

Hacia una web semántica social

Por **Jorge Morato**, **Sonia Sánchez-Cuadrado**, **Anabel Fraga** y **Valentín Moreno-Pelayo**



Jorge Morato es licenciado en ciencias biológicas por la Universidad de Alcalá de Henares. A partir del año 1991 empezó a trabajar en distintas empresas relacionadas con la documentación. Obtuvo el título de doctor en documentación por la Universidad Carlos III en 1999. Desde ese mismo año ha impartido clases en las licenciaturas de documentación, LADE y en la ingeniería en informática en asignaturas relacionadas con la ingeniería de la información, recuperación de información e ingeniería del software.



Sonia Sánchez-Cuadrado es licenciada en filología por la Universidad Complutense de Madrid y doctora en Documentación por la Universidad Carlos III de Madrid. Desde 2002 es profesora ayudante en el departamento de informática de dicha Universidad con docencia en la licenciatura de documentación e informática sobre modelado e ingeniería de la información. Actualmente se halla involucrada en distintos proyectos para la creación automática de ontologías.



Anabel Fraga es ingeniero superior en informática y doctorando en ingeniería informática en la Universidad Carlos III de Madrid. Desde 2005 es profesora ayudante en el Departamento de Informática de dicha universidad con docencia en la ingeniería técnica de informática de gestión y en la ingeniería superior en informática, en temas relacionados con ingeniería de información, arquitectura de software, ingeniería del software y reutilización. Actualmente trabaja en proyectos de generación de ontologías y reutilización de conocimiento.



Valentín Moreno Pelayo es matemático (especialidad computación), profesor del Departamento de Informática de la Universidad Carlos III de Madrid, investigador en distintos proyectos de inteligencia artificial y recuperación conceptual financiados por Cicyt, y en redes temáticas. Actualmente trabaja en aplicación de la minería de datos al procesamiento del lenguaje natural y a la creación automática de tesauros.

Resumen: *La evolución de internet hacia un escenario con mayor potencial y con usuarios cada vez más involucrados en su gestión y desarrollo, requiere de la incorporación de documentos con menor grado de ambigüedad semántica. La propuesta de la web semántica, a pesar de haber sido realizada hace casi una década, tiene actualmente una implantación modesta. Por el contrario, la denominada web 2.0, que se refiere a una evolución autónoma de internet hacia entornos colaborativos tiene un gran éxito. La solución propuesta por la web 2.0 ha superado las limitaciones inherentes a la utilización de herramientas de procesamiento del lenguaje natural y a las aproximaciones estadísticas. Por lo tanto, parece lógico analizar las aportaciones que estos enfoques pueden aportar a la web semántica para su implantación. En este artículo se discuten medios para dar el salto de la web social a una web semántica socializada.*

Palabras clave: *Web 2.0, Web semántica, Web semántica social, Entornos colaborativos, Folksonomías, Usabilidad.*

Title: **Toward a social semantic web environment**

Abstract: *The internet's evolution toward a scenario of greater potential, with users increasingly involved in its development and management, demands a lower level of semantic ambiguity in the documents that are provided. The proposed semantic web, a concept that appeared almost a decade ago, has had only modest impact. On the other hand, web 2.0, an autonomous evolution of the web toward a collaborative environment, has met with enormous success. The solution devised by web 2.0 has exceeded the limitations of natural language processing tools and statistical approaches. For that reason, it seems logical to analyze the potential contributions that web 2.0 concepts could make to further the development of the semantic web. Some ways of making the leap from a social web to a social semantic web are discussed.*

Keywords: *Web 2.0, Semantic web, Social semantic web, Collaborative environments, Folksonomies, Usability.*

Morato, Jorge; Sánchez-Cuadrado, Sonia; Fraga, Anabel; Moreno Pelayo, Valentín. "Hacia una web semántica social". En: *El profesional de la información*, 2007, enero-febrero, v. 17, n. 1, ppp. 78-85.

