

## Minería de textos: una herramienta útil para mejorar la gestión del bibliotecario en el entorno digital

MSc. Eleazar Botta-Ferret<sup>1</sup> y Lic. Jania E. Cabrera-Gato<sup>2</sup>

### Resumen

En la etapa actual de desarrollo de la humanidad, la explosión, asimilación y utilización intensiva del conocimiento ha conducido a lo que se ha denominado *Sociedad del conocimiento*, en la que la gestión de la información, la documentación y el conocimiento se perfila como un componente estratégico de primer orden. Se han abierto los horizontes hacia otras profesiones que están llamadas a cooperar con los profesionales de la información: diseñadores de sistemas, proveedores de datos, editores, vendedores, archivistas, ingenieros y especialistas en codificación de texto electrónico, cuyas opiniones y experiencias permitan desarrollar las interfaces correspondientes para facilitar la localización, manipulación, recuperación y uso de la información digital. Se han creado nuevas herramientas para facilitar el acceso al cúmulo de información que se genera diariamente. Una de las más utilizadas a nivel organizacional es la minería de texto (*Text Mining*) que ofrece a la organización la posibilidad de explorar grandes cantidades de textos, no organizados en forma de datos, establecer patrones y extraer conocimientos útiles.

*Palabras clave:* Gestión de la información, conocimiento, minería de textos, información, bibliotecarios, sociedad del conocimiento.

### Abstract

In the current stage of human development, the boom, assimilation and intensive use of knowledge has led to what we know as Knowledge Society, in which information management, documentation and knowledge are considered as vital strategic components. The horizons have been widened towards other jobs summoned to cooperate with information professionals: system designers, data providers, editors, salespersons, archivists, technicians, engineers and electronic text codification specialists, whose opinions and experiences make possible the localization, management, retrieval and use of digital information. New tools have been created to make easier the access to the great amount of knowledge generated daily. One of the most used is Text Mining, which allows the different organizations to explore a great number of texts that are not organized in data form, as well as to establish patterns and to extract useful knowledge.

*Key words:* Information management, knowledge, text mining, information, librarians, knowledge society.

Copyright: © ECIMED. Contribución de acceso abierto, distribuida bajo los términos de la Licencia Creative Commons Reconocimiento-No Comercial-Compartir Igual 2.0, que permite consultar, reproducir, distribuir, comunicar públicamente y utilizar los resultados del trabajo en la práctica, así como todos sus derivados, sin propósitos comerciales y con licencia idéntica, siempre que se cite adecuadamente el autor o los autores y su fuente original.

Cita (Vancouver): Botta Ferret E, Cabrera Gato JE. Minería de textos: una herramienta útil para mejorar la gestión del bibliotecario en el entorno digital. Acimed 2007;16(4). Disponible en: [http://bvs.sld.cu/revistas/aci/vol16\\_4\\_07/aci051007.htm](http://bvs.sld.cu/revistas/aci/vol16_4_07/aci051007.htm) [Consultado: día/mes/año].

Vivimos hoy un proceso cada vez más acelerado de renovación tecnológica. Esta rapidez tiene como consecuencia que, en los nuevos modelos de aprendizaje, no sea suficiente adquirir, asimilar y almacenar nuevos conocimientos. Esta explosión, asimilación y utilización intensiva del saber ha conducido a lo que se denomina actualmente *Sociedad del conocimiento*,<sup>1</sup> y esta no es una ilusión o una quimera inalcanzable; se debe avanzar hacia ella con decisión, seriedad y voluntad.

Para lograrlo, sin dudas, se necesita distinguir sus elementos más relevantes, a pesar de las dificultades que significa la coexistencia de diferentes visiones del mundo ante un cambio de paradigma ineludible y no poder explicar claramente sus características. Si esto se logra, podremos orientarnos para desenvolvernos mejor en estas nuevas condiciones, así como para enriquecerlo y adecuarlo a nosotros.<sup>2</sup>

Se impone entonces, la necesidad de crear nuevas estrategias de gestión que satisfagan las urgencias sociales actuales y que preparen al ser humano para enfrentar la nueva época y beneficiarse de su velocidad. “Toda estrategia de desarrollo debe ser ambientalmente sustentable, económicamente sustentada y socialmente incluyente, es decir, con una amplia y activa participación de la sociedad”.<sup>3</sup>

