# *Metadatos, folksonomías y taxonomías ¿Qué hay de nuevo en la representación y organización de la información?*

BLANCA RODRÍGUEZ BRAVO
*Universidad de León*

## 1. La representación y organización del documento digital/multimedia

Los principales cambios de las dos últimas décadas en el ámbito de la representación y organización de la información han derivado de la necesidad de controlar la avalancha de documentos digitales. Las peculiaridades de estos documentos han forzado adaptaciones; sin embargo, las novedosas etiquetas adjudicadas a dichas modificaciones hacen considerar que las transformaciones son de mayor enjundia de lo que realmente son. Los avances tecnológicos han cambiado, en cierta medida, el cómo se gestiona la información, no el por qué, ni el para qué.

El modelo de representación de la información tradicional -la catalogación, indización y clasificación- presenta como principales fortalezas su estabilidad y correcto mantenimiento, su interoperabilidad en el entorno bibliotecario, su entendimiento común en dicho ámbito y las muy complejas descripciones del contenido elaboradas que favorecen la precisión en la recuperación. Como debilidades cabe destacar que se trata de modelos poco flexibles y extensibles, que su capacidad de adaptación es limitada, que son costosos de crear y mantener e incomprensibles fuera del entorno bibliotecario.

En fin, la proliferación de documentos web, heterogéneos, descentralizados y de calidad variable ha forzado la adaptación de los mecanismos de control y organización de la información. Nos parece adecuado, por tanto, comenzar esta exposición realizando una aproximación a las características del documento digital.

En opinión de Schamber (1996) un documento digital tiene una serie de características que lo distinguen del documento impreso: es fácilmente manipulable, es enlazable interna y externamente, es rápidamente transformable, es intrínsecamente accesible, instantáneamente transportable e infinitamente replicable.

Estas características del documento digital nos permiten observar que la primera diferencia fundamental entre los documentos informáticos y los demás reside en que en aquéllos se produce una disociación entre el soporte y el contenido. Las características que Schamber establece hacen referencia a los contenidos; los soportes han perdido relevancia; de hecho, los mensajes que contienen se copian con facilidad en otro soporte, transformándose, como señala esta autora, en la tercera característica que establece. Son los contenidos informativos los que se pueden manipular, enlazar remotamente, recuperar, transportar y copiar. El soporte es necesario para que el documento exista, pero su participación en el significado del documento ha decrecido.

Acertadamente, por tanto, distingue Moreiro (1999) entre las características de los documentos digitales y las de sus contenidos. Por lo que se refiere a los documentos informáticos, se distinguen de todos los anteriores por las peculiaridades siguientes. Se necesitan equipos informáticos para crearlos y para consultarlos; estos equipos permiten manipular los contenidos fácilmente, y por esta razón son documentos siempre abiertos; asimismo, permiten acceder a sus contenidos a distancia, lo que facilita la distribución de la información. La difusión de sus contenidos a través de redes permite que audiencias dispersas y múltiples accedan simultáneamente a ellos, de ahí que se sitúen por encima de los límites espacio-temporales. Su normalización y control no es sencilla, y, por último, son impulsores más de la comunicación que de la permanencia.

Por su parte, los contenidos fijados a documentos electrónicos presentan las siguientes propiedades novedosas: se permite el acceso hipotético a todos los documentos existentes, sin diferencias de procesamiento originadas en la materialidad del soporte ya que aportan por un medio único la información que hasta ahora se consideraba propia de diferentes soportes: de texto, gráficos, imágenes, sonido y vídeo. La digitalización ha informatizado las diferencias y ha hecho que los documentos electrónicos puedan ser impresos, películas, sonido y gráficos al mismo tiempo. El proceso se ha digitalizado y vuelto multimedia.

Una de las características de los documentos digitales es, por tanto, su carácter multimedia. Desde una perspectiva informativo-comunicacional se concibe

el concepto multimedia como el agrupamiento sobre un soporte o un único modo de acceso de imágenes fijas o animadas, sonidos, textos y datos almacenados en forma digital. Es común también el término hipermedia para referirse al mismo concepto, aludiendo a la conexión de los multimedia por mediación de los procedimientos hipertextuales.

Otras peculiaridades de los documentos electrónicos son las siguientes. Se comunican en tiempo real y con menor coste; son de fácil manejo, con posibilidad de manipular y recomponer sus mensajes; su control físico y la integridad de su contenido es problemática; son accesibles a todo tipo de audiencias por lo que su distribución se ha volatilizado; se diluye la distinción entre acceso y posesión, préstamo y venta; se distribuyen sin manifestarse como copias, por lo que es patente la dificultad de proteger los derechos de autor; la recuperación de la información se vuelve interactiva, el usuario tiene un papel más activo en la localización, identificación y uso de los contenidos multimedia; su permanencia es imprevisible, por un lado porque depende tanto de la tecnología utilizada para su creación, que el envejecimiento de ésta se convierte en un grave problema para su preservación y mantenimiento, por otro lado porque la fácil manipulación ya mencionada de estos documentos los convierte en más dinámicos y menos estables, se pueden copiar sin pérdida de calidad, el concepto diplomático de garantía de autenticidad ha quedado desfasado. Finalmente, mediante estructuras lógicas, los documentos electrónicos pueden enlazarse con otras informaciones que no están físicamente conectadas y para ello se sirven del hipertexto.

