

Tareas y métodos de trabajo colaborativo en documentación informativa

José Antonio Moreiro González*

RESUMEN

La deseable ruptura de la brecha digital existente entre los diferentes países de nuestro planeta y en el interior de éstos, ha buscado soluciones provenientes de las propias tecnologías de la información que la sustentan. En el caso de los sistemas documentales de los medios de comunicación social, se puede destacar la búsqueda, no privativa, de instrumentos caracterizados por el libre acceso a la información y por establecer lazos de cooperación. Se reflexiona aquí sobre las ventajas e inconvenientes que muestran tres recientes instrumentos del análisis y la gestión del contenido, en aplicación especial pero no exclusiva, a los archivos de prensa: el *software* libre para la gestión documental de periódicos digitales, las *folksonomías* por las que se indizan cooperativamente los contenidos y finalmente, el empleo de ontologías.

Palabras clave: documentación informativa, redacciones digitales, gestión de contenidos integrados, trabajo colaborativo, *folksonomías*, *software* libre, ontologías.

COLLABORATIVE WORKING METHODS AND TASKS IN INFORMATIVE DOCUMENTATION

ABSTRACT

The desirable rupture of the existing digital gap among different countries and inside them has looked for solutions from the information technologies themselves. In the case of the mass media documentary systems it is possible to emphasize the non privative search of instruments characterized by the free access to the information and by establishing cooperation links. This article discusses about the advantages and disadvantages shown in three recent instruments of content analysis and management, in special but nonexclusive application to the press archives: free software for the documentary management of digital newspapers, the folksonomies by which the contents are cooperatively indexed, and finally, the use of ontology.

Key Words: Informative documentation; Digital newsroom; Integrated contents Management; Collaborative work; Folksonomies; Open source software, Ontology.

* Departamento de Biblioteconomía y Documentación. Universidad Carlos III de Madrid.
Correo electrónico: joseantonio.moreiro@uc3m.es
Fecha de recepción: enero 26 de 2006.
Fecha de aprobación: enero 31 de 2006.

INTRODUCCIÓN

La actividad profesional se ha visto modificada por la presencia ineludible de las tecnologías de la información en los medios de comunicación social. Las rutinas del trabajo han envejecido a causa de una información que ahora es de dimensión planetaria, a la que es fácil acceder, que transmite al instante las noticias con gran economía de medios y que atiende a servir la información de acuerdo con el perfil pedido por el usuario. Una de las consecuencias más interesantes de esta transformación tecnológica en los medios ha sido la aparición y desarrollo, en tiempo muy escaso, de un nuevo mercado de servicios en línea para las noticias de archivo¹ o las listas de hemerotecas de periódicos en línea que contienen hasta 19 millones de páginas de 1440 periódicos, como es el caso del *Newspaper Archives Online*². De manera que el servicio de documentación se ha convertido en algunos medios en un recurso fundamental para el trabajo en las redacciones, al tiempo que se ha transformado en centro que almacena muchos productos comercializables.

Se intenta abordar aquí, desde una perspectiva de análisis valorativo, el proceso de transformación que, debido al uso generalizado de las tecnologías digitales en las redacciones y emisoras, ha motivado que se hable de Ciberperiodismo como referencia a una nueva actividad que genera nuevos productos para la prensa, la radio o la televisión, en muchas ocasiones sin poderlos asignar a un medio concreto. En este ambiente, la documentación se ha vuelto un instrumento imprescindible para conocer los recursos de información existentes y saber qué uso

darles (Morán, 2005). Las hemerotecas han retomado así protagonismo como producto informativo, pues sus contenidos, ofrecidos por la web, se han vuelto de interés inmediato para estudiantes, investigadores, historiadores, empresarios, aficionados y desde luego, para los propios periodistas. El suministro de tecnología para construir y gestionar hemerotecas que ofrecen herramientas de acceso, publicación y pago, es ahora un interesante campo de negocio como lo demuestran las líneas abiertas por *NewsViews Solutions* que suministra *software* libre para archivar y distribuir textos, fotos y gráficos² o la del *ActivePaper Archive* usado por muchos editores de periódico para integrar, preservar y vender archivos actuales e históricos.

LA GESTIÓN DE CONTENIDOS INTEGRADOS

Es destacable el impacto que ha tenido la introducción de las nuevas tecnologías de la información en el modo en que la información se produce. El hecho de contar con plataformas en las que se integran los aspectos, hasta ahora diferenciados, de la elaboración, gestión y publicación de los contenidos, incluyendo la recepción de información proveniente, de agencias, la elaboración de contenidos propios y la maquetación, integrando estas funciones a la gestión de la documentación, es un modo de actuar muy similar al seguido por los *datawarehouses* en las empresas. Evidentemente, estos cambios han supuesto de inmediato modificaciones en la producción y el consumo de información (Moreira, 2000):

1 Aparecen como claros ejemplos de estos nuevos servicios la *NewsLibrary: News Research made easy* de NewsLibrary, <http://www.newslibrary.com>, que recoge noticias de 791 periódicos y de otras fuentes, por lo que ofrece el archivo más completo de cada una de esas publicaciones encontrando los contenidos con mucha rapidez, o The British Online Newspaper Archive de la British Library, <http://www.uk.olivesoftware.com/>

2 Para una mayor información remitase a: *NewspaperArchive.com*, en <http://www.newspaperarchive.com/DesktopDefault.aspx> [Consultado el 05.12.05].

- ◆ Gran parte de la información se produce, consulta y conserva en soporte digital, lo que favorece el procesamiento técnico común de los anteriores soportes diversificados para prensa, radio y televisión.
- ◆ Cada vez es más necesario conseguir información especializada que sirva para alguna de las facetas de la compleja vida actual.
- ◆ El envejecimiento casi instantáneo de la información obliga a que su acceso se haga de forma urgente y fiable.
- ◆ La propia red es un nuevo medio que integra a todos los demás y que facilita la comunicación instantánea entre los diversos centros y entre los redactores de cada uno de los medios.
- ◆ Algunas actividades informativo-documentales se basan en la interactividad que permite el nuevo marco tecnológico mezclando las características propias de la creación informativa con las del usuario, como es el caso de los *weblogs* (Fernández, 2002).
- ◆ El crecimiento de los documentos multimedia ha favorecido especialmente la comunicación audiovisual.

