

Indización social y control de vocabulario

Noelia Navoni y Pablo González

II Encuentro Nacional de Catalogadores

“La Cooperación y las Normas para la Organización y Tratamiento de la Información en las Bibliotecas Argentinas”

25, 26 y 27 de Noviembre de 2009

Biblioteca Nacional, Buenos Aires, Argentina

Resumen

Las folksonomías, surgidas dentro del marco de reflexiones sobre la web 2.0, son herramientas de representación y recuperación de información digital. Tanto en blogs como en wikis, como en colecciones de objetos digitales o de marcadores, los usuarios pueden asignar palabras (etiquetas o tags) para describir un recurso del sitio. Este proceso de tagging o etiquetado da como resultado un índice de etiquetas – nube de etiquetas – que sirve como herramienta de búsqueda y acceso a recursos.

En una folksonomía se parte de la premisa de distribución y desnormalización del trabajo de indización: todos los usuarios pueden “indizar” todos los recursos utilizando todas las palabras que lleguen a su mente.

Esta dimensión colectiva o colaborativa es la que permite nombrar este proceso como indización social. Tiene a su favor simplicidad en el uso y posibilidades constantes de expansión y actualización. También presenta problemas de representación de información en lenguaje natural o sin herramientas de control de vocabulario

Si queremos hacer uso de esta opción a la hora de brindar algún tipo de acceso a nuestras colecciones, tendremos que diseñar sistemas y procurar aplicaciones que integren las folksonomías y las combinen con los vocabularios controlados ya existentes

1. Folksonomías

1.1 Definición

Derivado de *taxonomía*, el término *folksonomy* es un neologismo que se le atribuye a Thomas Vander Wall. Taxonomía procede del griego "taxis" y "nomos": *Taxis* significa clasificación y *nomos* (o *nomía*), ordenar, gestionar; por su parte, "folc"

proviene del alemán "pueblo" (Volk). En consecuencia, de acuerdo con su formación etimológica, folksonomía (folc+taxo+nomía) significa literalmente "clasificación gestionada por el pueblo (o democrática)".

Según el mismo Vander Wall (2007), “ una folksonomía es el resultado del etiquetado de información y objetos (cualquier cosa con una URL), personal y libre, para nuestra propia recuperación. El etiquetado se realiza en un contexto social (compartido y abierto a los demás). La folksonomía es creada desde el acto de etiquetar por la persona que consume la información”.

Por ello, aclarando el concepto de "colaboración" relacionado a las folksonomías, Spiteri (2007) precisa que “las folksonomías son creadas en un entorno donde las personas no puedan colaborar activamente en la creación y asignación de etiquetas, aunque sin duda pueden acceder y utilizar etiquetas asignadas por los demás, permitiendo así el uso de etiquetas comunes”.

“La indización social representaría un nuevo modelo de indización, en el que son los propios usuarios o consumidores de los recursos los que llevan a cabo su descripción...La descripción de cada recurso se obtendría por agregación, es decir, un mismo recurso sería indizado por numerosos usuarios, dando como resultado una descripción intersubjetiva y por tanto más fiable que la realizada por el autor del recurso, e incluso por un profesional” (Hassan Montero, 2006)

Por lo tanto el *tagging* (etiquetado) tiene dos dimensiones interrelacionadas: la personal (cada usuario confecciona su propio índice personal de tags o *personomía*) y la colectiva (los usuarios comparten sus tags, generando mediante una colaboración implícita un índice global de tags o *folksonomía*) (Hassan Montero 2006)

De esta forma (Hernández Quintana 2008), los usuarios se convierten en indizadores de sus propias necesidades y colocan sus propias palabras clave, los *tags* (etiquetas), categorías o metadatos libremente seleccionados o escogidos, sin mayores indicaciones formales ni atados a convenciones internacionales.

1.2 Aplicaciones

Las folksonomías son utilizadas principalmente en sitios de marcadores sociales y también en otras herramientas sociales dentro del contexto de aplicaciones de la web 2.0 (Noruzi 2006).