Hemos asistido en los últimos tiempos a un fenómeno trascendental, la migración de los sistemas textuales y audiovisuales a los nuevos entornos digitales. Precisamente la nueva categoría de documentos "multimedia" surge de la combinación de documentos textuales y audiovisuales posibilitada por los entornos informáticos. Los documentos digitales combinan los dos canales emisores (visual y acústico) y la trilogía de códigos disponibles (textual, icónico y musical). Los CD-ROMs o DVDs interactivos representan la primera versión digital del libro tradicional, pero el desarrollo de Internet y la Web ha propiciado una alianza entre las aplicaciones multimedia y las redes en línea cuyo alcance supera las anteriores formas documentales disponibles.

Dado que el documento multimedia se ha materializado en el documento digital, es decir, ha sido viable gracias a la tecnología digital, vamos a abordar las peculiaridades del tratamiento de los documentos digitales, que en muchos casos son multimedia. Evidentemente, ambos conceptos no son idénticos. Por un lado,

en los documentos digitales con soporte intangible que circulan por Internet, el peso de lo textual es todavía muy grande, hay documentos sólo textuales como los hay sonoros o audiovisuales. Por otro lado, el término multimedia no se aplica en exclusividad a la integración de códigos bajo un mismo soporte o con un único medio de acceso, sino también a la unión de libros, vídeos, casetes, cdroms, dvds, etc. con un mismo contenido y finalidad.

La aparición del ordenador multimedia ha popularizado este concepto y ha provocado un salto cualitativo y cuantitativo en la integración de los medios. La revolución de la WWW, entorno de comunicación multimedia e hipertextual, hipermedia, ha convertido esta realidad en omnipresente.

Por lo que se refiere a los documentos multimedia, son documentos en que se combinan, bajo las herramientas interactivas usuales, diferentes elementos comunicativos, texto, archivos sonoros, vídeo digital, etc. En los documentos multimedia interactivos el medio modifica el mensaje, lo que origina una clase de documentos distintiva a efectos de representación documental.

El análisis documental del documento multimedia habrá de considerar, por un lado, cada uno de los niveles o códigos comunicativos separadamente, atendiendo a sus peculiaridades, y por otro, los tres niveles conjuntamente, observando las transformaciones que experimentan como resultado de la combinación de códigos. Dada la inestabilidad de estos documentos los productos del análisis de contenido se integrarán entre los metadatos, lo que favorecerá su localización.

La problemática principal que presenta el documento digital no reside, no obstante, en la metodología a aplicar para la representación documental, que seguirá las pautas establecidas para el tratamiento de los distintos códigos de transmisión de contenidos: textual, sonido, imagen fija o en movimiento, sino en su organización lógica y autodescripción para permitir su localización y óptima recuperación por parte de los motores de búsqueda.

En el documento digital han desaparecido las limitaciones espaciales y temporales intrínsecas al resto de los documentos. Ahora la preocupación se orienta a que puedan ser localizados en la ilimitada selva digital donde los contenidos se atomizan en un mosaico de elementos cuyo sentido es reconstruido libremente por el usuario gracias al hipertexto. En este contexto surge la preocupación por la autodescripción y el concepto de metadato, noción que incluye información sobre el contenido y el contexto de los documentos digitales.

Evidencia Pinto (1999) que la colección de recursos multimedia web de Internet no se ha diseñado para permitir la recuperación organizada de la información. Sus contenidos están almacenados en una gran base de datos descentralizada, heterogénea, abierta, en evolución y sin grandes filtros de validación. De ahí que los mayores esfuerzos de los implicados en la nueva industria de la información se concentren en el tratamiento de este inmenso caudal informativo para dotarlo de una estructura que facilite su automatización.

La mayor carencia actual de la Red es un sistema universal de etiquetado, representación y estructuración de la información que permita la búsqueda y el procesamiento automático más adecuado de cualquier documento web.

La opción más utilizada para buscar información en la Red es el uso de los motores de búsqueda, cuya eficacia en cuanto a pertinencia de los resultados es baja en comparación con los sistemas de recuperación de la información tradicionales, debido, entre otros factores, a la falta de un mínimo tratamiento documental de la mayoría de los recursos de información almacenados en servidores web.

Para solventar esta situación han ido apareciendo diversas iniciativas dirigidas a la localización, identificación y descripción de los documentos digitales en Internet basadas en el uso de esquemas normalizados de metadatos. Los metadatos ayudan a la identificación, descripción y localización de los recursos en red, por medio de la estructuración de los datos de una forma similar a la que tradicionalmente se venía empleando con los restantes documentos, (véanse las normas ISBD, la norma ISAD, o el formato MARC), la diferencia principal estriba en que la representación de los documentos se realiza en el momento de su creación.

La necesidad de describir los documentos en su origen y de manera distribuida, es decir, no sólo como tarea de los profesionales de la información, es evidente. Los documentos digitales pueden describirse utilizando las normas de descripción conocidas, la ISBD(ER) y el Formato MARC, sin embargo, la proliferación, la inestabilidad y la diversa calidad de los contenidos que se crean en la web hace inviable para los centros de información el asumir dicha tarea en su totalidad. De ahí el interés de que los recursos se describan en el momento de su creación mínimamente y que sean los autores o editores quienes la efectúen.

