de cada usuario nunca debe ser guardada (Nicholson, 2003). Al igual que la mayoría de los trabajos de análisis de datos, la fase de creación del almacén es el más lento de todo el proceso, incluso que el propio minado de datos (Berry & Linoff 1997), por lo que es primordial empezar por un campo de estudio bastante definido y acotado, para mas tarde ir progresando a un análisis mas global de la situación.

5.1.4. Seleccionar herramientas apropiadas para el análisis

Para el análisis de los datos procedentes de una base de datos bibliotecaria se pueden utilizar tres herramientas posibles la primera de ellas es el Proceso Analítico en Línea (On-Line Analytical Processing), el cual permite a los usuarios o responsables administrativos poseer una herramienta a tiempo real, interactiva y flexible que funcione en base a consultas e informes sobre datos específicos de la organización, por ejemplo, monitorizar la cantidad de revistas en línea utilizadas en un tiempo determinado.

Otras de las herramientas posibles de ser utilizadas son los Sistemas de Información de Gerencia (Management Information Systems), el cual provee a los encargados una

manipulación básica de los datos, esta herramienta normalmente vienen incluidas en el paquete de los Sistemas Integrados de Bibliotecas (Integrated Library System), se destaca que para potenciar el uso del Sistema de Información de Gerencia se debe adaptar este al almacén de datos, así tendrá un mejor alcance y proporcionara la información que la biblioteca realmente necesita saber. Así los sistemas de notificación podrían monitorizar, por ejemplo, la actividad baja o alta en un punto determinado de la biblioteca proporcionando información en tiempo real al encargado de la biblioteca para distribuir el personal de la misma según lo requiera la situación.

La otra opción de análisis es la minería de datos, la cual según Nicholson (2003) inspira el nombre de bibliomining. Esta técnica consta de dos categorías principales a tener en cuenta para la extracción del conocimiento; la predicción y la descripción, siendo el primera de ellos la forma que podemos pronosticar la forma que evolucionaran ciertos patrones de comportamiento de nuestra biblioteca y el segunda la manera de como podemos tener una situación actual del área de estudio determinada para el análisis.

5.1.5. Analizar y disponer de los resultados

Una vez obtenidos los resultados, los informes, modelos y las posibles ayudas extraídas para la toma de decisiones estas deben ser validadas. La forma más común de estas validaciones es probar los datos sobre una nuestra de datos que no fueron utilizados para construir el modelo. A pesar de lo anterior, en el caso del bibliomining la técnica más importante es la validación que puede hacer un experto en la materia, por lo tanto el bibliotecario que este mejor familiarizado con el sistema o el contexto de la biblioteca. El análisis de este profesional se puede considerar crucial ya que si este considera que el o los patrones encontrados en los resultados del modelo no son los apropiados, se debería proceder a explorar nuevamente la secuencia de los datos.

5.2. Beneficios de la aplicación del bibliomining

En general los beneficios de aplicar técnicas de minería de datos en sistemas automatizados de bibliotecas posibilitan llevar de mejor forma la gran tarea de apoyar a la

gestión de las mismas, justificar sus servicios, o la adquisición de material, todo esto a través del conocimiento de los usuarios y el personal de las bibliotecas.

La biblioteca ha pasado de ser un lugar en donde sólo se depositan libros a tener un espacio predominante en las actividades de muchas organizaciones comerciales o públicas. Con la masificación de Internet las bibliotecas tuvieron la oportunidad de concebir a las bibliotecas digitales, así grandes cantidades de usuarios pueden hoy en día visitar la biblioteca sin estar físicamente en ella. Las bibliotecas debido a este fenómeno se encuentran con nuevas perspectivas, la primera es que las bibliotecas tienden cada vez a servir a las organizaciones en las cuales están insertas; las académicas sirven a las comunidades universitarias, las públicas a comunidades locales y las corporativas a entes privados (Nicholson, 2003). Otros de los elementos es que las bibliotecas desempeñan un papel preponderante dentro de sus instituciones como depósito y abastecedoras de recursos de información. Estos elementos ponen en evidencia que el bibliomining aporta beneficios; primero por que puede comprobar qué usuarios necesitan información, en que formato y en segundo lugar de que manera son respondidas esas necesidades del usuario por parte de la organización. Por lo tanto el bibliomining permite descubrir además, los patrones que tienen significado en el contexto de la institución en que esta inmersa la biblioteca.