“Las folksonomías comenzaron como una simple herramienta para los usuarios de Internet utilizada para organizar archivos en la web. Desde entonces, se han utilizado en intranets u otros sistemas corporativos de información, así como en bases de datos profesionales y en bibliotecas” (Weller 2007)

Los catálogos de las bibliotecas públicas son un campo de aplicación de folksonomías. Los usuarios de un opac podrán describir recursos de la colección a través de etiquetas que se integrarán en un índice o nube de etiquetas: estas muestran la popularidad de las

etiquetas para un documento o dentro de toda la folksonomía (las etiquetas que se aplican con mayor frecuencia se presentan en tamaños de letra más grande).

“De esta manera se crean términos de índice que representan distintos puntos de vista, proponiendo puntos de partida alternativos para investigaciones” (Weller 2007)

Steve Museum es un proyecto que está investigando la incorporación de folksonomías en los catálogos de museos (Wyman 2006). Participan Susan Chun, The Metropolitan Museum of Art, Rich Cherry, Guggenheim Museum, Doug Hiwiler, The Cleveland Museum of Art, Bruce Wyman, Denver Art Museum, E.E.U.U y Jennifer Trant, Archives & Museum Informatics, Canadá. Con la doble finalidad de mejorar el acceso y crear una nueva experiencia con las comunidades de interés se apuesta a la folksonomía como una herramienta facilitadora. En este momento se puede encontrar en la red un catálogo de prueba. Desde este sitio se invita a los visitantes a describir los objetos de interés con etiquetas. Los objetos son etiquetados con tantas etiquetas como detalles, pensamientos y sentimientos encuentren los usuarios. En el proyecto de investigación del grupo “Text, Tagging and Trust to Improve Image Access for Museums and Libraries” se estudia la integración de la indización automática, la indización social y un sistema de red de confianza para la valoración de los términos por parte de gente relacionada con la materia.

1.3 Características

Las características de las folksonomías, de forma general, son las mismas que las que tienen todas las herramientas de la Web 2.0: participación e inteligencia colectiva. Los usuarios publican, almacenan y comparten contenidos utilizando software social.

Algunas características específicas de las folksonomías son (Hernández Quintana 2008):

- Contextualidad: varían dependiendo de la conducta de la búsqueda de información (individual o colectiva) y de los contextos de comprensión, asimilación y tratamiento de los recursos.
- Orientación para y desde el usuario: reflejan directamente el lenguaje de los usuarios (en un sentido amplio).
- Adaptabilidad: el vocabulario cambia rápidamente hacia las necesidades léxicas de los usuarios.
- Economía: no hay costos significativos para la actualización y el mantenimiento.
- Comunicabilidad: preserva la naturaleza comunicativa de las estrategias discursivas de los recursos, objetos digitales y demandas informativas.
- Desprofesionalización: populares, no hay entrenamiento y cualquier usuario puede como autodidacta elaborar sus etiquetas son mayor esfuerzo ni costo cognoscitivo.

- Regeneración: capacidad de cambiar los términos de forma inmediata, sin necesidad de normas ni organismos que lo aprueben.
- Personalización: cada usuario puede generar etiquetas desde el estricto interés personal.
- Negociación: la conducta de un usuario puede relacionar e influenciar en un dominio a otros usuarios que utilizan el servicio.
- Cooperación: disminuye las barreras de cooperación bibliotecaria.
- Uso no anticipado: algunas etiquetas pueden tener un uso inesperado, que modifique la navegación en un campo del conocimiento.
- No discriminativa: no se ajustan a una determinada tipología documental. Tampoco prioriza una categoría, una clase o un término.
- No lucrativas: no ambiciona convertirse en sistemas hegemónicos y mucho menos especulativos.

