

INTRODUCCIÓN

La Web, un sistema de documentos de hipertexto enlazados contenidos en Internet ha marcado un nuevo hito en la historia. En 2008 ingenieros de Google¹³ anunciaron que el popular motor de búsqueda había descubierto 1 billón de URLs únicas, si miramos las estadísticas actuales¹⁴, en abril de 2010 el número estimado de páginas Web contenidas en la Web es de por lo menos 20 mil millones. Todo esto se suma a la "Web 2.0" cambiando la forma en que la gente maneja información e ingresa contenidos nuevos.

Los contenidos relevantes de la Web son aquellos que los usuarios considere relevantes, es por esto que los motores de búsqueda más populares indexan páginas y de acuerdo a palabras claves usadas por el usuario recuperan información. Estos resultados en muchas veces no concuerdan con el significado que realmente el usuario les quería dar. Esto pasa por que las consultas están compuestas por palabras que sólo la gente puede entender, pero a las que los computadores no pueden acceder u obtener el contenido. Todo esto se mejora ayudando a los computadores de alguna forma a manejar los datos efectivamente.

Tim Berners-Lee sugiere una Web semántica para hacer entender a los computadores (procesable por programas) como a la gente, el significado de los contenidos por medio de un significado general que se pueda intercambiar en todos los contenidos de la Web. Esto es, el uso de una ontología "Una especificación explícita de conceptualización"¹⁵ que defina el significado y las relaciones de conceptos específicos.

¹³ "We knew the web was big". Disponible en: <http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>. Consultado: Abril 2010.

¹⁴ "The size of the World Wide Web". Disponible en: <http://www.worldwidewebsite.com/> Consultado: Abril 2010.

¹⁵ Gruber, T. "A Translation Approach to Portable Ontology Specifications" Knowledge Acquisition, vol. 5, no. 2, 1993, p. 199-220.

Reconociendo entonces la importancia y necesidad de una Web semántica y de las ontologías, aparecen varios estudios que consideran y proponen diferentes métodos para recuperar información basados en ontologías y de esta manera hacer de la Web semántica una realidad. Toda esta información se encuentra distribuida en diferentes sitios y para poder empezar a desarrollar alguna tecnología nueva en este campo es preciso investigar ampliamente todos los temas relacionados. Queremos ahorrar un tiempo grande en esta búsqueda de conocimiento y con tal objetivo definido, presentamos este trabajo.

Se empezará introduciendo la red semántica y las ontologías, así como las metodologías para construirlas y sus usos. Se mostrarán algunos lenguajes utilizados en ontologías. Principalmente haremos un estudio primario sobre la literatura existente en relación con las técnicas de recuperación de información basadas en ontologías, se explorarán las diferentes técnicas de recuperación que se basan en algoritmos complejos, ranking, personalización de usuario y expansión de consultas.

1. OBJETIVOS

1.1 OBJETIVO GENERAL

Presentar el estado del arte sobre técnicas utilizadas para la recuperación de información basadas en ontologías.

1.2 OBJETIVOS ESPECÍFICOS

- Identificar diferentes lenguajes de búsqueda existentes, utilizados por el usuario para la recuperación de información basada en ontologías.
- Identificar los diferentes lenguajes utilizados en la representación de ontologías, usados para realizar búsquedas en ontologías.
- Proporcionar una preparación integral sobre el actual estado del arte en técnicas de recuperación de información basadas en ontologías.
- Identificar los modelos formales de redes semánticas más utilizados en la recuperación de búsquedas basadas en ontologías.
- Identificar los debates y problemas no resueltos en la frontera del conocimiento que incluye las técnicas de recuperación de información basadas en ontologías.