En relación a este último rasgo, resulta llamativa la nueva organización funcional y creativa de los medios que nacieron en ambientes totalmente digitalizados, de modo especial, los sistemas de noticias de algunas televisiones de reciente aparición, como pueden ser algunas de las cadenas autonómicas españolas. En estas nuevas emisoras, todos los puestos de periodistas trabajan sobre plataformas de texto, audio y video de alta resolución,

desde las que elaboran por completo sus noticias: escriben el texto, manejan el *software* de edición, usan servidores de video o unidades de almacenamiento compartido donde pueden encontrar las imágenes que necesitan (García, 2005). Estas imágenes pueden provenir de la propia emisora, de diferentes agencias de noticias, o de redes de emisoras de Televisión consorciadas. Tras emitirse, las noticias son archivadas con un cuidado control documental que permitirá su uso posterior mediante la integración con todos los servidores de emisión.

La principal consecuencia de trabajar con una redacción digital es la aparición de una nueva figura que asegura el buen funcionamiento de todo el sistema digital. Se trata del gestor de contenidos, figura que reúne los rasgos del experto en conocimientos, junto a los de canalizador de la información, con habilidad para saber qué información existe y para encontrarla y con aptitud para adaptarse a la evolución continua de las nuevas tecnologías. El gestor de contenidos trabaja con objetos media y con cada uno de los metadatos que los definen e identifican. El gestor controla, supervisa y mantiene cada uno de los servidores del sistema que aseguran la viabilidad de la emisión, y programa las tareas cotidianas como son, establecer fecha, horas de ingesta, transferencias entre servidores, espacio disponible para el trabajo, transferencias de archivo, material utilizado por programas al igual que transferencias diarias entre servidores, etc.

En su figura, en definitiva, coinciden las actividades que permiten crear, gestionar, y emitir informaciones, pero también las de almacenamiento, acceso y reutilización. De esta forma, la documentación informativa se ha adaptado para responder a unos medios en cambio continuo:

- ◆ Participando en la gestión de los ilimitados recursos de información.
- ◆ Estrechando la proximidad que existe entre producir y tratar noticias o programas (García , 2005).
- ◆ Contribuyendo a orientar y a formar a los usuarios.
- ◆ Promocionando, ofertando y vendiendo productos informativos.
- ◆ Haciéndose un experto en organizar los flujos internos de información, e incluso en seleccionar las entradas y salidas de la misma.
- ◆ Al tiempo que habilitándose como experto en conocimiento y como tecnólogo.
- ◆ Aunque, conservando la función de hacer accesible la información correcta para utilizarla con el fin de elaborar nuevas informaciones.

Internet se ha transformado en el recurso de mayor utilización en los medios tanto a la hora de generar información como a la de obtenerla y gestionarla por parte de los servicios de documentación; también de hacerlo en multifunción, pues se accede y propaga la información mediante textos, imágenes y sonidos.

Nos centraremos ahora en tres aplicaciones que aportan soluciones a las necesidades surgidas últimamente en la gestión de los contenidos dentro de los sistemas de documentación informativa:

- ◆ El empleo de *software* libre.

- ◆ La elaboración cooperativa de lenguajes de indización libres o *folksonomías*.
- ◆ La elaboración de lenguajes de indización controlados mediante ontologías.

SOFTWARE LIBRE PARA LA GESTIÓN DOCUMENTAL DE PERIÓDICOS DIGITALES

Si podemos apreciar la década de 1990³ como el momento en que aparecieron las estructuras abiertas sobre las cuales se integran los sistemas (Unsworth, 2004), la década actual corresponde con lo abierto (*open source, open archives, open systems, open access*). En ella se amplía el valor del proceso de elaboración de los programas realizados socialmente a través de una comunidad de programadores enlazados mediante Internet, en un desarrollo cooperativo de enorme eficacia. Es también novedosa la actitud favorable a reutilizar módulos aplicables a la solución de funciones concretas con código proveniente de otros programas, economizando esfuerzos, garantizando resultados y, de paso, integrando sistemas.

El «*free (software)*» no es un «*software gratis*», si no «libre», es decir franco, disponible y de uso expedito. Es un *software* que viene con autorización para que cualquiera pueda usarlo, copiarlo y distribuirlo, ya sea de modo literal o con modificaciones, gratis o retribuido, pero siempre el código fuente debe estar disponible para uso, desarrollo y distribución comerciales (Sturman, 2004: 257-270). El *software* abierto se caracteriza por distribuirse junto con su código fuente, posibilitando así su adaptación a unas necesidades concretas⁴.

³ Fueron el momento del *e-learning, e-mail, e-business, e-publishing, e-trade*.

⁴ Lo más frecuente es que los programas abiertos se distribuyan con una Licencia Pública General (GPL = *General Public Licence*) que autoriza no solo al uso ilimitado del *software*, sino incluso a su copia, distribución y modificación, con la condición de que cuantas alteraciones se hagan deberán mantenerse bajo la misma licencia. Las licencias disponibles constan en *Open Source Initiative* (<http://www.opensource.org>).

De todo el planteamiento anterior se puede apreciar que dos son los elementos principales: *el libre acceso a la información* y la *cooperación*, que han cuajado en múltiples proyectos para crear condiciones más fáciles de acceso libre a la información y para establecer lazos de cooperación más sólidos mediante las redes de comunicación, en especial de Internet (Tramullas, 2005), siempre con la finalidad de ayudar a romper la brecha digital (Arunachalam, 2005: 175-178).

Un recorrido por las herramientas que se utilizan para la creación y gestión de los contenidos que tienen licencia de *software* libre permite observar la presencia de algunas destinadas a atender las diversas prestaciones e indicaciones del periodismo digital (Castellanos, 2004: 571-576). Todas ellas basan su sistema de diseño en:

- ◆ Un servidor web.
- ◆ Un intérprete del lenguaje de programación.
- ◆ Un gestor de bases de datos.