Por estos motivos, una estrategia socialmente incluyente presupone una sociedad altamente participativa, culta, en sentido amplio; una sociedad donde todos y cada uno de sus miembros disfruten de iguales oportunidades para acceder a una enseñanza de alta calidad, donde el desempleo no exista o sea mínimo y donde no estén presentes enormes e injustas desigualdades sociales.<sup>4</sup>

Pero en esta nueva sociedad del conocimiento, cada día más cercana gracias al avance de la nanotecnología, la biotecnología, las tecnologías de la información y la ciencia cognitiva, que han alcanzado una posición de avanzada sobre una sólida y novedosa base científica, el tiempo es escaso. Tanto la información como el conocimiento se generan cada vez a mayor velocidad y, en ocasiones, se acumulan en enormes volúmenes en espera de un procesamiento adecuado.

Por otra parte, se considera que han quedado abiertos los horizontes hacia otras profesiones que están llamadas a cooperar con los profesionales de la información: diseñadores de sistemas, proveedores de datos, editores, vendedores, archivistas, ingenieros y especialistas en codificación de texto electrónico, cuyas opiniones y experiencias permitan desarrollar las interfaces correspondientes que faciliten la localización, manipulación, recuperación y uso de la información digital.<sup>5</sup> Las tecnologías de la información y el conocimiento (TIC), que comprenden diferentes ciencias y áreas del conocimiento como la computación, las telecomunicaciones y la microelectrónica, que intervienen en el procesamiento y transmisión de la información y la comunicación, incluyen herramientas para la gestión del conocimiento capaces de buscar, recuperar, filtrar, personalizar e, incluso, analizar datos estructurados o no.

Analizar el lugar de la información y el conocimiento en las actuales condiciones del desarrollo de las TIC y caracterizar la minería de textos como herramienta para la gestión del conocimiento de elevada utilidad para los profesionales de la información y los bibliotecarios, son los propósitos de esta contribución.

## Gestión de la información y el conocimiento

Entre muchas definiciones del término información hemos seleccionado aquella que plantea que: “es la agrupación de datos organizados y presentados en un contexto”;<sup>6</sup> que puede ampliarse al decir que existe en tres órdenes:

- Como representación, estímulos; a partir de apropiación interna del individuo y construcción negociada, validada.

- Como elemento vital - funcional de la actividad intelectual, a la que se accede sólo como datos informativos que producen acciones - reacciones en nuestro estado intelectual.
- Como eslabón del procesamiento para realizar tareas de investigación.

No obstante estos conceptos clásicos, otros autores reconocen la necesidad de adquirir una nueva visión de la información como recurso que genera valor para el negocio, sobre la base de la creación de productos más competitivos o de procesos más eficaces.<sup>7</sup> Debe superarse, entonces, la visión tradicional de la información como un medio de control y de coordinación interna de las diferentes funciones de la empresa.

De modo que puede afirmarse que la información es un mensaje significativo que se transmite de la fuente a los usuarios es la expresión material del conocimiento con fines de uso. Ella está destinada a resolver determinados problemas. Debe estar, entonces, disponible públicamente y servir para el desarrollo individual y corporativo. Se encuentra presente en todos los niveles de actividad y ramas de la economía, la política y la sociedad.<sup>8</sup> Vivimos en una sociedad del conocimiento, estamos sepultados por la información: “nunca antes hubo tanta información transmitida por tantos medios al mismo tiempo [...]”; sin embargo, el concepto de información no coincide de ningún modo con una comprensión bien elaborada del conocimiento”.<sup>9</sup>

En su dimensión básica se define el conocimiento como “una verdad justificada”,<sup>10</sup> acepción aplicable a cualquier contexto y derivada de las disquisiciones filosóficas realizadas a través de la historia cuya connotación se precisa en sus raíces epistemológicas y en el método para adquirir y concebir dicha verdad justificada. En términos más concretos, “el conocimiento surge cuando una persona considera, interpreta y utiliza la información de manera combinada con su propia experiencia y capacidad”.<sup>11</sup> En consecuencia, se puede afirmar que el conocimiento está determinado por la interpretación que las personas realizan con la información disponible, condicionada por el contexto en el que se desenvuelven y la experiencia que poseen.<sup>12</sup>

