

Bases de datos: El formato electrónico y la calidad de la respuesta

Marta Ligia Pomim Valentim

RESUMEN

La calidad de la respuesta, factor fundamental del trabajo con las bases de datos, es directamente proporcional al respeto que se le da a los diferentes formatos documentales que constituyen las bases de datos. Contemplar las diferentes tipologías documentales es un aspecto clave que incide en la calidad de la respuesta de las bases de datos a sus clientes. Este artículo muestra un estudio del formato utilizado en dos bases de datos y cómo el formato puede afectar la calidad y la recuperación de la información.

ABSTRACT

The quality of the answer, key element of database's handling, is directly proportional to the respect given to the different document formats which constitute a database; this factor is indispensable to assure the clients obtain the required answers. This article gives the results of a study carried on about the format used in two databases as well as how this format can affect the quality of information retrieval.

Sistematización de la información

En la década del sesenta, las instituciones que procesaban la información, la seleccionaban, la analizaban y la ponían a disposición de los interesados de forma impresa a través de los conocidos *abstracts* e *index*, se dieron cuenta de la necesidad de utilizar el recurso de la informática para agilizar el acceso a ellas.

Esas instituciones se unieron al desarrollo de la informática en la época en que se iniciaba la democratización del uso de las microcomputadoras en los Estados Unidos y en Europa (década del setenta), y empezaron a procesar la información de forma automatizada. A este nuevo formato se le dio como nombre bases de datos, y hasta hoy es conocido así.

Las bases de datos surgen justamente por la necesidad de obtener la información con más rapidez, más calidad, más relevancia y forma seleccionada.

Las bases de datos surgen justamente por la necesidad de obtener la información con más rapidez, más calidad, más relevancia y forma seleccionada.

Actualmente, en el mercado se presentan distintas categorías de bases de datos:

- *Bibliográficas*: Tienen la referencia bibliográfica y a veces el resumen del documento; en su mayoría son artículos de revistas, *papers*, trabajos científicos presentados en congresos, simposios, seminarios, etc. Patentes, informes técnicos.

Las bases de datos bibliográficas tienen su origen, por tanto, en los antiguos *abstracts* e *index*.

- *Referenciales*: Tienen los datos referenciales sobre las personas, instituciones públicas y privadas, productos, materiales, leyes o datos de cadastros. En este caso, las bases de datos referenciales tienen su origen en los anuarios, catastros, guías, directorios.
- *De fuentes*: Tienen datos originales e integrales a texto completo del documento, o sea, el interesado puede obtener electrónicamente el documento completo en el momento de la consulta/investigación. Este tipo de base de datos se inicia con la prensa diaria, los informes y balances financieros y ya ofrecen también, revistas e informes técnicos.

Las bases de fuentes se subdividen en tres grupos, a saber:

- *Núméricas*: Tienen documentos completos sobre investigaciones de opinión, de consumo, estadísticas sobre la población. Este grupo se caracteriza por presentar los datos numéricos a través de tablas, gráficos, etc.
- *Textuales*: Tienen documentos completos sobre artículos de revistas, noticias de periódicos, programas de computación, normas técnicas, patentes, etc.
- *Textuales y numéricas*: Tienen documentos completos sobre empresas, es decir, balances, informes anuales, etc. Se diferencian del primer grupo por la mezcla de los datos numéricos y textuales.

Las categorías de las bases de datos presentadas anteriormente pueden encontrarse en formato electrónico (en línea/Internet) u óptico (*offline*/CD-ROM). Estas tecnologías de la información, tanto la electrónica como la óptica, tienen como fin agilizar el acceso a la información que conforma estas bases de datos. Sin embargo, uno de los factores que mantiene insatisfechos a los usuarios es la cuestión de la calidad de la respuesta de la información recuperada durante una búsqueda, es decir, cuando un usuario busca en una base de datos un asunto de su interés, la mayor parte de las veces, menos de 50% de lo que se recuperó es lo que realmente se va a utilizar o es relevante para él.