## 2. Los metadatos

Los metadatos son descripciones estructuradas de un objeto de información que tienen como finalidad hacer útiles los datos. Para que los objetos digitales sean entendibles y recuperables por los motores de búsqueda necesitan un mecanismo para explicitar qué son y de qué tratan, quién garantiza la calidad/fiabilidad del contenido y para qué puede o debe utilizarse. Este mecanismo son los metadatos.

Se han definido los metadatos de forma redundante como datos sobre datos o información sobre información. Ilustrativa es la conceptualización que realiza Garduño (2000): (…) *los metadatos, considerados como el conjunto de elementos que pueden generar una semántica internacionalmente aceptada con el propósito de representar la información digital, evitar su dispersión a través de una sistematización apropiada y asegurar su recuperación. De manera general se puede señalar que los metadatos representan, por una parte, datos acerca de recursos informativos disponibles en redes, y contienen, por la otra, los elementos útiles para facilitar la identificación y localización de recursos digitales. En consecuencia los metadatos deben considerar el contenido, la condición, la cualidad y la calidad, entre otras características de la información digital*.

El establecimiento de metainformación no es algo nuevo; en los primeros lenguajes de marcado ya existe la posibilidad de introducir materias o de identificar al autor de un documento, como es el caso del lenguaje HTML. La novedad reside en la tendencia a aumentar el número de campos que informen sobre el documento, normalizando el tipo de informaciones y el etiquetado de las mismas.

Factores fundamentales en el éxito de Internet han sido los lenguajes HTML, SGML y XML que permiten la incorporación de estos metadatos en forma de etiquetas o *tags* que sirven para marcar las características fundamentales de los documentos digitales. La norma HTML no proporciona una lista cerrada de metaetiquetas, sino un método para declarar *tags* con el propósito de representar la información sobre el documento.

Se ha impuesto el acuerdo tácito entre la comunidad de creadores de páginas web en torno a la creación de tres etiquetas: <*author*>, <*keywords*> y <*description*>. Como afirma Codina (2000), cada vez hay más motores de búsqueda que entienden esas tres etiquetas y que les dan una triple utilidad: en primer lugar, suelen otorgar una mayor relevancia a los documentos cuyas palabras están en

esa sección, en segundo lugar, permiten una búsqueda más similar a la búsqueda por campos y, en tercer lugar, cuando un motor de búsqueda o un directorio encuentra la etiqueta utiliza el texto que contiene como resumen del documento, en lugar de intentar generarlo de manera automática.

Indica Méndez (2002) que las bibliotecas digitales y los *subject gateways* son el contexto ideal de recursos de información electrónicos organizados por medio de los metadatos, siendo éstos el sustento de sistemas informativos digitales de calidad. Gracias a ellos *los objetos en una colección pueden distribuirse, pero la colección se presenta cohesionada lógicamente, porque las estructuras de metadatos están asociadas a ella y soportan una navegación coherente*.

Los modelos basados en metadatos para la estructuración y representación de la información presentan una serie de fortalezas: son sistemas flexibles y extensibles, soportan la adaptación a entornos concretos, son modelos ampliamente aceptados fuera del entorno bibliotecario, presentan diversos grados de complejidad y son menos costosos de crear y mantener que los sistemas de representación de la información tradicionales. Como debilidad se puede mencionar que existen varios estándares, que la interoperabilidad no se halla bien resuelta y que favorecen más la exhaustividad en la recuperación que la precisión. Asimismo, todavía están en proceso de adopción en el ámbito bibliotecario.

No todos los metadatos son metadatos descriptivos utilizados para representar recursos de información. Existen metadatos administrativos, de preservación, técnicos y de uso.

Por lo que se refiere a los estándares de metadatos más conocidos se puede mencionar sistemas de metadatos de propósito general y sistemas de metadatos de propósito específico. De propósito general es el *Dublin Core Metadata Element Set* (DCMES) que puede usarse en todos los dominios, para todo tipo de recursos y puede operar conjuntamente con otras soluciones específicas http://dublincore.org/. Cabe mencionar también *Resource Description Framework* (RDF) http://www.w3.org/TR/rdf-schema/ que es más que un mero formato de metadatos. Se trata de un sistema que proporciona una infraestructura para la descripción de recursos basada en XML. RDF propone un modelo de datos coherente y un marco sintáctico para representar términos de metadatos, su semántica, relaciones de distintos modelos (MARC, DC, TEI, etc.)

Entre los modelos de metadatos de propósito específico podemos mencionar el *Text Encoding Initiative* (TEI) http://www.tei-c.org/index.xm, el Functional Requirements for Bibliographic Records (FRBR), *Encoded Archival Description* (EAD) http://www.loc.gov/ead, *Metadata Object Description Schema* (MODS) http://www.loc.gov/standards/mod y *Metadata Encoding and Transmission Standard* (METS).

Abordaremos ahora brevemente los modelos citados comenzando por apuntar las principales características de *Dublin Core*, intento multidisciplinar e internacional de especificación de un conjunto estándar de metadatos que sirvan para identificar con mayor precisión el contenido de los documentos y recursos web. El *Dublin Core* es la forma abreviada para el *Dublin Metadata Core Element Set*, proyecto cooperativo de ámbito internacional, promovido por OCLC y NCSA (*National Center for Supercomputer Applications*).