Consideraremos, en primer lugar, la utilización de plataformas especialmente diseñadas para cubrir las necesidades de los *periódicos y revistas digitales*. En consecuencia, sus prestaciones están orientadas al control de los procesos de edición, creación y publicación de contenidos; más que al desarrollo de aplicaciones o portales (Melero, 2004). De especial interés van a resultar en los próximos años, los sistemas destinados a la publicación de periódicos diarios, pero también de revistas de información general y de las científicas para dar soporte al proceso de envío, revisión y publicación:

- ◆ **Cofax** (<http://www.cofax.org/>): sistema basado en la web para la gestión y publicación de conteni-

dos de textos y multimedia. Se trata de un marco de módulos y de herramientas de software para gestionar y distribuir el contenido de noticias provenientes de diversas fuentes. Se diseñó para simplificar la presentación de periódicos en la Web y para agilizar la publicación en la Web en tiempo real. Es software de código abierto desarrollado entre KnightRidder.com y Philly.com junto a un grupo de desarrolladores de código abierto. Se hizo usando Java, bases de datos y tecnologías XML. Se asemeja, en cierta manera, a productos comerciales como *Vignette StoryServer*, *FutureTense Internet Publishing System*, y *Pantheon Builder*, pero su puesta en marcha, mantenimiento y uso son mucho más baratos.

- ◆ **NewsViews Solutions:** *software* libre para archivar y distribuir textos, fotos y gráficos; ofrece la publicación integrada de las noticias, permitiendo a los redactores escribir, corregir y publicar tanto la edición impresa como en pantalla, así como archivar al mismo tiempo. Este *workflow* editorial integrado incluye el seguimiento de las múltiples generaciones de una noticia durante su ciclo vital⁵.
- ◆ **ePrints:** aunque no es exactamente un sistema de publicación de periódicos se incluye aquí por su interés para el ámbito académico y su posibilidad de ser adaptado para hacer publicaciones autoeditadas. Si lo traemos a colación es debido a que suministra *software* libre (GNU) para la gestión en línea de literatura de investigación científica mediante auto-archivo hecho por autores e instituciones⁶.

Por su parte, los **entornos de colaboración** se corresponderían con la clásica definición de

⁵ Consulte: <http://www.newsviewsolutions.com>

⁶ Consulte: <http://www.eprints.org/>

groupware: herramientas para trabajo en grupo. Su objetivo es ofrecer las prestaciones necesarias a grupos de usuarios especializados para que puedan realizar trabajos y proyectos en común. En este tipo de entornos, los flujos de trabajo, los usuarios, los puntos de control y los documentos entregables (*deliverables*) son los contenidos claves a gestionar. Se trata pues, de una especialización evidente de la gestión de contenidos que se relaciona con la gestión de proyectos. Un entorno colaborativo muy conocido y extendido es *wiki*, sobre todo destinado a la elaboración de documentos para comunidades de intereses comunes. Existen gran número de plataformas *wiki* que han aumentado progresivamente sus prestaciones, de manera que unos autores las sitúan entre las herramientas para portales, mientras otros las colocan en este apartado:

- ◆ eGroupware (<http://www.egroupware.org/>)
- ◆ phpCollab (<http://www.php-collab.org/>)
- ◆ Wiki (<http://www.wiki.org/>)

Finalmente consideraremos los *blogs* o *bitácoras*⁷, fenómenos bastante recientes, pero ya consolidados; por cuyo medio se le ha concedido auténtico potencial democrático a la red. Su existencia confunde las actividades de creación y las de uso (Pareja, 2002). En ellos, se da un modelo de gestión de contenidos simplificado, pues su flujo de trabajo ha sido hasta ahora, sencillo. Si bien comienza a observarse la presencia progresiva de contenidos multimedia. Son una muestra de la evolución de la prensa escrita tradicional que cuenta con la tecnología y las posibilidades audiovisuales. Las herramientas de publicación de *weblogs* son empleadas por muchos medios digitales, pues facilitan el proceso de publicación. Además, en el caso de los periodistas, la bitácora muestra sin intermediarios la pericia

comunicativa de su autor. A pesar de que los grandes servidores de *blogs* dan al usuario la solución ya desarrollada, hay herramientas para *blogs* con tecnologías LAMP/WAMP, que se descargan y funcionan con licencia libre⁸. Debe indicarse, en cambio, una tendencia a aumentar las prestaciones de los *blogs*, intentando dotarlos de parecidas características a las disponibles en las herramientas para portales:

- ◆ La plataforma semántica avanzada de publicación personal y preocupada con la estética, los estándares de la web, y la usabilidad. **WordPress**, (<http://wordpress.org/>) es libre y gratis al mismo tiempo, está construido sobre PHP y MySQL y con licencia GLP. Es el sucesor oficial de b2/cafelog, por lo que se trata de un *software* reciente, pero cuyas raíces y desarrollo se remontan al año 2001, por lo que se trata de un producto maduro y estable.

LAS FOLKSONOMÍAS, INDIZACIÓN COLABORATIVA DE CONTENIDOS

Denominamos con el neologismo *folksonomía* a la práctica de indizar en colaboración mediante el uso libre de palabras clave. El empleo de las *folksonomías* se incardina dentro de las actuaciones destinadas a conseguir la web semántica, donde cada página dispone de metadatos, en los cuales se describe su contenido⁹. En la elaboración de una *folksonomía* intervienen de forma espontánea personas que cooperan a la hora de organizar la información en categorías. Precisamente aquí reside su principal ventaja, la de ser una arquitectura de la información que se construye de manera social y en colaboración. Se permite así, que cada internauta asigne libremente palabras clave personales

7 Un *weblog*, también conocido como *blog* o bitácora, es un sitio web periódicamente actualizado que recopila cronológicamente textos y/o artículos de uno o varios autores donde el más reciente aparece primero, con un uso o temática en particular, siempre conservando el autor la libertad de dejar publicado lo que crea pertinente. Los *weblogs* usualmente están escritos con un estilo personal e informal (<http://es.wikipedia.org/wiki/Weblog>).