DOI: 10.3145/epi.2008.ene.09

Recursos de interés

- NaturalFinder.
<http://demos.bitext.com/MSN/>
- Vivisimo.
<http://vivisimo.com/search?tb=vivisimocom&query=&x=15&y=11>
- CleverSearch.
<http://wdok.cs.uni-magdeburg.de/clever-search/>
- Wikipedia. Web 2.0.
http://en.wikipedia.org/wiki/Web_2.0#Characteristics_of_Web_2.0
- Swoogle.
<http://swoogle.umbc.edu/>
- Tabulator: async javascript and semantic web.
<http://dig.csail.mit.edu/2005/ajar/release/tabulator/0.8/tab.html>
- SKOS Core guide.
<http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20050510/>
- Published subject indicators for modelling thesaurii.
<http://www.techquila.com/psi/thesaurus/>
- Zthes-specifications for thesaurus representation, access and navigation.
<http://zthes.z3950.org/>
- Metadata Authority Description Schema (MADS).
<http://www.loc.gov/standards/mads/>
- Simile-My Piggy Bank.
http://simile.mit.edu/wiki/Piggy_Bank
- W3C: annotea project.
<http://annotest.w3.org/>
- Comprehensive listing of 250 semantic web tools (updated), AI3 adaptive information.
<http://www.mkbergman.com/?p=291>
- Searchtools.com. XML query and searching resources.
<http://www.searchtools.com/info/xml-resources.html>
- OntoSelect.
<http://olp.dfki.de/OntoSelect/>
- Ontaria.
<http://www.w3.org/2004/ontaria/>
- Semwebcentral. SemWebCentral: open source tools for the semantic.
<http://www.semwebcentral.org>
- Webkb. WebKB home page.
<http://www.webkb.org/>

Internet y problemas que afronta su evolución

LAS CARACTERÍSTICAS QUE DIFERENCIAN A INTERNET respecto a los repositorios documentales tradicionales y la trayectoria que ha sufrido la web en los últimos años hacen necesario incorporar planteamientos novedosos para poder ofrecer nuevas funcionalidades. A continuación se pormenorizan algunas de estas diferencias y tendencias:

a. Internet se caracteriza por la existencia de enormes volúmenes de documentos electrónicos, contabilizados en miles de millones.

Estos poseen una calidad variable, una temática multidisciplinar y una audiencia no predefinida.

b. La naturaleza de los documentos no es siempre textual. Los principales sistemas de indización automática se basan en la identificación de las palabras que forman parte del texto del documento. Sin embargo, este sistema falla en el caso de documentos no textuales como imágenes, archivos de audio o vídeo. En consecuencia, se requiere de un mecanismo alternativo de indización, dado que tradicionalmente los motores de búsqueda tienen dificultades para recuperar este tipo de materiales.

c. El extenso volumen de documentos en lenguaje natural expresado mediante html. Su escasa semántica, dada su función principalmente visual, hace que las aplicaciones informáticas no puedan acceder de forma correcta al contenido, ya que éste no se encuentra estructurado. En efecto, la mayor parte de estos documentos carece de estructura salvo por un escaso número de elementos, como el título o las etiquetas *<meta>*.

d. La autoorganización mediante diferentes mecanismos, por ejemplo, los hiperenlaces, la incorporación de páginas a directorios, la estandarización de elementos de di-

seño, o la utilización de metadatos (*tags*, *<meta>*, etc.). Este aspecto de la web en relación a su evolución ya fue descrito por **Flake** (2003).

e. El auge actual de aplicaciones informáticas para tratar la información contenida en los miles de millones de documentos existentes tiene fuertes restricciones debidas a algunas características de la web actual, como:

- La existencia de documentos no textuales.

- Las limitaciones semánticas del html.

- La falta de normalización de la estructura de los materiales presentes en internet.

- Los problemas de polisemia y sinonimia del lenguaje natural.

- En la actualidad, las respuestas a las problemáticas anteriores se producen de manera parcial y local; es decir, se practican soluciones optimizadas para casos concretos y con un grado de reutilización escaso, debido en parte a la poca interoperabilidad entre los formatos de los documentos de distintos repositorios.

f. La web ya ha alcanzado una fase de madurez, como se puede comprobar consultando algunos indicadores como:

- La simplificación de las herramientas para creación de páginas web. Hoy es posible crear un gran sitio web con escaso conocimiento informático, por ejemplo mediante herramientas de gestión de contenido (*CMS*, *Content Management Systems*) como *Joomla*, o aplicaciones colaborativas como *DokuWiki*, o *Moodle*.

- Los usuarios ya no son meros recuperadores-lectores de páginas web, sino que reclaman un papel más activo para crear y compartir recursos.

- Algunos organismos se han convertido en referentes, afectando sus decisiones a la organización de

la información. Un caso claro, es el *W3C*, ya que sus propuestas son frecuentemente aceptadas como estándares de facto.