Si bien, la mayoría de las bibliotecas automatizadas utilizan con frecuencia bases de datos relacionales para almacenar información referente al usuario, de los recursos, de las transacciones e incluso registros bibliográficos buscados, hoy en día la tendencia marcada por la biblioteca digital aumenta la oferta de los recursos de información en formato. En este sentido Nicholson (2003) destaca que la mayoría de las actividades siguen siendo iguales que la biblioteca tradicional. Las bibliotecas digitales no tienen una función de la circulación; si los usuarios necesitan algo, lo guardan en soportes digitales o bien lo imprimen desde su propia estación de trabajo. Por otro lado la biblioteca digital permite rastrear y registrar a través de sus sistemas de información la visita completa de un usuario. Así los datos correspondientes a esta información almacenados en las bases de datos de bibliotecas tradicionales y digitales representan los patrones del comportamiento de dos sectores importantes de la biblioteca: el personal y los usuarios.

No importa que los usuarios de la biblioteca sean el público de una comunidad local o el personal interno de una corporación grande como puede ser una universidad, en ambos

casos al comprender los procesos de búsqueda, préstamo, y los patrones relacionados con el comportamiento pueden indicar si han obtenido los recursos de información que necesitaban, qué recursos de la información encuentran más útiles, o bien las posibles necesidades o requerimientos futuros de los usuarios.

Por último cabe destacar que, proporcionando la información de cómo funciona y la utilidad de la biblioteca como unidad, el bibliomining puede proporcionar la justificación para la ayuda financiera e institucional permanente para la gestión de la biblioteca.

5.3. Aplicaciones de bibliomining en bibliotecas

Biblioteca actual posee muchas herramientas informáticas la cual le permitiría utilizar el bibliomining con el fin de proporcionar una comprensión más profunda de las fuentes individuales comentadas con anterioridad. A pesar de esto se puede descubrir mucha más información cuando las fuentes se utilizan conjuntamente con una otra. La mayoría de estas fuentes de datos contienen los campos que se pueden utilizar para cruzarlos con otras fuentes.

5.3.1. Información del usuario

Los usuarios de los servicios de la biblioteca son una de las áreas más importantes de la mayoría de este tipo de organizaciones. La mayoría de las bibliotecas existen para responder a las necesidades de información de usuarios, y por lo tanto, entender esas necesidades es crucial para el éxito de cualquier biblioteca. El comportamiento del usuario que es examinado en un nivel individual puede ayudar en entender al individuo, pero esto dice a muy poco al bibliotecario sobre el conjunto. Examinar los comportamientos de un grupo grande de usuarios para conseguir patrones puede permitir que la biblioteca tenga una idea mejor de las necesidades de información y por lo tanto modificar y mejorar los servicios de la biblioteca en base a los requerimientos particulares de los usuarios. Por último cabe destacar que el bibliotecario depende de las encuestas para recopilar la información del usuario, pero con bibliomining, pueden descubrir patrones similares sin perder el tiempo del usuario, recursos humanos o dinero para confeccionar dichos exámenes (Estabrook, 1996).

5.3.2. Toma de decisiones

El bibliomining se puede utilizar para ayudar a los encargados de biblioteca a supervisar su organización y tomar decisiones estratégicas. Por ejemplo, mientras que el comportamiento del usuario se puede conocer a través del ILS (Integrated Library System), el comportamiento del personal de la biblioteca también puede ser descubierto al analizar como trabajan estos con las bases de datos. Mientras que la supervisión del personal y su funcionamiento puede ser un concepto incómodo para muchos bibliotecarios, los presupuestos cada vez más apretados y las la constante justificación de servicios requieren tener constantemente en cuenta el funcionamiento interno de la biblioteca. Las investigaciones han demostrado que las medidas objetivas en evaluaciones de funcionamiento puede mejorar realmente la imparcialidad y la eficacia de esas evaluaciones (Stanton, 2000). Cabe destacar que los conceptos sugeridos aquí no se piensan como sustituto de métodos más tradicionales de medición del desempeño del personal de la biblioteca pero si como una herramienta de apoyo más.