Los metadatos que especifica *Dublin Core* tienen la forma de etiquetas que se escriben como código HTML en la sección <HEAD> de los documentos web; pero, a diferencia de las *tags* señaladas antes, la norma *Dublin Core* propone un conjunto de 15 metadatos sobre los que existe suficiente acuerdo internacional para confiar en que pueden ser un instrumento de gran capacidad para describir documentos y recursos digitales en Internet. Su finalidad es facilitar el descubrimiento de recursos a través del uso de motores de búsqueda y bases de datos.

Uno de los objetivos principales del *Dublin Core* es proporcionar un método estándar para que sean los propios organismos creadores o editores de documentos HTML quienes describan los recursos de manera que sea fácil realizar después búsquedas mucho más precisas de lo que permiten los motores de búsqueda actuales.

Los 15 metadatos principales Dublin Core son los siguientes:

- Elementos relacionados principalmente con el contenido del recurso: título, tema, descripción, fuente, lengua, relación y cobertura.
- Elementos relacionados principalmente con el recurso cuando es visto como una propiedad intelectual: autor, editor, colaboradores y derechos.
- Elementos relacionados principalmente con la "instanciación" del recurso: fecha, tipo de recurso, formato e identificador.

El *Dublin Core* es uno de los pilares esenciales de la web semántica. Sus principales fortalezas derivan de su simplicidad y de su independencia sintáctica que le permite integrarse en la nueva estructuración de la información (XML/RDF).

Por lo que se refiere a MODS se trata de un esquema de metadatos complejo en XML orientado a las bibliotecas. Toma como punto de partida el formato MARC21 del que selecciona sus veinte elementos principales.

En cuanto a METS, es un estándar para codificar metadatos descriptivos, administrativos y estructurales de documentos incluidos en una biblioteca digital, expresados usando el lenguaje XML del Consorcio World Wide Web.

TEI, FRBR y EAD son esquemas con propósitos más concretos destinados a la descripción de documentos literarios, registros catalográficos e instrumentos de descripción archivística.

A modo de conclusión cabe destacar la variedad de esquemas de metadatos y la todavía no óptima compatibilidad entre ellos; de ahí las limitaciones que desde el punto de vista funcional presentan los buscadores para localizar recursos digitales, como ya se ha señalado. La solución ha venido de la mano de la recolección de metadatos y, principalmente, del protocolo OAI-PMH (*Open Archives Initiative-Protocol for Metadata Harvesting*). El protocolo citado es uno de los pilares que han permitido el desarrollo de los repositorios de información digital en acceso abierto (Rodríguez y Alvite, 2007).

## 3. Las folksonomías

Una folksonomía es un sistema de indización generado colaborativamente o indización social agregada, como la denomina Hassan (2006), dirigida a la descripción de recursos web. El *tagging* o etiquetado libre se ha impuesto como una nueva modalidad de indización en lenguaje natural, especialmente a través de las herramientas y recursos de la web social. Como señala Noruzzi (2006) el rasgo más característico de las folksonomías es que los responsables del proceso de etiquetado son, generalmente, los usuarios de los recursos y/o sus creadores. Las etiquetas o *tags* empleadas sirven a la mejora de la eficacia en la búsqueda de recursos porque el contenido es categorizado utilizando un vocabulario familiar y accesible para el usuario común.

El término folksonomía, que es el resultado de la fusión entre *folks* y taxonomía, fue acuñado por Vander Wal, y se puede entender como la organización de contenidos web (taxonomía) realizados por cualquier persona (*folks*). Se trataría de un sistema de asignación de palabras clave popular o social, donde no intervienen los profesionales de la información.

Folksonomías, *tagging* o etiquetado se presentan como alternativas novedosas para la organización y clasificación de la información en el contexto de la web 2.0 o web social, en la que el usuario se ha transformado de consumidor pasivo en un activo "prosumidor" de información, o persona que es productora y consumidora de un mismo producto (Rodríguez Yunta, 2009). Este fenómeno se identifica con una democratización en el ámbito de la información y el conocimiento, dado que la folksonomía se crea por agregación de información sin ningún punto de partida previo, y sin dirigismo por parte de expertos.

Las etiquetas son palabras o frases que los usuarios utilizan para representar el contenido de un sitio web o de una página web. Las *tags* son simples etiquetas para recursos web seleccionadas para ayudar a los usuarios en la recuperación posterior de dichos objetos de contenido. Sirven, además, para agrupar recursos relacionados. No existe un conjunto de categorías fijas, ni un vocabulario controlado predeterminado. Un usuario puede asignar las denominaciones en función de su punto de vista, necesidades e intereses.

Los servicios de folksonomías indican quién ha etiquetado cada recurso y proporcionan acceso a todos los recursos descritos por la misma persona. Ello facilita que los usuarios puedan establecer conexiones con otros usuarios interesados en la misma temática. De este modo, una comunidad de usuarios puede terminar estableciendo una única estructura de palabras clave para definir los recursos de un área concreta. Las listas de etiquetas establecidas son descriptivas de los intereses de cada cual así como de su modo de organizarlos. Por tanto, otros usuarios pueden localizar recursos web de su interés y personas con afinidades similares.

