Metodología y Documentación Científica

Trabajo Final

Adam Westerski

1. **Titulo del trabajo a investigar.**
   Estudio de usabilidad en el desarrollo de aplicaciones y agregación de información (mashups) en la área de Linked Data y la Web Semántica.

2. **Introduccion.**

   Web semántica

   La definición de la web semántica y sus campos de aplicación ha evolucionado desde la contribución inicial de Tim Berners-Lee en 1998 [1]. La investigación ha progresado enormemente y se intentan adaptar los conceptos originales a la situación actual. La definición más simple del W3C [2] afirma que la *web semántica es una infraestructura basada en metadatos para posibilitar razonamiento sobre la web*. Aunque es una explicación sencilla, permite capturar los aspectos clave de este tema.

   La web actual ha sido creada para ser comprendida de una forma sencilla por parte de humanos. La iniciativa de la web semántica intenta proporcionar los medios necesarios para que las computadoras puedan comprender el contenido de la web. De acuerdo con el tutorial del W3C [2], la web semántica no busca desarrollar sofisticados algoritmos de inteligencia artificial. Aunque la inteligencia artificial podría tener cierta conexión, no es el tema principal de la web semántica. La idea clave es menos compleja y ambiciosa que crear máquinas inteligentes, sino que se busca que las máquinas razonen con el contenido de los recursos mediante el análisis de metadatos. Actualmente, la mayoría de la información almacenada en la web está almacenada en una forma que sólo es entendible por humanos y carece de orden alguno. Los usuarios utilizan lenguaje natural para describir el contenido y lo hacen en múltiples idiomas y maneras. La enorme diversidad de metadatos hace que esto sea muy complicado de analizar.

   Uno de los objetivos de la web semántica sobre el nivel de metadatos está en crear un método de descripción universal para anotar recursos, proponiéndose el uso de ontologías para lograr esta meta. Su tarea es definir conceptos y relaciones que se refieren a un determinado dominio.

   Dichas descripciones podrán ser usadas en un número de maneras diferentes para razonar sobre las relaciones entre recursos anotados, siendo un proceso que consiste en elegir dos recursos (posiblemente muy complejos) y ser capaces de determinar si son similares, partes entre sí, o quizá conectados de alguna otra manera. De hecho, éste es el significado real del proceso anteriormente comentado mediante el que las computadoras pueden entender el contenido de la web. Tal habilidad ofrece un amplio abanico de posibilidades a la web.

   Web de datos

   La web de datos, también conocida como web de datos enlazados, es el próximo paso en la evolución de la idea de web semántica. No pretende ser la tendencia prevista para la evolución de la web semántica, sino su estado actual.

   Existen muchos proyectos maduros de web semántica e incluso estándares ya publicados. En todos ellos se pretende construir la base de la nueva web, pero, por regla general, no son excesivamente populares ni ampliamente aceptados. Esta situación se suele achacar a la excesiva complejidad de las tecnologías o a los dominios de aplicación, en algunos casos poco realistas. La originariamente sobrevalorada inteligencia artificial para la web no se ha hecho realidad, por lo que se busca una solución más simple como paso previo para alcancar una nueva web inteligente. El primer paso es conocido como la web de los datos. Uno de los conceptos claves en esta tendencia consiste en proporcionar no sólo las anotaciones para la web, sino también los medios para interconectar los recursos anotados. El resultado de la actividad inicial de la web semántica es una situación en la que se dispone de metadatos y anotaciones de

recursos web, pero los datos no están interrelacionados y por lo tanto no pueden ser utilizados en la práctica. Ciertos sitios web o grupos de interés particulares tienen recursos anotados, pero fuera de estas comunidades la información no es utilizable o no se tiene interés por que así sea. Por ello, el éxito de la iniciativa de la web de datos está determinado por el grado de adopción de la comunidad de Internet; no únicamente grupos aislados, sino audiencia masiva. El objetivo está en introducir paradigmas fáciles de usar para la anotación de contenidos web [3][4] y aplicaciones funcionales que aprovechen dichas anotaciones (como motores de búsqueda semánticos [5][6][7][8] o portales basados en web semántica como Twine [9]).

En algunos casos, la web de datos se conoce como web 3.0 porque profundiza en la idea principal del éxito de la web 2.0: las redes sociales. Un gran número de las principales actividades de la web de datos pone más énfasis sobre los usuarios de Internet, las actividades de los usuarios y las sociedades virtuales nacidas en torno a la Internet moderna. Existe una gran discusión e investigación sobre anotaciones en entornos colaborativos que dominan el entorno de la Internet contemporánea. Al final, la meta sigue siendo la misma que la de la web de datos, de forma similar a la anteriormente mencionada web semántica, que pretende proporcionar una web más amigable para el usuario mediante contenidos comprensibles por computadoras y mediación entre máquinas y humanos.