Posibles soluciones para la nueva internet

Para conseguir que la Red sea un repositorio con mayores funcionalidades, se han venido realizando diferentes intentos, que se pueden englobar en tres tipos: el procesamiento del lenguaje natural (PLN); la utilización de la estadística y la minería de datos; y la creación de sistemas de organización del conocimiento. Además de su descripción, a continuación se plantean algunos de los problemas que presentan estos enfoques:

a. Dotar a las aplicaciones, sobre todo buscadores, de capacidades de PLN. Aunque con evidentes atractivos, se deben afrontar distintos problemas:

- La ambigüedad terminológica inherente al lenguaje natural, agravada por el hecho de que los documentos de la web pertenecen a múltiples dominios.

- La falta de estructuración de los contenidos dificulta la interpretación por los sistemas automáticos. Además, en algunos formatos como las tablas, la expresión de los conceptos se hace visualmente en vez de utilizar estructuras lingüísticas.

- Las estructuras lingüísticas en un entorno tan heterogéneo no son previsibles. Diferentes sintaxis con los mismos términos pueden expresar el mismo concepto en unas ocasiones, o ideas diferentes en otras. Este problema, unido a la falta de estructuración de los documentos y la ambigüedad semántica, hace que sea muy complicado desarrollar una recuperación tipo pregunta-respuesta: que ante una pregunta el sistema ofrezca una respuesta, y no un conjunto de documentos que probablemente la contendrán.

- La brevedad de las consultas, frecuentemente formadas por uno o dos términos, invalida el análisis por contexto.

- Se trata de enfoques muy dependientes del lenguaje, que obtiene peores resultados en repositorios multilingües. Hay que tener en cuenta que los mejores rendimientos se obtienen cuando a la aplicación de conocimiento lingüístico se le suministran corpus etiquetados, colecciones de prueba y frecuencias de uso, y que todo este conocimiento varía según la zona geográfica, el dominio y la disponibilidad de recursos en el idioma objetivo.

- Es una aproximación con costes computacionales evidentes.

- Carece de mecanismos efectivos para mejorar la interoperabilidad entre bases de datos o para su procesamiento por aplicaciones informáticas; por ejemplo, para realizar inferencias a partir de un conjunto de documentos.

A pesar de todos estos inconvenientes, actualmente las aplicaciones web integran mecanismos de PLN. Un ejemplo es *Google*, que incorpora identificación de palabras vacías y lematización de términos no ambiguos. Así, si se escribe “nació Juan” encuentra documentos con los términos “nacido” o “nace”. Otro ejemplo se puede ver en la demo de *NaturalFinder* para buscar en lenguaje natural con el buscador *MSN*.

b. La utilización de estadísticas y de minería de la web tiene actualmente gran popularidad. Estas técnicas son frecuentemente utilizadas para la identificación de clusters de términos. Este enfoque, con un gran auge actual en los motores de recuperación web, tiene bastantes problemas, algunos comunes con el PLN, entre los que destacan:

- La escasa mejora que facilitan para optimizar la comprensión que tienen las máquinas de los documentos no estructurados.

– Los problemas ya mencionados de ambigüedad del lenguaje.

– Funcionan mejor en dominios bien representados en internet pero no en otros muy específicos o con escasa presencia de documentos.

Como en el caso anterior se puede ver el funcionamiento de los clusters en numerosas aplicaciones web, que varían desde las técnicas de clasificación documental (Steinbach, 2000) a la incorporación de estos recursos a motores de recuperación. Ejemplos en recuperadores se pueden ver en *Google* con la opción “Ud. quiso decir” o en algoritmo de posicionamiento; o en *Vivísimo*, que crea una taxonomía de forma automática a partir de la frecuencia de concurrencia de los términos de las páginas.

c. La utilización de sistemas de organización del conocimiento. Aunque parece ser la solución que más éxito tiene, no ha podido eludir la perenne dialéctica entre la utilización de lenguaje libre o controlado, o si se quiere, ir de lenguajes menos a más estructurados, de acuerdo con la graduación que se muestra en este ya tradicional esquema:

– Palabras-clave independientes, que se usan en indización libre tanto por extracción como por asignación, uno de cuyos tipos son las folksonomías.

– Listas de palabras, como los glosarios, de nombres o los diccionarios.

– Categorizaciones y clasificaciones: con el objetivo de crear conjuntos temáticos, por ejemplo las clasificaciones bibliotecarias (Zeng, 2003).