Partiendo de este punto, surge la tesis de que a pesar de que las bases utilizan un lenguaje documentario y lo controlan, por ejemplo, a través de tesauros, la recuperación de la información, la mayor parte de las veces, no resulta satisfactoria para el usuario. Esto

indica que el tratamiento de la información debería dividirse en dos partes igualmente importantes: la primera parte, sin duda alguna, sería la utilización de un lenguaje documentario, que es lo que se está haciendo; la segunda parte sería la referente al formato documental, o sea, en las bases de datos bibliográficas es usual el suministro de varias tipologías documentales: artículos de revistas, patentes, anales de eventos, normas técnicas, etc. Cada una de estas tipologías documentales tiene una estructura propia y exige una lectura e identificación de contenidos diferentes a las otras.

Siguiendo este razonamiento, las bases de datos bibliográficas crearían diferentes formatos en su estructura, teniendo en cuenta realmente el contenido de las diferentes tipologías documentales. Las bases de datos desarrollarían un formato para libros, otro para artículos de revistas, otro para patentes, otro para normas y así sucesivamente, de forma que respetándose la estructura básica de cada tipología documental, sería posible extraer integralmente el contenido documental y se mejoraría ampliamente la calidad de la respuesta, ya que a la vez que se respeta la estructura documental, es mayor la posibilidad de que toda la información relevante esté a disposición del usuario.

Tratar la información con calidad significaría procesar el análisis documentario, estableciendo una relación profunda con el formato documental, es decir, relacionar el análisis documentario con la estructura del documento, buscando el contenido.

Otra cuestión importante es el control del lenguaje documentario mediante instrumentos como los tesauros. La recuperación de la información es mejor en las bases de datos que trabajan una terminología especializada que en las que no se trabaja con el lenguaje.

Es más factible presentarle al usuario un resultado eficaz cuando la estrategia de búsqueda se apoya en un tesoro, porque la elaboración se hace a través del lenguaje utilizado por la base de datos a la cual se accedería.

Cabré [1] defiende la necesidad de trabajar con una terminología normalizada para hacerle frente a la comunicación entre los especialistas de una misma área, ya que la apertura de mercados internacionales exigirá una mayor comunicación entre las empresas, las personas, los gobiernos, etc.

El tesoro es uno de los instrumentos que puede utilizarse para alcanzar este objetivo, pero él trabaja con descriptores, unidades preferenciales, que no

siempre indican su definición, la mayoría de las veces presenta sólo sus relaciones con el universo de la especialidad.

Lariviere afirma que el tesoro es un “intermediario entre los documentos y los usuarios [2, p. 458]” justamente por el hecho de que el usuario no encuentra la definición del descriptor y sí su relación con el universo especializado a través de términos más genéricos, específicos y relacionados; actualmente la terminología presenta los términos como unidades monorreferenciales, o sea, unidades significantes de la lengua.

Así, se sugiere que el tesoro debe ser uno de los instrumentos para controlar el lenguaje utilizado en las bases de datos, pero, en su estructura, además del término más genérico, de los más específicos y de los relacionados, debería incluir también la definición del descriptor, convirtiéndose de esta forma en un tesoro terminológico.

Análisis documentario/tratamiento de la información

Para comenzar, un pequeño levantamiento teórico sobre el tema demuestra que el análisis documentario transforma la información en un producto documentario para seleccionar y recuperar. Lara define la comunicación documentaria como un “proceso que abarca la codificación y decodificación de contenidos informacionales, o sea, el tratamiento y la recuperación [3, p. 4]”. Para producir bases de datos es fundamental comprender bien lo que significa el tratamiento y la recuperación de la información, pues de lo contrario es imposible acceder a ellas.