En el ámbito del desarrollo de la colección se considera a la información que no es utilizada como un gasto inadecuado, por lo tanto el bibliomining puede proporcionar información hasta el momento oculta en cuanto a cómo se justificaron y adquirieron esos recursos para la biblioteca. Buscando correlaciones entre los documentos de poco uso y los títulos, el editor, el vendedor, el plan de la aprobación, la fecha, el formato o soporte, los documentos en el almacén de los datos, los encargados podrían descubrir áreas problemáticas en la propia colección o la organización de la biblioteca.

El bibliomining también puede ayudar a tomar decisiones sobre el funcionamiento de los servicios bibliotecarios. Si se consigue por ejemplo conseguir información sobre tiempos diarios de transacciones, estos pueden ser utilizados para construir patrones del comportamiento en el área de circulación. Observando estos patrones, los encargados de biblioteca pueden optimizar el número de los miembros de personal que se necesitan en el área de circulación.

En el área de adquisición, la selección y el precio del vendedor se pueden analizar sin problema alguno. Por ejemplo, si un miembro del personal utiliza constantemente a un

vendedor más costoso cuando existen otras alternativas más económicas, el bibliotecario de adquisiciones deberá estar implicado con políticas más claras sobre la selección del vendedor. Los patrones que buscan en el tiempo toma entre el recibo de un libro y la colocación de ese trabajo sobre el estante puede traer a la luz algunas áreas que necesiten la atención.

Si en un proyecto de de bibliomining se considera combinar los datos de las adquisiciones y de las bases de datos de circulación, se puede intentar buscar los patrones que predicen, por ejemplo, el bajo uso de algún material en especial, por lo tanto estos resultados pueden ayudar en la selección apropiada de vendedores y de planes de compras.

Como con todos los proyectos de bibliomining, obtener más información favorece que los analistas descubran patrones significativos. Por lo tanto, es importante asegurar el registro individual en el sistema del personal implicado con la creación de registro de usuarios, circulación, y otras interacciones con el sistema. Con esta información, los encargados de biblioteca pueden investigar y solucionar problemas con el fin de hacer que el poco dinero del presupuesto se pueda literalmente estirar lo más posible.

5.3.3. Informes y Reportes

Comúnmente las bibliotecas no son una institución independiente ya que casi siempre pertenece o responde a instituciones de ámbito más global. En este sentido las bibliotecas pueden encontrar en el bibliomining una forma de generar informes o reportes que, por ejemplo, responden a las conductas de los usuarios o de la comunidad en que están insertas. Por otra parte esta es una buena forma para lo encargados de las bibliotecas de justificar el financiamiento para sus bibliotecas un ejemplo de ello es permitir que los encargados de biblioteca justifiquen el coste para los cursos de educación de usuarios o los materiales de instrucción para este servicio. Los encargados de las bibliotecas muchas veces se encuentran con quejas sobre las políticas bibliotecarias dentro del centro, el bibliomining, a través de sus reportes puede proporcionar la justificación basada en datos para sostener las políticas para tales discusiones.

En definitiva la utilización del bibliomining puede proporcionar información sobre los grupos de usuarios que utilizan la biblioteca. En el caso de las bibliotecas de digitales se pueden

utilizar la información de los registros extraídos vía web (Liddy, 2000). Por otro lado si se logra identificar a grupos de individuos con intereses comunes o bien con conocimientos avanzados en un área del conocimiento, el bibliomining puede servir de apoyo para los sistemas de gerencia del conocimiento. Esta capacidad puede ser particularmente valiosa dentro de las organizaciones grandes donde los esfuerzos de investigación y del desarrollo se dispersan sobre localizaciones múltiples. Las búsquedas utilizadas y las preguntas de referencia que los usuarios solicitan son la materia prima para encontrar patrones esta ha sido un área explorada por los investigadores Harter & Hert (1997).

5.4. Consideraciones previas al implantar un proyecto de bibliomining

Para el desarrollo de un proyecto de minería de datos y de la disciplina en si misma existen elementos externos que no permiten un completo y efectivo análisis de los datos, uno de ellos es el trabajar con datos personales. En el caso de la biblioteca los datos de los usuarios y el aislamiento de los mismos puede presentar posiblemente un problema menos complicado en bibliotecas corporativas y especiales, sin embargo, en la mayoría de los países las leyes de protección de datos imposibilitan el uso de aislado de los datos de los ciudadanos. De hecho, muchos sistemas automatizados de las bibliotecas de los Estados Unidos ocultan o desechan automáticamente los expedientes correspondientes al área de circulación después de que el material se haya devuelto a la biblioteca.