Existen múltiples epistemologías sobre el conocimiento; en el libro titulado *Personal Knowledge*,<sup>13</sup> sus autores distinguen entre conocimiento explícito y tácito. Con respecto a la dimensión ontológica, *Spender* (1996) considera dos niveles: individual y colectivo, que combinados con los tipos de conocimientos antes esbozados, resultan en nuevas categorías:<sup>14</sup>

	DO	Individual	Social
TC			
Explícito		Consciente	Objetivo
Implícito		Automático	Colectivo

(TC): tipos de conocimiento.  
(DO): dimensiones ontológicas.

“El conocimiento es un recurso estratégico para el desarrollo económico y social contemporáneo; la información es el elemento básico principal en el proceso de adquisición, generación, gestión y transmisión del conocimiento”.<sup>15</sup>

En este sentido, otros autores indican que para los nuevos modelos de negocio, la gestión de la información, la documentación y el conocimiento se perfila como un componente estratégico de primera magnitud. La orientación tradicional del carácter táctico de los proyectos de gestión de la información o de la documentación cambia cuando se considera un verdadero componente de la estrategia empresarial. Este cambio no sólo afecta a las organizaciones que se crean para actuar específicamente en el ámbito del *e-business* —tan cuestionadas ahora por la crisis de las puntocom— sino que, al menos en el ámbito teórico, es una realidad hasta en las empresas

más "tradicionales".<sup>16</sup>

Las razones para la introducción de la gestión de la información y el conocimiento están bien establecidas; por tanto, la biblioteca, como organización moderna, debe ser capaz de garantizar el acceso de los sectores más excluidos al desarrollo en ella, a partir del fortalecimiento de la capacidad de aprender de estos grupos y el acceso al capital intelectual y al grueso de la información

## Desafíos del bibliotecario moderno

En el pasado, e incluso recientemente, los bibliotecarios se han dedicado a la orientación bibliotecaria, formación de usuarios e instrucción bibliográfica. Ahora tienen que preocuparse por suministrar a estudiantes y a otros colectivos información útil y destrezas informáticas para que puedan manejarse con eficacia en la sociedad de la información. Los diversos dominios que interactúan con esta información no poseen el tiempo para la consulta y la asimilación de los grandes volúmenes de información existentes. Este fenómeno no sólo atañe a los usuarios como consultores de la información, sino también a los especialistas de la información. Estos, evidentemente, han cambiado sus funciones y actitudes y son cada vez menos lo que se dedican a la labor de extracción o de realización de resúmenes.

Ayudar al usuario a tener éxito en la búsqueda y utilización de la información en un entorno electrónico es un proceso extremadamente complejo. Ahora, los usuarios de las bibliotecas públicas y académicas desean recolectar todo lo necesario en un único sitio de manera eficaz y efectiva, en un entorno sin límites de acceso. No quieren leer manuales complicados, estudiar las complejidades de las bases de datos, o conocer complejos mecanismos de búsqueda. Para tratar esta situación, los bibliotecarios tienen que desarrollar sistemas bibliotecarios diseñados para facilitar la formación en las destrezas de acceso e información en el momento en que se necesiten.

Una tendencia actual es la creación de catálogos locales, bibliotecas virtuales o de las también llamadas pasarelas temáticas (*subject gateways*) con recursos tomados de Internet, que resulten punteros en cada temática, que funcionen como buscadores para los usuarios potenciales de una organización.<sup>17</sup>

La tendencia al crecimiento acelerado del volumen de información disponible en el Web no se puede obviar, lo que plantea nuevos problemas y retos para la recuperación de la información. Los motores de búsqueda pueden desempeñar un papel esencial en la viabilidad de los sistemas de información basados en Internet, siempre que existan aplicaciones que puedan analizar y evaluar la relevancia de la información para el usuario.<sup>18</sup> Combinar las habilidades de los especialistas en información y de los informáticos puede ayudar a organizar este caos existente en el ciberespacio. Además, si se considera que el contenido en el Web se encuentra mucho más disperso que en una colección estándar, es comprensible la necesidad de que las habilidades de clasificación y de selección de los sistemas bibliotecarios se complementen con la automatización de las tareas de indización, clasificación y almacenamiento de la información.<sup>17</sup>

