

debe únicamente a la representación dinámica de los contenidos, sino a la interacción entre usuarios, formando redes sociales, y ofreciendo una nueva tipología de servicios (blogs, wikis, etc.).

La evolución de la Web no termina aquí. El enorme tamaño que ha alcanzado, a la vez que es una de las claves de su éxito, hace que algunas tareas, principalmente de búsqueda y presentación de contenidos concretos, sean ineficientes, requiriendo en ocasiones un tiempo excesivo y/o siendo demasiado laborioso obtener el resultado deseado. Desarrollar programas que mejoren su funcionamiento es complicado, ya que es muy difícil reproducir en una máquina la capacidad de una persona para comprender los contenidos de la Web tal y como están codificados actualmente.

De esta necesidad surge el etiquetado en la Web, y con este concepto el de Web *Semántica*. Se trata de una corriente, promovida por el propio inventor de la Web (Tim Berners-Lee²) y presidente del consorcio W3C³, cuyo último fin es lograr que las máquinas puedan entender, y por tanto utilizar, lo que la Web contiene. Esta nueva Web estaría poblada por agentes o representantes software capaces de navegar y realizar operaciones en lugar de los usuarios para ahorrar trabajo y optimizar los resultados. Para conseguir esta meta, la Web Semántica propone describir los recursos de la Web con representaciones procesables y entendibles, no sólo por personas, sino por programas que puedan asistir, representar, o reemplazar a las personas en tareas rutinarias o inabarcables para un humano. Las tecnologías de la Web semántica buscan desarrollar una Web más cohesionada, donde sea aún más fácil localizar, compartir e integrar información y servicios, para sacar un partido todavía mayor de los recursos disponibles. Sobre la Web Semántica se basa el concepto de Web 3.0.

En este trabajo nos centraremos en revisar las dos principales tendencias de etiquetado colaborativo en la red, las ontologías y las folksonomías. Después de esta introducción, veremos brevemente en el capítulo 2 qué es la Web Semántica. En el tercer capítulo introduciremos qué función tiene el etiquetado de la información y describiremos ambos conceptos que nos ocupan. En el capítulo número 4 haremos la comparativa entre ambas tendencias, presentando en el quinto capítulo las conclusiones.

2. Web Semántica

La Web Semántica es una Web extendida con metadatos, dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida. Al dotar a la Web de más significado y, por lo tanto, de más semántica, se pueden obtener soluciones a problemas habituales en la búsqueda de información gracias al uso de una infraestructura común, mediante la cual, es posible compartir, procesar y transferir información de forma sencilla. Esta Web extendida y basada en el significado, se apoya en lenguajes universales que resuelven los problemas ocasionados por una Web carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante.

Frente a la semántica implícita, el crecimiento caótico de recursos, y la ausencia de una organización clara de la Web actual, la Web semántica aboga por clasificar, dotar de estructura y anotar los recursos con semántica explícita procesable por máquinas. La figura 2 ilustra esta propuesta. Actualmente la Web se asemeja a un grafo formado por nodos del mismo tipo, y arcos (hiperenlaces) igualmente indiferenciados. Por ejemplo, no se hace distinción entre la página personal de un profesor y el portal de una tienda on-line, como tampoco se distinguen explícitamente los enlaces a las asignaturas que imparte un profesor de los enlaces a sus publicaciones. Por el contrario en la Web semántica cada nodo (recurso) tiene un tipo (profesor, tienda, pintor, libro), y los arcos representan relaciones explícitamente diferenciadas (pintor – obra, profesor – departamento, libro – editorial).

² Página en wikipedia sobre Tim Berners-Lee: http://es.wikipedia.org/wiki/Tim_Berners-Lee