Otra de las trabas esta relacionada con la anterior y de los beneficios del uso de la minería de datos como es el de permitir trabajar con grandes volúmenes de datos. Así en el caso de las bibliotecas esos grandes volúmenes de datos pueden aumentar su valor si son intercambiados o cruzados con otros datos de bibliotecas que estén bajo el alero de una misma institución. Pero si las leyes del estado prohíben compartir la información del usuario con terceros, entonces el compartir estos grandes volúmenes de datos también puede ser prohibido. Por lo tanto, antes de comenzar un proyecto de bibliomining, los bibliotecarios deben asegurarse de que sus actividades bibliomining bajen dentro de los límites de la actividad legal (Nicholson, 2003).

Dentro de las consideraciones éticas en el uso del bibliomining podemos destacar el seguimiento que le hacemos al usuario para conocer sus comportamientos. Por ejemplo: ¿es ético analizar los datos de un conjunto de usuarios para saber que libros utilizó, que cosas

consultó en los OPACS o que libros buscó en las estanterías?. Si bien esta es una práctica que lleva desde hace décadas haciéndose, las bibliotecas se conocen como un lugar para efectuar consultas o una investigación de forma reservada. Así los usuarios realizan su investigación confiando en que la biblioteca resguarde la confidencialidad con el fin de que no sea observado el desarrollo de la misma. Por lo tanto, aunque es legal aplicar estas herramientas, también cabe considerar estas razones éticas al momento de aplicar las técnicas de bibliomining.

Existen algunas formas de sortear estos inconvenientes, la más fácil de ellas es desarrollar políticas de aislamiento de los datos a utilizar, pero sin duda una de las mejores y más acertadas soluciones es consultar e informar previamente al usuario sobre la utilización que se hará con los datos analizados. Esta opción es la más tediosa ya que requieren el uso de mayor tiempo y dinero en consultas por teléfonos, e-mails o persona a persona, pero sirve para informar de lo que se está haciendo y también como método de asegurarse que el grupo de usuarios por analizar es el que ha aprobado su participación en el proyecto.

Nicholson (2003) analiza otra opción que considera aún más segura, esta no considera examinar ninguna información que conecte al usuario en particular a un historial de préstamo. Sin embargo esta propuesta reduciría la calidad y la personalización del bibliomining, pero puede ser necesario en algunos casos para evitar utilizar la información ética y/o legal. El mismo Nicholson asegura que las instituciones públicas pueden encontrar que ésta es la única solución al problema, mientras que las bibliotecas privadas pueden tener más flexibilidad en lo que pueden hacer. Cualquiera que sea el caso de cada biblioteca, cada una de ellas debe ponderar su situación legal y ética con el fin de manejar mejor su biblioteca y de servir mejor a sus usuarios.

6. Conclusiones

Ha llegado un momento en el que disponemos de tanta información que nos vemos incapaces de sacarle provecho. Los datos tal cual se almacenan no suelen proporcionar beneficios directos. Su valor real reside en la información que podamos extraer de ellos: información que nos ayude a tomar decisiones o a mejorar nuestra comprensión de los fenómenos que nos rodean. Así el proceso de KDD que incluye a la minería de datos utiliza

técnicas y modelos para descubrir conocimiento a partir de información que esta oculta en bases de datos.

Hasta ahora, los mayores éxitos en la minería de datos se pueden atribuir directa o indirectamente a avances en bases de datos (un campo en el que los ordenadores superan a los humanos). No obstante, muchos problemas de representación del conocimiento y de reducción de la complejidad de la búsqueda necesaria (usando conocimiento a priori) están aún por resolver.