8 Incluye un grupo de herramientas cuyo acrónimo es LAMP (WAMP en versión Windows): Linux, Apache, MySQL, y PHP.

9 Se atribuye el origen del término a Thomas van der Wall, como resultado de combinar las palabras *folk* (gente, popular) y taxonomía (Gestión *-taxis-* de la clasificación *-nomos-*), de forma que *folksonomía* viene a ser etimológicamente «clasificación gestionada popularmente», o clasificación cooperativa. Respecto a los metadatos. Méndez, E. *Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales*. Gijón: TREA, 2002.

mediante etiquetas (*tags*) en las que se categoriza todo tipo de contenidos con el objetivo de prepararlos para su recuperación, desde enlaces de noticias hasta fotografías, pasando por artículos especializados, canciones, entradas en un *blog*; en fin, todo tipo de información y de objetos con la única condición de que dispongan de una URL.

Precisamente, la utilidad viene dada por la posibilidad de indizar un mismo asunto de manera diferente, asignándole etiquetas sinónimas (multívocas) o utilizando la misma etiqueta para asuntos diferentes (polisemia), pues de esta forma se generan inmediatamente relaciones entre conceptos que muestran similitud. De esta forma, en la era de la globalización que supone Internet, el conocimiento se organiza localmente, lo que significa que ninguna categoría puede ser considerada universal, y se hace mediante la libre participación a la hora de asignar etiquetas para identificar y clasificar información digitalizada.

Así pues, la principal novedad que aportan las *folksonomías* proviene del uso compartido de etiquetas que genera un sistema de categorización no jerárquico, pues por encima de clasificar los contenidos importa contar con un instrumento que ofrezca la posibilidad de emplear libremente etiquetas semánticas para efectuar la indización¹⁰. De forma que nos movemos dentro de unas categorías que evolucionan según la participación de los usuarios, en vez de contar con rígidas jerarquías predefinidas¹¹. Los metadatos así obtenidos, reflejan no solo la naturaleza de lo indizado, además ayudan a los usuarios a encontrar recursos de for-

ma automática de acuerdo con sus intereses, debido a que:

1. Están hechos por cualquier persona (Weiss, 2005).
2. Para crearlos se requiere manejar muy pocas reglas.
3. Se producen sin interés propio.
4. Las ventajas de los usuarios crecen con la agregación.
5. No se rompen cuando hay datos incompletos.
6. No están diseñados *a priori*, lo que les hace más flexibles.
7. Se presentan en forma de conjuntos y no de jerarquías.
8. No pertenecen a nadie, pues nadie las centraliza ni controla¹².
9. Son relevantes para los propósitos e intereses de un sitio web.

Entre los sitios web donde las *folksonomías* se hacen patentes, son muy conocidos *Del.icio.us social bookmarks* y *Flickr* (de Yahoo)¹³. Los creadores de *Del.icio.us* se definen como gestores sociales de sitios web favoritos. Su producto permite clasificar mediante palabras clave los sitios para poder recuperarlos después e intercambiar esta información con otros usuarios. A través de una palabra clave propuesta por un usuario, se puede navegar por los contenidos etiquetados. Se capta pues, el conjunto de términos que unos usuarios emplean para clasificar un contenido¹⁴.

Por su parte, *Flickr* es un sitio web sobre fotografía en el que los usuarios pueden compartir fotos mediante un sistema de organización colaborativa que

10 Permite que alguien marque una página web que sea de su interés, que organice estas páginas por categorías usando etiquetas libremente y que comparta sus descubrimientos con otras mentes curiosas. Jon Udell publicó una serie de artículos acerca de *del.icio.us* en su *blog*: weblog.infoworld.com/udell/ [consultado el 29.07.05]. *How Do You Use del.icio.us?* <http://www.primidi.com/2004/11/08.html#a1019> [consultado el 29.09.05].

11 Para mayor información consulte: *Folksonomie*. 2004. Many2many. A group weblog on social software. www.corante.com/many/archives/2004/08/25/folksonomy.php [consultado el 08.08.05]. En Many2many aparecen constantemente noticias sobre *folksonomías*.

12 Desde la teoría general de sistemas la *folksonomía* se comprende como una herramienta generada de abajo para arriba (*bottom-up*), pues nadie define la reglas a seguir. Habla de estas herramientas: Bertalanffy, V. *Perspectivas en la teoría general de sistemas: estudios científico-filosóficos*. Madrid: Alianza, 1992.

13 Tutorial de *Flickr* en español: <http://pop-page.com/recursos/tutorial-flickr-en-espanol>

14 Otra iniciativa similar a *Del.icio.us* pero quizás menos popular, es *CiteULike.com*. *A free online service to organise your academic papers*. www.citeulike.org/ [consultado el 31.12.05].

las dispone para la consulta¹⁵. Es el propio usuario el que asigna palabras clave a cada fotografía de acuerdo con lo que su contenido le sugiere.

En esos dos ejemplos se puede observar que existe una distinción entre (Vander Wal, 2004):

- **Broad folksonomy** o **folksonomía genérica**: en que diferentes usuarios etiquetan el mismo objeto y cada uno puede hacerlo con sus propias etiquetas o vocabulario, tal como sucede con *Del.icio.us*, por eso se las conoce también como relacionadas. En este caso son sitios web los que se etiquetan y comparten.
- **Narrow folksonomy** o **folksonomía específica**: en que una o pocas personas etiquetan la información. Así sucede con *Flickr*, cuyo uso está determinado por el hecho de poder alojar fotografías propias y etiquetarlas para que luego otros puedan buscar mediante esas etiquetas. De manera que en esta *folksonomía* se clasifica según la forma en que los usuarios etiquetan sus contenidos.

Parece obvio que las folksonomías se usan porque resultan efectivas para los usuarios a la hora de ordenar su información. Además, al tratarse de usuarios inexpertos en la organización del conocimiento, el tiempo que emplearían usando complejos sistemas de clasificación jerárquicos se reduce mucho con este método.

Dado que quienes organizan la información en los *blogs* y los defensores del uso social del *software* suelen ser los principales usuarios, las *folksonomías* producen resultados que reflejan con exactitud el modelo de conceptos que tiene la gente¹⁶. Se trata

de unos sistemas simples y emergentes, que se alejan de la manera tradicional de clasificar desde arriba hacia abajo, por lo que su ventaja proviene de la capacidad de emparejar las necesidades verdaderas de los usuarios y la lengua, no de buscar su precisión (Quintarelli, 2004). Por lo que se refiere al empleo de las *folksonomías* en periodismo, beneficia tanto a los periodistas como a los lectores. Los periodistas pueden usar sus propias palabras para clasificar los contenidos que les interesan, buscar sitios favoritos o encontrar lo último que haya aparecido. Mientras que los lectores y los gabinetes de prensa están así capacitados para montar su propio periódico con los temas que les interesan¹⁷.