³ Consorcio World Wide Web: <http://www.w3.org/>

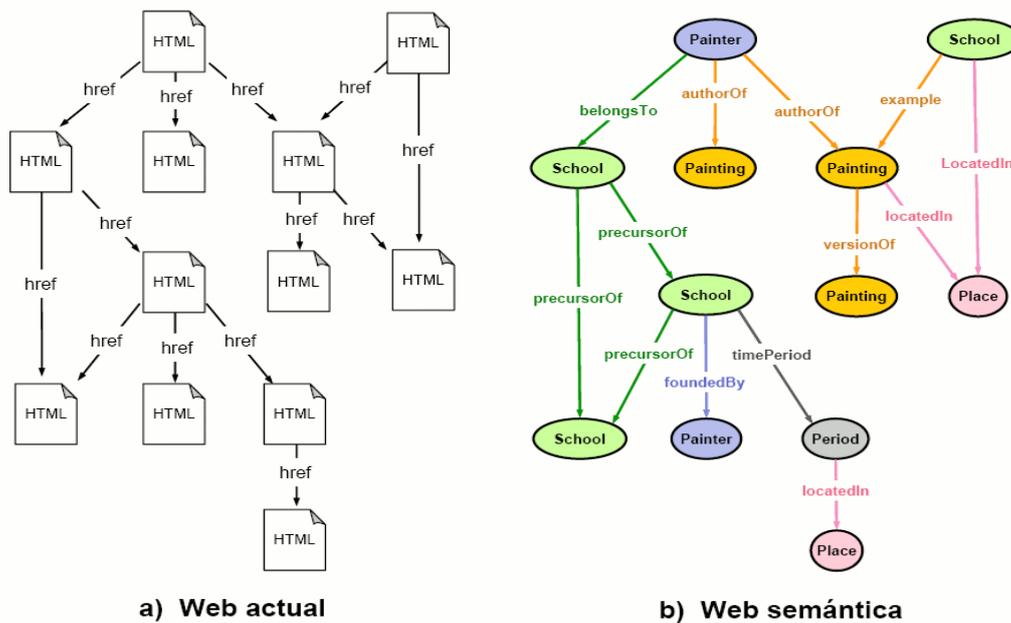


Figura 2: Estructura de la Web actual y de la Web Semántica.

La Web Semántica no cambia los principios que han hecho un éxito de la Web actual, como son la descentralización, compartición, compatibilidad, máxima facilidad de acceso y contribución, o la apertura al crecimiento y uso no previstos de antemano. Pero en la Web tradicional, como ya se ha comentado, el problema es alcanzar un entendimiento entre las partes que han de intervenir en la construcción y explotación de la Web: usuarios, desarrolladores y programas.

La idea es que la Web Semántica esté formada (al menos en parte) por una red de nodos tipificados e interconectados mediante clases y relaciones definidas, compartidos por sus distintos autores. Aunque no sea el objetivo de este trabajo, vamos a presentar brevemente los principales componentes de la Web Semántica. Éstos son los metalenguajes y los estándares de representación XML, XML Schema, RDF, RDF Schema y OWL. La OWL (Web Ontology Language Overview) describe la función y relación de cada uno de estos componentes de la Web Semántica:

- XML aporta la sintaxis superficial para los documentos estructurados, pero sin dotarles de ninguna restricción sobre el significado.
- XML Schema es un lenguaje para definir la estructura de los documentos XML.
- RDF es un modelo de datos para los recursos y las relaciones que se puedan establecer entre ellos. Aporta una semántica básica para este modelo de datos que puede representarse mediante XML.
- RDF Schema es un vocabulario para describir las propiedades y las clases de los recursos RDF, con una semántica para establecer jerarquías de generalización entre dichas propiedades y clases.
- OWL añade más vocabulario para describir propiedades y clases: tales como relaciones entre clases (por ejemplo, disyunción), cardinalidad (por ejemplo, "únicamente uno"), igualdad, tipologías de propiedades más complejas, caracterización de propiedades (por ejemplo, simetría) o clases enumeradas.

3. Etiquetado de información

Como hemos visto, la Web Semántica necesita información extra, en particular, que la información sea catalogada o etiquetada. El objetivo principal del etiquetado de información es construir una representación fidedigna del concepto que está siendo marcado; este proceso se puede separar en dos pasos: el análisis conceptual y la traducción. El análisis conceptual consiste en decidir de qué trata el concepto a etiquetar y qué es relevante en particular del mismo. Por otra parte, la traducción consiste en encontrar un conjunto de

términos o etiquetas apropiados que representen lo esencial del análisis conceptual.

El etiquetado de información puede ser considerado una “ciencia”, hay numerosos expertos que estudian la metodología para generar la meta-información necesaria para la Web Semántica, multitud de estudios y diversos eventos en los que se tratan estos temas.

