

Editorial

Las tecnologías de bases de datos al servicio de la metría de la información y del conocimiento científico

DOI: <http://dx.doi.org/10.22201/iibi.24488321xe.2017.nesp1.57883>

En la actualidad es impensable que una organización pueda hacer uso eficiente de sus datos sin que éstos se encuentren almacenados en una base de datos. Sin embargo, éstas no siempre se crean de la mejor manera, ni utilizando un software especializado. Es común tener los datos almacenados en repositorios sin estructura, control, ni protección; principalmente debido a que son el resultado de una evolución sin supervisión de personal con conocimiento de las bases de datos y de las características con que se deben crear, en muchas ocasiones se ha pasado de conjuntos de datos personales a repositorios de información sin una estructura adecuada. Por lo tanto, se desconoce el riesgo que conlleva la falta de protección de los datos, y la complejidad de manipulación de los mismos para satisfacer la necesidad de información a partir de esos datos.

En la evolución natural de las bases de datos y sus aplicaciones, se tiene a los almacenes de datos (del inglés *data warehouse*). Éste es un tipo particular de base de datos en la que se integran los datos provenientes de diferentes fuentes (sean o no bases de datos) relacionados con un tema de interés para las organizaciones, por ejemplo, las ventas, una vez almacenados los datos no es posible eliminarlos y, por lo tanto, el almacén de datos mantiene la historia de aquéllos.

La razón de ser de los almacenes de datos es permitir medir la productividad de una organización desde diferentes ángulos, llamados dimensiones, por ejemplo, una medida sería la cantidad de alumnos titulados y las dimensiones podrían ser, entre otras, por periodo de tiempo, carrera, universidad, país.

También se considerarían las medidas de acuerdo con las diferentes dimensiones.

En el campo de la metría de la información y del conocimiento científico, campo en el que se incluyen las especialidades métricas conocidas como bibliometría, informetría, biblioteconometría, archivometría y la cienciometría, resultan de gran utilidad los almacenes de datos por la naturaleza misma de los datos con que se trabaja, así, al tener registrada la productividad científica, se obtienen fácilmente indicadores por áreas geográficas (por ejemplo, clasificadas por continente, país, ciudad), áreas económicas, por áreas de conocimiento, tiempo, etc. Suena bien, sin embargo, se requiere de mucho trabajo para recopilar la información e integrarla en el formato adecuado. Es un trabajo arduo, pero la recompensa es grande, al no tener la necesidad de consultar todas las fuentes de datos originales, cada vez que se requiera y hacer programas para satisfacer preguntas particulares.

Otra línea de evolución de las bases de datos es la minería de datos, cuyo propósito es extraer información oculta en las bases de datos. Esta información está oculta porque no era el objetivo original al crear la base de datos. En este caso, se trabaja con ésta sin crear una adicional, sólo es necesario que los datos estén “limpios”, que es el principal problema en esta aplicación, pues resulta común que los datos no estén completos o sean inconsistentes, entre otros problemas. Una vez más, esto se debe principalmente a la poca importancia que se da a las bases de datos, por lo que no se utilizan los mecanismos necesarios para evitar tales problemas en los datos. Estas restricciones no son difíciles de implementar en una base de datos, almacenada en un sistema que las manipule, el problema nuevamente es el desconocimiento del área.

Como parte de la minería de datos, es de particular interés la minería de textos por el amplio rango de aplicaciones dentro de las tareas que cotidianamente se realizan en la investigación informétrica. Se diría que el objetivo de la minería de textos es convertir datos textuales (texto) en información de gran calidad o conocimiento que permita realizar acciones, esto con el fin de minimizar el esfuerzo empleado en las tareas que consumen datos textuales. En muchas ocasiones, aún se realiza el análisis de textos manualmente, siendo que se han desarrollado

muchas técnicas para ejecutar esta tarea, aunque sus aplicaciones han sido en áreas distintas al análisis que se realiza con la información bibliográfica.

Entre las principales tareas realizadas en la minería de textos, destacan el minado y análisis de tópicos, así como el minado y análisis de opiniones y sentimientos. Ambas tareas se realizan tomando como base los textos escritos, ya sean elementos estáticos (como los artículos científicos) o información dinámica (como la que se genera cada vez que se realiza una consulta a un sistema de información disponible en la Internet).

El minado de tópicos permite identificar el tema principal que se trata en un texto. Este análisis se realiza en diferentes niveles, ya sea dentro de una oración o dentro de un artículo de investigación. El análisis de tópicos se ha utilizado en gran cantidad de aplicaciones, de particular interés es su uso en el análisis de artículos de investigación para identificar las áreas en que se ha realizado investigación en distintas épocas.

El minado y análisis de opiniones y sentimientos permite identificar y extraer estos elementos, presentes en gran cantidad de documentos. Las técnicas desarrolladas permiten identificar al sujeto que detenta una opinión en un texto, así como el ente sobre el que se opina (objetivo de la opinión) y la opinión en sí. También es posible determinar el contexto en el que se emite la opinión. Tomando en cuenta todo lo anterior, se deduce el sentimiento asociado a esa opinión, ¿es una opinión positiva o negativa? Las principales aplicaciones realizadas se encuentran en los sistemas de recomendación, principalmente los orientados a promover productos comerciales, por ejemplo, los distintos modelos de teléfonos celulares. Se usan estas técnicas en diversos campos, como sistemas para auxiliar en la toma de decisiones, sistemas que tratan de entender a sus usuarios, por ejemplo, para conocer sus preferencias, o en las encuestas que permiten identificar nichos de mercado. La aplicación de estas técnicas está abierta a una gran cantidad de campos.

Cabe señalar que, para realizar las tareas del minado de textos, se requiere de trabajo previo. En particular, es necesario llevar a cabo actividades para la recuperación de información, ya sea como parte del preprocesamiento, así como en la determinación de las fuentes de conocimiento.

La tarea de recuperación de información cobra relevancia en esta época en la que la cantidad de información disponible aumenta a pasos agigantados, ya no se diga cotidianamente, sino minuto a minuto. En un momento determinado, de toda la información disponible sobre un tema, ¿cuáles documentos son relevantes para la actividad en que se trabaja? Precisamente, el objetivo principal de la recuperación de información es determinar el conjunto de documentos relevantes. Además, también permite clasificar esos documentos, de acuerdo a su relevancia, para que el escrutinio de la información se haga sobre el conjunto de datos que aporta mayor calidad a los resultados.

Los datos son la materia más importante en cualquier organización, de su calidad dependen los resultados que se obtengan. Se ha avanzado mucho en el desarrollo de herramientas que facilitan la extracción de información en los almacenes de datos. Actualmente, también se cuenta con herramientas muy poderosas para la minería de datos, en especial para la minería de textos. Sin embargo, aún es frecuente encontrar fuentes de información (bases de datos, repositorios de datos, etc.) que requieren de mucho trabajo previo a la aplicación de estas tecnologías. Es una labor ardua, pero los beneficios que se obtienen justifican esa inversión de recursos humanos y materiales.

*Amparo López Gaona
Salvador López Mendoza*

Para citar este texto:

López Gaona, Amparo y López Mendoza, Salvador. 2017. "Las tecnologías de bases de datos al servicio de la metría de la información y del conocimiento científico". *Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información* (Número Especial de Bibliometría): 11-14.
<http://dx.doi.org/10.22201/ibi.24488321xe.2017.nesp1.57883>