Según Kobashi, “...la actividad documentaria no es neutra, porque sus productos están comprometidos con el proceso social de producción, circulación y consumo de bienes materiales y simbólicos [4, p. 12]”. Así, el producto documentario o la información documentaria estará debidamente “tratado (a)”, de forma “que respondan de manera pertinente a una pregunta dada [...] la información documentaria debe elaborarse mediante metodologías que garanticen la equivalencia de sentido entre el texto fuente y su representación [4, pp. 24-25]”. Comprender la necesidad de la equivalencia de sentido entre el contenido y su representación es fundamental para que una búsqueda obtenga un resultado satisfactorio.

El tesoro debe ser uno de los instrumentos para

controlar el lenguaje utilizado en las bases de datos, pero, en su estructura, además del término más genérico, de los más específicos y de los relacionados, debería incluir también la definición del descriptor, convirtiéndose de esta forma en un tesoro terminológico.

Para Gardin, según Guimarães, el análisis documentario es un “conjunto de procedimientos efectuados con la finalidad de expresar el contenido de los documentos científicos de manera que se facilite la recuperación de la información [5, p. 164]”.

Según Guimarães, para varios autores como Smit, Cunha, Kobashi, Tálamo, Amaro y Lara, el análisis documentario se caracteriza por dos etapas: la primera, el análisis-síntesis unido a la identificación de las informaciones principales del discurso” y la segunda, la representación documentaria unida a la conversión de esas ‘informaciones principales en informaciones documentarias’ a través de un proceso de traducción haciendo uso de un lenguaje documentario [5].

En el caso específico de las bases de datos, la mayoría utiliza el análisis documentario ofrecido por los propios editores. Como ejemplo, podemos citar las revistas especializadas que ya envían los artículos con el resumen y las palabras clave. Para los documentos sin este tipo de procedimiento documentario, las bases “tratan” la información haciendo uso del análisis documentario, es decir, una vez seleccionado el documento que ha de formar parte de las bases de datos, se realiza el análisis documentario.

Algunas bases de datos controlan el lenguaje a través de sus propios tesauros, otras sólo coordinan los términos listados en diccionarios en la propia base de datos.

Partiendo de esto, se pretende en este estudio observar el tratamiento de las informaciones sistematizadas en las bases de datos, verificar si respetando realmente la estructura documental (este punto se refiere a los

elementos internos y externos) la recuperación de la información tendrá mayor calidad de respuesta.

La idea es comparar bases bibliográficas comunes, o sea, aquellas que mezclan diferentes tipologías documentales, sistematizando las informaciones en un único formato, con las bases bibliográficas que trabajan con una única tipología documental, supuestamente más adecuada a la estructura/gradación del documento y verificar la calidad de la respuesta obtenida.

Formato versus contenido

Para comenzar es necesario entender la circulación de informaciones y documentos como sugiere Kobashi: “a) colección de documentos; b) tratamiento de los documentos y c) diseminación de los documentos”, así como, el ciclo documentario que según la autora “presupone dos operaciones básicas y complementarias: la fabricación de informaciones documentarias y la recuperación de la información [4, p. 15]”.

La primera etapa está ligada a la selección, análisis y representación del documento y la segunda etapa está unida a la pregunta del usuario y la vinculación de la pregunta a la representación del documento, o sea, las dos operaciones están absolutamente relacionadas y dependen una de otra.

La estructura formal de las bases de datos contempla generalmente dos grandes áreas de la representación documentaria: a) representación descriptiva y b) representación temática.

Según Kobashi, el documento “se somete a procesos de tratamiento diferenciados [4, p. 19]”, la representación descriptiva tiene como fin dar información sobre los datos llamados físicos del documento, por ejemplo, autor(es), título, año de publicación, etc. Así, se genera un producto documentario muy conocido y utilizado que es la referencia bibliográfica. La representación temática tiene como fin dar información sobre los contenidos temáticos del propio documento, el asunto principal, los asuntos secundarios y otros asuntos afines relevantes, generándose así algunos productos documentarios también muy conocidos y utilizados como son el resumen y la indización.