Otra propiedad de las *folksonomías* es que incluyen cada una de las palabras propuestas, sin dejar ninguna fuera. No hay una autoridad central que imponga su visión, de forma que cada voz tiene su lugar. Este aspecto implica que tenemos también la opción de derivar hacia ideas no entroncadas con la corriente principal, que pueden emerger del interés de una parte pequeña de la población (Porter, 2004).

Marcar con etiquetas de carácter social puede considerarse «un puente de baja inversión entre la clasificación personal y la clasificación compartida» (McMullin, 2005). De hecho, la clasificación facetada ofrece más flexibilidad y una aproximación semántica. Sin embargo, las facetadas no podían ser la respuesta correcta cuando la gente está obligada a clasificar el contenido que ella misma produce, como sucede hoy. Si el número de términos y de sus combinaciones se amplía, las clasificaciones facetadas multiplican el número de las decisiones requeridas para clasificar un documento, lo que agrega un

¹⁵ <http://www.flickr.com/> [consultado el 31.12.05].

¹⁶ Por lo general, los *bloggers* han clasificado sus escritos en categorías, a ello se añade en su aplicación, que es la gente la que clasifica, no los buscadores. Quizás por ello las *folksonomías* tienen la ventaja de permitir aumentar el flujo de lectores a un *blog*. Estaría aquí la razón del éxito de *Technorati*, que ha sabido convertir las clasificaciones espontáneas en etiquetas para poder agrupar las materias exploradas por miles de usuarios de *blogs*, y cuyo manual de uso está disponible en: <http://consumingexperience.blogspot.com/2005/02/technorati-tags-introduction.html>.

¹⁷ Así el periódico colectivo *Common Times* que utiliza etiquetas asignadas y recomendaciones escritas por los propios lectores: <http://www.commontimes.org/> [consultado el 30.12.05]. Nació con la idea de que la primera página se conformase con las noticias más consultadas. En su grupo, se sitúan otras iniciativas gratis y libres como *Common Tunes*, donde se obtiene música; *Common Bits*, con discursos y recursos multimedia sobre política y a la vez es gratuito; y *Common Flix*: para intercambio de videos.

significativo coste cognitivo a la clasificación. A menudo, esta inversión cognitiva es demasiado elevada para los *bloggers*, que prefieren describir sus envíos con palabras claves libres. De modo que, por encima de su posible eficacia en la recuperación, el uso de las *folksonomías* tiene como valor principal su carácter social y cooperativo, pues por su medio los servicios de almacenamiento y acceso a la información crean nexos entre los usuarios y cumplen una tarea que, de otro modo, quedaría desatendida (Vander Wall, 2005).

Entre las perspectivas de uso es conveniente hablar del lugar llamado a ocupar por las *folksonomías* en las empresas puesto que éstas son generadas por los usuarios y es por lo tanto, conveniente ponerlas en marcha; ya que suponen una alternativa económica y más útil que las tradicionales taxonomías institucionales o los vocabularios controlados. Las *folksonomías* pueden además, resultar útiles para distribuir las tareas de gestión entre quienes trabajan en una empresa o institución. De forma que su presencia ha traído mayor capacidad a los sitios web, con efectos sociales inmediatos, nuevas oportunidades de negocio y el establecimiento de relaciones semánticas entre los documentos de la web.

LAS FOLKSONOMÍAS POR COMPARACIÓN CON LOS LENGUAJES CONTROLADOS

Los vocabularios controlados no son aplicables ni económicamente extensibles a la mayoría de los casos donde podrían utilizarse metadatos. Su construcción, mantenimiento y aplicación resultan a menudo demasiado costosos tanto en tiempo de desarrollo como en el tiempo que necesita el usuario del sistema para aprender el esquema de la clasificación. De manera que las *folksonomías* son una solución cuando la indización o clasificación tradicionales no son viables (Shirky, 2006).

Las *folksonomías* son sistemas de indización abiertos que emergen de forma espontánea, desordenada

y sin coordinar, frente a los sistemas tradicionales de indización, los cuales son jerárquicos y tienen unas categorías prefijadas de forma rigurosa. Mientras que en los nuevos servicios la libertad al usuario origina su propia etiqueta, después emerge la *folksonomía* del uso colectivo. Así, en *Flickr*, lo que comienza siendo una etiqueta personal se convierte en una categoría en la que se agrupan cientos de fotos en torno al mismo asunto, que además el propio servicio de *Flickr* explota. A partir de una de las fotos puede saltarse a otras de la misma etiqueta pertenecientes a otras decenas de internautas. En otros sistemas, partiendo de una etiqueta se puede llegar a decenas de artículos sobre el mismo tema que han sido guardados con etiquetas coincidentes asignadas por otros. De este modo resulta una alternativa fácil y espontánea a la categorización jerárquica tradicional, pues los metadatos se asignan con rapidez y facilidad, sin necesidad de recorrer ninguna estructura jerárquica, por lo general de difícil comprensión para quienes no sean expertos en el dominio reflejado en ella.

Es evidente el valor de muchos de los términos propuestos como candidatos a integrarse en vocabularios controlados, ya que se trata de términos o conceptos comúnmente empleados por los usuarios de Internet, aunque esos usuarios no sean profesionales de la información, a diferencia de lo que sucede con los tesauros y sus creadores.

La aplicación de las *folksonomías* presenta aún muchas debilidades. La *folksonomía* no ofrece un lenguaje controlado, pues no tiene jerarquías ni otro tipo de asociaciones más allá de los posibles *clusters* de recursos que se puedan formar según las clasificaciones o palabras clave de los usuarios. Así pues, al carecer de un lenguaje controlado, la ambigüedad es muy alta. Para un usuario, un término tiene un significado o puede expresar una realidad completamente distinta a la de otro usuario, pues no hay guías de uso para etiquetar los contenidos.