La minería de textos es una de las herramientas que ofrecen perspectivas, desde un punto de vista elemental, que podrían explotarse por los servicios de las instituciones de información a la vez que sus presupuestos metodológicos pudieran relacionarse con el desarrollo de nuevos métodos para resumir información extraída de Internet. Nuevos enfoques basados en la integración de la minería de textos con la gestión del conocimiento pueden ofrecer mejores soluciones a la gestión de la información ejercida por el bibliotecario.<sup>18</sup>

## Características de las herramientas para el análisis de la información

Los volúmenes de datos, informaciones y conocimientos que se almacenan en bases de datos y en los que se han dado en llamar grandes almacenes de datos o *warehouse*, hoy —como resultado

del desarrollo de las TIC— permiten una exploración dirigida al descubrimiento del conocimiento almacenado en ellos.

Existe cierta tendencia a identificar como sinónimos la minería de datos y el descubrimiento de conocimientos en bases de datos, que —de forma abreviada— se refieren con las siglas KDD (del inglés *Knowledge Discovery in Data Bases*), la convergencia del aprendizaje automático, la estadística, el reconocimiento de patrones, la inteligencia artificial, las bases de datos, la visualización de datos, los sistemas para el apoyo a la toma de decisiones, la recuperación de información y otros muchos campos.<sup>15</sup>

El KDD se define como “la extracción no trivial de información implícita, desconocida, y potencialmente útil de los datos.<sup>19</sup> Existe una distinción clara entre el proceso de extracción de datos y el descubrimiento del conocimiento. Bajo sus convenciones, el proceso de descubrimiento del conocimiento toma los resultados como vienen de los datos (proceso de extraer tendencias o modelos de los datos) cuidadosamente y, con precisión, los transforma en información útil y comprensible. Esta información no es típicamente recuperable por las técnicas normales, pero se descubre mediante el uso de técnicas de inteligencia artificial. KDD puede usarse como un medio de recuperación de información, de la misma manera que los agentes inteligentes realizan la recuperación de información en el Web. También puede utilizarse como una base para las interfaces inteligentes del mañana, agregando un componente del descubrimiento del conocimiento a una máquina de bases de datos o integrando KDD con las hojas de cálculo y presentaciones. Abarca el procesamiento completo —pre-procesamiento, minería, posprocesamiento— de la información,<sup>20</sup> y sería de utilidad invaluable para facilitar la recuperación y entrega de información que realizan los bibliotecarios (tabla 1).

Tabla 1. Descripción de las principales herramientas utilizadas en el proceso de descubrimiento del conocimiento o KDD<sup>20</sup>

Herramienta	Descripción
Data Warehousing	Es una técnica para consolidar y administrar datos de variadas fuentes con el propósito de responder preguntas de negocios y tomar decisiones. La definición más conocida fue propuesta por Inmon [MicroSt96] en 1992: “Un DW es una colección de datos orientados a temas integrados, no-volátiles y variante en el tiempo, organizados para soportar necesidades empresariales”.
Data Marts	Es un pequeño Data Warehouse, para un determinado número de usuarios, para un área funcional, específica de una compañía. También se puede definir que un Data Marts es un subconjunto de una bodega de datos para un propósito específico. Su función es apoyar a otros sistemas para la toma de decisiones.
Minería de Datos	Es un mecanismo de explotación, consistente en la búsqueda de información valiosa en grandes volúmenes de datos. Está muy relacionada con las bodegas de datos que proporcionan la información histórica con la cual los algoritmos de minería tienen la información necesaria para la toma de decisiones. Según <i>Fallad</i> y sus coautores (1996): “La minería de datos es un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos”. Según <i>Molina</i> y sus colaboradores (2001): “Es la integración de un conjunto de áreas que tienen como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten un sesgo hacia la toma de decisión”.