La indización genera a su vez diversos productos documentarios con diferentes denominaciones como el índice, la pista de contenido, unitérmino, encabezamiento del asunto, etc., y su denominación está relacionada con el sistema utilizado para extraer el contenido del documento. En este estudio, en

cuanto a la indización, interesa tan sólo el producto índice o palabras clave.

El ciclo documentario presentado por Kobashi, más específicamente en la operación de fabricar la información documentaria, se divide en dos etapas de trabajo. La primera es la descomposición del texto, que a su vez tiene dos aspectos, el análisis y la síntesis del documento, o sea, extracción de los temas relevantes y la segunda, la estructuración del texto mediante la representación documentaria, o sea, la representación de estos tópicos relevantes ofrecidos a los usuarios en formatos documentarios, es decir, el resumen y el índice.

La descomposición de un documento desde el punto de vista del análisis documentario resulta muy importante porque es en este momento en el que se produce la selección de lo que se juzga relevante y representativo del texto-documento en cuestión.

Las bases de datos, principalmente las bibliográficas, trabajan con diversos tipos de documentos, por ejemplo: artículos de revistas; ponencias de congresos, simposios, etc.; libros; patentes y otros; sin embargo, el formato utilizado para coleccionar y sistematizar los datos es el mismo para todos los tipos.

También con bastante frecuencia se observa que las bases de datos que trabajan con un único tipo de documento tienen resultados de búsqueda más eficaces y con más calidad. Por ejemplo, una base de datos como la INPADOC o WPI o U.S. Patents obtienen mejor calidad de respuesta. La única tipología tratada es la patente, y esto es lo que puede estar generando esa diferencia en la calidad de respuesta.

En el área científico-tecnológica podemos tomar como ejemplo documentos que presentan una forma determinada como las patentes y las normas técnicas, documentos más fáciles de reconocer gracias a su estructura o gradación de la exposición del contenido y también al propio lenguaje que utilizan.

Van Dijk afirma que una tipología de texto-documento tiene una estructura propia independiente del contenido que nos permite reconocer un orden lógico en las diversas partes de un texto-documento. Esta estructura que el autor denomina superestructura se define como un “tipo de esquema abstracto que establece un orden global en un texto y se compone de una serie de categorías [6, pp.142-143, pp. 168-169]” vinculadas a características pragmáticas y semánticas del texto.

| | |
|--|---|
| Dialog Accepción Number: Title: Autor: Corporate Source/Autor Affiliation: Patent Number: Journal Name: Journal Announcement: Document Type: Language: Abstract: Descriptors: Identifier: Alloy Class Code: Section Headings: | MA Accepción Number:Patent Application Date: Publication Year: Element Symbol: Alloy Class Name: ISSN/ISBN: |
|--|---|

Fig. 1. Metadex. Formato o gradación del contenido.

La proposición de Van Djik, estudiada y trabajada por Kobashi, demuestra que es posible observar en una tipología documental una superestructura que permite seleccionar a través de estas demarcaciones textuales lo que se debe seleccionar en términos temáticos en un documento.

A los productores de las bases de datos se les propone establecer una gradación/superestructura para cada tipología documental existente en la base, como artículos de periódicos, libros, patentes, normas técnicas, informes, etc. El análisis documental se elaboraría a través de la gradación/superestructura de cada tipología documental, extrayéndose los contenidos relevantes y secundarios de un texto de forma adecuada.

Es importante aclarar en este momento que en ninguna hipótesis se habla de la necesidad de producir bases de datos con tipologías documentales únicas. Al contrario, resulta irrelevante si una base de datos trabaja con una única tipología documental o con varias tipologías documentales. Sin embargo, se insiste en la importancia de que las bases de datos trabajen con diferentes tipologías respetando el formato-gradación-superestructura de los documentos para que la respuesta tenga mayor calidad y más eficacia.

Observando la realidad

En la mayor parte de las bases de datos, tanto internacionales como nacionales, se ofrecen dos productos documentarios: el resumen y los descriptores (índices/palabras clave).