Las etiquetas o palabras clave, pueden ser extraídas de la información que está siendo catalogada, o pueden ser asignadas por un “etiquetador”. Hay diversas maneras de etiquetado de información, tanto automáticas como manuales. Los sistemas de etiquetado automático difieren según sus algoritmos: por ejemplo, pueden tratar de encontrar palabras clave en la información que están catalogando considerando las palabras más repetidas con ciertas restricciones, o pueden asignar las mismas etiquetas que otras informaciones similares que puedan encontrar. El etiquetado colaborativo también tiene diversas líneas, y entre ellas encontramos las dos que nos ocupan en este trabajo y que serán descritas en las dos siguientes secciones: las folksonomías y las ontologías.

3.1. Folksonomía

El concepto de *folksonomía* se aplica a la clasificación colaborativa de la información mediante etiquetas simples, es decir, sin restricciones ni relación entre ellas. Surge de las comunidades de usuarios, en el momento en que éstos cooperan para ordenar/clasificar la información y/o objetos (fotografías, enlaces, artículos, canciones,...) de forma libre, por lo que su etiquetado obedece a una decisión de los propios usuarios, en contraposición a una lógica jerárquica, por lo que todas tienen la misma importancia.

anime art birthday birthdays books christmas
college comics daily dreams family fandom fanfic fanfiction
fic fiction food friends fun funny games geek general
harry potter health holiday holidays home house hp
humor icons internet life link links lj lost love
lyrics lytdybr me meme memes movie
movies music nanowrimo news party
personal photo photography photos pics pictures
poetry politics poll quiz quizzes quotes random
rant rants real life religion rl school sex shopping
sick snow survey tech thoughts travel tv uni update
weather weekend work writing wtf xmas ЖЖ жизнь
картинки кино книги личное лытдыбр музыка мысли

**Figura 3: Listado de palabras usadas para categorizar conceptos.
La fuente es proporcional al número de apariciones.**

El término folksonomía ha sido atribuido a Thomas Vander Wal⁴, y deriva de *taxonomía*. Taxonomía procede del griego *taxis*, que significa clasificación, y *nomos* (o nomía), que significa ordenar, gestionar. La raíz de la palabra folksonomía, *folc* (aunque haya derivado a folk), proviene del término alemán *Volk* que significa pueblo. Por tanto, folksonomía (folc+taxo+nomía) significaría “clasificación gestionada por el pueblo”.

La folksonomía está ganando importancia en Internet y en las nuevas aplicaciones con componentes sociales que surgen en la red. Algunos ejemplos de los sitios sociales más representativos donde se utilizan este tipo de etiquetas compartidas son: delicious (sitio sobre enlaces), Flickr (fotos), Tagzania (lugares), flog

⁴ Web oficial de Thomas Vander Wal: <http://www.vanderwal.net/>

(lugares), CiteULike (repositorio para que los académicos guarden enlaces personales a artículos especializados, libros,...), GenieLab (servicio de recomendación musical que funciona de forma parecida con canciones), y Tagcloud (clasifica fuentes RSS de noticias), Technorati (clasifica en folksonomías las anotaciones de los *blogs*) ó 43 Things (los usuarios catalogan los deseos que quieren cumplir en su vida). Gmail también dispone de un sistema de categorización mediante etiquetas, pero no se comparten, por lo que no cumple el concepto de las folksonomías.

Diversos conceptos van asociados a o están relacionados con folksonomía, por ejemplo: colaboración, participación, inteligencia colectiva, intercreatividad, democracia, espontaneidad, caos, descoordinación.

Otra característica social del concepto radica en la unión entre usuarios que pueden crear las folksonomías. Muchos sistemas sugieren posibles etiquetas en base a lo que han marcado otras personas anteriormente, o te permiten consultar lo que otro usuario tiene guardado y con qué marcas.

2.2 Ontologías

El otro término que nos ocupa en este trabajo es ontología. En filosofía, la ontología (que proviene del griego *οντος*, genitivo del participio del verbo *εμι*, ser, estar y *λόγος*, ciencia, estudio, teoría) se ocupa de la definición del ser y de establecer las categorías fundamentales o modos generales de ser de las cosas a partir del estudio de sus propiedades, estructuras y sistemas.

