



UNIVERSIDAD CATÓLICA ANDRÉS BELLO  
FACULTAD DE INGENIERÍA  
ESCUELA DE INGENIERÍA INFORMÁTICA

# Desarrollo de una ontología y un buscador utilizando tecnologías de Web Semántica

TRABAJO ESPECIAL DE GRADO  
presentado ante la  
UNIVERSIDAD CATOLICA ANDRES BELLO  
como parte de los requisitos para optar al título de  
INGENIERO EN INFORMÁTICA

REALIZADO POR **Israel Fermín Montilla**

PROFESOR GUIA **Wilmer Pereira**

FECHA Caracas, septiembre de 2011

# AGRADECIMIENTOS

...

A mi tutor, Dios, Familia, escuela, jaladita de bola a la universidad y de más cursilerías que uno pone aquí

...

*Israel Fermín Montilla*

# SINOPSIS

Toda la paja en una página.

# INTRODUCCIÓN

Copiar y pegar lo mismo de la propuesta

## PROBLEMA

### 1.1. Planteamiento del Problema

La World Wide Web (WWW), ha cambiado radicalmente la manera como las personas se comunican entre sí, la forma como la información se distribuye y como se diseminan los mensajes y los modelos de negocio de muchas empresas alrededor del mundo[1]. Basta con revisar las páginas web de las grandes empresas a nivel nacional e internacional, como Polar, Apple Computer o Microsoft para darse cuenta de que ya no se enfocan sólo en tener presencia en radio y televisión, sino también en las distintas redes sociales que han surgido a través de la Web 2.0.

Ciertamente Internet, desde su aparición en 1975, ha evolucionado, ya los sitios web no son páginas con texto estático e imágenes en las que sólo un subconjunto de los usuarios es capaz de publicar contenido, mientras que otros sólo leen y reciben información. Ahora se cuenta con páginas y aplicaciones web complejas, en las que todos los usuarios están en capacidad de generar y publicar contenido. Esto se conoce como Conocimiento Colectivo [2], de esta manera,

con cada vez más usuarios publicando contenido en la web, la cantidad de información ha crecido exponencialmente, ya para 2009, el estimado de usuarios de Internet en América Latina y el Caribe era de 175,8 millones de personas aproximadamente [3]. Lo que implica un enorme volumen y tráfico de datos en la WWW pues todos esos usuarios intercambian, publican y consultan información permanentemente.

Todos estos usuarios poseen necesidades de información esperando a ser satisfechas. La información para satisfacer a todos y cada uno de los usuarios, está disponible en línea, pero resulta difícil acceder a ella si no se sabe dónde se encuentra, es decir, si no se conoce su URL. Es por ello que fueron creados los buscadores como Yahoo, Google, Altavista y Bing, por mencionar algunos de los más populares todos estos funcionan buscando las palabras claves que proporciona el usuario en los documentos que ha indexado, es decir, son buscadores basados en palabras clave. De esta manera, si el usuario introduce, por ejemplo, “Internet”, el buscador retornará todos los documentos conocidos que contienen esa palabra, bien sea en el título o en el cuerpo del texto. Pero cuando la búsqueda involucra más de una palabra, las cosas pueden complicarse un poco, por ejemplo: si el usuario introduce “Internet de Venezuela” los buscadores basados en palabras clave omiten los conectores pues aparecen en todos los documentos y sitios web, por lo que la búsqueda sería “Internet Venezuela” y el resultado de la búsqueda contendría todos los recursos en los que aparezcan ambas palabras o alguna de las dos.

Pero qué ocurre cuando el usuario desea realizar búsquedas más especializadas, por ejemplo “Aerolíneas que viajan a Valera”, probablemente el buscador nos de lo que estamos buscando, pero requiere un trabajo adicional de búsqueda por parte del usuario pues los resultados obtenidos contendrían páginas de “Aerolíneas”, páginas que contienen la palabra “viajan”, y noticias sobre “Valera”, localidad del Estado Trujillo, o sobre personas con el apellido “Valera”. Esta resulta ser la principal desventaja de los buscadores actuales basados en palabras clave o “key words”, cuando se desea hacer una búsqueda basada en un tópico específico o enmarcada en un contexto dado, los resultados de la búsqueda no son muy cercanos a lo que el usuario desea buscar pues los buscadores no son capaces de interpretar el significado detrás de las palabras clave que componen la consulta realizada.