– Grupos de relaciones, basados en las establecidas entre términos y conceptos, de estructura más compleja, entre los que se sitúan los tesauros, las redes léxicas (por ejemplo, *WordNet*), los *topic maps*, y las ontologías. Como se discutirá en el siguiente punto, hoy por hoy, es la opción que más aceptación tiene entre los usuarios. Ejemplos de su empleo se encuentran en las listas de sinónimos (se puede comprobar con la búsqueda “~auto”: encuentra términos como “car”, “automobile”, etc.), o con la aplicación de la red léxica *WordNet* a la expansión de consultas en *CleverSearch*.

Iniciativas para la evolución de internet

Hasta el momento se han planteado dos grandes iniciativas para incorporar a internet los sistemas de organización del conocimiento: la web 2.0 y la web semántica (tabla 1):

a. La evolución de la Red hacia la web 2.0, es decir, la creación de una red social que concede a los usuarios mayor control sobre los recursos. **Dale Dougherty** de *O'Reilly Media* creó el término web 2.0 en una conferencia con **Craig Cline** de *MediaLive*. Durante la charla sobre la evolución de la Red se percataron que habían surgido numerosos servicios caracterizados por un grado de colaboración y la voluntad de compartir recursos por parte de los usuarios. El término se afianzó en la *Conferencia web 2.0* en 2004 y está basado en una arquitectura de participación que anima a los usuarios a dotar a las aplicaciones de valor añadido mediante el uso que hacen del recurso, por ejemplo mediante la descripción con *tags*.

	Web 2.0	Web semántica
Origen	Constatación de la evolución natural de la web	Propuesta de Tim Berners para evolucionar la web
Implantación	Muy alta	Escasa (Palacios , et al., 2006)
Coordinación	No existe	Centralizada, sobre todo por el W3C
Foco	Personas	Aplicaciones informáticas
Primeras menciones	2003, primera conferencia 2004	1999 (Berners-Lee , 1999)
Expresión	Lenguaje libre, expresado mediante folksonomías, palabras clave denominadas etiquetas (<i>tags</i>), con problemas de sinonimia y polisemia	Lenguaje controlado, mediante lenguajes para expresión de ontologías, <i>KOS</i> y vocabularios de metadatos
Algunas características	<ul style="list-style-type: none"> - Descripción de los recursos para mejorar su distribución gratuita, se comparten conocimiento y desarrollos - Arquitectura de colaboración - Usabilidad alta - Un recurso es más útil cuanto más uso tenga 	<ul style="list-style-type: none"> - Utilización de un lenguaje estandarizado con sintaxis uniforme y semántica no ambigua - Interoperabilidad: intercambio de información entre cualquier repositorio - Usabilidad escasa

Tabla 1. Algunas diferencias entre la web 2.0 y la web semántica

“La web semántica responde a un entorno centralizado, a diferencia de la web 2.0”

b. La propuesta de la web semántica o web 3.0. Es decir, ampliar la audiencia de los documentos de la web, en un principio dirigido a su lectura por personas, para que sea interpretado por aplicaciones informáticas, mejorando de este modo la calidad de la recuperación y la interoperabilidad entre distintos repositorios.

Problemas de las soluciones propuestas

Ambas iniciativas tienen ventajas e inconvenientes. En este apartado analizaremos las soluciones propuestas por cada una y las problemáticas que presentan.

1. Solución aportada por la web 2.0

Internet tradicionalmente ha presentado deficiencias en la falta de semántica en la descripción de algunos recursos. La solución 2.0 a estas carencias son las folksonomías. Se trata de etiquetar recursos mediante términos de descripción procedentes del lenguaje natural. Han alcanzado gran popularidad en los últimos tiempos, probablemente por la sencillez en su aplicación, ya que no utilizan un lenguaje controlado que deba asimilarse. Entre sus ventajas destacan:

– Permitir que un usuario cualquiera, especialmente quien publica un recurso, por ejemplo una foto o un vídeo, pueda añadir términos descriptivos sin un entrenamiento previo sobre un determinado vocabulario controlado. La facilidad para añadir términos provoca que un gran volumen de usuarios se animen a etiquetar recursos a un coste nulo.

– Suprimen los costes derivados de la construcción, consenso terminológico y familiarización de un vocabulario controlado con la dificultad que supondría para el dominio de internet caracterizado por la heterogeneidad temática.