Por ello, trata de escribir o proponer las categorías y relaciones básicas del ser o la existencia para definir las entidades y de qué tipo son. Las entidades comprenden los objetos, las personas, los conceptos, las ideas, las cosas, etc. En cierto modo, reflexiona sobre las concepciones de la realidad, sus relaciones y características.

El término ontología en informática hace referencia a la formulación de un exhaustivo y riguroso esquema conceptual dentro de uno o varios dominios dados; con la finalidad de facilitar la comunicación y la compartición de la información entre diferentes sistemas y entidades. Por tanto, es una jerarquía de conceptos con atributos y relaciones, que define una terminología consensuada para definir redes semánticas de unidades de información interrelacionadas. Una ontología proporciona un vocabulario de clases y relaciones para describir un dominio, poniendo el acento en la compartición del conocimiento y el consenso en la representación de éste. Por ejemplo, una ontología sobre arte podría incluir clases como Pintor, Cuadro, Estilo o Museo, y relaciones como autor de un cuadro, pintores pertenecientes a un estilo artístico u obras localizadas en un museo.

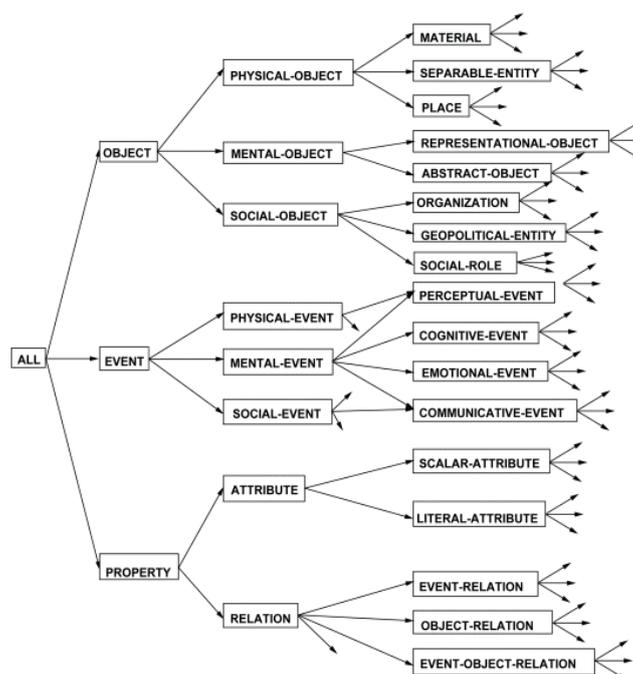


Figura 4: Etiquetado estructurado en las ontologías.

Por ejemplo, una vez establecida una ontología sobre cuadros y pintura, un museo virtual puede organizar sus contenidos definiendo instancias de pintores, cuadros, etc., interrelacionándolas y publicándolas en la Web semántica. La adopción de ontologías comunes es clave para que todos los que participen de la Web semántica, contribuyendo o consumiendo recursos, puedan trabajar de forma autónoma con la garantía de que las piezas encajen. Así por ejemplo varios museos podrían colaborar para dar lugar a un gran meta-museo que integre los contenidos de todos ellos. Un programa que navegue por una red como ésta puede reconocer las distintas unidades de información, obtener datos específicos o razonar sobre relaciones complejas. A partir de aquí sí podemos distinguir entre un cuadro pintado por un artista y un retrato de un artista.

En general, los casos reales en los que encontramos el uso de ontologías son lugares científicos y de recursos muy específicos, por ejemplo:

- Open Medical Ontologies⁵, el cual es un sitio común para ontologías de las áreas médica y biológica. Encontramos un ejemplo, the Gene Ontology⁶, el cuál provee de un diccionario cerrado para describir los genes del organismo y atributos de los mismos.
- BioDash⁷, donde se encuentra un prototipo para la ayuda al desarrollo de un grupo de medicamentos conjuntando conocimiento de diversas fuentes.

3. Folksonomía vs. Ontologías

¿Por qué las folksonomías funcionan? Realmente es difícil decidir si las folksonomías son mejores o peores que otros métodos, puesto que no hay definida una métrica que las pueda comparar, pero lo que está claro es que son una manera muy efectiva y exitosa de organizar la información. He indicado efectiva, que no eficiente, ya que como veremos después tiene sus limitaciones frente a la rigurosa clasificación definida por las ontologías. En esta sección intentaremos hacer una comparativa de ambos sistemas, aunque como hemos visto en las definiciones previas, el concepto de ontología no sólo abarca el etiquetado de los conceptos, sino que también define sus relaciones.