Las bibliotecas han recopilado datos sobre sus colecciones y usuarios por años, pero no han sabido utilizar esos datos para una toma de decisión mejor. El bibliomining que integra en esfuerzos actuales de la investigación y de la evaluación permitirá a encargados y a investigadores de las bibliotecas una idea más completa de los recursos contenidos en sus organizaciones y cómo están siendo alcanzados por los usuarios. Así tomando un acercamiento más activo basado en usos de minería de datos, de la visualización de los mismos, y de la estadística, estas organizaciones de información pueden conseguir una visión más clara de las necesidades de entrega y de la gestión de la información.

El desarrollo de la tecnología de Minería de Datos aplicada en bibliotecas contiene una serie de elementos que la hacen operable, sin embargo, existen algunos factores que pueden crear un descrédito o complicaciones para su desarrollo. Uno de ellos es el aspecto legal de la protección de datos de los usuarios de las bibliotecas ya que mientras que el bibliomining puede ayudar a bibliotecarios en la dirección de su biblioteca, las leyes de cada país pueden lograr que el bibliomining sea muy difícil de poder ejecutar. Para ello, si el bibliomining llega a ser más popular, es posible que los bibliotecarios, las bibliotecas, y las organizaciones profesionales puedan trabajar para permitir que este tipo de análisis administrativo sea posible.

Una de las ventajas de la minería de los datos es que puede aprovecharse de cantidades grandes de datos y la información descubierta con el uso de técnicas bibliomining da a biblioteca el potencial de ahorrar el dinero, proporcionar programas más apropiados, resolver más de las necesidades de información del usuario, observar problemas de su colección, y sirve como fuente de información más eficaz de sus usuarios.

El Bibliomining puede proporcionar justificaciones basadas en los datos para las decisiones difíciles, de las peticiones de financiamiento que los encargados de las bibliotecas deben hacer. Finalmente, el bibliomining puede informar a los procesos y a los productos la gerencia del conocimiento que han crecido en importancia dentro de organizaciones contemporáneas.

7. Bibliografía

- BANERJEE, K. (1998). Is data mining right for your library?. *Computers in Libraries*, 18(10), 28-31.
- BERRY, M. and Linoff, G. *Data Mining Techniques For Marketing, Sales, and Customer Support*. New York: John Wiley & Sons. 1997.
- CHAKRABARTI, S. Data mining for hypertext : A tutorial survey. *SIGKDD explorations*, 12 1-11, 2000
- CHAPMAN, Pete et al. *CRISP-DM 1.0 : Step-by-step data mining guide*. [En Línea]. 2000. Disponible en : <http://www.crisp-dm.org/CRISPWP-0800.pdf> [Consultado: 01/06/2006]
- CHAU, M. Y. (2000). Mediating off-site electronic reference services: Human-computer interactions between libraries and Web mining technology. *Fourth international conference on knowledge-based intelligent engineering systems & allied technologies* (Vol. 2, pp. 695-699). Piscataway, NJ: IEEE.
- CLEVELAND, W. S. *Visualizing Data*, AT&T Bell Laboratories, Murray Hill, NJ, Hobart Press, Summit NJ, 1993.
- DOSZKOCS, T. E. Neural networks in libraries: The potential of a new information technology. [En Línea] 1991. Disponible en <<http://web.simmons.edu/~chen/nit/NIT'91/027-dos.htm>> [Consultado: 10/05/2006]
- EDELSTEIN, Herbert. *Introduction to Data Mining and Knowledge Discovery*, Third Edition. Two Cows Corporation. USA. 1999.
- ESTARBROOK, L. Sacred trust or competitive opportunity: Using patron records. *Library Journal*, 121(2), 48-49. 1996.
- ETZIONY, Oren. The World-Wide Web: Quagmire or Gold Mine. *Communications of the ACM*. Vol.39, No.11, November 1996 .pp. 65-68
- FAYYAD, Usama M. et al. *Information visualization in data mining and knowledge discovery*. San Francisco : Morgan Kaufmann , 2002. 407 p.