Entre sus debilidades se encuentra también la imprecisión derivada de utilizar etiquetas que solo

permiten que se empleen unitérminos.¹⁸ Igual sucede con el empleo indistinto de singulares y plurales, así en *Flickr* se admite la búsqueda y el etiquetado según plurales o singulares, con recuperaciones dispares, pues se trata de distintas fotografías que hacen referencia a un mismo concepto, pero que se buscan por términos diferentes.

SITIOS WEB RELACIONADOS

CITEULIKE: www.citeulike.org
DEL.ICIO.US: <http://del.icio.us>
FEEDSTER: <http://feedster.com>
FLICKR: www.flickr.com
IFAVORITOS: www.ifavoritos.com
TAGCLOUD: www.tagcloud.com¹⁹
GENIELAB: <http://genielab.com>
PUBSUB: <http://pubsub.com>
TECHNORATI: www.technorati.com
UPCOMING: <http://www.upcoming.org>²⁰

ONTOLOGÍAS PARA ARCHIVOS DE PRENSA

La hemeroteca de un periódico no es sólo una herramienta de trabajo fundamental para los redactores, ahora es una línea de negocio cuando sus productos se ofrecen a través de Internet. La adecuada gestión de los complejos contenidos que almacena, lleva a buscar soluciones provenientes de las tecnologías de la información de aspiración semántica (Castells, 2004).

Las ontologías prestan aquí adecuados servicios en especial a la hora de describir la información almacenada y de explorar y visualizar los contenidos (Gil, 1999). Las recuperaciones de carácter semántico son uno de los motivos por los que se aspira a la llamada web semántica. Precisamente las ontologías

se proponen como herramienta de búsqueda conceptual en repertorios enormes de información. Frente a la limitación de las recuperaciones por palabras clave, las ontologías estudian las categorías que existen entre los objetos de un dominio.

Las ontologías sirven para etiquetar con precisión los contenidos de la web, establecer sus propiedades y sus relaciones con otros conceptos, y transformarlos así en términos, de forma que, cuando un agente de *software* encuentre una palabra, estará en condiciones de interpretar su significado. Esta parte de la conceptualización de las «ontologías» según el WC3 es equivalente a lo que en documentación se ha venido conociendo desde hace décadas como «tesauro» (Alonso, 2004).

Una ontología representa la información dentro de un espacio de referencia, lo que no es nuevo si pensamos quizá en las taxonomías o los tesauros antes de contar con Internet. Las ontologías aportan vocabulario controlado para indicar el contenido y lo hacen manteniendo la jerarquía de clases y las reglas de relación. Buscando la precisión de las búsquedas, la información que contiene un *corpus* documental puede representarse como una ontología de forma que se obtenga respuesta como entidades de la ontología.

Así pues, una ontología es una jerarquía de conceptos con atributos y relaciones, que proporciona un vocabulario consensuado para definir redes semánticas de unidades de información interrelacionadas.

La ontología nace de la representación lógica de un lenguaje que es capaz de expresar las relaciones existentes entre las entidades de un dominio dado (Sowa, 1999: 492-493), presentando estas posibilidades de conformación:

18 Gene Smith enumera algunos inconvenientes de las *folksonomías* en su *Folksonomy: social classification*, http://atomiq.org/archives/2004/08/folksonomy_social_classification.html [consultado el 30.12.05].

19 TagCloud es un clasificador automatizado de fuentes RSS de noticias (RSS es un formato de archivo derivado de XML y usado para la distribución de contenidos). RSS es una tecnología que permite suscribirse a los sitios que se le indiquen y ver las noticias o temas que se van publicando en ellos, allí extrae las palabras clave del contenido y las lista de acuerdo con su predominio dentro de las alimentaciones de RSS. Haciendo clic sobre el enlace de la etiqueta, aparecerá una lista de todos los resúmenes de artículos asociados con esa palabra clave.

20 Para compartir fotografías y llevar la agenda de acontecimientos.

- ◆ Ontología informal: catálogo de tipos que pueden estar definidos o indefinidos por declaraciones en una lengua natural.
- ◆ Ontología formal: colección de nombres para conceptos y tipos de relación organizada en orden parcial mediante la relación tipo - subtipo. A su vez se subdivide en:
 - ◆ Ontología axiomática: distingue los subtipos por axiomas y definiciones sugeridos (indicados) por un lenguaje formal.
 - ◆ Ontología basada en prototipos: distingue los subtipos por comparación con un elemento típico o prototipo de cada subtipo.

Se recomienda seguir estos pasos para construir una ontología:²¹

1. Determinar el dominio y ámbito de la ontología.
2. Determinar la intención de uso de la ontología.
3. Reutilizar ontologías o vocabularios controlados existentes.
4. Enumerar los términos cabecera del dominio.
5. Definir la jerarquía de clases.
6. Crear las instancias.

Igual que sucede en la construcción de tesauros, para desarrollar las ontologías, los diversos ámbitos de especialización tienen que definir los conceptos de su área. Los tesauros reflejan los conceptos de un campo de interés o dominio, que explicitan en términos y establecen entre esos términos relaciones de equivalencia, jerarquía y asociación. Para cada concepto fijan un término unívoco, concediéndole así un significado explícito, que evita las di-

ferencias de interpretación propias de la subjetividad o la posibilidad de establecer equivalencias entre términos que, en esencia, signifiquen lo mismo. Los propios tesauros, cuando existen en un dominio, sirven como cantera terminológica sobre la cual se puede desarrollar una ontología, por lo que pueden considerarse base o parte de la ontología, junto a diccionarios de términos, vedemécums, glosarios, etc. Para construir una ontología hay que hacer un gran esfuerzo de normalización conceptual y terminológica.

El periodismo atiende, en su actividad diaria, a todos los temas posibles en cualquier campo del conocimiento, lo que obliga a parcializar obligatoriamente cada uno de los dominios a representar, pues en caso contrario se corre el peligro de pretender representar enciclopédicamente el conocimiento. La creación de una base de conocimiento para estas entidades, que cubra completamente el ámbito informativo potencial de un medio de comunicación, excede la capacidad de cualquier organización, empresa o agencia que se propusiera semejante tarea.