Web Mining	Según <i>Kosala</i> y sus colaboradores (2000): "Consiste en aplicar las técnicas de minería de datos a documentos y servicios del Web. Todos los que visitan un sitio en Internet dejan huellas digitales (direcciones de IP, navegador, etc.) que los servidores automáticamente almacenan en una bitácora de accesos (Log). Las herramientas de Web Mining analizan y procesan estos Log para producir información significativa. Debido a que los contenidos de Internet consisten en varios tipos de datos, como texto, imagen, vídeo, metadatos o hiperligas, se utiliza el término multimedia Data Mining (minería de datos multimedia) como una instancia del Web Mining (Zaiane y otros, 1998) para tratar ese tipo de datos. 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 del Web Mining.
Text Mining	Debido a que el 80 por ciento de la información de una compañía está almacenada en forma de documentos, las técnicas como la categorización de texto, el procesamiento de lenguaje natural, la extracción y recuperación de la información o el aprendizaje automático, entre otras, apoyan al Text Mining (minería de texto). En ocasiones se confunde el Text Mining con la recuperación de la información (Information Retrieval o IR) (Hearst, 1999). Esta última consiste en la recuperación automática de documentos relevantes mediante indizaciones de textos, clasificación, categorización, etcétera. Generalmente, se utilizan palabras clave para encontrar una página relevante. En cambio, el Text Mining se refiere al examen de una colección de documentos y el descubrimiento de información no contenida en ningún documento individual de la colección; en otras palabras, trata de obtener información sin haber partido de algo. (Nasukawa y otros, 2001).

Fuente: Bressán G. 2003. Almacenes de datos y minería de datos. 2003. Disponible en: <http://www/banners/interstitial.html?http://exa.unne.edu.ar/> [Consultado: 22 de mayo de 2007].

En el caso de la minería de textos, puede decirse que es una herramienta capaz de abarcar una amplia gama de dominios, desde aquellos de la recuperación y extracción de información, presentación, resumen de multidocumentos, minería de datos aplicada a textos, etcétera. Es un término que en dependencia del autor que lo emplee se restringen o se amplían sus esferas de aplicación.<sup>21</sup>

En una revisión más reciente sobre el tema, se plantea que la *minería de texto* consiste en la búsqueda de regularidades o patrones que se encuentran en un texto, a partir de técnicas de *aprendizaje automático*; por tanto, se considera como una de las muchas ramas de la lingüística computacional.<sup>20</sup> Como proceso se ocupa del descubrimiento del conocimiento que no existe en el texto, pero que surge al relacionar el contenido de varios textos y se divide en varias etapas.

Entre las herramientas desarrolladas para extraer información, y que intentan inferir relaciones que no aparecen de forma implícita en esa información, pueden citarse: *TextAnalyst*, *twURL*, *T-LAB*, *LexiQuest Mine*, *Text Miner* y *Weka* (tabla 2).

Tabla 2. Descripción de las principales técnicas utilizadas en la minería de textos

Técnica	Descripción		Observaciones
	Etapas	Descripción de la etapa	
Técnicas clásicas	Pre-procesamiento	Es el proceso mediante el cual los textos se transforman en algún tipo de representación estructurada que facilite su análisis.	Como se puede observar, todas las etapas están muy interrelacionadas, así entonces, la primera etapa condiciona el descubrimiento de los patrones que la minería de texto puede realizar.
	Representación	La representación depende de la técnica de pre-procesamiento utilizada y determinará el algoritmo de descubrimiento a utilizar.	