1.4 Clasificación

Fundamentalmente (Hernández Quintana 2008 y Weller 2007) existen dos tipos de folksonomías: las estrechas (narrow folksonomies) - las etiquetas pueden ser asignadas a un documento una sola vez por el autor- como Flickr; y las generales (broad folksonomies) - el documento es etiquetado por varias personas- como es el caso de Delicious. También suelen nombrarse las personomías, más cercanas a las estrechas.

Los usuarios describen no sólo lo que el recurso es o de lo que el recurso trata, sino también aquello que piensan hacer con el recurso o algo que el recurso provocó en ellos. Aportan además elementos referenciales como nombres propios, lugares, datos cronológicos, etc.

Weller (2007) identifica funciones posibles de las etiquetas:

- Identificar sobre qué o quién trata un documento.
- Determinar qué es el documento.
- Identificar quien posee el documento.
- Identificar cualidades o características.
- Auto referencia.
- Organización de tareas.

Puede establecerse una tipología de tags de acuerdo al criterio de “clasificación” del usuario (Ros-Martín, 2008):

- tags basadas en el contenido temático,
- tags basadas en el contexto de producción o almacenamiento del recurso,
- subjetivas (opinión y emociones del usuario),
- tags de atributos del recurso no derivados del contenido,

- tags organizacionales (etiquetas que describen materiales personales de los usuarios o etiquetas recordatorias de tareas)

1.5 Ventajas

Moreiro González (2007) nos señala las siguientes ventajas de las folksonomías:

- simplicidad de utilización: no requieren de aprendizaje ni capacitación de una herramienta o reglas. Sin costes cognitivos.
- economía: gran rentabilidad debido a su bajo costo. Cuantos más usuarios cooperen más ventajas se obtienen.
- adecuación al entorno web: es la única solución para indizar los enormes volúmenes de información en la red, sobre todo en información no textual (videos, fotos, audio, etc.).
- ejecución de consultas: las búsquedas suelen ser más específicas ya que los usuarios asignan etiquetas concretas. Permite recuperar material multimedia.
- simplicidad en la gestión: escaso mantenimiento. Los términos se incorporan instantáneamente. Actualización permanente.
- flexibilidad: no hay precoordinación o un lenguaje definido a priori.

También es importante destacar que las folksonomías son más realistas en cuanto a los intereses de los usuarios on line y no tienen límite para la asignación de etiquetas (Hernández Quintana 2008).

2. Folksonomías y sistemas de representación del conocimiento

Las folksonomías son un tipo de índice de palabras clave que cuenta con los problemas ya conocidos de la representación de información en lenguaje natural o sin herramientas de control de vocabulario: sinonimia, polisemia, falta de conexión entre términos relacionados, términos vacíos de significado, unitérminos, etc.

Como herramienta de búsqueda y recuperación de información, tiene a favor que sus términos cuentan con autoridad del usuario,

“Esta justificada la inclusión de un término en un índice sólo si es de interés para el usuario del servicio de información... Satisfacer el requisito de autoridad de los usuarios implica que los términos se han de obtener de los usuarios potenciales de un servicio de información y representar sus intereses concretos” (Lancaster, 1995)

Además de contemplar dos de los principios de elección de términos de índice enumerados por Cutter:

- uso - “el encabezamiento elegido debe representar un uso común, o el uso que una clase de lector le da a la materia que representa el encabezamiento” (Taylor, 2000) -
- foco en el lector – “el encabezamiento, en su formulación y estructura, debería ser el que el lector buscaría en el catálogo, si conocemos o podemos presumir lo que el lector está buscando” (Taylor, 2000) -.