La elaboración de resúmenes generalmente implica algunas operaciones ejecutadas con frecuencia por los documentalistas, por ejemplo: análisis del texto, comprensión del texto, condensación del texto y transposición hacia el lenguaje documentario, que puede seguir una norma o regla establecida nacional o internacionalmente.

Kobashi afirma que el “resumen se caracteriza por ser un texto que representa al original pero bajo la forma de un nuevo texto condensado [...] el análisis y la comprensión del texto a los fines documentarios requiere estrategias adecuadas de lectura [4, pp. 108]”; ahora, según la autora, el “...’tema’ es un elemento en torno al cual se estructura el mensaje, es su núcleo informativo [4, p. 110]”. Completando el pensamiento, la autora prosigue planteando que “el texto, en sentido amplio, designa una unidad de comunicación organizada sintagmáticamente y dotada de cohesión y coherencia [4, p. 110]”.

De esta forma, el producto documentario “resumen” puede y debe extraerse de la gradación-superestructura textual propuesta anteriormente, respondiendo a la cohesión y coherencia necesarias para recuperar la información.

Para establecer una relación entre la proposición para el tratamiento de las informaciones en las bases de datos y la realidad, se llevó a cabo una investigación en dos bases de datos con diferentes características: Metadex: *Metals Science (Materials Information)* de tipo bibliográfico, contempla diferentes tipologías documentales y *Derwent World Patents Index*

| | |
|--|------------------------------|
| Dialog Accepción Number: Title: Index Terms: Patent Assignee: Author (Inventor): Number of Patents: Number of Countries: Patent family: Patent Code County: Patent Number: Patent Date: Derwent Week: Application Country Code and Number: Application Date: Language:Pages: International Patent Classification: | Derwent Accepción Number: |
|--|------------------------------|

Fig. 2. Derwent World Patents Index . Formato o gradación del contenido.

(*Derwent Information*) de tipo bibliográfico, contempla una única tipología documental: patentes.

Se seleccionaron estas dos bases a partir de los siguientes criterios: a) bases de datos producidas por instituciones de renombre; b) bases de datos utilizadas por usuarios de cualquier parte del mundo; c) bases de datos que procesan el análisis documentario y controlan el lenguaje; d) bases de datos con diferentes características: una trata diferentes tipologías documentales con un formato único-gradación de contenido, por tanto, superficial (Metadex), y una que trata una única tipología documental con un formato único-gradación del contenido, por tanto más abarcadora (WPI).

Primeramente se analizó la manera en que las dos bases de datos coleccionan las informaciones, observando qué contenidos asume la gradación del contenido o el formato utilizado (Fig. 1 y 2).

Observando los formatos o gradaciones del contenido utilizados por las dos bases de datos se constata una significativa diferencia entre la base de datos dedicada a una única tipología documental (WPI) y la que trabaja con diferentes tipologías documentales (Metadex).

La base de datos que trabaja con una sola tipología documental (WPI) utiliza un formato o gradación del contenido que contempla todas o casi todas las informaciones del documento de patente, respetando la estructura natural o la superestructura del documento, conforme afirma Van Djik [6], lo que no ocurre con la base de datos que trabaja con diferentes tipologías documentales.

Este tipo de bases de datos, por su naturaleza, se utiliza con un único formato o gradación de contenido que obligatoriamente contempla muchas tipologías documentales. Así, el formato-gradación acaba por perder informaciones importantes de varios tipos de documentos, pues no permite seleccionar con profundidad lo que se debe seleccionar en términos temáticos en un documento.

