

Folcsonomías: el valor agregado de la indización social en el Web

Folksonomies: the added value of social indexing in the Web

Obdulia María Santovenia González

Licenciada en Información Científico - Técnica y Bibliotecología. Instituto Cubano de Radio y Televisión (ICRT). La Habana, Cuba.

RESUMEN

Se estudian las folcsonomías como una nueva forma para la indización de los recursos de información en el Web. Se definen brevemente los términos indización, lenguajes documentales, análisis documental y folcsonomías. Se muestra la evolución de los lenguajes documentales desde sus primeras manifestaciones hasta las folcsonomías, un valor agregado de la indización social. Se caracterizan las folcsonomías como expresión del lenguaje y se explican las ventajas y desventajas de su utilización.

Palabras clave: Indización social, lenguajes documentales, folcsonomías.

ABSTRACT

The folksonomies are studied as a new form for the indexing the information resources in the Web. The terms indexing, documentary languages, documentary analysis and folksonomies are briefly defined. The development of the documentary languages is shown from its first manifestations to the folksonomies, which are considered an added value of the social indexing. The folksonomies are described as a language expression, and the advantages and disadvantages of their use are explained.

key words: Social indexing, documentary languages, folksonomies.

La organización del conocimiento, como campo de estudio, se ocupa de la naturaleza y la calidad de los procesos que organizan el conocimiento. Comprende asimismo, los sistemas para organizar el conocimiento, los documentos, sus representaciones y los conceptos.¹

En particular, la organización del conocimiento abarca actividades como la descripción documental, la clasificación y la indización.

La indización ha sido tradicionalmente uno de los temas mas tratados en el campo de las ciencias de la información, debido a su importancia en la recuperación de la información.

El desarrollo de las tecnologías de la información y las comunicaciones y la aparición de las denominadas redes sociales en Internet han forzado la representación del conocimiento y el acceso a la información en formato electrónico. En este sentido, se han asumido un grupo de nuevas tecnologías y herramientas denominadas Web 2.0, que permiten crear, compartir contenidos e interactuar con otros usuarios, que otorgan una nueva dimensión social y participativa a la red.

La Web 2.0 posee aplicaciones para facilitar la colaboración y fomentar la sabiduría de las masas, entre ellas: Wiki, Blog, Podcast y RSS. Una de estas aplicaciones, los marcadores sociales, facilitan almacenar, clasificar, organizar y compartir en línea, enlaces a recursos valiosos. Los marcadores sociales permiten la llamada indización social en la cual son sus propios usuarios los que asignan etiquetas a los recursos de información. Esta forma de indización, que se diferencia totalmente de la tradicional, puede realizarse por cualquier individuo con acceso a los recursos de información y carece de normas; tampoco sigue un procedimiento formal y presenta un alto grado de subjetividad.

La indización social tiene como resultado de valor agregado lo que se conoce como folcsonomías, llamadas a convertirse según la opinión de algunos, en un nuevo paradigma en la evolución y desarrollo de los lenguajes de indización y por tanto de la organización y representación de la información en el ambiente digital.

INDIZACIÓN Y ANÁLISIS DOCUMENTAL

El proceso de indización se fundamenta en el análisis documental y busca identificar elementos (palabras, conceptos, frases, inferencias) en un recurso que permitan la representación y recuperación de la información que contiene. El análisis documental es el proceso mediante el cual se estudian los documentos y aprehenden sus contenidos esenciales. Ello permite posteriormente describir y representar su esencia; así como recuperar la información que portan.

Ahora bien, la aparición en el Web de los marcadores sociales, sistemas que permiten añadir enlaces a un sitio público, sobre la base de un sistema de etiquetas

en lenguaje natural, permite a cualquier usuario con acceso a la red, describir, organizar y compartir recursos en ella.²

Ello, necesariamente, ha generado un nuevo modelo de indización denominado indización social. En este, es el propio usuario quien realiza la descripción/representación durante el proceso de consulta del recurso de información.