- FAYYAD, Usama M.; PIATETSKY-SHAPIRO, Gregory; Smyth, Padhraic. [En línea] 1996. The KDD Process for Extracting Useful Knowledge from Volumes of Data. Communications of the ACM, Disponible en: <http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf> [Consultado: 21/05/2006]
- FURNAS, G. W.; BUJA, A. Prosections views: dimensional inference through sections and projections. [En Línea] 1994, Journal of Computational and Graphical Statistics, 3(4), 323-385. Disponible en: <http://citeseer.ist.psu.edu/furnas94prosection.html> [Consultado: 22/05/2006]
- GAMBERGER, D et al. Data Mining Server. Laboratory Information System. [En Línea]. 2001 Rudjer Boskovic Institute. Disponible en : http://dms.irb.hr/tutorial/tut_intro.php [Consulta: 28/05/2006]
- GONZALEZ, Gustavo et al. [En Línea] 2005. Preprocesamiento de bases de datos masivas y multi-dimensionales en minería de uso web para modelar usuarios: comparación de herramientas y técnicas con un caso de estudio. Disponible en: http://eia.udg.es/~gustavog/esp/publicaciones/cedi2005_gustavo_sonia_published.pdf [Consultado: 25/05/2006]
- HARTER, S. P. and HERT, C. A. Evaluation of information retrieval systems: Approaches, issues, and methods. In M.E. Williams (Ed.), Annual review of information science and technology (Vol. 32, pp. 3-94). 1997. Medford, NJ: Information Today.
- HERNANDEZ, José et al. Introducción a la minería de datos. Madrid : Pearsons, 2004. 656 p.
- INMON, William H. (1996a) The data warehouse and data mining. Communications of the ACM Volume 39, Issue 11 (November 1996) Pages: 49 - 50
- INMON, William H. (1996b) Building the Data Warehouse. 2ª ed. New York: John Wiley & Sons, 1996
- KOENIG, M. E. D. Bibliographic Information Retrieval Systems and Database Management Systems. Information Technology and Libraries, 4, 247-272.
- KOPANAKIS, I. and THEODOULIDIS, B. Visual data mining modeling techniques for the visualization of mining outcomes. Journal of Visual Languages and Computing, 14(6), 543-589. 2003
- LIDDY, Elizabeth. Text mining. Bulletin of the American Society for Information Science. 13-14. November/December 2000,
- NICHOSLON, Scott The Basis for Bibliomining: Frameworks for Bringing Together Usage-Based Data Mining and Bibliometrics through Data Warehousing in Digital Library Services. [En Línea] 2005] Information Processing and Management 42(3):pp. 785-804 Disponible en: <http://dlist.sir.arizona.edu/886/01/nicholson2.pdf> [Consultado: 30/052006/]

- NICHOLSON, Scott. and STANTON, J. [En Línea] 2003. Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. In Nemati, H. & Barko, C. (Eds.). Organizational data mining: Leveraging enterprise data resources for optimal performance. Hershey, PA: Idea Group Publishing. 247-262. Updated on 2/16/04 . Disponible en : <<http://www.bibliomining.com/nicholson/odmcom.html> > [Consultado: 01/06/2006]
- NICHOLSON, Scott. [En Línea] 2003 The Bibliomining Process: Data Warehousing and Data Mining for Library Decision-Making. *Information Technology and Libraries* 22 (4). Disponible en: <<http://www.bibliomining.com/nicholson/biblioprocess.htm>> [Consultado: 29/05/2006]
- NUTTER, S. K. Online systems and the management of collections: Use and implications. *Advances in Library Automation Networking*, 1, 125-149. 1987
- PETERS, T. Using transaction log analysis for library management information. *Library Administration & Management* 10 (1), 20-25. 1996.
- PRITCHARD, A. Statistical Bibliography or Bibliometrics?. *Journal of Documentation*, v. 25, n. 4, p. 348-349, Dec. 1969.
- PYLE, Dorian. *Business Modeling and Data Mining*. Morgan Kaufmann, 2003.
- SCHULMAN, S. Data mining: Life after report generators. *Information Today*, 15(3), 52. 1998
- STANTON, J. M. Reactions to employee performance monitoring: Framework, review, and research directions. *Human Performance*, 13, 85-113. (2000).
- ZAIANE O. R. et al. [En Línea] 1998. MultimediaMiner: A System Prototype for Multimedia Data Mining, in Proc. Int'l Conf. on Management of Data (ACM-SIGMOD'98), pp. 581-583, Seattle, 1998. Disponible en : <<http://www.cs.ualberta.ca/~zaiane/postscript/3m98.pdf> > [Consultado: 15/05/2006]
- WITTEN, Ian H. and Frank, Eibe. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco, 2ª edition, 2005.