

La arquitectura del contenido virtual

JOSÉ ALBERTO GONZÁLEZ SOLANO
Universidad Autónoma del Estado de México

INTRODUCCIÓN

El aumento de la producción de información en formato digital, el crecimiento del número de objetos virtuales así como el valor de esta clase de contenido, hacen necesario el desarrollo de mecanismos que incrementen la eficiencia en la recuperación de estos recursos por aquellos usuarios que la demandan. Una correcta identificación y descripción de los contenidos en un entorno virtual suministra la información necesaria para ser recuperados, así como para ser tratados. Esta función es cumplida por los metadatos.

Para Méndez Rodríguez (2002), los metadatos son los “elementos o estructuras de organización de la información que, asignados a cada objeto de información electrónica, la clasifica, la categoriza y la describe. Son, en definitiva, datos sobre datos, información estructurada sobre la información distribuida”.

El papel de los metadatos sigue tres vertientes que en la actualidad son de gran importancia en la gestión de la información: la actualización de contenidos, su preservación y las migraciones. Para el Consejo Internacional de Archivos (CIA), la preservación digital reside en el hecho de lograr que los documentos sean y permanezcan “auténti-

cos, completos, accesibles y comprensibles, procesables y potencialmente reutilizables”, en donde la correcta preservación se puede asegurar si “los metadatos indican el nombre y la versión del formato de cada archivo y la identificación del software con que ha sido creado y con que se ha realizado la última modificación”.

Los metadatos no sólo cumplen una función de representación del documento, sino que también ofrecen información sobre cómo gestionarlo y su trayectoria documental. La relación que existe entre el contenido informativo y los metadatos que lo describen debe estar basada en razones de coherencia funcional y flexibilidad.

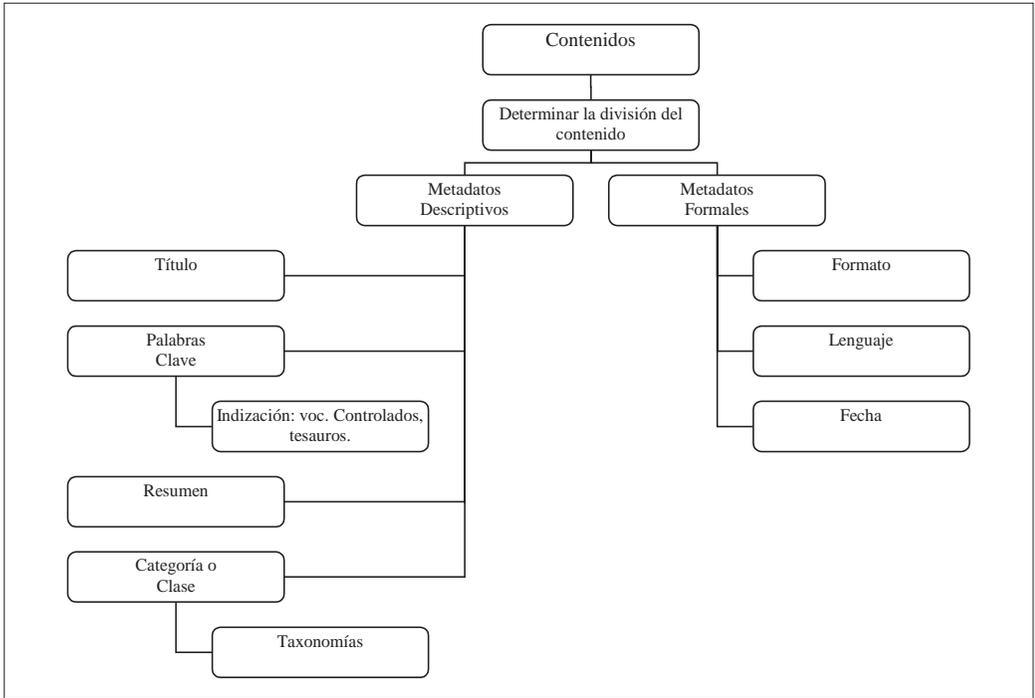
LOS METADATOS EN LA ORGANIZACIÓN DE LOS CONTENIDOS

El arquitecto de la información tiene la función de facilitarle el acceso a sus usuarios mediante la organización de grandes cantidades de contenido. El objetivo real de esta actividad, por lo tanto, radica en el hecho de que el usuario pueda ser capaz de satisfacer sus principales necesidades de información al tiempo que recuperar los contenidos con un esfuerzo razonable, en términos de tiempo y viabilidad para el propio usuario. En otras palabras, se trata de un proceso mediante el cual se pretende organizar un entorno que por naturaleza está desorganizado, con el fin de que la información sea recuperable, localizable y accesible.