Como solución, las folksonomías se adecuan a las necesidades que internet requiere, ya que permiten:

– Etiquetar de forma económica un medio que cuenta por miles de millones los recursos disponibles.

– Adecuar las etiquetas a un término concreto, a pesar del entorno multidisciplinar y multifuncional. Pueden ajustarse a múltiples perfiles de usuarios, ya que son ellos mismos los que las incorporan según sus intereses.

– Etiquetar documentos no textuales como vídeos y fotos que de otra manera no serían correctamente recuperados por los *crawlers* de los motores web basados en indexación automática de textos.

– Facilidad para comprender el empleo e interpretación de las *tags* de la folksonomía.

– Mayor implicación de los usuarios en la gestión de los recursos.

Pero a pesar de esta simbiosis, presenta determinadas carencias debidas a que no permiten:

– Eliminar la ambigüedad terminológica.

– Una interpretación correcta por parte de las aplicaciones informáticas.

– Expandir o restringir consultas.

– Suprimir los problemas para fusionar dos folksonomías.

– La incorporación de inferencias mediante la representación de información.

– Asegurar una mejor calidad de la descripción del recurso.

Además, muchos de los términos descriptivos ocasionan ruido al no tener sentido únicamente más que para un usuario concreto (por ejemplo: “Juan”, “casa de Pepe”, etc.).

2. Solución aportada por la web semántica

Uno de los pilares sobre los que se asienta es la creación de sistemas de organización del conocimiento como las ontologías. Para que la web semántica sea una realidad se necesita una sintaxis normalizada, conseguida mediante xml-rdf (W3C, 2005) y una semántica no ambigua, obtenida mediante vocabularios de metadatos, es decir, lenguajes controlados.

Sus ventajas más destacadas son:

– Interoperabilidad: algunos autores la consideran como un proyecto para crear un mediador universal para el intercambio de información. Esto es factible mediante la creación de documentos interoperables con una semántica no ambigua para las aplicaciones informáticas de la *www*. Es decir, se trata de convertir la web en una gran base de datos.

– Recursos semánticos: se necesita que el conocimiento semántico sea expresado en documentos escritos en un lenguaje orientado al modelado de conocimiento como *rdf*. Las ontologías tienen un rol importante en la web semántica porque soportan el conocimiento semántico para ser utilizado por las aplicaciones informáticas resultando información con menos ruido y mayor pertinencia para el usuario.

Lamentablemente, esta propuesta planteada en el año 1999 (Berners-Lee, 1999) no ha tenido el éxito esperado. Una búsqueda en *Google* muestra que, en junio de 2007, existen 2.770.000 documentos con extensión *rdf*, 41.500 con *OWL*, 2.330 *XTM*, 4.540.000 *RSS* y 212.000 *ATOM*. Es decir 7,5 millo-

nes de documentos en un conjunto de 10.000 millones. Con *Swoogle*, un recuperador especializado en la web semántica, los resultados no son mucho mejores: 158.000 documentos semánticos conteniendo el término rdf y 2.323.857 en el caso de *OWL*.

“La web semántica es ya una propuesta de 1999 que lamentablemente no ha tenido el éxito esperado”

Las causas son diversas, pero las principales a tener en cuenta son:

- Falta de legibilidad de los lenguajes rdf y *OWL*, lo cual supone un cuello de botella para que los expertos validen las ontologías (**Gómez-Pérez**, 2004). En 2005 **Mika** subrayó la importancia que tienen los usuarios (denominada “dimensión social” en el artículo original) para la aceptación de la web semántica, la cual tiene diferentes grados de complejidad en la creación de recursos, inversamente proporcional a la proximidad al usuario. En la figura 1 se muestra gráficamente este hecho: existe una tendencia que provoca que cuando se incrementa la complejidad en la representación semántica se produce una disminución en la dimensión de contacto con el usuario. Esta dimensión social engloba diferentes afectos como usabilidad, legibilidad o necesidad de conocimientos previos para su interpretación.

“La web semántica tiene diferentes grados de complejidad, la cual es inversamente proporcional a la proximidad al usuario”

- Escasez de herramientas que faciliten la creación de documentos semánticos mediante formularios usables, un ejemplo de un entorno más amigable se puede ver con *Protégé* o con *Tabulator*.

- La migración de folksonomías a folkontologías es un tema aún por desarrollar, aunque ya existen estudios (**Damme**, 2007; **Matsuo**, 2006). Básicamente, este proceso se hace mediante herramientas estadísticas (**Bagelman**, 2006) y lingüísticas, o incluso proponiendo una normalización para asignar etiquetas (**Xu**, 2006).