En resumen, los sistemas basados en folksonomías permiten:
- Almacenar los recursos personales preferidos o marcadores
- Analizar la historia de los *bookmarks* de los usuarios y crear grupos de usuarios con los mismos intereses
- Recomendar recursos que sean los comúnmente preferidos.

Por tanto, algunos sistemas de base folksonómica como los sistemas colaborativos de marcadores del tipo *Delicious* http://delicious.com/, sugieren etiquetas a considerar en la representación de recursos. Otros sistemas que utilizan folksonomías son *Flicker* http://www.flickr.com/ y *You Tube* http://www.youtube.com

A diferencia de los Favoritos de navegadores como *Internet Explorer*, los sistemas de folksonomías como *Delicious* permiten a los usuarios crear o eliminar asociaciones entre etiquetas y recursos web añadiendo, reemplazando o borrando etiquetas o *bookmarks*. La ventaja de archivar los marcadores por este sistema es que una vez que el marcador de un usuario se encuentra en la web, es accesible desde cualquier ordenador y no sólo desde el del usuario.

Otras ventajas del etiquetado libre serían las propias de todo sistema de indización en lenguaje natural: simplicidad, transparencia, establecimiento de pesos por popularidad y aparición inmediata de nuevos términos, además de razones de economía de escala, entre otras (Rodríguez Yunta, 2009).

El etiquetado hereda, así mismo, todos los problemas tradicionales de los vocabularios no controlados (Ros, 2008). Entre las desventajas de las folksonomías se puede mencionar la baja precisión en la recuperación que caracteriza a todos los sistemas que no utilizan un lenguaje documental en la indización. Indica Noruzzi (2006) que los principales problemas del etiquetado mediante folksonomías son: la polisemia, la sinonimia, los plurales y la profundidad o especificidad del etiquetado. A estos obstáculos se pueden añadir la ausencia de normas para la construcción de términos compuestos, y la presencia de etiquetas de tipo afectivo o subjetivo. Como afirma Rodríguez Yunta (2009) supone la apuesta por un sistema de recuperación basado en la serendipia, muy lejos del intento de construcción de sistemas que aseguren cierto equilibrio entre exhaustividad y pertinencia.

Como posibles mecanismos para la mejora, propone Hassan (2006) que se apliquen soluciones invisibles para el usuario final, como el empleo de procedimientos propios de la indización automática: ponderación por frecuencias de uso, por autoridad, desambiguación del significante en función del contexto, etc. En este sentido servirían de ayuda, asimismo, las agrupaciones de etiquetas y la introducción de fórmulas de *clustering*. Subraya también Rodríguez Yunta (2009) la necesidad de contar con herramientas de control del vocabulario, necesidad puesta de relieve igualmente por Spiteri (2007), quien pone de manifiesto la conveniencia de que existan recomendaciones básicas para la redacción de las

etiquetas, superar el uso masivo de unitérminos y encontrar soluciones para la desambiguación de entradas polisémicas.

Las folksonomías han modificado la metodología de los sistemas de organización del conocimiento convirtiéndolos en un proceso distribuido y descentralizado, eliminando el concepto de jerarquía y facilitando la indización web y la localización de recursos de interés para los usuarios. Se prevé que los sistemas que utilizan folksonomías evolucionen e introduzcan, paralelamente, el uso de vocabularios controlados que permitan a los motores de búsqueda presentar los resultados en *clusters* y representar cada uno con los términos o etiquetas que hayan tenido la máxima frecuencia o aceptación. Tendrán que ser capaces, así mismo, de recomendar etiquetas a otros usuarios: *Muchos usuarios que utilizan la etiqueta "Open Access" utilizan también "OA"*.

Rodríguez Yunta (2009) asevera que las herramientas documentales tradicionales, como los tesauros, pueden servirse de estos nuevos sistemas de indización popular. El principal efecto práctico de la comparación entre las folksonomías y los tesauros radica en la detección de nuevas entradas que pueden enriquecer el lenguaje controlado, en especial con fines de su utilización para la recuperación de documentos a partir de su resumen o de su texto completo. Con este objetivo el tesauro deberá incluir todas las formas que pueden expresar los conceptos correspondientes a su ámbito temático.

## 4. Las taxonomías

Las taxonomías son sistemas para organizar el contenido en sitios web, intranets o portales con el fin de facilitar la navegación y el descubrimiento de recursos de información. Hay que considerarlas como estructuras predeterminadas que se usan para dividir un área temática, o el contenido de un sitio web, y esta área temática en otras áreas más pequeñas y así sucesivamente con el fin de lograr una organización a partir de determinadas propiedades o características. Para su realización se requiere un análisis conceptual que diferencie esas propiedades o características.

Se trata de jerarquías semánticas. Así, indica la norma ANSI (2005) que las taxonomías incorporan las relaciones de equivalencia y de jerarquía. Exigen que sus componentes estén organizados y sus características definitorias son su finalidad, prioriza la exploración y, por lo tanto, su entorno de aplicación, el entorno digital.