El diseño de la arquitectura de la información, o lo que viene siendo lo mismo, la organización de los contenidos, puede resumirse en el esquema de la página siguiente.

CONTENIDOS Y SU DIVISIÓN

Los contenidos son todos aquellos extractos de información o materiales informativos que deben ser organizados y clasificados por el arquitecto de la información; éstos pueden ser, por ejemplo, imágenes, texto, video, sonido, etcétera. La primera etapa en el proceso de organización del contenido tiene que ver con el análisis que se hace de la información con el propósito de dividirla o descompo-



nerla en partes, una actividad a la que muchos autores se han referido como la fase de granularidad. En este sentido, el nivel mínimo de granularidad lo podemos entender como aquel grado de descomposición de un contenido en el que una u otra parte del mismo no pierde su significado comunicativo, es decir, el grado en el que físicamente resulta imposible seguirlo descomponiendo. Así por ejemplo, habrá casos en los que una imagen, pese a ser separada de un contenido siga estando vinculada al mismo contexto para el cual fue creada; y por el contrario, también existe la posibilidad de que una imagen separada de su contenido pueda utilizarse en distintos contextos por su asociación con diferentes objetos informativos.

De esta manera, una vez que se realiza la separación en partes de un solo contenido se procede a la siguiente tarea, que es la de catalogar o describir las respectivas unidades, o sea, elaborar los metadatos

que caractericen cada unidad de contenido mediante una serie de pares **atributo-valor/es**.

METADATOS

Los creadores de documentos multimedia, así como también de objetos en formato digital realizan registros detallados donde se almacenan datos que describen a cada uno de estos documentos o a cada uno de estos objetos. Algunos ejemplos de metadatos son los siguientes:

- ❖ Título del objeto o del documento
- ❖ Autor
- ❖ Fecha de publicación
- ❖ País
- ❖ Ciudad
- ❖ Ubicación de las copias existentes
- ❖ Descripción de los contenidos, o sea, resúmenes.

Así, de los tipos de metadatos mencionados los últimos son de particular importancia para recuperar la información en cuanto que facilitan la búsqueda de los documentos en función de sus contenidos. De esta manera caemos en la cuenta de que la exactitud de los metadatos es esencial para la precisión con la que un documento puede ser localizado; asimismo la proporción de documentos relevantes recuperados depende enteramente de que los metadatos estén completos, por lo que su creación debe ser muy minuciosa y rigurosa.

Por otra parte debemos considerar el hecho de que muchos archivos de tipo multimedia almacenan metadatos en forma de texto libre, de manera que la recuperación de los documentos consista en localizar aquellos en los cuales el texto contenga una determinada palabra clave, siguiendo las metodologías explicadas para las búsquedas en bases de datos textuales. Sin embargo, las búsquedas basadas en metadatos en formato de texto libre implican muchas veces importantes factores, entre los que se pueden mencionar los siguientes:

- ❖ Por lo general, el proceso de describir documentos multimedia mediante metadatos en formato de texto libre se realiza de manera manual, lo que supone de antemano una fuerte inversión de tiempo.
- ❖ El contenido o los contenidos que no hayan sido descritos no pueden ser visibles al efectuar las búsquedas.
- ❖ Debe pensarse en el uso futuro que se hará del documento por parte de los creadores de los metadatos, sobre todo para que pueda asegurarse que todas las características importantes del mismo están siendo debidamente reseñadas.
- ❖ En ciertos tipos de medios, tales como el sonido, las descripciones tienden a ser subjetivas, lo que implica un serio inconveniente en la recuperación de la información deseada.

Debido precisamente a las limitaciones que el texto libre ofrece, en la actualidad ha surgido una línea de investigación dirigida a la inclusión de metadatos en otro tipo de formatos, como por ejemplo el audio y el video. No obstante, en dirección paralela ha surgido otra línea de investigación que busca desarrollar herramientas automáticas de generación de metadatos e incluso el indexado de documentos. En este sentido, cabe mencionar que por el momento no existen sistemas automáticos de creación de metadatos ni mucho menos índices con esta característica, pues por ahora en alguna etapa del proceso se hace necesaria la intervención de un operador humano. Otra área de importante crecimiento en la actualidad es la estandarización del formato de los metadatos, sobre todo cuando se requiere del intercambio de documentos con sus respectivos metadatos entre dos o más archivos.