Casos como el anteriormente expuesto se ven a diario entre los estudiantes de carreras afines a las Ciencias de la Computación, como lo es la Ingeniería Informática. Muchas veces se trata de buscar información que aparece en la web bajo otro nombre o, por ejemplo, al buscar conocimientos complejos como “cálculo de rutas óptimas”, se consiguen resultados muy avanzados para el nivel de experiencia que se tiene en esa área, y no se desglosa los resultados en tópicos que son necesarios para dominar el objeto de la búsqueda, por ejemplo, resultados sobre el manejo de matrices de adyacencia y resultados sobre algoritmos ya existentes que tienen dicha utilidad como Dijkstra y Bellman-Ford. De esta manera, el estudiante puede tener una guía para poder atacar el tópico de su interés.

Por todo lo dicho con anterioridad, se plantea el desarrollo de una aplicación con capacidad semántica que sirva como herramienta de consulta en materia de Ciencias de la Computación e Ingeniería Informática.

Resulta de interés resaltar que todo lo relativo a la Web Semántica y sus estándares y tecnologías, se encuentran actualmente en definición y en proceso de desarrollo, por ello, se debe tener presente que la Web Semántica se encuentra en un nivel experimental, de igual manera que el desarrollo de este trabajo.

## 1.2. Objetivo General

- ➡ Desarrollar una ontología y un prototipo de buscador sobre dicha ontología utilizando tecnologías y estándares de Web Semántica.

## 1.3. Objetivos Específicos

1. Construir una Base de Conocimientos.
2. Evaluar de manera cualitativa distintos Motores de Inferencia y seleccionar el que mejor se adapte a las necesidades del proyecto.
3. Definir las reglas de inferencia a ser aplicadas sobre el vocabulario creado.
4. Construir una infraestructura básica de búsqueda para realizar las pruebas necesarias sobre la ontología.
5. Desarrollar las interfaces necesarias para realizar consultas y mostrar los resultados.
6. Implementar un procedimiento semiautomático para la evolución de la ontología.

## 1.4. Alcance

Desarrollar una ontología y un buscador con capacidad semántica para realizar búsquedas utilizando ontología desarrollada.

## 1.5. Limitaciones

- ➡ La ontología se realizará únicamente sobre tópicos referentes al área de Ciencias de la Computación e Ingeniería en Informática.



- ▣ Debido a la enorme cantidad de temas existentes dentro del área de Ciencias de la Computación e Ingeniería Informática, se cubrirá el dominio de temas en extensión.
- ▣ En general, se profundizará a tres (3) niveles en todos los temas cubiertos.
- ▣ Dado que el vocabulario a ser utilizado no es estándar sino que es producto del presente trabajo, la realización de búsquedas masivas en Internet no será un tema pertinente a este trabajo.

## 1.6. Justificación

En la última década (2000 – 2010), la cantidad de usuarios de Internet a nivel mundial se ha incrementado en un 446% [3], esto da una idea acerca de la importancia que tiene la World Wide Web como herramienta de comunicación, búsqueda e intercambio de información entre las personas en la actualidad. La WWW, ha cambiado la manera como las personas interactúan entre si, extendiendo la forma también como se presta servicio de educación, mediante plataformas de e-learning. La web ha evolucionado y se ha vuelto tan confiable que en la última década la tendencia de muchas grandes empresas ha sido "subir todo a la nube", es decir, aprovechar plataformas de Cloud Computing para tener sus servicios accesibles en cualquier momento y en cualquier parte del mundo.

Dada la importancia de la WWW a nivel mundial, día a día se ven iniciativas para hacer más placentera la experiencia de los usuarios en la red, así como nuevas tecnologías y estándares para optimizar y mejorar su funcionamiento, al punto que gracias a tecnologías como el Really Simple Syndication (RSS), no es necesario visitar constantemente un sitio web para leer sus actualizaciones u obtener el último episodio de nuestro Podcast favorito. Todo esto no sólo le hace la vida más fácil al usuario, sino que optimiza el funcionamiento de la web.