La representación, intersubjetiva, se obtiene por agregación, en tanto el mismo recurso se indiza por varios usuarios.

El grado de exhaustividad y la consistencia de la indización, con frecuencia, son mayores que el que se obtiene con la indización tradicional. Los productos de esta colaboración se expresan en las denominadas folcsonomías.

El tratamiento conceptual y el fundamento teórico de la indización social son aún prematuros en tanto son insuficientes los acercamientos al tema entre especialistas reconocidos en el campo de las ciencias de la información, si bien existen variadas contribuciones sobre ella.

LENGUAJES DOCUMENTALES

El producto, tanto del análisis documental como de la indización, se expresan por medio de una clase de lenguajes, llamados en las Ciencias de la Información, documentales.

Los orígenes de lo que en la modernidad se conoce como lenguajes documentales, se remontan a fines del siglo XIX, con los aportes que al tema hicieron entre otros *Cutter* y *Dewey*; pero si nos atenemos al contenido y no a la denominación del concepto, en alguna forma, existieron desde las primeras bibliotecas, desde el momento en el que se hizo necesario organizar una gran cantidad de volúmenes de forma que pudieran localizarse cuando así se requiriera.³ De esta forma, pueden establecerse 3 periodos en la evolución de los lenguajes documentales:

- Embrionario: Abarca desde la Edad Antigua hasta el siglo XVI aproximadamente; toma como referente las clasificaciones que agrupaban los fondos bibliográficos por afinidades temáticas.
- Precientífico: Comprende los siglos XVII y XVIII. Los cambios en las corrientes de pensamiento europeas evidenciaron la necesidad de encontrar sistemas de clasificación para organizar las ciencias. En este contexto, el concepto de lenguaje documental se aproximó a su concepción moderna.
- Científico: Se produce a partir del siglo XIX cuando ocurre una evolución importante en el concepto de lenguaje documental, el cual pasa de ser un elemento subordinado a otros intereses (ordenar libros en los catálogos o bibliografías) a conformar un cuerpo teórico independiente que se consolida en el siglo XX con un concepto moderno. Surgen distintos sistemas de clasificación y los tesauros, vinculados fundamentalmente con el desarrollo tecnológico y la aparición de las computadoras.

Un lenguaje documental, según *Pinto Molina*, comprende el conjunto de términos _controlados o de la lengua natural-, que emplea un sistema de recuperación documental para cumplir con sus objetivos esenciales. Tal vez, la solución más sencilla en este sentido, sería el uso de la lengua natural para indizar, si bien esto, es causa de fenómenos tan conocidos como la sinonimia y la polisemia. De ahí que se crearan lenguajes artificiales, aunque basados en la lengua natural, con estos propósitos.⁴

Ahora bien, siempre que un lenguaje pueda utilizarse para representar el contenido esencial de un documento puede considerarse como un lenguaje para la indización y ello incluye a la lengua natural. En estas últimas, se ubican las folcsonomías.

El concepto de folcsonomía deriva etimológicamente de taxonomía y es producto de una simbiosis: *folk* (pueblo) deriva del alemán, *taxis* (clasificación) y *nomos* (gestión, administración), que proviene del griego. El término fue acuñado por *Thomas Van der Wal*, en el año 2004 y ello por sí solo muestra su juventud.

Existen diversas definiciones, entre ellas:

- Son los metadatos creados por los usuarios, afirma *Mathes A.*⁵
- Según *Flickr*, es una forma de representar un recurso de información, que desarrollan sus propios usuarios. Es un cuadrado de palabras (etiquetas) impresas con tamaños de letras variables, que responden al número de votos recibidos. Este tamaño representa su fuerza. Cuanto mayor fuerza presente la palabra, más citas recibió de la comunidad de usuarios del recurso. El usuario tiene la opción de navegar por las representaciones realizadas por el total de los usuarios. La folcsonomía es una representación social en la que las personas que emplean un mismo código (vocabulario) esperan reencontrarse con un objeto de interés. (*Hernández Quintana A.* Fundamentos de la organización y representación de la información y el conocimiento. 2009. Observaciones no publicadas).
- Es un modo de etiquetar objetos mediante el vocabulario propio de sus usuarios de manera que sea fácil de encontrar nuevamente un recurso de información de interés (*Hernández Quintana A.* Fundamentos de la organización y representación de la información y el conocimiento. 2009. Observaciones no publicadas).
- Es una representación gestionada por los usuarios de un recurso. Se habla entonces de una categorización cooperativa por medio de etiquetas simples en un espacio de nombres planos (y abiertos), sin jerarquías ni relaciones de parentesco determinadas. En otras palabras, se trata de un modo en que los usuarios añaden palabras clave a cualquier tipo de fichero, documento u objeto en la red (fotografía, video, audio, Web....) sobre la base de su propio vocabulario con vistas a simplificar su búsqueda.⁶

Las folcsonomías emplean el lenguaje natural; pero este evoluciona en la medida en que se desarrollan los fenómenos antropológicos y sociales inherentes al ser humano. Por ende, estas pueden ser simples o presentar toda la complejidad e imprecisión de las lenguas y las fuentes en que se originan. Pretenden, al igual que la lengua natural, la comunicación entre los seres humanos.

En las folcsonomías, las personas colaboran para organizar la información en categorías. Se basan en una arquitectura social de cooperación en la que cada usuario escoge las formas de representación que le facilitan su recuperación posterior. Se asignan libremente entonces, una serie de palabras clave de uso

personal. Estas etiquetas son capaces de categorizar cualquier tipo de contenido, información u objeto.

Al realizarse la representación en un lenguaje natural no controlado, personal, es inevitable que en esta se coloquen etiquetas que responden a motivos exclusivamente personales. Por ello, una parte importante de las etiquetas asignadas a un recurso tendrán solo sentido para un determinado usuario y carecerán de significación colectiva alguna. Cuanto mayor sea el número de usuarios que etiquetan un recurso, mayor será este problema.

En las folcsonomías, la representación, al contrario de lo que sucede con las formas tradicionales, compuestas por categorías fijas y en su mayoría jerárquicas, es libre y se aconseja asignar múltiples etiquetas a cada elemento que se pretende clasificar. Y por ello, los propios usuarios llegan a calificarla como "caótica y compulsiva".

Existen diversos tipos de etiquetado:⁷

- Etiquetado egoísta: El recurso se etiqueta con términos que sirven exclusivamente a los fines de un usuario en particular. Las etiquetas carecen de significado fuera del contexto personal. Por ejemplo, "para Juan", "pendiente de leer", etcétera.
- Etiquetado "amiguista": El recurso se etiqueta para compartirlo con otras personas del círculo social al que pertenece quien asigna la etiqueta (amigos, compañeros de trabajo u otros). Se emplean términos consensuados y sin un significado preciso para quienes se encuentran fuera del referido círculo. Por ejemplo, "práctica 6", "informe", "tesis", entre otras.
- Etiquetado altruista: El recurso se etiqueta con el propósito de compartirlo con todos los usuarios que sea posible. Se eligen etiquetas más representativas, conocidas y generalmente aceptadas. Por ejemplo, "música", "clásicos", "años 60", "*The Beattles*" y otras similares.
- Etiquetado populista: El recurso se etiqueta para hacerlo más atractivo y que obtenga un número mayor de visitas. Se emplean etiquetas que sean populares y llamativas a una mayoría. Por ejemplo, "muy bueno", "*bestseller*", "super-interesante".

Las folcsonomías poseen ciertas características generales como:

- El uso del lenguaje natural no controlado para el etiquetado.
- La realización de la indización por los creadores y usuarios de los recursos.
- La indización de un mismo recurso por varios usuarios.
- La elaboración de una descripción intersubjetiva.
- La carencia de límites con respecto a la asignación de etiquetas, algo que incrementa la exhaustividad del lenguaje y que lleva a una representación más consistente, pero que lo convierte en un instrumento de un tamaño desproporcionado.