Al observar el formato o gradación del contenido utilizado en la base de datos Metadex, se percibe que las tipologías documentales libros y artículos de revistas, están retratadas en el formato, como contrapartida, la tipología documental patentes, casi no aparece en el formato o gradación del contenido. De esta forma, no consigue recuperar la mayor parte de las informaciones del documento —en este caso el documento de patente—, obligando esta tipología documental a encuadrarse en un formato o gradación del contenido patrón, que no satisface.

| | | |
|---|----|-----------------------------|
| 1778067 | MA | Accesión Number : 90-621097 |
| Multilayer plain Bearing | | |
| Mahrus, D; Carubelli, A | | |
| Metal Leve | | |
| Patent: GB2224788, UK 16 Oct. 1989 | | |
| 15 May 1990 | | |
| Journal Announcement: 9010 | | |
| Document Type: PATENT | | |
| Language: ENGLISH | | |
| Abstract: A multiplayer plain bearing comprises a backing layer, an optional intermediate layer, a diffusion or bonding layer, and a sliding layer, the sliding layer of laminar structure formed by the alternate deposition of layers of a softer material and layers of a harder material. The softer material may be Cu-Sn-Pb alloy and the harder material may be Cu or a Cu-Sn alloy. A corrosion-protection layer may be included. Descriptors: multiplayer plain bearing; plate friction; bonding layer; sliding layer; laminar structure; corrosion; deposition. | | |

Fig. 3. Metadex. Resultado de la investigación.

Para analizar con mayor rigor la importancia del formato o gradación del contenido, en consecuencia, los temas extraídos del documento para una futura recuperación, se elaboró una investigación en el sistema de información Dialog, más específicamente en las dos bases de datos anteriormente citadas, o sea, Metadex y WPI.

La estrategia de investigación solo utilizó el número de patente GB2224788 para su realización tanto en la base WPI como en Metadex. El objetivo era introducir un único dato-información respecto a un documento de patente y, a partir de él, obtener informaciones relevantes. El número de la patente también se seleccionó por estar contemplado en términos de formato o gradación del contenido en las dos bases de datos. La investigación se realizó a partir de estos criterios y los resultados se muestran en las figuras 3 y 4.

El resultado de la investigación que presenta la base de datos WPI es muy superior al resultado de la investigación de la base de datos Metadex. Se verifica que mediante el formato o gradación del contenido utilizado en WPI, dedicado a la tipología documental

| 008238823 | | WPI AccNo.: 90-125824/17 | | |
|--|---------|--------------------------|--------|------|
| XRPX Acc No.: N90-097464 | | *Image available* | | |
| Multilayered friction bearing for IC engines - has support plate, with friction layer of laminated structure | | | | |
| Index Terms: MULTILAYER FRICTION BEARING IC ENGINE; SUPPORT PLATE FRICTION LAYER LAMINATE STRUCTURE | | | | |
| Patent Assignee: (META-) METAL LEVE SA | | | | |
| Author (Inventor): MAHRUS D; CARUBELLIA | | | | |
| Number of Patents: 005 | | | | |
| Patent Family: | | | | |
| CC | Number | Kind | Date | Week |
| DE | 3934141 | A | 900419 | 9017 |
| GB | 2224788 | A | 900516 | 9020 |
| BR | 8805486 | A | 900605 | 9027 |
| AT | 8902385 | A | 901115 | 9049 |
| US | 5056936 | A | 911015 | 9144 |
| Priority Data (CC, No., Date): BR 885486 (881017) | | | | |
| Applications (CC, No., Date): DE 3934141 (8910212); GB 8923294 (891016); US422915 (891017) | | | | |
| Abstract (Basic): DE 3934141 | | | | |
| <p>The multi-layered friction bearing for internal combustion engines incorporates a support-plate (1) intermediate layer (2) and a diffusion or binding layer (3) and a friction layer (4).The friction layer (4) has a laminated structure of alternate soft (4a) and hard (4b) material. The softer material (4a) consists of a copper-tin-lead alloy and the herder material (4b) of copper or copper-tin alloy.</p> <p>USE / ADVANTAGE - The multi-layered friction bearing is especially for the connecting rods and crank shafts of internal combustion engines.</p> <p>Abstracts (US): 9144 US 5056936</p> <p>The multiplayer plain bearing of the type comprising a backing layer, and one of a diffusion and bonding layer adjacent to the backing layer. A sliding layer is adjacent to the one of a diffusion and bonding layer. The sliding layer has a laminate structure of separate alternated layers of a softer material and a harder material. The laminate structure has at least two layers of softer material and at least one layer of harder material positioned between two layers of softer material.Both the layer of the laminate structure are close to the backing layer and the layer of the laminate structure farthest from the backing layer being formed of softer material. The sliding layer has a thickness from 2 to 30 microum.</p> <p>USE - For internal combustion engines.</p> <p>Derwent Class: Q62; Q51</p> <p>International Patent Classification: F16C-033/12; F01C-021/02; F16C-009/00; C22C-009/02.</p> | | | | |