Evidencia Currás (2005) que, etimológicamente, taxonomía procede de los términos griegos *taxis*, ordenación, y *nomos*, norma. Aristóteles fue uno de los primeros en utilizar este término para designar esquemas jerárquicos orientados a la clasificación de objetos científicos. El botánico Linneo designó con el término taxonomía la clasificación de los seres vivos en agrupaciones jerárquicamente ordenadas de más genéricas a más específicas (reino, clase, orden, género y especie). A partir de esta concepción clásica, se desarrolló la taxonomía como un subcampo de la biología dedicado a la clasificación de organismos de acuerdo con sus diferencias y similitudes.

En opinión de Grove (2003), los principios que proporcionaban una guía rigurosa para la construcción de taxonomías eran la base lógica, la observación empírica, la estructura jerárquica basada en la herencia de propiedades, la historia evolutiva, y la utilidad pragmática. Las fuentes terminológicas de la lengua general todavía recogen el significado especialmente orientado al entorno de las ciencias experimentales, como demuestra el artículo que incorpora el Diccionario de la lengua española, 22ª ed. (2001).

En su concepción clásica, vinculada a las ciencias experimentales, la taxonomía aplica un criterio monojerárquico en el establecimiento de los sistemas de clasificación, es decir, cada una de las agrupaciones o clases que lo componen sólo puede ocupar un lugar en la estructura jerárquica.

A principios de los años 90 del siglo XX, el concepto de taxonomía se incorpora a otros ámbitos del conocimiento, como la psicología, las ciencias sociales y la informática, para designar los sistemas de acceso a la información que intentan establecer coincidencias entre la terminología del usuario y del sistema. Los primeros especialistas que desarrollaron sistemas de organización de contenidos para la Web formaban parte del área de consultoría en gestión del conocimiento, y procedían de ámbitos próximos a la informática y la ingeniería (gestión de contenidos y arquitectura de la información). No conociendo la tradición de los lenguajes documentales del ámbito de la documentación, asignaron el término taxonomía para los sistemas que desarrollaban. Este término se mantiene en uso actualmente para designar los sistemas de organización de contenidos en el contexto de Internet, aunque la teoría y la práctica de los lenguajes documentales se han venido aplicando de forma intensiva en este contexto.

Centelles (2005) realiza un trabajo de identificación y confrontación de los rasgos semánticos con que se define. Para ello, lleva a cabo una amplia búsqueda

de definiciones en todos los ámbitos de estudio, desarrollo y/o aplicación del término taxonomía, descartando únicamente aquéllas elaboradas a partir de una concepción clásica del término. El análisis de las definiciones muestra que éstas inciden sobre cuatro variables: el lugar que ocupa la taxonomía en el ámbito de los sistemas de organización del conocimiento; el contexto informativo en que se aplica la taxonomía; las finalidades que persigue; y el modelo estructural con que se interrelacionan los elementos que la componen.

Más de la mitad de las definiciones restringen el ámbito de aplicación de las taxonomías: algunas a entornos digitales y, más específicamente, al desarrollo de sitios web; otras a entornos corporativos y, más concretamente, de empresas. En menos casos convergen ambos criterios de restricción del significado de taxonomía; muestra de esta corriente es la definición propuesta por Gilchrist, Kibby y Mahon (2000), que ha logrado una considerable aceptación en la bibliografía especializada:

*A correlation of the different functional languages used by the enterprise to support a mechanism for navigating, and gaining access to the intellectual capital of the enterprise by providing such tools as portal navigation aids, authority for tagging documents and other information objects, support for search engines, and knowledge maps and possibly, a knowledge base in its own right.*

Las definiciones que vinculan las taxonomías al entorno digital destacan, como finalidades prioritarias, la mejora de la navegación y el desarrollo de sistemas de búsqueda basados en la exploración y en la recuperación. Las definiciones que vinculan las taxonomías al entorno corporativo destacan el valor estratégico de las taxonomías en áreas como la gestión del capital intelectual y, en general, del conocimiento.

Desde el punto de vista estructural, la mayoría de las definiciones consideran que las taxonomías se caracterizan por la aplicación de la relación jerárquica entre los elementos que organiza. En los casos en que la definición de taxonomía se orienta a su posición en el marco de los vocabularios controlados (Fast, Leise y Steckel, 2003 y ANSI, 2005) las definiciones asignan a la taxonomía una posición central determinada por la aplicación de las relaciones de equivalencia y de jerarquía.

Sirva de ejemplo la definición de Moreiro (2009b): *es la clasificación o categorización de un conjunto de objetos de forma jerárquica. Se establece entre*

*ellos una relación esquemática de generalización-especialización, a partir de una semántica simple que puede mostrarse preferentemente mediante estructuras arborescentes*.

A la luz de las propiedades mayoritariamente aceptadas en las definiciones formuladas en los ámbitos de estudio, desarrollo y/o aplicación, Centelles (2005) propone la siguiente definición: *una taxonomía es un tipo de vocabulario controlado en que todos los términos están conectados mediante algún modelo estructural (jerárquico, arbóreo, facetado...) y especialmente orientado a los sistemas de navegación, organización y búsqueda de contenidos de los sitios web*.

Zhonghong, Chaudhry y Khoo (2006) establecen las similitudes y diferencias entre las taxonomías, los tesauros y las clasificaciones a la vez que fijan sus rasgos principales.