En el sector periodístico se han desarrollado diferentes estándares para los vocabularios controlados, entre los que cabe mencionar:²²

- ◆ NewsML: <http://www.newsml.org>
- ◆ NITF (News Industry Text Format): <http://www.nitf.org>.
- ◆ XMLNews: <http://www.xmlnews.org>
- ◆ IPTC, jerarquía de clasificación por temas [IPTC Subject Reference System and NewsML Topicsets: <http://www.iptc.org/metadata>
- ◆ PRISM (Publishing Requirements for Industry Standard Metadata): <http://www.prismstandard.org/>²³

21 En *Publishing Requirements for Industry Standard Metadata (PRISM)*. Updated PRISM Specification Keeps Pace with New Information Technologies. <http://www.prismstandard.org/>. Con el apoyo de W3C se han desarrollado diversos lenguajes y estándares para la definición de ontologías, como XML; RDF; DAML+OIL; y OWL.

22 Extraído de Castells, P.; Pulido, E. *et al. Neptuno: tecnologías de la web semántica para una hemeroteca digital*, en <http://griho.udl.es/publicacions/2004/neptuno-interaccion04.pdf>

23 En 1999 IDEAlliance (International Digital Enterprise Alliance: consorcio de empresas editoriales y de fabricantes de *software* de publicación formado, entre otros, por Adobe, Quark, Condé Nast y Time Inc) fundó PRISM, un estándar XML que facilita manejar el contenido digital. Desde entonces, el marco de metadatos de PRISM se ha extendido ampliamente. En la actualidad incluye:

- Aplicaciones de RDF y de XML/RDF para PRISM.
- Documentación para los metadatos de PRISM basada en modelos de XML/RDF.
- Prácticas actualizadas para trabajar en un ambiente de PRISM/XMP.
- Nuevos metadatos incluidos en el espacio PRISM Dublin Core.
- Referencias y bibliografía sobre los estándares y especificaciones actuales.

Tanto NewsML como NITF (News Industry Text Format) son estándares basados en XML para representar y gestionar noticias durante todo su ciclo de vida, incluyendo la creación, intercambio y consumo de las mismas; pero mientras que NewsML se utiliza para representar noticias como paquetes multimedia, NITF se ocupa de la estructura de los documentos. XMLNews es un subconjunto de NITF y está basado en RDF. Incluye una serie de marcadores (como localización, personaje o fecha) que permiten anotar noticias para facilitar la búsqueda de información.

CONCLUSIONES

Como de su nombre puede deducirse, el ámbito del *software* social es cada día de mayor interés. El etiquetado compartido, la programación colectiva, la investigación sobre mapas *folksonómicos*, o el *Social Bookmarking*, están generando continuas aplicaciones y estudios que buscan soluciones alternativas a las propuestas por las corrientes más generales planteadas hacia el logro de la web semántica que desde esta perspectiva, podría también denominarse web social.

Existen plataformas especialmente diseñadas para crear y gestionar los contenidos de periódicos y revistas digitales a partir de licencias de *software*, con excelentes prestaciones en el control de los procesos de edición, creación, publicación y archivo. Por otra parte, la metodología cooperativa y social empleada para desarrollar el *software* abierto puede usarse también para desenvolver otras tareas técnicas biblioteconómicas como normas abiertas de catalogación, sistemas de clasificación o tesauros. Se ha abierto un espacio de grandes posibilidades para los desarrollos cooperativos descentralizados, extendiendo y ampliando las capacidades de las redes.

La gestión de contenidos, y las herramientas de *software* libre que facilitan su puesta en marcha, ofre-

cen soluciones para los servicios que deben ofrecer, en el entorno digital, las unidades de información. Algunas propuestas son de carácter general, otras son específicas. Además, es innegable la existencia de un creciente mercado de productos y soporte para estas herramientas. Sin embargo, y a modo de conclusión, es necesario delinear varios factores críticos que deben ser tenidos en consideración.

Frente a la rigidez analítica de los grandes motores de búsqueda, el uso de etiquetas asignadas socialmente en las *folksonomías*, permite acercar los metadatos existentes en páginas ya etiquetadas a los conceptos que son de interés para los usuarios. Estableciéndose de esta forma una búsqueda personalizada, pues está capacitada para conocer la proximidad existente entre un concepto indizado y las necesidades de un usuario.

La indización *folksonómica* tiene la ventaja de no ser jerárquica. Además, al usar diversas etiquetas por cada concepto y describir una misma etiqueta varios conceptos, permite establecer con facilidad relaciones de semejanza para navegar por etiquetas relacionadas que sagazmente relacionan páginas de bitácoras con elementos de música en formato mp3, entre otras posibles, y encontrar en ellas la información deseada sin recurrir a una estrategia de búsqueda complicada

Una *folksonomía* es muy útil para los usuarios por permitir relacionar palabras de búsqueda, y por sugerir búsquedas asociadas a partir de la proximidad de las palabras o a través de intersecciones de varias búsquedas cruzadas. Las *folksonomías* son también interesantes desde el punto de vista de la organización personal de datos, fotos y todo tipo de documentos.