Fig. 4. WPI. Resultado de la investigación.

“patentes” se obtiene mayor número de informaciones, tanto cuantitativa como cualitativamente; es un resultado más abarcador y más exhaustivo que el de la base de datos Metadex.

La base de datos WPI, además de ofrecer la familia de patentes de la patente solicitada inicialmente a modo de investigación —información esencial para

que un investigador pueda conocer toda la ramificación del documento de la patente original, así como los países en los que se depositaron los pedidos de patente y en qué fechas, etc.—, ofrece otros datos como son: la clasificación internacional de patentes, el uso y sus aplicaciones. El resumen es más

completo, ofrece el resumen básico y también el resumen de las patentes depositadas en otros países.

El número de origen del documento de patente y los números con que se depositaron en otros países, son informaciones esenciales para acompañar y monitorear estas patentes, o ambos.

Los descriptores también se definen a partir de la elaboración del resumen, y se establece una secuencia numérica en el texto para su identificación. En el caso específico del resumen, WPI logra organizar con mayor claridad los aspectos relevantes para la recuperación, justamente porque tiene una gradación o superestructura del documento mejor establecida.

El formato que ofrece Metadex es tan incompleto si se compara con el de WPI, que se puede percibir a simple vista.

Conclusión

Comparando los formatos utilizados por las bases de datos Metadex y WPI se llega a la conclusión de que debe respetarse el formato/gradación/estructura de las diferentes tipologías documentales.

En relación con el tratamiento de la información, los productores de las bases de datos deben priorizar formatos o gradaciones del contenido específico para cada tipología documental, o sea, abarcar la heterogeneidad de los documentos, trabajando la tipología documental por separado, elaborando formatos o gradaciones del contenido diferenciados, de manera que se respete la estructura del texto y se retiren los contenidos temáticos, para que la base de datos no pierda información y le ofrezca al usuario informaciones lo más completas posible.

En este pequeño estudio comparativo se percibe claramente que para el documento de patente, la tipología documental seleccionada, la base de datos que respetó la estructura/gradación/formato documental, le brindó al usuario informaciones más completas del documento, y la que no respetó la estructura/gradación/formato documental, sólo le ofreció al usuario informaciones básicas.

La calidad de la respuesta de la base que trabajó con formato más adecuado la tipología documental, es muy superior. Además, hizo posible que el usuario tuviera contacto con informaciones que supuestamente él desconocía, como la familia de patentes y la clasificación internacional de patentes.

En la elaboración de las bases de datos se debe trabajar con metodologías coherentes que sin perder de vista al usuario final, faciliten el

tratamiento de la información, faciliten la investigación y el acceso a la información. Respetar la estructura/gradación/formato documental es muy importante.

El acceso y la obtención de información (pagados o no) transita por el tratamiento y el tratamiento es un factor básico para la calidad de la respuesta a los usuarios de esa tecnología de la información tan conocida, las bases de datos.

Evidentemente, el tratamiento de la información no abarca solo el formato. El análisis documentario, el control del lenguaje, entre otros aspectos, son tan importantes como el propio formato, pero la estructura/gradación/formato documental es la guía para extraer los contenidos inherentes al documento, por lo que cuanto más se respete el formato documental, mejor será la recuperación de la información relevante y mayor la calidad de la respuesta.