Las folksonomías tienen limitaciones, pero también tienen importantes ventajas que hacen útil este tipo de sistemas. La primera y más clara ventaja es el coste reducido para los usuarios que usan el sistema, en términos de tiempo y esfuerzo. Esto es debido a que la clasificación no recae en complejas categorías ni esquemas jerárquicos, sino en etiquetas simples, decididas por los propios usuarios. Esta forma de funcionar nos devuelve otra ventaja, el hecho de que el usuario esté totalmente involucrado en el sistema hace que participe de forma activa en él, y en muchos casos, a la hora de realizar el etiquetado se reciben sugerencias sobre los términos que han usado otros usuarios de la comunidad. Por tanto se crea una red de comunicaciones, feedbacks, y dependencias.

Quizás, una de las virtudes más importantes de las folksonomías sea que directamente refleja el vocabulario de los usuarios. Si el listado de palabras dentro de un directorio de recursos fuese cerrado, quizás no cubriese perfectamente conceptos relacionados con la medicina y a la vez conceptos relacionados con el hip-hop, o en un momento dado, alguna corriente cultural novedosa que pudiese surgir. Por tanto, el vocabulario utilizado se adapta a las características de los usuarios y los intereses comunes que pudiesen tener, que en general serán distintos de los creadores del contenidos o los profesionales que han creado el sistema.

Por contra, también disponen de varias limitaciones y debilidades que radican directamente en los vocabularios abiertos. La ambigüedad de las etiquetas puede surgir, ya que los usuarios pueden aplicar la misma etiqueta en distintos sentidos. Por ejemplo, al buscar en delicious por “filtrado”, obtenemos diversos resultados, unos relacionados con el filtrado de tráfico p2p, otros con el bloqueo de información para menores, con la seguridad en internet y filtrado MAC...

Los acrónimos representan otra área con gran ambigüedad. Por ejemplo, AVE puede ser usado para Avenue,

⁵ Sitio Web de Open Medical Ontologies: <http://www.obofoundry.org/>

⁶ Sitio Web de Gene Ontology: <http://www.geneontology.org/>

⁷ Sitio Web de BioDash: <http://www.w3.org/2005/04/swls/BioDash/Demo/>

Average, Audio Video Entertainment, Advertising Value Equivalent, etc. en inglés, pero también Alta Velocidad Española, en español. De todas formas, examinando el contexto del término el usuario puede decidir qué resultado era el esperado, aunque en ocasiones le lleve un tiempo encontrarlo.

Otro caso problemático es el uso de espacios o varias palabras en las etiquetas. Algunos sistemas como *delicious* no permiten espacios en las etiquetas, pero sin embargo, *Flickr* sí los acepta. Al no haber ninguna convención entre usuarios, algunos decidirán usar para catalogar una foto de la Catedral de Logroño como "Catedral Logroño", otros "Catedral/Logroño", otros "CatedralLogroño" y otros usarán dos etiquetas "Logroño" y "Catedral". Estos dos sitios, Flickr y delicious, no hacen distinción entre letras mayúsculas y minúsculas, pero si hiciesen distinción daría lugar de nuevo al uso de distintas etiquetas para marcar mismos conceptos.

Por último, otro caso por el que los sistemas basados en diccionarios abiertos pueden resultar menos eficaces es debido a los sinónimos. En general no hay control de sinónimos en los mismos, por lo que los usuarios pueden catalogar el mismo elemento mediante etiquetas con el mismo significado, como "mac", "macintosh", "apple", todas ellas relacionadas con los ordenadores de Apple. También el uso indistintamente de las formas en singular y plural de las etiquetas supone un problema, además de las formas en masculino y femenino.