La Web Semántica es una iniciativa del World Wide Web Consortium (W3C) que pretende "extender la web, dotándola con un mayor significado para que cualquier usuario pueda conseguir respuesta a sus preguntas en Internet de manera rápida y sencilla" [4]. La Web Semántica viene a resolver el problema de acceso a la información en internet, ya que al añadirle semántica a la Internet, se puede buscar, transferir y compartir información de una manera más sencilla, permitiéndole al usuario, delegar estas tareas en software de manera confiable.

Por todo lo anteriormente mencionado, cualquier proyecto o investigación destinado al mejoramiento y optimización de la WWW y, más aún, dentro del campo de la Web Semántica, tiene pertinencia en el presente, ya que es un tema novedoso y que es objeto de investigación en muchas Universidades importantes alrededor del mundo. Desde el año 2001, cuando Tim Berners-Lee escribió un artículo para la revista Scientific American describiendo las posibilidades de la

Web Semántica, muchos grupos de investigadores han tomado este tema como tópico de investigación. Actualmente, en Venezuela, se llevan a cabo numerosas investigaciones en este campo, siendo la Universidad Simón Bolívar (USB) la que presenta mayor actividad, teniendo un grupo de investigadores dedicados a este tema, entre los que destacan la Prof. María Esther Vidal y la Prof. Soraya Abad-Mota, ambas especialistas y ponentes internacionales en Web Semántica.

En la Universidad Católica Andrés Bello (UCAB), no existe ninguna investigación en esta área, lo que da aún más pertinencia a un Trabajo de Grado en esta área, pues permitiría a la UCAB incursionar en este campo, contribuyendo al desarrollo de una tecnología de software nueva y a nivel internacional y abriría las puertas a una nueva línea de investigación dentro de la Escuela de Ingeniería Informática de la UCAB, permitiendo la realización de nuevos trabajos más avanzados acerca de este tema.

Para este proyecto, no sólo será necesario implementar estándares ya creados pues, como se explicará más adelante, casi la totalidad de los estándares y tecnologías que construyen la Web Semántica, aún cuando ya están siendo utilizados, se encuentran en desarrollo todavía. Será necesario también elaborar un procedimiento de traducción de la consulta del usuario a lenguaje de consultas sobre RDF (SPARQL), este vendría siendo, si se quiere, el elemento de mayor dificultad para lograr el buen funcionamiento y desempeño del prototipo de buscador que se desea desarrollar. Además, el desarrollo de un proceso de inclusión de nuevos conceptos a la ontología añade un factor más de dificultad que hace que sean necesarias las habilidades y conocimientos de un Ingeniero en Informática para resolver problemas que no sólo tienen que ver con programación o codificación de algoritmos (que es una de las capacidades del Ingeniero en Informática), sino con la definición y construcción de los mismos para resolver un problema computacional específico de manera eficaz y eficiente.

Por otra parte, el desarrollo de una ontología acerca en materia de las Ciencias de la Computación e Ingeniería Informática, permitiría organizar el conocimiento disponible en esos campos y, además, disponer de un repositorio dónde consultar dicho conocimiento, con el valor agregado de que las búsquedas se realizarían con sentido semántico, lo que da una mayor seguridad al usuario de que los resultados del buscador tienen coherencia con la consulta realizada

y la información deseada. El producto de este trabajo, podría implantarse como una herramienta de consulta electrónica para los estudiantes de carreras afines a la Ingeniería Informática en la UCAB. Posteriormente, la herramienta, podría ser extendida a otros temas mediante otros trabajos similares e incluso, podría portarse a otras carreras o a otras universidades.

## MARCO TE'ORICO

Copiar y pegar lo mismo de la propuesta, pero extender

## MARCO METODOLÓGICO

Copiar y pegar lo mismo de la propuesta

## DESARROLLO

Desarrollo de la tesis, debería reflejar el cumplimiento de los objetivos

## CAPÍTULO 5

# RESULTADOS

Exponer los resultados del desarrollo



## CONCLUSIONES Y RECOMENDACIONES

### 6.1. Conclusiones

...

### 6.2. Recomendaciones

...