	Descubrimiento	Son algoritmos que, a partir de una representación estructurada de la información, son capaces de descubrir regularidades en los textos.	
Grafos conceptuales	<p>Un grafo conceptual es un grafo bipartito que tiene dos tipos de nodos, conceptos y relaciones conceptuales. Los grafos se comparan utilizando conocimiento del dominio como diccionarios de sinónimos y jerarquías de conceptos. Se realiza una operación de intersección entre dos grafos para dar un resumen de ambos y a dicho resumen se le valora con una puntuación que indica el grado de similitud entre ambos textos.</p> <p>La agrupación de dos o más grafos permite descubrir la estructura oculta de la colección de textos. Para agrupar los grafos, se pueden utilizar técnicas de agrupamiento como las estrategias colaborativas, el agrupamiento en k medias o <i>Comweb</i>.</p>		Las técnicas de grafos conceptuales aportan mayor semántica. Todas estas técnicas se basan en suministrar a los algoritmos, un conjunto de ejemplos a partir de los cuales se generan las agrupaciones.
Programación lógica inductiva	Esta técnica permite introducir conocimiento <i>a priori</i> del dominio en forma de definiciones mediante predicados relacionados. Requiere no sólo de un conjunto de entrenamiento con ejemplos sino también de las relaciones descubiertas por el diseñador y basadas en las <i>cláusulas de Horn</i> .		La ventaja que posee es su capacidad de representación basada en una lógica de segundo orden, que permite generalizar conceptos y descubrir definiciones de conceptos de forma automática. Normalmente se utiliza <i>Prolog</i> para programar las herramientas.
Programación genética	<p>Es un método de generación automática de programas para computadoras con inspiración evolutiva; sobre la base de programas muy simples, mediante el cruce de unos con otros y procesos de mutación aleatoria, se generan programas más y más aptos para la realización de la tarea que se le asigna. La aptitud de los programas se mide de forma numérica mediante una función denominada <i>de fitness</i>. Algunas extensiones de la programación genética permiten describir nuevas primitivas a partir de las primitivas inicialmente descritas.</p> <p>Básicamente, la idea consiste en introducir como primitivas las relaciones expresadas como <i>cláusulas de Horn</i> y utilizar un sistema que permita utilizar las denominadas ADFs, que no son más que evoluciones paralelas de otras primitivas que se pueden utilizar en la definición principal, para generar definiciones muy resumidas de los conceptos. La función <i>de fitness</i> deberá medir el número de ejemplos que se cubren con la definición de cada uno de los individuos generados en cada generación, que se pondera con la longitud de la definición para guiar al algoritmo a soluciones sencillas y el descubrimiento de conceptos intermedios.</p>		Esta técnica ha tenido problemas con las definiciones recursivas, debido a que la programación genética tiene problemas de eficiencia con primitivas recursivas. Si se conoce que la naturaleza de la definición a encontrar es recursiva, probablemente sean mejor solución las técnicas anteriormente descritas.

Fuente: Sagrado Olivenza I. Minería de texto. Recuperación y organización de la información. 2007. Disponible en: <http://mineriainformacion.com/recuperacion-informacion.html> [Consultado: 22 de mayo de 2007].

En la literatura sobre el tema, se refiere que las aplicaciones de la minería de textos se utilizan principalmente para:<sup>18</sup>

- Extraer información relevante de un documento.
- Agregar y comparar información automáticamente.
- Clasificar y organizar documentos según su contenido.
- Organizar depósitos para búsqueda y recuperación.
- Clasificar textos e indizarlos en el Web.

Pero la realidad es que son muchos los casos de organizaciones que han aplicado la minería de textos con el afán de alcanzar objetivos más reales y ambiciosos (en el mejor sentido) que estos. Por ejemplo, el uso que le dan algunas empresas para identificar el contenido de los correos electrónicos que les envían los clientes y redirigirlos a los departamentos apropiados, incluso si el sistema es capaz de identificar el contenido de una consulta frecuente, envía una respuesta estándar, sin necesidad de intervención humana. En el campo de la vigilancia tecnológica y en la "*Business Intelligence*", para bucear en las bases de datos textuales y seguir la evolución de la competencia. Además de que se apunta también la posibilidad de usar esta técnica en la investigación de mercados en el Web, sobre la base de recoger estadísticas sobre la utilización de determinados conceptos y temas en la red con el objetivo de estimar la demografía y las curvas de demanda de productos asociados con ellos. Todas esas aplicaciones son perfectamente transferibles a la gestión de información que ocurre en el interior de las bibliotecas y están llamadas a redimensionar la función de la entidad, así como del bibliotecario, tanto hacia adentro como fuera de esta.

## Consideraciones finales

La información constituye, en las condiciones actuales, un recurso económico altamente valorado no sólo por sus propiedades intrínsecas, sino porque permite, además, el perfeccionamiento del empleo del resto de los recursos de las organizaciones. De ahí que la gestión de la información y del conocimiento, como estadio superior, cobre cada día mayor importancia en la elevación de su competitividad.