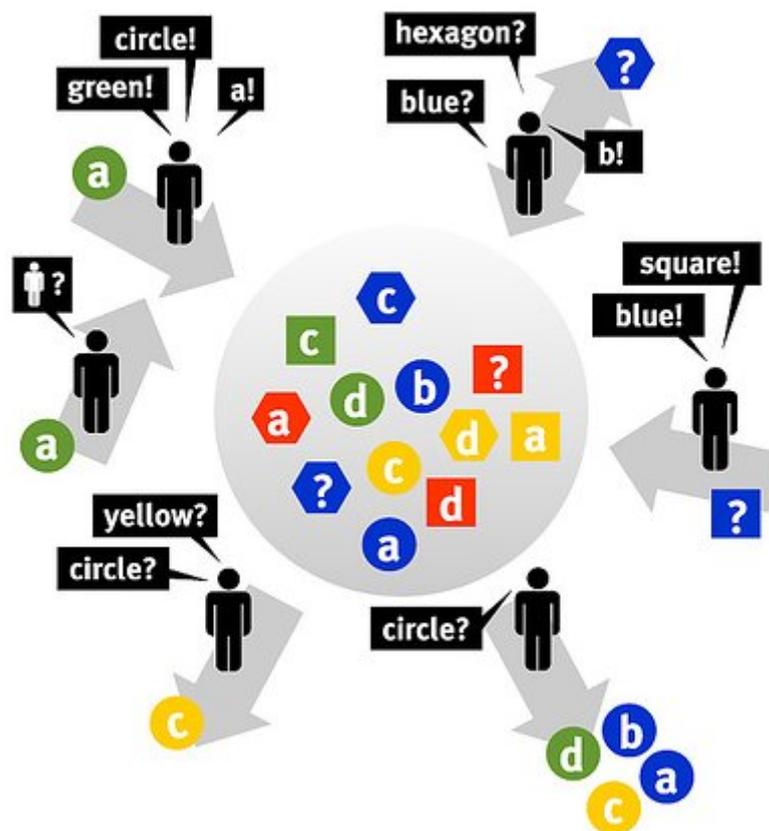


Figura 5: Folksonomías, etiquetado colaborativo y social.

Las folksonomías, al ser un etiquetado realizado por los propios usuarios, se suelen centrar en recursos sociales, por contra las ontologías resultan más interesantes para catalogar recursos de otro nivel, quizás más científico o profesional. Es el caso de muchos registros de educación, bibliotecas o documentación técnica de empresas, los cuales tienen un número de categorías cerradas y dentro de ellas ciertos términos prefijados. Estos sistemas tratan de clasificar y gestionar los contenidos en base a un conjunto finito de categorías establecidas siguiendo ciertos criterios predefinidos. Pero este sistema presenta un problema, durante la clasificación es necesario pasar por el filtro de un grupo de especialistas encargados de etiquetar cada uno de los recursos. Es un trabajo lento pero evita el ruido, la sinonimia o los falsos positivos de las folksonomías. En cambio, la implicación por parte del resto de usuarios, es escasa debido a la falta absoluta de participación.

Al respecto de los falsos positivos, se podrían considerar un fallo o una desventaja de las folksonomías, pero en ciertos ámbitos pueden resultar interesantes, e incluso importantes. Mientras que con el uso de vocabularios cerrados se pretenden centrar y acotar los resultados de la búsqueda, los falsos positivos que dan lugar a la sinonimia y las etiquetas no cerradas en las folksonomías, permiten que se navegue por información que no se esperaba. En muchas ocasiones se puede encontrar resultados interesantes, que en caso contrario no se hubiesen hallado, ya que el sistema podría haber devuelto "0 resultados". Hay que señalar que otra vez aparece la diferencia entre el usuario que busca algo muy concreto y relevante para un ámbito, por ejemplo profesional, y el usuario que busca recursos variados relacionados con un concepto.

Después de este intento de comparativa entre ambos sistemas, creo que ninguno de los dos cubre las necesidades de los usuarios en todos los ámbitos. Y que una opción interesante sería la combinación de ambas metodologías, por eso en la siguiente sección se intenta revisar esta propuesta.

3.1. Folksonomías más ontologías

Iba a haber llamado a esta sección *Estrategias mixtas*, pero creo que con el título elegido se hace mayor hincapié en la idea de mostrar que quizás el mejor sistema sea una mezcla de los dos estudiados.