BIBLIOGRAFÍA

- Agirreazaldegui, T. *El uso de la documentación audiovisual en los programas informativos diarios de televisión*. Bilbao: Universidad del País Vasco, 1996.
- Alonso, J. *Estándares Web del W3C*. Noviembre 2004 <http://www.w3c.es/Presentaciones/2004/11-GiraEstandares-JA/> [consultado el 30.12.05].
- Amat, C. «Factores condicionantes de las operaciones documentales en medios de comunicación» *Revista Española de Documentación Científica* 19 2 (1996): 188-201.
- Arquero, R. y García-Ocho, L. *La hemeroteca de prensa*. Gijón: TREA, 2005.
- Arunachalam, S. «Science on the periphery: bridging the information divide». *Handbook of quantitative Science and Technology Research. The use of publication an Patent Statistics in studies of S&T Systems*. Dordrecht, Kluwer, 2004. Trata los temas de *Open Acces* entre las páginas 175 y 177 y de *Open Archives* en las páginas 177 y la 178.
- Bustamante, E. «Televisión digital: el reto de los contenidos y servicios». Delgado, C. y Rodríguez, A. (eds.). *Jornadas sobre publicación electrónica. Un nuevo espacio de comunicación*. Madrid, Universidad Carlos III de Madrid (1999): 15-17.
- Castellanos, N. y Sánchez, J. «Mobile access to digital libraries via generic interfaces». *Avances en Ciencias de la Computación*. Colima: 2004: 571-576.
- Castells, P. et al. «Neptuno: Tecnologías de la Web Semántica para una hemeroteca digital» IX Encuentros Internacionales sobre Sistemas de Información y Documentación. IBERSID 2004. Zaragoza. <http://griho.udl.es/publicacions/2004/neptuno-interaccion04.pdf>.
- Castells, P. y Pulido, E. et al. *Neptuno: tecnologías de la web semántica para una hemeroteca digital*, en <http://griho.udl.es/publicacions/2004/neptuno-interaccion04.pdf>
- Embley, D. et al. «Ontology-Based Extraction and Structuring of Information from Data-Rich Unstructured Documents». <http://www.cs.wisc.edu/~smithr/pubs/cikm98.pdf> (consultado el 28.10.2005)
- Fernández, I. «Herramientas del periodismo digital, en Pareja». *Guía de Internet para periodistas*. Madrid: CINDOC, 2002.
- Fernández-García, N.; Sánchez Fernández, L. «Building an Ontology for NEWS Applications» *Mcilraith, Sheila A., Plexousakis, Dimitris, Van Harmelen, Frank. The Semantic Web – ISWC 2004: Third International Semantic Web Conference, Hiroshima, Japan, November 7-11, 2004*. Proceedings.
- Fuentes, M. *La información en Internet*. Barcelona: CIMS, 1997.
- García de Torres, E. y Pou Américo, M. «Características de la comunicación digital». *Manual de redacción ciberperiodística*. Barcelona: Ariel, 2003.
- García J. «Televisión digital interactiva. La importancia de los contenidos». *Actas de ponencias y comunicaciones. III Congreso Internacional de Ciberperiodismo*. Madrid: Fundación Antonio de Lebríja, 2005.
- Gil, Q. *Autocrítica del periodismo digital*. Revista electrónica Sala de Prensa 13, noviembre 1999. Disponible en Sala de prensa. web para profesionales de la comunicación iberoamericanos, noviembre de 2000. <http://www.saladeprensa.org/Índiceporautor> [consultado el 25.11.05].
- Gómez-Pérez, A. *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web*. London: Springer, 2004.
- http://www.olivesoftware.com/ap_aparchive.html [Consultado el 05.11.05].

- Iksal, S. y Garlatti, S. *Adaptive Special Reports for On-line Newspapers*. Proceedings of the Workshop on Personalization Techniques in Electronic Publishing on the Web: Trends and Perspectives, 2002. <http://www.dimi.uniud.it/~mizzaro/AH2002/proceedings/pdfs/3iksal.pdf>. (Consultado el 31.10.2005).
- Khan, L., Mcleod, D. y Hovy, E. *Retrieval Effectiveness of an Ontology-Based Model for Information Selection*. <http://www-scf.usc.edu/~csci586/paper/latifur-vldb.pdf>. (Consultado el 28/10/2005)
- McMullin, J. *The cognitive cost of classification*, 2005. <http://www.interactionary.com/index.php?cat=7>. [consultado el 18.08.05].
- Melero, R. y Pérez-Agüera, J. «Plataforma digital de revistas científicas electrónicas españolas. Relación con el movimiento *Open Access*». *Scripta Nova. Revista electrónica de Geografía y Ciencias Sociales*, 2004, 8 nº 170. <http://www.ub.es/geocrit/sn/sn-170-74.htm> [Consultado el 09.12.05].
- Morán, P. *Nuevos discursos informáticos en la era de Internet, en Actas de ponencias. II encuentro Global de Ciberperiodismo*. Madrid: Fundación Antonio de Lebrija, 2005: 47-51.
- Moreira, J. *Manual de Documentación informativa*. Madrid: Cátedra, 2000.
- Moreiro, J. (coord.). *Manual de Documentación informativa*. Madrid: Cátedra, 2000.
- Pareja, M. «Introducción al periodismo digital». *Guía de Internet para periodistas*. Madrid: CINDOC, 2002.
- Pareja, V. M. (coord.). *Guía de Internet para periodistas*. Madrid: CINDOC, 2002.
- Porter, J. *Controlled vocabularies cut off the long tail*, 2004. http://bokardo.com/archives/controlled_vocabularies_long_tail/. [consultado el 08.08.05].
- Quintarelli, E. *Folksonomies: power to the people*, 2005. <http://www.iskoi.org/doc/folksonomies.htm>. [consultado el 08.08.05].
- Shirky, C. *Folksonomies + controlled vocabularies*. http://www.corante.com/many/archives/2005/01/07/folksonomies_controlled_vocabularies.php [consultado el 11-01-06].
- Sowa, F. *Knowledge Representation: logical, philosophical and computacional foundations*. Brooks Cole Publishing, Pacific Grove, 1999: 492-493.
- Sowa, J. F. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. London: Brooks/Cole, 2000.
- Sturman, R. «Il software *open source* per la gestione integrata delle biblioteche: una nuova risorsa» *Bollettino AIB 3 (Associazione italiana biblioteche)* 2004: 257-270.
- Thomas Vander Wal. 2004. *Explaining and Showing Broad and Narrow Folksonomies*: <http://www.vanderwal.net/random/entrysel.php?blog=1635> [consultado el 23.11.05].
- Tramullas, Jesús.- Herramientas de software libre para la gestión de contenidos [on line], en *Hipertext.net 3*, 2005. <http://www.hipertext.net/web/pag258.htm> [Consultado el 24.12.05].
- Unsworth, J. «The next wave: liberation technology». *The Chronicle of Higher Education: The Chronicle Review*, 2004. <http://www.iath.virginia.edu/~jmu2m/liberation.html> [Consultado el 04.10.05].
- VanderWal, T. *Off the Top: Folksonomy Entries*, en <http://www.vanderwal.net/random/category.php?cat=153> [consultado el 11.11.05].
- Weiss, A. «The power of collective intelligence» *netWorker 3*, 2005.
- Zafra, J. y Cerezo, J. *El impacto de Internet en la prensa*. Madrid: Fundación Auna, 2003 (Cuadernos de la Sociedad de la Información Nº 3).