Comunidades en línea semánticamente enlazadas (Semantically-Interlinked Online Communities, SIOC)

En el contexto de esta propuesta y de la aplicación de las tecnologías de la web semántica en las redes sociales, SIOC [10] es una iniciativa muy importante. Propone anotaciones de los recursos de la comunidad de Internet, tales como blogs, portales, listas de correo, etc. SIOC define cómo y dónde deben ser incluida la información automáticamente procesable. Específicamente se señala la manera de describir cada una de las mencionadas redes sociales. Además, proporciona los medios para interconectar todas las fuentes de información. Por ejemplo, con la ayuda de SIOC un mismo usuario puede ser identificado a través de muchos sistemas y vinculado a recursos tales como blogs y listas de correo al mismo tiempo. Finalmente, con el uso de los anteriormente mencionados motores de búsqueda semántica, las aplicaciones pueden extraer todos estos datos y presentarlos en un formato legible por humanos (por ejemplo después de un inicio de sesión de usuario en el foro puede ver su blog y las entradas de listas de correo de otros portales a través de todo Internet). SIOC ha sido publicado en el W3C [11].

Motores de búsqueda semánticos

La iniciativa SIOC proporciona anotaciones pero no los medios para organizarlas o usarlas. Actualmente, SIOC es principalmente utilizado en algunas comunidades seleccionadas, pero se da la circunstancia de que la cantidad de información es masiva incluso aun siendo producida por un número muy pequeño de fuentes. Es evidente que es imposible controlar dicho flujo de información con utilidades sencillas, por lo que para indexar, buscar y navegar eficientemente, se han propuesto diversos motores de búsqueda semántica.

En general esta rama de investigación de la web semántica se supone que busca proporcionar tecnologías del lado servidor que proporcionarían servicios de búsqueda similares como Google para la web actual. La diferencia es que los motores de búsqueda de la web semántica no pretenden ser utilizados por humanos, sino por otras aplicaciones. Muchos investigadores ven estos motores como el puente entre las aplicaciones de usuarios y la enorme cantidad de información almacenada en la web de datos [12]. Algunos de los motores de búsqueda semántica más populares son Sindice[5], Semantic Web Search Engine(SWSE)[6], Watson[7] y Falcons[8].

Redes sociales basadas en la web semántica

La iniciativa para crear aplicaciones de fácil uso que utilicen datos de la web Semántica está todavía en sus primeras fases. Sus aplicaciones se perciben como la última pieza que falta para hacer realidad la web de datos y la web semántica. Los elementos anteriormente mencionados (la iniciativa de la web de datos, los estándares de anotaciones como SIOC o los motores de búsqueda como Sindice) sólo construyen las bases. Salvo que a los usuarios de Internet se les proporcione auténtico valor, nunca aprovecharán las posibilidades de lo que en el mundo investigador se le pueda llamar web semántica.

Entre los numerosos productos y demostradores, existe una aplicación de redes sociales llamada Twine [9]: un portal social centrado en los datos. Al contrario que otros, no pretende construir una comunidad digital basada en las conexiones sociales existentes en la vida real (como es el caso de Twitter o Facebook), sino que anima a sus usuarios a crear redes sociales desde cero basados en determinadas áreas de interés. Twine utiliza datos publicados de la web semántica y también produce datos automáticamente procesables. Esto posibilita que los usuarios importen de forma semiautomática la información de sus intereses al portal, al tiempo que proporciona los medios para introducir los datos en otras aplicaciones y ser usada en otros entornos personales.
Aunque inicialmente ha sido criticada y considerada como sobrevalorada, Twine y otras aplicaciones continúan creciendo y pretenden dar un paso más en acercar la web semántica a los usuarios.

Herramientas para la creación de mashups utilizando información semántica

Las herramientas encontradas para la creación de mashups a partir de información semántica están todavía en una etapa inicial, siendo su desarrollo muy inmaduro todavía. Como la mayoría de las herramientas relacionadas con tecnologías semánticas, las aplicaciones son resultado de las investigaciones que se están realizando, muy lejos de ser herramientas comerciales para un usuario final.
La herramienta de este tipo más relevante es DERI Pipes:

DERI Pipes [13], una aplicación web inspirada en la herramienta Yahoo Pipes [14], que permite realizar diferentes transformaciones a partir de datos que se encuentran en formatos RDF, XML, JSON o microformatos. Durante la creación de un mashup utilizando esta herramienta, el usuario puede conectar las diferentes fuentes haciendo uso de operadores que transformarán los datos para producir la salida deseada. Dependiendo del formato de los datos importados se utilizarán diferentes lenguajes de consulta (SPARQL o XQUERY). De una forma parecida al mencionado Yahoo Pipes, todas las interacciones se realizan de una forma gráfica: arrastrando y conectando los diferentes componentes. Los mashups creados por los usuarios pueden ser salvados en el servidor de DERI Pipes para poder ser consumidos como un servicios REST. También permite la opción de exportarlos para su ejecución local a través de una consola de comandos.