- La necesidad de incorporar técnicas semiautomáticas para la creación de sistemas de organización del conocimiento basadas en PLN y minería de datos (**Sánchez-Cuadrado**, 2007), ya que la carencia de estos recursos, junto con la lentitud en su creación dificulta la implantación de la web semántica.

- Presencia de duplicidades en los vocabularios de metadatos y ontologías, que provoca la desconfianza y confusión del usuario que no sabe cuál es el idóneo o más generalizado. Como ejemplo están los vocabularios de metadatos para

expresar tesauros. Actualmente existen, entre otros: *SKOS-Core* del W3C, *PSI* de los *topic maps*, *Zthes* y *MADS*.

Discusión. Hacia un nuevo conjunto de soluciones para un nuevo concepto de internet

Como consecuencia, parece evidente que la única forma de popularizar la web semántica sería mediante una aproximación a la web social, para implicar a los usuarios en la creación y gestión de documentos semánticos. Actualmente, se están realizando distintas propuestas de conseguir una mayor usabilidad de los documentos semánticos, obtener una mayor implicación de los usuarios, y suministrarles recursos que les ayuden a mejorar la localización y el grado de popularidad de los contenidos.

- Una opción prometedora en esta dirección son los enfoques híbridos que combinan ambas tendencias. Un ejemplo es *Piggy Bank* (figura 2), una aplicación que captura localmente las etiquetas de los rdf visitados, para organizarlos en una ontología local, y que puede ser compartida en el *Semantic Bank*

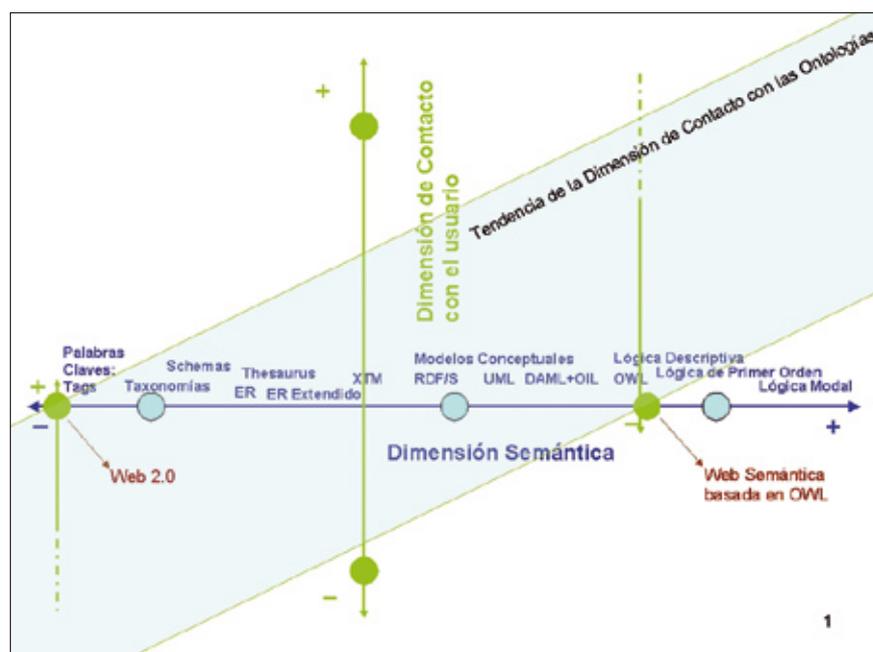


Figura 1. Relación entre la complejidad de la web semántica y la proximidad al usuario

“La única forma de popularizar la web semántica sería mediante una aproximación a la web social para implicar a los usuarios en la creación y gestión de documentos semánticos”

(Senso, 2007). Esta herramienta permite poner *tags* a las url seleccionadas para recuperarlas posteriormente. Un proyecto muy similar ha sido propuesto por el W3C bajo la denominación de *Annotea*, instalable mediante un plugin en *Firefox* denominado *Annozilla*. El problema de ambas aproximaciones son las inherentes a las *tags*, es decir, la sinonimia y la polisemia.