“Las etiquetas se ajustan estrechamente con las directrices NISO para la elección y forma de vocabularios controlados. Las etiquetas representan mayormente sustantivos, con muy pocos adjetivos no calificativos o adverbios. Las etiquetas representan los tipos de conceptos recomendados por NISO y se ajustan al estándar utilizado; y conforme a las normas establecidas por la ortografía. La mayoría de las etiquetas respetan los términos estándar; en pocas instancias se utilizan términos no estandarizados, como son el argot o la jerga. Brevemente, la estructura de las etiquetas está dentro de las normas establecidas y reconocidas para construcción de vocabularios controlados.” (Spiteri, 2007)

La naturaleza múltiple de las etiquetas y la falta de control de su terminología las vuelve incapaces de representar consistentemente un dominio o contexto específico y no brindan además estructuras sólidas de navegación entre conceptos.

En las antípodas de las folksonomías se encuentra otro sistema de representación del conocimiento, de gran aplicación también dentro del marco de reflexiones sobre la web –en este caso, web semántica-; se trata de las ontologías:

“Una representación explícita y formal de una conceptualización compartida, que implica una perspectiva sobre cierta realidad y que se constituye en la estructura conceptual de una base de conocimiento. Su objetivo final es el compartir el conocimiento que representa” (Arano, 2005).

“Las ontologías son diseñadas por expertos y deben ser utilizadas para hacer explícito y unívoco el significado de los documentos, no sólo para comunicación interpersonal, sino también para las interacciones humano – computadora y entre computadoras. Son conceptualizaciones formales de un área del conocimiento, expresadas en sistemas de conceptos (clases), casos (individuos) y relaciones entre ellos” (Weller 2007).

Si bien las ontologías no son vocabularios controlados propiamente dichos, en el sentido de que no son herramientas que estén diseñadas para la representación y el acceso temático al contenido de recursos de información, comparten con estos más de una característica. Al igual que un vocabulario controlado, una ontología es

- un conjunto limitado de términos puestos en relación
- pertenecientes a un campo disciplinar o contexto específico

Su objetivo es hacer unívoca y explícita la terminología que recubre un dominio, y en pos de lograr este objetivo amplía las posibilidades de representación de los sistemas ya existentes. Los tesauros por ejemplo, plantean las relaciones clásicas entre términos:

- relaciones de equivalencia, para términos intercambiables en un contexto determinado
- relaciones jerárquicas (genero-especie; todo-parte; categoría-caso)
- relaciones asociativas entre descriptores ni equivalentes ni jerárquicos

La inclusión de etiquetas nudos en los tesauros permite aclarar cual es el sentido y la naturaleza de la relación entre términos, en los casos de jerarquía y asociación al haber más de una lógica de relación involucrada (ANSI/NISO Z39.19 2005).

De acuerdo a Weller, las ontologías parten de relaciones jerárquicas como base estructural para diseñar un dominio pero van más allá al

- permitir definir y modelar libremente relaciones entre conceptos de un mismo ámbito
- y hacer explícitas todas las interrelaciones posibles, volviendo transparente su estructura conceptual.

“La ontología es una representación formal y explícita de la estructura conceptual del campo sobre el que se trabaja. Este recurso lingüístico incluye como mecanismo de inferencia a la herencia, que implica una economía en la codificación de la información: los conceptos superiores transmiten sus características a los conceptos inferiores” (Arano 2004)

Son recursos lingüísticos que funcionan como bases de conocimiento para una comunidad al contener información sobre áreas, objetos y contextos, y pueden ser utilizadas en aplicaciones finales de representación y recuperación de información como estructura de navegación ínter conceptual a la hora de indizar o buscar información.

La tendencia observada en el desarrollo de sistemas de recuperación de información, basados o no en el control terminológico, es la retroalimentación entre las herramientas utilizadas y el aprovechamiento de la información almacenada en un recurso para la creación o el desarrollo de un recurso superador:

“El uso de información conceptual y lingüística almacenada en otros tipos de recursos (por ejemplo, en una ontología) permite el enriquecimiento a nivel estructural de los tesauros, y posibilita el incremento de la amigabilidad de estas herramientas de cara a su utilización por parte de un usuario final” (Arano, 2005)