Con respecto a su finalidad, cabe destacar que las taxonomías se dirigen específicamente a la organización de contenidos en el contexto de la organización del conocimiento. No se limitan a describir contenidos, reflejan los objetivos, los procesos y el personal de la organización.

Por lo que se refiere a su estructura, las taxonomías básicamente se componen de dos elementos: estructura semántica y etiquetas. Chaudhry y Goh (2005) señalan que los rasgos clave de las taxonomías son una estructura hecha de categorías y relaciones que las conectan y que facilitan a los usuarios clasificar asuntos dentro de una jerarquía. La estructura jerárquica es la columna vertebral de las taxonomías.

En cuanto a las funciones de las taxonomías Zhonghong, Chaudhry y Khoo (2006) citan las siguientes: si bien pueden ser usadas en entornos muy variados, se las asocia con términos como navegación, intranets y portales, dado que fundamentalmente han demostrado su efectividad como sistemas de navegación utilizados en diferentes iniciativas web.

Presentan paralelismos con los sistemas tradicionales de clasificación y con los tesauros. Si los rasgos principales de las taxonomías son su estructura jerárquica y las etiquetas para denominar conceptos representados mediante términos, la primera característica la comparten con las clasificaciones y la segunda con los tesauros. Se considera que los tres sistemas tienen en común componentes, relaciones y terminología, si bien sus aplicaciones difieren. Asimismo, cabe

indicar que las taxonomías son posteriores a las clasificaciones y los tesauros, es decir, son una evolución de los anteriores.

No obstante, también difieren en muchos aspectos. Las diferencias más significativas entre taxonomías, esquemas de clasificación y tesauros, por lo que se refiere a su alcance y utilidad, pueden atribuirse al entorno para el que han sido creadas. Las clasificaciones y los tesauros, sobre todo las primeras, se usan en un entorno más general, bibliotecas y centros de información, sirven a grupos de usuarios mayores y se limitan a gestionar contenidos.

Las clasificaciones fueron creadas en la comunidad bibliotecaria y usadas para clasificar y ubicar colecciones de temáticas definidas de antemano. Los tesauros, por su parte, han sido creados en un entorno *online* y utilizados para indizar contenidos temáticos de documentos, además de para facilitar la búsqueda del usuario. Las relaciones que establecen entre su vocabulario son más ricas que las de las taxonomías. Clasificaciones y tesauros se encuentran más vinculados a la representación y organización de la información de documentos tangibles y a la comunidad académica, aunque se utilicen también en el entorno Web.

Las taxonomías se crean para trabajar con recursos digitales en la Web y no se limitan a tratar sus materias. Principalmente se utilizan con el fin de organizar los sitios web y para ello categorizan los recursos con vistas a la navegación. No se circunscriben a los contenidos documentales ni al entorno bibliotecario. Las taxonomías más que en los contenidos se focalizan en los usuarios.

Es preciso considerar que el contenido en un sitio web de una organización es complejo. Por ejemplo, en el entorno organizacional el contenido puede clasificarse por funciones, productos, departamentos, servicios, lugares y personal, además de por materias.

Asimismo, las taxonomías se caracterizan por su estructura dinámica y sus etiquetas intuitivas. Pueden tener una estructura multidimensional o facetada. Por el contrario, las clasificaciones clásicas se limitan a una estructura unidimensional.

De igual modo, los esquemas de clasificación y los tesauros son lentos de actualización y no pueden reaccionar con rapidez para cubrir nuevas áreas de interés. Por tanto se relacionan mal con la naturaleza dinámica del entorno orga-

nizacional y/o con los recursos web. Por el contrario, las taxonomías se adecuan a este contexto. Son flexibles y fáciles de modificar.

Las taxonomías crean sus estructuras jerárquicas basándose en un contexto dado y en unos usuarios determinados. Eligen etiquetas intuitivas en lugar de notaciones para dirigir la navegación del usuario. Las *tags* son más fáciles de comprender por el usuario y permiten una distribución en categorías flexible. Además, dichas clases pueden organizarse de modo alfabético o sistemático. Esta flexibilidad organizativa facilita la localización y navegación por los recursos, así como el mantenimiento sencillo de las estructuras jerárquicas.

Cabe subrayar que las taxonomías están presentes en todos los esquemas, tesauros, modelos conceptuales y ontologías (Moreiro, 2009b).

## 5. A modo de conclusión

A modo de conclusión señalaremos que los metadatos están al servicio de la descripción y recuperación de documentos web. Representan sus principales características, datos de identificación, de contenido, de utilización y de preservación. Su finalidad es facilitar la recuperación por medio de los buscadores. El paralelismo con la tradición de representación y organización de la información lo estableceríamos con el concepto amplio de catalogación que utilizan los bibliotecarios aludiendo a todos los datos que forman parte de un registro bibliográfico.

Las folksonomías son un sistema de representación del contenido de recursos web; en concreto se trata de indización en lenguaje libre, fundada en palabras y conceptos, y realizada de manera distribuida y colaborativa por los usuarios con la finalidad de compartir recursos.

Por último, las taxonomías son un sistema de organización de contenidos aplicado a portales y sedes web. Se trata de un modelo de indización controlada en la que la organización jerárquica es la columna vertebral. Se asemejan a las tradicionales clasificaciones bibliográficas pero no utilizan notaciones sino etiquetas, lo que las aproxima también a los tesauros.

