

Tecnologías para la Web Semántica

Trabajo Práctico N°2

La Web Semántica

Darién Julián Ramírez

1. Ejercicio 1

Lea el artículo "The Semantic Web" de Tim Berners-Lee, James Hendler y Ora Lassila. Responda las siguientes preguntas:

- ¿Por qué consideran los autores que en la actualidad las computadoras no tienen capacidad de procesar la semántica de manera confiable? ¿Qué aporte tendrá la Web Semántica en este sentido?

La mayor parte del contenido de la Web hoy, es diseñado para que los seres humanos lo lean, no para que los programas de computadoras puedan manipularlo de manera significativa. Los ordenadores pueden analizar las páginas web con facilidad para el diseño y las rutinas "poner un encabezado aquí." "hay un enlace a otra página", pero en general, las computadoras no tienen manera de procesar la semántica.

La Web Semántica traerá una estructura para el contenido significativo de las páginas web, creando un entorno en el que los agentes de software página a página puedan llevar a cabo sofisticadas tareas para los usuarios. Tal agente no sólo sabrá que en la página hay palabras clave como "tratamiento, medicina, físico, terapia", sino también que el Dr. Hartman trabaja en esta clínica los lunes, miércoles y viernes, y que el script tiene un intervalo de fechas en formato aaaa-mm-dd y devuelve el horario de las citas. Y conocerá todo esto sin necesidad de inteligencia artificial a la escala de Hal o la de C-3PO.

- Según los autores, ¿Cuál es la propiedad esencial de la WWW? ¿Cuál es la de la Web Semántica?

La propiedad esencial de la World Wide Web es su universalidad. El poder de un enlace de hipertexto es que cualquier cosa se puede vincular a cualquier cosa. La tecnología web, por lo tanto, no debe discriminar entre el garabateado y el pulido rendimiento, entre información comercial y académica, o entre culturas, idiomas, medios de comunicación, etc.

La información varía a lo largo de muchos ejes. Uno de ellos es la diferencia entre la información producida principalmente para el consumo humano y la principalmente dirigida a las máquinas. En un extremo de la escala, tener todo, desde el televisor a la poesía. En el otro extremo tenemos bases de datos, programas y salida de sensores.

Hasta la fecha, la web se ha desarrollado más rápidamente como medio de documentación para las personas en lugar de para datos e información que se puede procesar automáticamente. La Web Semántica pretende compensar esto.

Al igual que Internet, la Web Semántica será lo más descentralizado posible.

- Según se muestra en el ejemplo desarrollado en el artículo, ¿Cómo se lograría obtener la correcta combinación de información distribuida en diferentes páginas web? Ejemplifique.

La correcta combinación de información distribuida en las diferentes páginas web se lograría mediante agentes semánticos.

Se le pide al agente una consulta con un doctor a través del navegador web. Los agentes recuperan información del tratamiento de la persona otorgado por los agentes del médico, buscan información de proveedores, y se analizan los que están en plan de seguros de la persona dentro de un radio de 20 millas de su casa y con una calificación excelente o muy buena en los servicios de calificación confiables. Entonces comienzan a tratar de encontrar un turno entre los horarios de citas disponibles (suministrados por los agentes de los proveedores individuales a través de sus sitios web).

- ¿Qué tecnologías semánticas (lenguajes y herramientas) menciona el texto?

Dos tecnologías importantes para el desarrollo de la Web Semántica ya están funcionando: eXtensible Markup Language (XML) y el Resource description Framework (RDF).

El XML permite a todo el mundo crear sus propias etiquetas ocultas como que se anotan en las páginas Web o en secciones de texto en una página. Los scripts, o programas, pueden hacer uso de estas etiquetas en formas sofisticadas, pero el lector de scripts tiene que saber qué el para que usa cada etiqueta el escritor de scripts. XML permite a los usuarios agregar estructura arbitraria a sus documentos, pero no dice nada sobre la definición de las estructuras.

El significado es expresado por RDF, que codifica en conjuntos de triples, siendo cada triple como el sujeto, el verbo y objeto de una oración elemental. Estos triples se pueden escribir usando etiquetas XML. En RDF, un documento hace afirmaciones de qué cosas (personas, páginas web o lo que sea) tienen propiedades (como *es una hermana de*, *es el autor de*) con ciertos valores (otra persona, otra página Web). Esta estructura resulta ser una forma natural de descripción de la gran mayoría de los datos procesados por máquinas, el sujeto y el objeto son identificados por un identificador de recurso universal (URI), tal como se utiliza en un enlace de una página Web (URL, Uniform Resources Locators, son el tipo más común de URI). Los verbos también son identificados por URI's, lo que habilita a cualquiera a definir un nuevo concepto, un nuevo verbo, definiendo un URI para ello en algún lugar de la Web.