3. Aplicaciones de folksonomías en el acceso a colecciones de información digital

Los autores consultados presentan enfoques respecto de

- utilizar folksonomías como adiciones a los sistemas existentes, ejerciendo sobre ellas algún tipo de control (3.1)
- combinar las folksonomías con sistemas controlados ya existentes (3.2)

3.1 Aplicar técnicas de indización automática a folksonomías (Hassan Montero 2006)

- ✓ ponderación de etiquetas mediante frecuencia de uso
- ✓ relación entre etiquetas mediante coocurrencias de palabras
- ✓ eliminación de etiquetas vacías
- ✓ desambiguación del significado en función del contexto

3.2. Integración entre folksonomías y ontologías

“Ontologías y folksonomías no deberían excluirse mutuamente, pues presentan características complementarias que, explotadas de forma conveniente, podrían generar sinergias productoras de más valor: apoyo mutuo y suma de ventajas:

Las aplicaciones basadas en folksonomías se benefician de la naturaleza dinámica y extensible de las mismas y de su potencial para canalizar la colaboración con los usuarios: mecanismo sencillo de indización.

Las aplicaciones basadas en ontologías explotan su rigor, siendo capaces de ofrecer respuestas basadas en el razonamiento lógico y una estructuración bien definida del conocimiento: aportan formalismo al back-end de la aplicación” (Morfeo Project 2007).

Los distintos sistemas se integran para intercambiar ventajas entre sí

3.2.1 Enriquecimiento de folksonomías mediante ontologías

“Las ontologías pueden aplicarse tras la escena del etiquetado social proporcionando recomendaciones para etiquetas relacionadas explícitamente que puedan añadirse a las palabras que ya ha introducido un usuario...Otra opción es usar ontologías, u otras representaciones estructuradas de conocimiento, para los mecanismos de ampliación de consultas, a través de etiquetas relacionadas, dentro de las plataformas de etiquetado social” (Weller 2007).

3.2.2 Enriquecimiento de ontologías mediante folksonomías

“Las folksonomías proporcionan información sobre el vocabulario empleado para etiquetar los documentos y puede capturar el lenguaje en uso actual. Una comparación entre etiquetas sociales y términos de un vocabulario controlado para un dominio dado puede ser realizada. Esto ayuda a actualizar los sistemas existentes y a evaluar la oportunidad, perceptibilidad e idoneidad de un sistema de representación del conocimiento diseñado por profesionales. Los términos de la folksonomía pueden utilizarse como sugerencias para nuevos términos (conceptos o casos) controlados” (Weller 2007).

Esta autora también propone la creación de nuevas ontologías sobre la base de evaluar las etiquetas, sin hacer conscientes a los usuarios de dicho proceso.

“La ingeniería de ontologías en colaboración es un tema de creciente interés. Diferentes editores de ontologías están en desarrollo para aprovechar la colaboración del usuario de diferentes maneras” (Weller 2007).

4. Conclusiones

En cuanto a las folksonomías, según Hernández Quintana (2008), los profesionales de la información tendrían tres tareas básicas a desarrollar: una educación de los usuarios para el mejoramiento de las etiquetas, una participación inteligente de diseñadores y administradores de los sitios, y un mejoramiento de los sistemas de información en general para aceptar este tipo de etiquetado. Aunque el desafío mayor es el cambio hacia una cultura profesional capaz de personalizar y generalizar; que atienda a particularidades y a colectividades; que no se proponga exclusiones como norma; que trabaje con el entretejido significativo de las redes a través de términos en lenguaje natural.

Para Noruzi (2006) las folksonomías no sólo son un cambio en la metodología (la distribución y descentralización del trabajo) sino también necesitan un cambio rápido en la forma en que se clasifica y se organiza la información. Se aparta de los conceptos de "jerarquía" de los esquemas de organización del conocimiento, facilitando el descubrimiento fortuito del conocimiento y la indización en la web.