Los atributos pueden ser formales (los cuales tienen que ver con la administración), como por ejemplo, el lenguaje, el formato, la fecha de creación, el tamaño del documento, etcétera, que cumplen con una función relacionada con la gestión de los contenidos; o bien, pueden ser descriptivos, los cuales tienen el objetivo de facilitarle la recuperación de la información al usuario. Ejemplo de ello es la siguiente tabla:

III Encuentro de Catalogación y Metadatos

ATRIBUTO	VALOR/ES
Fecha	30/08/1998
Título	Diseño Participativo de Sistemas de Información
Palabras clave	Acceso abierto, repositorios, derechos
Categoría	Recurso en línea

La naturaleza propia de los contenidos es la que determinará, al final de una previa evaluación, la cantidad de metadatos que se deberán emplear.

PALABRAS CLAVE

Las palabras clave se encargan de resumir los conceptos más relevantes de un contenido. La polisemia y la sinonimia representan el primer gran inconveniente de las palabras clave, puesto que reflejan la ambigüedad propia del lenguaje natural. Un término –por ejemplo– puede representar más de un concepto distinto, o bien, puede ser representado por varios términos diferentes. Por esta razón se hace necesario emplear herramientas o mecanismos que aseguren la constancia de los términos empleados para esta clase de metadatos que ayuden a controlar el vocabulario empleado. Un vocabulario controlado es un índice de términos que unívocamente establecen relaciones precisas entre ellos y los conceptos que se representan; asimismo, un índice comprende una actividad relacionada con la catalogación y la descripción de los recursos mediante las palabras clave, las cuales forman parte de un vocabulario controlado o de un índice terminológico.

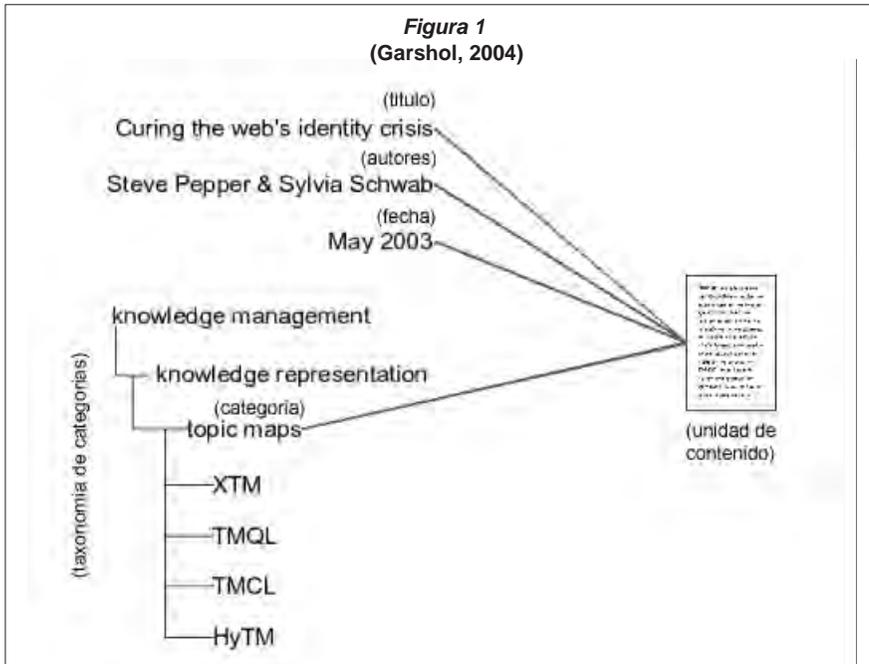
El uso de vocabularios controlados representa una serie de ventajas entre las que se pueden destacar las siguientes:

1. Se puede alcanzar un nivel alto de constancia de las palabras clave asignadas de modo tal que cada concepto tenga una correspondencia unívoca con un determinado término, con lo que se eliminan los obstáculos de polisemia y sinonimia.
2. El usuario puede recuperar de manera eficiente la información mediante los sistemas de búsqueda y navegación.

LAS TAXONOMÍAS

Mientras que las palabras clave asociadas con una unidad de contenido representan sus conceptos más significativos, la categoría describe el tema principal que engloba el contenido. Las categorías y sus relaciones se organizan en lo que en el ámbito de la arquitectura de la información se conocen como taxonomías de categorías.