- Según su opinión, ¿Cómo pueden estas tecnologías semánticas ayudar a coordinar, descubrir y organizar la información y el conocimiento?

El hecho de que estas tecnologías permitan la utilización de etiquetas que referencien el contenido de la página y permitan establecer diferentes tipos de relaciones con ellas ayuda a que las búsquedas no se bases sólo en la coincidencia de palabras clave si no de encontrar páginas con contenido relevante a la búsqueda en cuestión y mostrarla ordenada mediante algún criterio en particular.

- Explique su punto de vista acerca de la practicidad y futuro de estas herramientas y procedimientos.

Las estrategias que utilizan estas herramientas es buena y promete buenos resultados en una primera instancia pero para lograr establecer una web semántica como la que se describe en el artículo será necesario innovar las herramientas y procedimientos existentes y crear nuevos métodos.

2. Ejercicio 2

Lea la entrevista a Tim Berners-Lee en la revista BusinessWeek. Responda las siguientes preguntas:

- ¿Por qué considera Berners-Lee que el gobierno fue uno de las primeras instituciones en adoptar la filosofía de la Web Semántica?

DARPA (Defense Advanced Research Projects Agency) tuvo sus propios problemas graves con grandes cantidades de datos de todas las fuentes diferentes sobre todo tipo de cosas. Así, vieron la Web Semántica justamente como algo dirigido directamente a resolver los problemas que tenían en un a gran escala. DARPA entonces financió algunos de los primeros desarrollos.

- ¿Cómo considera Berners-Lee que la Web Semántica puede impulsar el descubrimiento de la cura para enfermedades? En su opinión, ¿esto es posible?

Cuando una compañía farmacéutica mira una enfermedad, toman los síntomas específicos que están conectados con proteínas específicas dentro de una célula humana. Así que el arte de encontrar la droga es encontrar el producto químico que interferirá con las cosas malas que suceden y animan las cosas buenas que suceden dentro de la célula, que implica comprender la genética y todas las conexiones entre la proteínas y los síntomas de la enfermedad. También requiere mirar todas las otras conexiones, ya sea las regulaciones sobre el uso de la proteína y cómo se ha utilizado antes. Se tiene la información reguladora gubernamental, los datos de los ensayos clínicos, los datos genómicos y proteómicos que están todos en diferentes departamentos y diferentes piezas de software. Un científico que está pasando por ese proceso creativo de lluvia de ideas para encontrar algo que posiblemente podría resolver la enfermedad tiene que de alguna manera mantener todo en su cabeza en al mismo tiempo o ser capaz de explorar todos estos diferentes ejes de una manera conectada. La Web Semántica es una tecnología diseñada para hacer específicamente eso: abrir fronteras entre los silos, para permitir a los científicos explorar las hipótesis, las cosas se conectan en nuevas combinaciones que nunca antes se habían soñado.

En mi opinión, si es posible implementar la web semántica, es posible realizar toda esta organización relacionada con la cura de enfermedades. Por lo tanto, no me parece una idea muy descabellada, además las palabras son claras "...puede impulsar el descubrimiento..."

- ¿Cómo resume Berners-Lee el lenguaje de la Web Semántica?

El lenguaje de la Web Semántica, en su corazón, es muy, muy simple. Se trata sólo de las relaciones entre las cosas.

3. Ejercicio 3

Investigue sobre buscadores semánticos. Ejemplifique.

Un buscador semántico es aquél que realiza el rastreo atendiendo al significado del grupo de palabras que se escriben y no basándose en las actuales etiquetas. En pocas palabras, un buscador inteligente. Aunque todavía su uso es incipiente, ya existen algunos que ofrecen resultados reseñables e interesantes y ya son referencia para el futuro de la búsqueda de información: WOLFRAMALPHA, SWOOTTI, IDEAS AFINES, BUSCADOR EXPERIMENTAL DEL BOPA, HAKIA, LEXXE, KARTOO, RETRIEVR, MNEMOMAP, QUINTURA, BLINKX, SWOOGLE, GNOSS, DUCK DUCK GO, MEANING TOOL, NATURAL FINDER, ASKWIKI, POWERSSET, SEMANTIC WEB SEARCH, ETC.

Referencias

- [1] Berners-Lee T., Hendler J., Lassila O., *The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, Scientific American, 2001.
- [2] *Entrevista Q&A with Tim Berners Lee*, BusinessWeek, 2007