Son numerosos y múltiples los acercamientos a una definición de la herramienta de gestión del conocimiento de minería de texto. Se trata de un proceso que se ocupa del descubrimiento de conocimiento que no existe en el texto, y que surge al relacionar el contenido de varios textos, a partir de técnicas de aprendizaje automático de regularidades o patrones que se encuentran en ellos, con la particularidad de que estos son textos no estructurados, información almacenada en forma textual no estructurada: informes, correos electrónicos, actas de reuniones, etcétera.

Utilizada en su gestión por no pocas organizaciones de sectores tradicionales o no, la minería de textos ha generado como resultado un aumento en la eficacia de sus modelos predictivos de y ha permitido ahorros de tiempo y dinero; así como la mejora de la capacidad de respuesta a las necesidades de sus clientes.

De modo que es inevitable pensar que con los niveles alcanzados por las tecnologías informáticas de apoyo y el desarrollo de las herramientas útiles para el descubrimiento y procesamiento de la información en las cada vez mayores fuentes de acumulación de datos, textos e informaciones, se agilizará y perfeccionará el proceso de gestión del conocimiento que ocurre intra e inter bibliotecas.

## Referencias bibliográficas

1. Conceição P, Heitor MV, Veloso F. Introduction: Knowledge, Technology and Innovation Systems for Inclusive Development. En: Conceição P, Gibson DV, Heitor MH, Sirilli G, Veloso F (eds.). Knowledge for Inclusive Development. New York C: Quorum Books. 2001.
2. Faloh R. Introducción. Gestión de la innovación. Una visión actualizada para el contexto Iberoamericano. La Habana: Academia. 2006.
3. Sachs I. Desenvolvimento incluyente, sustentable, sustentado. Rio de Janeiro: Garamond. 2004.
4. Sáenz T, Souza MC. Innovación tecnológica y sustentabilidad. En: Gestión de la innovación. Una visión actualizada para el contexto Iberoamericano. La Habana: Academia. 2006.
5. Torres Pombert A ¿Catalogación en el entorno digital?: una breve aproximación a los metadatos. Acimed. 2006;14(5). Disponible en: [http://bvs.sld.cu/revistas/aci/vol14\\_5\\_06/aci09506.htm](http://bvs.sld.cu/revistas/aci/vol14_5_06/aci09506.htm) [Consultado: 22 de mayo de 2007].
6. Murray P. Gestión-Información-Conocimiento. Biblio. 2002;4(14):5-6.