Las bases de datos deben pensar estratégicamente en la globalización de la información. La venta de la información a través de medios electrónicos y ópticos, es una realidad. El usuario comienza a tener un patrón de calidad de esa industria de la información. El acceso y la obtención de información (pagados o no) transita por el tratamiento y el tratamiento es un factor básico para la calidad de la respuesta a los usuarios de esa tecnología de la información tan conocida, las bases de datos.

Referencias

- 1) Cabré, M. T. Terminologie ou terminologies?: spécialité linguistique ou domaine interdisciplinaire? *Meta* 36(1):55-63, 1991.
- 2) Larivière, L. Vers un produit unifié en terminologie et en documentation le thésaurus terminologique. *Meta* 34(3):457-467, 1989.
- 3) Lara, M. L. G. A representação documentária: em jogo a significação.

Disertación de maestría. São Paulo, ECA/USP, 1993. 132 p.

- 4) Kobashi, N. Y. *A elaboração de informações documentárias: em busca de uma metodologia*. Tesis de Doctorado. São Paulo, ECA/USP, 1994. 195 p.
- 5) Guimarães, J. A. C. *Análise documentária em jurisprudência*. Tesis de Doctorado. São Paulo, ECA/USP, 1994.
- 6) Vandjik, T. A. *La ciência del texto*. Barcelona, Paidós, 1989.

Bibliografía

Albertin, Y. *Problemes techniques poses aux producteurs de base de donnees lors de l'implantation de leur base sur un "seveur"*. Paris, CNAM, 1980.

Cerpa, N. Pre-physical data base design heuristics. *Information & Management*. 28:351-59, 1995.

Cintra, A. M. M., M. de F. G Tálamo, M. L. G de Lara y N. Y. Kobashi. *Para entender as linguagens documentárias*. São Paulo, Polis/APB, 1994. 72 p. (Coleção Palavra-Chave).

Cunha, I. M. R. (comp.). *Análise documentária: considerações teóricas e experimentações*. São Paulo, FEBAB. 1989. 191 p.

Fayol, M. Le résumé: un bilan provisoire des recherches de psychologie cognitive. En Charolles, M. et Petitjean, A. *Le resumé de texte: aspects linguistiques, sémiotiques, psycholinguistiques et automatiques*. Paris, Klincksieck, 1991. pp.163-182.

Gardin, J. C. Document analysis and linguistic theory. *The Journal of Documentation* 29(2):37-169, June 1973.

Gomes, H. E. O indexador face às novas tecnologias de informação. *Trans-in-formação (Campinas) 1(2):161-171, maio/agosto, 1989*.

Lara, M. L. G. de. Linguagens documentárias, instrumentos de mediação e comunicação. *Revista Brasileira de Biblioteconomia e Documentação (São Paulo) 26(1/2):72-80, janeiro/junho, 1993*.

Lee, H. Justifying database normalization: a cost/benefit model. *Information Processing & Management 31(1):59-67, 1995*.

Salimon, A. M. *Proposta de integração entre bases de dados e recuperação de informação através de linguagem natural*. Disertación de maestría. São Carlos, UFSCAR, 1993. 100 p.

Smit, J. W. (comp.). *Análise documentária: a análise da síntese*. Brasília, IBICT, 1987.

Smit, J. W. Documentação e lingüística: interrelação e campos de pesquisa. *Revista Brasileira de Biblioteconomia e Documentação (São Paulo) 11(1/2), 1978*.

Spink, A. Term relevance feedback and mediated database searching: implications for information retrieval practice and systems design. *Information Processing & Management 31(2):161-171, 1995*.

Recibido: 16 de abril del 2001.

Aprobado: 20 de abril del 2001.

Marta Ligia Pomim Valentim

Universidade Estadual de Londrina
Centro de Educação, Comunicação e Artes
Departamento de Ciências da Informação
Campus Universitário
86.051-970 - Londrina - PR
Brasil
Correo electrónico: <valentim@uel.br>.