A parte de DERI Pipes, siendo la referencia en este tipo de herramientas, se pueden encontrar otras iniciativas como Banach [15] y SPARQLMotion [16]. La primera de ellas ha sido creada como parte del proyecto SMILE, extendiendo las capacidades del almacén de datos Sesame implementando diversos operadores que pueden ser combinados con la información de Sesame.

SPARQLMotion, en cambio es más parecido a DERI Pipes, sin embargo, esta herramienta es un producto comercial. La gran diferencia con DERI Pipes, es que SPARQLMotion es una aplicación de escritorio que puede ser integrada con aplicaciones de la mismas compañía.

3. Aspectos de trabajo de investigación.

3.1.Objetivos.

- reconocer las fuentes más populares de Linked Data
- reconocer el tamaño de la nube de Linked Data y medir su ritmo de crecimiento
- proporcionar medidas de frecuencia de uso de los diferentes tipos de Linked Data
- proporcionar un estudio de los entornos y las utilidades que son usadas para la creación de mashups
- reconocer las áreas en que se construyen mashups y los objetivo que cumplen
- reconocer los mashups más exitosos y las aplicaciones de la Web Semántica que funcionan al margen del campo
- proponer mejoras en el desarrollo de mashups

3.2.Hipótesis.

A pesar de las enormes cantidades de Linked Data y de la actual tendencia a producir mas, las aplicaciones de Web Semántica actual no puede usar el potencial de Linked Data. La razón de este estado no es la falta de los datos sino la falta de aplicaciones prácticas.

3.3.Método experimental.

Investigar aplicaciones actual y orígenes del datos y analizarlos problemas mas frecuentes. Comparar metadatos de Linked Data (y Web Semántica) con otros esquemas de anotaciones Web e iniciativas de metadatos exitosos (p.e. RSS en comparación con SIOC ontológica; REST vs. WSMO/SOA). Señalar la diferencia y enumerar las ventajas y las desventajas. Presentar los aspectos de Linked Data que sólo se pueden utilizar en teoría y no funcionan en la práctica. Medir el tamaño de las nubes de Linked Data y mirar qué tipos de los datos crecen más rápido, y cuáles son más populares. Las variables sobre el tamaño y el crecimiento de Linked Data se pueden medir con ayuda de aplicaciones en línea como motores de búsqueda semánticos y servicios de información como pingthesemanticweb.com.

3.4.Variables.

- Tamaño de las nubes de Linked Data (cantidad de triplas)
- Cantidad del proyectos/iniciativas exitosos de Web Semántica o Linked Data usados al margen del campo (p.e. en el contexto de proyectos que son presentado en áreas seleccionadas - conferencias mas importantes para Web Semántica u otras)
- Grado de crecimiento de Linked Data (semanal, mensual)

3.5.Cronograma.

Fase	Actividades	Tiempo aproximado
Análisis inicial	Búsqueda y análisis de literatura sobre estudio de usabilidad	2 meses
	Determinar las principales líneas de investigación en el área de Linked Data y Web	2 meses

Fase	Actividades	Tiempo aproximado
	Semántica	
	Revisión bibliográfica en cada área de interés	2 meses
	Análisis del estado del arte en cada área de interés	2 meses
	Análisis del estado del arte y análisis bibliográfico en áreas similares (sobre metadatos)	1 mes
Ejecución	Selección de nubes de Linked Data representativas en diferentes áreas (p.e. Dbpedia, SIOC metadata)	1 meses
	Diseño de una manera de observar el crecimiento de los datos y métricas para describirlo apropiadamente	1 mes
	Observar cambios de datos	2 meses
Análisis	Análisis de nubes de Linked Data y observación de resultados	1 mes
	Análisis de usabilidad de metadatos en las actuales nubes de Linked Data basado en comparación de trabajos relacionados en diferentes dominios y Web Semantica/Linked Data	1 mes
	Revisión de aplicaciones creadas en diferentes áreas que usan metadatos	1 mes
	Determinar las carencias de las aplicaciones y las necesidades de los usuarios	1 mes
	Determinar la relación entre la cantidad (u otras características) de datos y los problemas de la usabilidad de las aplicaciones de Web Semántica	1 mes
Conclusiones y resultados	Evaluación del conocimiento extraído	1 mes
	Proponer cambios y aplicaciones	1 mes

3.6.Búsquedas.