7. Goñi Zabala JJ. La información en la estrategia de las organizaciones: un recurso por desarrollar. 2002. Disponible en: <http://www.hacienda.go.cr/centro/datos/Articulo/La%20Informaci%C3%B3n%20en%20la%20estrategia%20de%20las%20organizaciones.%20Art%C3%ADculo%20Gesti%C3%B3ndelconocimiento.com.doc> [Consultado: 19 de mayo de 2007].
8. Capote Marrero B, González Machón D, Rodríguez Durán E. La gestión de información como herramienta fundamental en el desarrollo de los centros toxicológicos. *Acimed*. 2003;11(2). Disponible en: <http://eprints.rclis.org/archive/00001790/01/gestion.pdf> [Consultado: 22 de mayo de 2007].
9. Kurz R. La ignorancia de la sociedad del conocimiento. 2007. Disponible en: <http://www.sindominio.net/oxigeno/archivo/conocimiento.PDF> [Consultado: 13 de mayo de 2007].
10. Nonaka I, Takeuchi H. *The knowledge – creating company*. New York C: Oxford University Press. 1995.
11. Mazo I, Ortiz de Frutos E. 1998. *KMAT como herramienta de análisis de gestión del conocimiento*. Clúster del conocimiento. Bilbao: PMP. 1998.
12. López MS, Cabrales F, Schmal R. Gestión del conocimiento: una revisión teórica y su asociación con la universidad. *Panorama Socioeconómico*. 2005;(30). Disponible en: <http://redalyc.uaemex.mx/redalyc/pdf/399/39903004.pdf> [Consultado: 21 de mayo de 2007].
13. Polanyi M. *Personal Knowledge*. Chicago: University Chicago Press. 1958.
14. Balbastre F. La autoevaluación según los modelos de la gestión de calidad total y el aprendizaje en la organización. Una investigación de carácter exploratorio. 2001. Disponible en: [http://www.tesisenxarxa.net/TESIS\\_UV/AVAILABLE/TDX-0312104-114550/balbastre.pdf](http://www.tesisenxarxa.net/TESIS_UV/AVAILABLE/TDX-0312104-114550/balbastre.pdf) [Consultado: 21 de mayo de 2007].
15. Febles Rodríguez J, González Pérez A. 2002. Aplicación de la minería de datos en la bioinformática. *Acimed*. 202;10(2). Disponible en: [http://scielo.sld.cu/scielo.php?pid=S1024-94352002000200003&script=sci\\_arttext](http://scielo.sld.cu/scielo.php?pid=S1024-94352002000200003&script=sci_arttext) [Consultado: 19 de mayo de 2007].
16. Bustelo Ruesta C, García-Morales Huidobro E. Tendencias en la gestión de la información, la documentación y el conocimiento en las organizaciones. *El Profesional de la Información*. 2001;10(12):4-7.
17. Rodríguez Perojo K, Ronda León R. La Web como sistema de información. *Acimed*. 2006;14(1). Disponible en [http://bvs.sld.cu/revistas/aci/vol14\\_1\\_06/aci08106.htm](http://bvs.sld.cu/revistas/aci/vol14_1_06/aci08106.htm) [Consultado: 25 de mayo de 2007].
18. Bordón L, D'Avanzo E. Perspectivas para la integración de la minería de textos y la gestión del conocimiento. *The IPTS Report*. No.85. 2004. Disponible en: <http://www.jrc.es/home/report/spanish/articles/vol68/ICT2S686.html> [Consultado: 25 de mayo de 2007].
19. Frawley W, Piatetsky Shapiro G, Matheus C. Knowledge Discovery in Databases: An Overview. *AI Magazine*. 1992 ;13(3):57-70.
20. Bressán GE. 2003. Almacenes de datos y minería de datos. 2003. Disponible en: <http://www/banners/interstitial.html?http://exa.unne.edu.ar/> [Consultado: 18 de mayo de 2007].
21. Sagrado Olivenza I. Minería de texto. Recuperación y organización de la información. 2007. Disponible en: <http://mineriainformacion.com/recuperacion-informacion.html> [Consultado: 18 de mayo de 2007].

Recibido: 31 de julio de 2007. Aprobado: 6 de agosto de 2007.

MSc. *Eleazar Botta Ferret*. Instituto de Investigaciones de Sanidad Vegetal. Calle 110 No. 514 entre 5ta B y 5ta F, Playa, Ciudad de La Habana, Cuba. Correo electrónico: [ebotta@inisav.cu](mailto:ebotta@inisav.cu)

<sup>1</sup>**Máster en Biología Vegetal. Investigador Agregado. Instituto de Investigaciones de Sanidad Vegetal.**

<sup>2</sup>**Licenciada en Derecho. Instituto de Información Científica y Tecnológica (IDICT).**

Ficha de procesamiento

Términos sugeridos para la indización

Según DeCS<sup>1</sup>

GERENCIA DE LA INFORMACIÓN; ANÁLISIS DE DATOS; BASES DE DATOS.  
INFORMATION MANAGEMENT; DATA ANALYSIS; DATABASES.

Según DeCI<sup>2</sup>

GESTIÓN DEL CONOCIMIENTO; ANÁLISIS DE DATOS; BASES DE DATOS.  
KNOWLEDGE MANAGEMENT; DATA ANALYSIS; DATABASES.

<sup>1</sup>BIREME. Descriptores en Ciencias de la Salud (DeCS). Sao Paulo: BIREME, 2004.

Disponible en: <http://decs.bvs.br/E/homepagee.htm>

<sup>2</sup>Díaz del Campo S. Propuesta de términos para la indización en Ciencias de la

Información. Descriptores en Ciencias de la Información (DeCI). Disponible en: <http://cis.sld.cu/E/tesauro.pdf>

[Índice Anterior](#) [Siguiente](#)