- Que no se precisa de ningún conocimiento previo para asignar una etiqueta.
- Que se trata de una clasificación popular.
- Que permite la navegación por el etiquetado.
- Que no se explicitan las relaciones entre las etiquetas dentro de las folcsonomías.
- Que las etiquetas pueden tener referencias personales con un significado único para el usuario que las colocó, que eleva la proporción de términos carentes de una significación colectiva.
- Que no existen reglas, esquemas o normas predefinidas para el control del etiquetado.
- Que los mismos términos pueden asignarse en plural o singular, se pueden emplear sinónimos y homónimos y pueden ocurrir otros fenómenos lingüísticos similares.
- Que el etiquetado resulta una madeja compleja y muchas veces imprecisa.
- Que existe la posibilidad de extraer relaciones semánticas subyacentes.
- Que carece de estructura jerárquica, de relaciones directas o de asociación entre sus términos.
- Que los contribuyentes a una folcsonomía no están obligados a emplear una terminología predefinida para clasificar los recursos.

En una [figura](#) como la siguiente, pueden apreciarse algunos de los elementos señalados:

Tags.

Here are the 150 most popular tags. The bigger the link, the more popular the tag.



Fuente: Tomado de Flickr

- Términos sin un significado colectivo: "2001", "2002", "7610", "bw", "me", etcétera.
- Uso de singulares y plurales: "animal", "animals", "cat", "cats", etcétera.
- Empleo de sinónimos: "newyork", "nyc", etcétera.

Entre las ventajas de las folcsonomías, desde el punto de vista de la indización, se encuentran:

- El empleo de categorías relevantes para el usuario del recurso.
- La reducción de la inconsistencia entre los indizadores: diferentes indizadores utilizan distintos términos para describir un mismo documento.
- La creación de relaciones de semejanza, que permiten navegar por etiquetas relacionadas.

Como desventajas, presentan:

- Polisemia.
- Sinonimia.

- Uso indistinto de plurales y singulares.
- Falta de exhaustividad en algunos etiquetados y profusión en otros, lo cual resta valor a la representación de la información.
- Dificultad en la especificidad.
- Etiquetado egocentrista.
- Necesidad de conocimientos avanzados de programación para diseñar y poner en funcionamiento un sistema que posibilite la indización social y construya folcsonomías.

CONSIDERACIONES FINALES

Bajo el paradigma tecnológico, la indización social debe afianzarse y convertirse en un nuevo modelo de indización, con la folcsonomía como una de sus manifestaciones. Sin embargo, debe trabajarse en resolver algunas de sus desventajas como la inclusión de términos sin significado y problemas lingüísticos como la sinonimia y la polisemia, con vistas a eliminar mediante diferentes mecanismos la ambigüedad y el ruido que dificulta la recuperación de la información contenida en los recursos indizados.

Una debilidad en el contexto nacional puede ser la condición necesaria de integrarse a un entorno compartido para poder acceder a los recursos, intercambiar información, asignar etiquetas, etc., una desventaja que alcanza igualmente a los profesionales de las ciencias de la información y dificulta el estudio y empleo de estas herramientas.

No obstante, el empleo de las folcsonomías, como forma de indización social, puede instrumentarse en intranets institucionales de centros de investigación y otros similares, que pueden lograr con el empleo de estas herramientas de organización de la información y el conocimiento una potenciación de sus recursos, a partir del incremento de la colaboración, la creación de comunidades científicas en línea, entre otras acciones dirigidas a resolver problemas de la organización.

El desarrollo de herramientas para corregir las desventajas señaladas a las folcsonomías; el estudio de formas novedosas para la representación de los recursos; así como la creación de lenguajes nuevos y más precisos, como herramientas para la recuperación de la información, entre otras vertientes, son un reto al que deberán responder con más y mejores investigaciones durante los próximos años los profesionales de las infociencias.