– Como ya se mostró en la figura 1, es necesario cambiar la trayectoria natural de la dimensión de contacto con el usuario mediante aplicaciones informáticas usables o la simplificación. La simplificación (Senso, 2007) se puede dar de diferentes formas, por ejemplo, prescindiendo de los lenguajes formatos a favor de expresar el contenido mediante documentos Xhtml (una solución denominada microformatos). Esta opción, dado su carácter de solución local, no consigue uno de los principales objetivos de la web semántica, la interoperabilidad. Por el contrario, la interoperabilidad no se ve mermada en determinados documentos semánticos que han tenido un gran éxito gracias a la simplicidad del vocabulario subyacente; un ejemplo es la sindicación de contenidos mediante *RSS*.

– La creación de nuevos recursos que faciliten la localización y popularización de los documentos semánticos: un ejemplo es el ya mencionado buscador *Swoogle*, o los también conocidos *OntoSelect* u *Ontaria*. Como en el caso de *PigggyBank*, *OntoSelect* es a su vez un

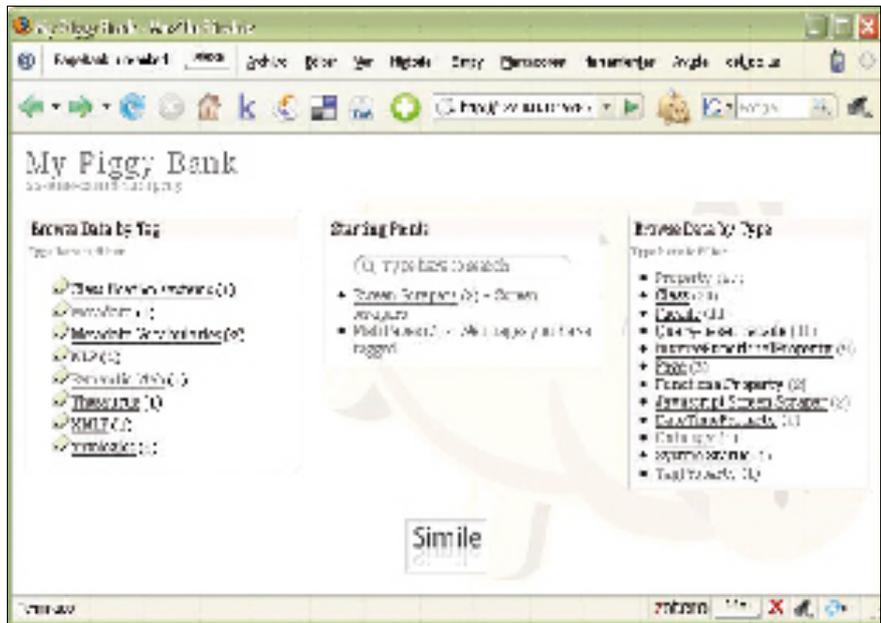


Figura 2. Visualización de la captura de una página con MyPiggyBank

repositorio de documentos semánticos.

Existen propuestas, en fase de experimentación, que incorporan varias de estas tecnologías. Estas soluciones combinan repositorios de documentos semánticos con buscadores sobre popularidad y ayudas para traducir las *tags* de los usuarios a un vocabulario genérico centralizado basado en recursos similares a *WebKB* (Palacios, 2006). Sería capaz de proporcionar estadísticas de popularidad de cada ontología, lo que evitarían problemas de confusión del usuario debido a duplicidades de vocabularios.

Consideraciones finales

La trayectoria de internet en los últimos años hace necesario incorporar nuevos planteamientos a la resolución de problemas clásicos para tratar el gran volumen de información de todo tipo en diferentes formatos y dirigida a un público heterogéneo. Por otro lado, los usuarios y las aplicaciones informáticas necesitan de un mayor contenido semántico para dar mayor funcionalidad a la web.

Las soluciones adoptadas han seguido varios caminos: el desa-

rollo de herramientas de procesamiento del lenguaje natural, la incorporación de técnicas estadísticas, y la creación de recursos terminológicos para dotar a los recursos de una semántica adicional.

Entre las propuestas para la creación de recursos terminológicos destacan las folksonomías y las ontologías, procedentes respectivamente de cada una de las dos grandes iniciativas planteadas hasta el momento: la web 2.0 y la web semántica. La primera de ellas ha logrado un éxito mayor debido a su proximidad al usuario y a su simplicidad, pero tiene el gran inconveniente de no poder subsanar problemas de interoperabilidad, sinonimia y polisemia. Por el contrario la web semántica no consigue alcanzar el mismo grado de implantación a causa de los mismos factores, poca legibilidad y complejidad, pero carece de los problemas de la web 2.0.