## 6. Referencias bibliográficas


- ANSI (2005). Z39.19-2005. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Bethesda, Maryland, NISO.

- CENTELLES, M. (2005). "Taxonomías para la categorización y la organización de la información en sitios web". *Hipertext.net*, III. <http://www.hipertext.net>

- CHAUDRY, A. S.; GOH, L. H. (2005). "Building taxonomies using organizational resources :a case of business consulting environment", *Knowledge Organization,* XXXII, 25-40.

- CODINA, L. (2000). "Evaluación, descripción y representación de recursos digitales" en Rovira, C. y Codina, L. (dir.). *Documentación digital 2000*. Barcelona, UPF.< http://docdigital.upf.es >

- CURRÁS, E. (2005). *Ontologías, taxonomías y tesauros: manual de construcción y uso*. Gijón, Trea.

- FAST, K.; LWISE, F.; STECKEL, M. (2003). "Controlled vocabularies: a glossothesaurus", *Boxes & Arrows*, October 27.http://www.boxesandarrows.com/view/controlled_vocabularies_a_glosso_thesaurus

- GARDUÑO VERA, R. (2000). "Paradigmas normativos para la organización documental en los albores del siglo XXI", *Investigación Bibliotecológica*, XIV, XXVIII, 115-149.

- GILCHRIST, A.; KIBBY, P.; MAHON, B. (2000). *Taxonomies for business: access and connectivity in a wired world*. London, TFPL.

- GROVE, A. (2003). "Taxonomy" en *Encyclopedia of library and information science*. 2nd ed., rev. and enlarg. New York, Marcel Dekker, 2770-2777.

- HASSAN MONTERO, Y. (2006)."Indización social y recuperación de información", *No Solo Usabilidad*, V <nosolousabilidad.com>.

• MÉNDEZ RODRÍGUEZ, E. (2002). *Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales*. Gijón, Trea.

• MOREIRO GONZÁLEZ, J. A. (1999). "La industria de los contenidos" en M. Caridad Sebastián (coord.). *La sociedad de la información: política, tecnología e industria de los contenidos*. Madrid, Centro de Estudios Ramón Areces, 311-331.

• MOREIRO GONZÁLEZ, J. A. (2009a). "Folksonomía", en J. M. Díaz Nafría, F. Salto Alemany y M. Pérez Montoro (coord.). *Glosario Bitrum*. http://sites.google.com/site/glosariobitrum/Home/folksonomia-folksonomy

• MOREIRO GONZÁLEZ, J. A. (2009b). "Taxonomía" en J. M. Díaz Nafría, F. Salto Alemany y M. Pérez Montoro (coord.). *Glosario Bitrum*. http://sites.google.com/site/glosariobitrum/Home/taxonomia

• NORUZZI, A. (2006). «Folsonomies: (Un)Controlled Vocabulary?", *Knowledge Organization*. XXIII, IV, 199-203.

• PINTO MOLINA, M.; GARCÍA MARCO, F. J.; AGUSTÍN LACRUZ, Mª C. (2002). *Indización y resumen de documentos digitales y multimedia: técnicas y procedimientos*. Gijón, Trea.

• PINTO MOLINA, M. (1999). "Tratamiento de los contenidos en la Sociedad de la Información", en Caridad Sebastián, M. (coord.). *La sociedad de la información: política, tecnología e industria de los contenidos*. Madrid, Centro de Estudios Ramón Areces, 267-288.

• RODRÍGUEZ BRAVO, B. (2010). Apuntes sobre representación y organización de la información. Gijón, Trea (en prensa).

• RODRÍGUEZ BRAVO, B. (2002). El documento, entre la tradición y la renovación. Gijón, Trea.

• RODRÍGUEZ BRAVO, B.; ALVITE DÍEZ, Mª L. (2007). "E-science and open access repositories in Spain" *OCLC Systems & Services*, XXXIII, IV, 363-371.

• RODRÍGUEZ YUNTA, L. (2009). "Etiquetado libre frente a lenguajes documentales. Aportaciones en el ámbito de Biblioteconomía y Documentación", en Lloret Romero, N. (ed.). *IX Congreso ISKO-España: Nuevas perspectivas para la difusión y organización del conocimiento*. Valencia, Universidad Politécnica de Valencia, 832-845.

• ROS MARTÍN, M. (2008). "Folksonomías o el etiquetado social", en *Comunidad de prácticas: web social para profesionales de la información*. Madrid, Sedic, http://comunidad20.sedic.es/?p=144

• SCHAMBER, L. (1996). "What is a document? Rethinking the concept in uneasy times", *Journal of the American Society for Information Science*, ILVII, IX, 669-671.

• SPITERI, L. F. (2007). "The structure and form of folksonomy tags: the road to the public library catalogue" en Rodríguez Bravo, B. y Alvite Díez, Mª L. (eds.). *La interdisciplinariedad y la transdisciplinariedad en la organización del conocimiento científico: Actas del VIII Congreso ISKO-España*. León, Universidad de León, 459-467.

• ZHONGHONG, W.; CHAUDHRY, A. S.; KHOO, C. (2006). "Potential and Prospects of Taxonomies for Content Organization", *Knowledge Organization*, XXXIII, III, 160-169.