1. IEEE library - “usability study” búsqueda
Link: http://ieeexplore.ieee.org/search/searchresult.jsp?SortField=Score&SortOrder=desc&ResultCount=25&maxdoc=100&coll1=ieeejrns&coll2=ieejrns&coll3=ieeecnfs&coll4=ieecnfs&coll5=ieeestds&coll6=preprint&coll7=books&coll8=modules&srchres=11&history=yes&queryText=((usability+study)%3CIN%3Emetadata)&oldqrytext=((usability+web)%3Cin%3Emetadata)&imageField.x=0&imageField.y=0&imageField=((usability+web)%3Cin%3Emetadata)&radiobutton=cit

Resultados interesantes:
- Ramalingam, S.; Iourinski, D.; “Using fuzzy functions to aggregate usability study data: a novel approach”
- Ricks, K.; Arnoldy, B.A.; “How to conduct your own usability study”
- Iourinski, D.; Ramalingam, S.; “Using Dempster Shafer theory to aggregate usability study data”
- Rosenbaum, S.; Kantner, L.; “Learning about users when you can’t go there: Remote attended usability studies”

2. Springer Link - “semantic web usability” búsqueda
Link: http://www.springerlink.com/content/?k=semantic+web+usability

Resultados interesantes:
- Ahmed Seffah, Michel Desmarais and Eduard Metzker, “HCI, Usability and Software Engineering Integration: Present and Future”
- Anthony Jameson , “Usability and the Semantic Web“

3. Google - “semantic web problems” búsqueda
Link: http://www.google.pl/search?hl=pl&q=semantic+web+problems&btnG=Szukaj&lr=

Resultados interesantes:
- “Problems of the Semantic Web”
  (www.semanticfocus.com/blog/entry/title/5-problems-of-the-semantic-web/)
- “Semantic Web: Difficulties with the Classic Approach – ReadWriteWeb”
  (www.readwriteweb.com/archives/semantic_web_difficulties_with_classic_approach.php)
- “Semantic Web roadmap”
  (www.w3.org/DesignIssues/Semantic.html)
- “Layering the Semantic Web: Problems and Directions”
  (204.178.16.26/who/pfps/publications/layering.pdf)

4. Google - “usability study in Semantic Web” búsqueda
Link: http://www.google.pl/search?hl=pl&q=semantic+web+usability+study&btnG=Szukaj+w+Google&lr=&aq=f&oq=

Resultados interesantes:
- “The Semantic Puzzle | Usability”
  (blog.semantic-web.at/tag/usability/)
- “User-Centered Design for the Semantic Web”
  (www.dfki.de/~jameson/iswc07-tutorial/ISWC07-tutorial-UCD.pdf)
- “A Flexible Dialogue System for Enhancing Web Usability”
  (http://data.semanticweb.org/conference/www/2009/paper/164/html)

- “Why Evaluating Semantic Web Applications is Difficult”
  (swui.webscience.org/SWUI2008CHI/vanOssenbruggen.pdf)

3.7.Bibliografía.

[1] T. Berners-Lee, Semantic Web Road map, (1998).
[2] I. Herman, Tutorial on Basic SW Technologies, (2004).
[3] RDFa Primier, http://www.w3.org/TR/xhtml-rdfa-primer/
[4] R. Cyganiak, H. Stenzhorn, R. Delbru, S. Decker, G. Tummarello, Semantic Sitemaps: Efficient and Flexible Access to Datasets on the Semantic Web. In Proceedings of the Proceedings of the 5th European Semantic Web Conference 2008.
[5] E. Oren, R. Delbru, M. Catasta, R. Cyganiak, H. Stenzhorn, G. Tummarello, Sindice.com: A document-oriented lookup index for open linked data. In International Journal of Metadata, Semantics and Ontologies, 3(1), 2008.
[6] A. Harth, A. Hogan, J. Umbrich, S. Decker, Building a Semantic Web Search Engine: Challenges and Solutions. In Proceedings of the 3rd XTech Conference, 2008, Dublin, Ireland, 2008.
[7] M. d’Aquin, M. Sabou, M. Dzbor, C. Baldassarre, L. Gridinoc, S. Angeletou, and E. Motta. Watson: A Gateway for the Semantic Web. Poster session of the European Semantic Web Conference, ESWC 2007.
[8] Y. Qu, G. Cheng, H. Wu, W. Ge, X. Zhang, Seeking Knowledge with Falcons. In Semantic Web Challenge, 2008
[9] Twine website, http://www.twine.com/
[10] SIOC project website, http://sioc-project.
[11] SIOC W3C submission website, http://www.w3.org/Submission/2007/02/
[12] M. d'Aquin. Building Semantic Web Based Applications with Watson. Developers' Track of the WWWW 2008 conference, Beijing China
[13] DERI Pipes website, http://pipes.deri.org/
[14] Yahoo Pipes website, http://pipes.yahoo.com/pipes/
[15] Simile Banach project website, http://simile.mit.edu/wiki/Banach
[16] Sparql motion website, http://www.topquadrant.com/sparqlmotion/