“En el caso en que los catálogos de bibliotecas decidan incorporar folksonomías, deberán considerar la creación clara de un escrito con recomendaciones para la elección y forma de las etiquetas que pueden ser incluidas en las siguientes áreas:

- La diferencia entre sustantivo cuantitativo y no cuantitativo, así como la explicación de como se usa el singular y el plural para la correcta recuperación.
- Una manera estándar para la construcción de etiquetas compuestas, por ejemplo, la inserción de un espacio entre los componentes del término, o el uso de un subrayado entre los términos.
- Un link hacia un diccionario on line reconocido y a Wikipedia para permitir a los usuarios que determinen los significados del término, para eliminar la ambigüedad entre homógrafos, y determinar si la forma completa puede ser preferible a la forma abreviada. Una explicación del impacto de las etiquetas ambiguas y homógrafos para la recuperación sería de utilidad.
- Unas políticas de uso aceptable” (Spiteri, 2007) .

En cuanto a la aplicación de folksonomías como formas de representación y acceso a una colección, Weller sostiene que “el principal desafío es encontrar los enfoques correctos, o una combinación de enfoques, para apoyar las aplicaciones concretas. Si esto se consigue, los proveedores de contenido, así como los usuarios privados y gestores corporativos de información, se beneficiarán de nuevos métodos de indización que son social y formalmente mejorados”.

Bibliografía

- Arano, Silvia (2004). La ontología: una zona de interacción entre la Lingüística y la Documentación. [Versión electrónica] *Hipertext.net*, 2.
- Arano, Silvia (2005). Los tesauros y las ontologías en la Biblioteconomía y la Documentación. [Versión electrónica] *Hipertext.net*, 3
- Entregable D.3.3: Estado del arte y propuesta de técnicas para la integración de ontologías y folksonomías [Versión electrónica] (2007 Junio 30) *Morfeo-MyMobileWeb*
- Hassan Montero, Yusef (2006). Indización Social y Resuperación de Información. [Versión electrónica]. *No solo usabilidad*, nº 5.
- Hernández Quintana, Ania R. (2008). Folksonomías: las más recientes evidencias ecológicas en la industria de la información. [Versión electrónica] *Ciencias de la Información*, 39(2), 15-22.
- Intitute of Museum and Library Service (2009). Steve: the museum social tagging proyect. Recuperado Noviembre 5, 2009 de <http://www.steve.museum/>
- Lancaster, F. (1995). El control de vocabulario en la recuperación de información. Valencia: Universitat de Valencia.
- Moreiro González, J. A. (2007). La representación de los contenidos digitales: de los tesauros automáticos a las folksonomías. [Versión electrónica] En *Actas del VI Workshop CALSI*.
- NISO (2005). Guidelines for the construction, format, and management of monolingual controlled vocabularies. [Versión electrónica] National Information Standards Organization.
- Noruzi, Allreza (2006) Folksonomies: (Un)controlled vocabulary? [Versión electrónica]. *Knoeledge Organization*, 33(4), 199-203.
- Ros-Martín, M. (2008). Folksonomías, mercado social y filtrado social de noticias. *Web social para profesionales de la información*. Recuperado Noviembre 10, 2009 de <http://eprints.rclis.org/13709/1/1.pdf>
- Spiteri, Louise F. (2007) The structure and form and folksonomy tags: the road to the public library catalogue. [Versión electrónica] *Information Technology and Library*, sept.
- Taylor, A. (2000). Wynar' s Introduction to Cataloging and Classification. Englewood, Co, Libraries Unlimited.

- Vander Wall, T. (2007). Folksonomies. Recuperado Octubre, 2009 de <http://74.125.93.132/searchq=cache:http://www.vanderwal.net/folksonomy.html>
- Weller, Katrin (2007). Folksonomies and ontologies two new players in indexing and knowledge representation. [Versión electrónica] *Online information*.
- Wyman, B. (2006) Steve.museum: An Ongoing Experiment in Social Tagging, Folksonomy, and Museums. [Versión electrónica] *Museums and the Web 2006*.