Una solución pasaría por la definición de ontologías básicas que hagan que los homónimos y sinónimos se fusionen en términos individuales para designar los objetos deseados. Estas ontologías podrían definirse manualmente, o de forma automática por inferencias léxicas o morfosemánticas, es decir, eliminando plurales (por ejemplo, chicos/chico -> chico), eliminación de sexo (chico/chica -> chico), etc. o por el contexto. El contexto da mucho juego. Es lo que ya hace Flickr con sus tag-bundles, definir ontologías que cataloguen, por su contexto, todas aquellas etiquetas que se refieren a una misma realidad. Esa misma etiqueta en otro contexto podría significar una cosa completamente distinta.

Otra opción más compleja sería superponer una ontología a una folksonomía. Así, tendríamos una folksonomía en la que determinadas etiquetas están estructuradas por expertos y reservadas para un uso determinado, es decir, una folksonomía con determinados vocabularios controlados. Determinadas etiquetas, las más genéricas o las utilizadas como primer filtro de manera más habitual, estarían reservadas y categorizadas de una manera más rígida que otras, para asegurar un uso adecuado de las mismas. Así, simplemente se fijarían unos criterios determinados y se mantendría la libertad de los usuarios para todas las demás. Estos usuarios deberían tener un margen de paciencia determinado, ya que no tendrían libertad para detallar todas las etiquetas de inicio libremente, pero tampoco sufrirían un esquema de categorización extremo.

Así, este sistema recoge las bondades de los predecesores, por una parte, la primera aproximación se realiza a unas categorías marcadas, pero después al no haber un diccionario totalmente acotado, permite catalogar de manera menos rígida que en las ontologías cerradas y ajustar más la descripción.

4. Conclusión

La Web ha ido evolucionando y a su vez creciendo en contenidos. Actualmente los buscadores devuelven muchos enlaces no relevantes, y en ocasiones encontrar lo que se necesita es complicado. La necesidad de dotar de mayor significado a la información en la red, conlleva el tener que marcar la misma con meta-información, desembocando en el futuro en la Web Semántica. Esta meta-información permitirá que las máquinas puedan comprender y catalogar los recursos. Así serán capaces de ser más eficaces en ciertas tareas que a los usuarios les puede llevar demasiado tiempo.

Hemos revisado dos sistemas de etiquetado colaborativo, las folksonomías y las ontologías. En las primeras, tal como indica su etimología, son los propios usuarios el núcleo de su éxito, ya que se encargan de catalogar libremente su propia información y compartir la misma con sus etiquetas. Las segundas engloban no solo el marcado de la información con estructuras de datos y diccionarios totalmente cerrados, sino también la relación entre los términos. Por su estricta forma, de inicio parecen más orientadas a catalogar información más científica, al contrario que las folksonomías, siendo necesario que expertos realicen las tareas.

Los problemas en las folksonomías surgen por inconsistencia, un origen diverso condiciona criterios diferentes en la asignación de etiquetas, lo que en ocasiones acaba derivando en sinonimias, homonimias, polisemias o, en general, la falta de homogeneidad y de coherencia. Para un mismo concepto numerosas etiquetas son utilizadas. Esto conlleva que a la hora de realizar la búsqueda, el sistema nos devuelva diversos resultados que no son los esperados, pero como se ha comentado en alguna ocasión resultan interesantes.

¿Es un sistema mejor que otro? Realmente no son comparables, cada uno tiene su ámbito, y en cada caso habría que estudiar si es mejorable. Dentro del documento se ha propuesto el uso de sistemas que combinen conceptos de ambos sistemas, marcando una primera aproximación de los conceptos en varias áreas determinadas, y posteriormente permitiendo a los usuarios incluir sus propias palabras claves. Estas palabras clave podrían ser revisadas por el sistema, para evitar palabras en singular y plural, masculino y femenino, etc.

¿Será posible modelar y catalogar toda la información que existe en la red? Es una pregunta que dejo abierta, ya que en mi opinión se podría lograr catalogar gran parte de la misma por parte de los propios usuarios, pero esto no aseguraría que este etiquetado sea totalmente correcto. Quizás fuese suficiente para aplicaciones sociales, pero no para otras que requieran la total certeza de que las etiquetas son correctas y fiables.

¿Deberían crear una nueva Red de información, basada en la Web Semántica y abandonando lo que hoy conocemos como Internet?



Figura 6: Preguntas, ¿sin respuestas?