“Se propone la creación de recursos semánticos más usables, simples y con potencialidades de búsqueda”

Como soluciones, en este documento se proponen la creación de recursos semánticos más usables, simples y con potencialidades de búsqueda. Una posibilidad sería la traducción de las folksonomías de los usuarios a ontologías, algo que se podría realizar mediante dos procedimientos:

– La incorporación de procesamiento del lenguaje natural y estadístico a la migración (folksonologías).

– La asistencia al usuario para alinear su vocabulario en lenguaje natural con otro en lenguaje controlado.

Bibliografía

- Begelman, G.; Keller, P.; Smadja, F.** “Automated tag clustering: improving search and exploration in the tag space”. En: *WWW2006*, 2006. <http://www.rawsugar.com/www2006/20.pdf>
- Berners-Lee, T.; Fischetti, M.** *Weaving the web*. 1st edition. San Francisco: Harper, 1999.
- Damme, C.**, et al. *Ontology: an integrated approach for turning folksonomies into ontologies*, 2007. <http://www.heppnetz.de/files/vandammehepp-siorpaes-folksonology-semnet2007-crc.pdf>
- Flake, G. W.; Pennock, D. M.; Fain, D. C.** “The self-organized web: the yin to the semantic web’s yang”. En: *IEEE intelligent systems*, 2003, July/Aug, pp. 72–86.
- Fumero, A.; Roca G.; Encinar, J.** *Web 2.0*. Colección Fundación Orange, 2007. http://www.fundacionauna.com/areas/25_publicaciones/indice_web2.asp
- Gómez-Pérez, A.**, et al. *Ontological Engineering*. London: Springer, 2004.
- Matsuo, Y.**, et al. “Spinning multiple social networks for semantic web”. En: *AAAI-06*, 2006.
- Mika, P.** “Ontologies are us: a unified model of social networks and semantics”. En: *Proceedings of the 4th International semantic web conference (ISWC 2005)*, 2005. <http://www.cs.vu.nl/~pmika/research/papers/ISWC-folksonomy.pdf>
- Palacios, V.; Morato, J.; Sánchez-Cuadrado, S.; Lloréns, J.; Moreiro, J. A.** “An improved methodology for semantic scheme qualification”. En: *The 1st international workshop: semantic information integration on knowledge discovery (SIK 2006)*, 2006.
- Sánchez-Cuadrado, S.** *Definición de una metodología para la construcción automatizada de un sistema de organización del Conocimiento*. Tesis Doctoral Universidad Carlos III, 2007.
- Senso, José A.** “Navegadores semánticos o semantizar el navegador”. En: *ThinkEPI*, 2007. <http://www.thinkepi.net/repositorio/navegadores-semanticos-o-semantizar-el-navegador/>
- Steinbach, M.; Karypis, G.; Kumar, V.** *A comparison of document clustering techniques*. Technical Report 00-034, Department of Computer Science and Engineering University of Minnesota, 2000. <http://glaros.dtc.umn.edu/gkhome/node/157>
- Getting into rdf & semantic web using N3, W3C*. <http://www.w3.org/2000/10/swap/Primer>
- Xu, Z.**, et al. “Towards the semantic web: collaborative tag suggestions”. En: *WWW2006*, 2006. <http://www.rawsugar.com/www2006/13.pdf>
- Zeng, Marcia Lei; Mai Chan, Lois.** “Trends and issues in establishing interoperability among knowledge organization systems”. En: *Journal of the American Society for Information Science and Technology*, v. 55, n. 5, pp. 377 – 395

Jorge Morato; Sonia Sánchez-Cuadrado; Anabel Fraga; Valentín Moreno-Pelayo, Departamento de Informática, Universidad Carlos III, Av. Universidad, 30. 28911 Leganés (Madrid).
jorge@ie.inf.uc3m.es
ssanchez@ie.inf.uc3m.es
afraga@ie.inf.uc3m.es
vmoreno@ie.inf.uc3m.es

Suscripciones

Renovar (o comenzar) la suscripción a “El profesional de la información” es ágil y sencillo.

Usted puede gestionar online su suscripción conectándose a esta web:

<http://www.elprofesionaldelainformacion.com/suscripciones.html>

Si lo desea puede comunicar con nosotros dirigiéndose a:

suscripciones@elprofesionaldelainformacion.com

o al teléfono: +34-609 352 954