De hecho, el término taxonomía proviene del campo de la biología, en donde se refiere al estudio de la clasificación de los seres vivos en estructuras jerárquicas. Así, el término fue adoptado por el campo de la arquitectura de la información en el que por lo general, y dependiendo del contexto en que se utilice, se le viene asignando ya sea uno u otro significado debido gracias al hecho de que aun en la actualidad suele ser un término bastante ambiguo y vago. Sin embargo, y procurando no perder el mismo sentido utilizado en la ciencia de la biología, una taxonomía no es sino una estructura jerárquica de elementos, o bien, de grupos de elementos. Para fines del presente trabajo, una taxonomía se entiende como la organización jerárquica del conjunto de categorías bajo las cuales se clasifican las partes del contenido. Un ejemplo de taxonomía es el que se observa en la siguiente figura, donde las líneas de arriba representan las relaciones jerárquicas entre categorías, y la línea de abajo el vínculo de pertenencia entre la parte del contenido y su categoría:



LA INTERACCIÓN DEL USUARIO CON EL ESPACIO DE INFORMACIÓN

La descripción y la categorización de los contenidos en línea están llamadas a facilitar la interacción entre el usuario y el espacio de información. Dicha interacción gira en torno a los sistemas de búsqueda y la navegación. El sistema de búsqueda es la herramienta más útil para el usuario al facilitarle la recuperación de la información que necesita. Su funcionamiento depende de la correcta indización de forma automática de la parte textual del contenido, la cual se realiza al contabilizar las frecuencias de ocurrencia o aparición de términos con el objetivo de fortalecer la capacidad de los mismos para representar el significado de la información. De esta forma, cada vez que el usuario realiza una consulta, el sistema de búsqueda equipara la consulta con los índices ponderados de términos o palabras clave, y le devuelve al usuario resultados que poseen mucho mayor pertinencia

con relación a los términos introducidos. Sin embargo, al contrario de la indización relativamente automática que por lo general es empleada en los sistemas de búsqueda actuales, la indización humana puede ser capaz de aumentar la recuperación de la información con base en la exhaustividad y la precisión. Mediante el empleo de la indización humana el sistema de búsqueda podría recuperar resultados en donde el usuario utilice términos que no necesariamente se encuentren en el cuerpo del contenido, sino entre las palabras clave asignadas al contenido por el arquitecto de la información. Para el caso de las fotografías u otra imagen en cualquier otro formato, la indización humana resulta particularmente útil ya que estos objetos de información carecen de partes textuales que puedan ser indexadas de forma automática.

CONCLUSIONES

De manera muy somera se ha intentado describir en este trabajo el proceso de catalogación (descripción), indización y clasificación de contenidos. Se puede comprobar que las posibilidades que ofrecen las herramientas de navegación basadas en metadatos son diversas y que deben emplearse de manera eficiente por el arquitecto de la información con el fin de facilitarles la tarea de recuperación a los usuarios.

A mayor grado de descripción de las partes de un contenido; es decir, a mayor número de metadatos aplicados, mayores serán las opciones de navegación que se le pueden ofrecer al usuario final.

En la actualidad el reto de cualquier profesional de la información /arquitecto de la información está en hacer accesibles los contenidos y objetos virtuales para que sean recuperables y reutilizables por los usuarios en el ámbito virtual. Es necesario crear una semántica que permita hacer entendible el contenido; es decir, que el contenido esté descrito de tal forma que pueda ser reconocido por cualquier sistema de búsqueda con el fin de que pueda ser recuperado de manera óptima.

BIBLIOGRAFÍA

- Baiget, Tomas (1997), “Metadatos-metadata”, en *El Profesional de la Información*, 6 (5), 34.
- Candás Romero, Jorge (2006), “El papel de los metadatos en la preservación digital”, en *El Profesional de la Información*, 15 (2), 126-136.
- Daudinot Founier, Isabel (2006), “Organización y recuperación de información en Internet: teoría de los metadatos”, en *ACIMED*, 14, (5).
- Díaz Ortuño, Pedro Manuel (2003), “Problemática y tendencias en la arquitectura de metadatos en web”, en *Anales de Documentación*, 6, 35-58.
- García Martínez, Ana María (2001), “Definición y estilo de los objetos de información digitales y metadatos para la descripción”, en *Boletín de la Asociación Andaluza de Bibliotecarios*, (63), 23-47.
- Garshol, Lars Marius (2004), “Metadata? Thesauri? Taxonomies? Topic Maps!”, en *Journal of Information Science*, 30 (4), 378-391.
- Méndez Rodríguez, Eva María (2002), *Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales*, Gijón, Trea.