REFERENCIAS BIBLIOGRÁFICAS

1. Hjørland B. What is knowledge organization (KO)? Knowledge Organization 2008; 35(2/3)86-101.

2. Núñez M. Aplicaciones de los marcadores sociales para la investigación. Disponible en: http://www.delfos.co.cu/boletines/bsa/PDF/3bookmarks_para_la_investigacion.pdf [Consultado: 2 de junio de 2009].
3. Historia de los lenguajes documentales. Disponible en: <http://www3.unileon.es/dp/abd/tesauro/pagina/conceptos/historia.htm> [Consultado: 23 de mayo de 2009].
4. Mániez J. La evolución de los lenguajes documentales. En: Vizcaya D (comp.). Fundamentos de la organización de la información: Selección de lecturas. La Habana: Félix Varela; 2004. p.180-195.
5. Spiteri LF. Structure and form of folksonomy tags: the road to the public library catalogue. Webology 2007; 4(2). Disponible en: <http://www.webology.ir/2007/v4n2/a41.html> [Consultado: 25 de mayo de 2009].
6. García Ferradanes F. Folksonomias: Sistemas emergentes de gestión del conocimiento. 2005. Disponible en: http://www.cibersociedad.net/recursos/art_div.php?id=94 [Consultado: 28 de mayo de 2009].
7. Tags, Folksonomies, Web 2.0., Tag Cloud. Disponible en: <http://www.ugr.es/~victorhs/gbd/docs/tagging.pdf> [Consultado: 33 de junio de 2009].

Recibido: 3 de septiembre de 2009.
Aprobado: 6 de septiembre de 2009.

Lic. *Obdulia María Santovenia González*. Instituto Cubano de Radio y Televisión (ICRT). Calle 23 No. 258 entre L y M, El Vedado. Ciudad de La Habana. Cuba.
Correo electrónico: yuya@icrt.cu

Ficha de procesamiento

Términos sugeridos para la indización

Según DeCS¹
RESUMEN E INDIZACIÓN COMO ASUNTO; ALMACENAMIENTO Y RECUPERACIÓN DE LA INFORMACIÓN; GERENCIA DE LA INFORMACIÓN; INTERNET.

ABSTRACTING AND INDEXING AS TOPIC; INFORMATION STORAGE AND RETRIEVAL; INFORMATION MANAGEMENT; INTERNET.

Según DeCI²

INDIZACIÓN; RECUPERACIÓN DE LA INFORMACIÓN; GESTIÓN DE LA INFORMACIÓN; RECURSOS DE INFORMACIÓN; INTERCAMBIO DE INFORMACIÓN; WWW.

INDEXING; INFORMATION RETRIEVAL; INFORMATION MANAGEMENT; INFORMATION RESOURCES; INFORMATION EXCHANGE; WWW.

¹BIREME. Descriptores en Ciencias de la Salud (DeCS). Sao Paulo: BIREME, 2009. Disponible en: <http://decs.bvs.br/E/homepagee.htm>

²Díaz del Campo S. Propuesta de términos para la indización en Ciencias de la Información. Descriptores en Ciencias de la Información (DeCI). Disponible en: <http://cis.sld.cu/E/tesauro.pdf>

Copyright: © ECIMED. Contribución de acceso abierto, distribuida bajo los términos de la Licencia Creative Commons Reconocimiento-No Comercial-Compartir Igual 2.0, que permite consultar, reproducir, distribuir, comunicar públicamente y utilizar los resultados del trabajo en la práctica, así como todos sus derivados, sin propósitos comerciales y con licencia idéntica, siempre que se cite adecuadamente el autor o los autores y su fuente original.

Cita (Vancouver): Santovenia González OM. Folcsonomías: el valor agregado de la indización social en el Web. Acimed 2009; 20(3). Disponible en: [Consultado: día/mes/año].