

# GENERACIÓN DE CONOCIMIENTO MEDIANTE EL ENFOQUE DE LA WEB SEMÁNTICA

**Dr. Víctor del Carmen Avendaño Porras**

Centro Regional de Formación Docente e Investigación Educativa

[victor.avendano@cresur.edu.mx](mailto:victor.avendano@cresur.edu.mx)

**Mtra. Iris Alfonso Albores**

Instituto de Estudios Superiores de Chiapas

[psic\\_iaa@hotmail.com](mailto:psic_iaa@hotmail.com)

## RESUMEN

En este trabajo se aborda el problema de la búsqueda de información en internet. Se estima que habrá alrededor de 25 mil millones de dispositivos conectados a Internet en 2015 y 50 mil millones en 2020 (Evans, 2011). Se necesita un número tan impresionante de dispositivos altamente distribuidos y heterogéneos que estén interconectados y en comunicación en distintos escenarios de forma autónoma.

Esto implica que proporcionar interoperabilidad entre los datos que se encuentra en la WEB es de gran importancia y uno de los requisitos más fundamentales para encontrar información direccionada. El conjunto de tecnologías desarrolladas en la Web Semántica, como las ontologías y los servicios web semánticos pueden ser utilizadas como soluciones principales para los fines de la generación de nuevo conocimiento.

## PALABRAS CLAVE

Web semántica, áreas del conocimiento, base de conocimiento, internet.

## ABSTRACT

This paper addresses the problem of finding information on the Internet. It is estimated to be approximately 25 billion devices connected to the Internet by 2015 and 50 billion in 2020 (Evans, 2011). An impressive number of highly

distributed and heterogeneous devices that are networked and communicate in different scenarios autonomously required.

This implies that provide interoperability between the data found on the web is very important and one of the most fundamental requirements to find address information. The set of technologies developed in the Semantic Web, including ontologies and semantic web services can be used as main solutions for the purpose of generating new knowledge.

## **KEY WORDS**

Semantic Web, knowledge areas, knowledge base, internet.

## **INTRODUCCIÓN**

La extensión de Internet con objetos interconectados, físicos o virtuales ha sido una tendencia creciente en los últimos años. Esto ha creado una gama de potenciales nuevos productos y servicios en muchos ámbitos diferentes como los hogares inteligentes, la salud electrónica, la automoción, el transporte y la logística, así como la vigilancia del medio ambiente.

Las investigaciones en todas las áreas del conocimiento han cobrado impulso recientemente y están apoyadas por la colaboración de la academia, la industria y los organismos de normalización en varias comunidades como las telecomunicaciones, la web semántica y la informática (Penrose, 1991).

Un objetivo primordial al gestionar y generar conocimiento es que dicho conocimiento se interconecten con el entorno real y cree bienes y servicios novedosos, así como amigables ambientes circundantes. Sin embargo,

generar conocimientos novedosos en un mundo tan cambiante e interconectado resulta complejo.

La diversidad, volatilidad y ubicuidad hacen que la tarea de la elaboración, integración e interpretación de los datos del mundo real sea muy difícil.

El volumen de datos en Internet y la Web ya ha sido abrumadora y sigue creciendo a un ritmo impresionante: todos los días alrededor de 2,5 trillones de bytes de datos se crea y se estima que se generó el 90% de los datos de hoy en día en los últimos dos años.

Más aún, los datos inmersos en bases de datos gubernamentales relacionados con diferentes eventos y sucesos pueden ser analizados y convertidos en conocimiento para la acción para darnos una mejor comprensión acerca de nuestro mundo físico y para crear más productos y servicios de valor añadido, por ejemplo: lecturas de medidores pueden ser utilizados para predecir y equilibrar el consumo de energía en las redes inteligentes; el análisis de la combinación de tráfico, registros de datos sensoriales de la contaminación, el clima, etcétera (Aguillo, 2001).

## **LA IMPORTANCIA DE LA WEB SEMÁNTICA**

Se estima que habrá alrededor de 25 mil millones de dispositivos conectados a Internet en 2015 y 50 mil millones en 2020 (Evans, 2011).

Se necesita un número tan impresionante de dispositivos altamente distribuidos y heterogéneos que estén interconectados y en comunicación en distintos escenarios de forma autónoma.

Esto implica que proporcionar interoperabilidad entre los datos que se encuentra en la WEB es de gran importancia y uno de los requisitos más fundamentales para encontrar información direccionada (Nunberg, 1998).

El conjunto de tecnologías desarrolladas en la Web Semántica, como las ontologías y los servicios web semánticos pueden ser utilizadas como soluciones principales para los fines de la generación de nuevo conocimiento. La importancia de la Web Semántica radica en la posibilidad de integrar los datos de una manera más eficiente y rápida; para ello hay 5 elementos clave:

- **Interoperabilidad de la Web Semántica**

La interoperabilidad semántica significa que las diferentes partes interesadas pueden acceder e interpretar los datos de forma inequívoca. Las redes tienen que intercambiar datos entre sí y con otros usuarios de Internet, proporcionar descripciones de datos inequívocos de una manera que puede ser procesada e interpretada por las máquinas y los agentes de software.

- **Integración de datos**

Los datos generalmente se originan a partir de un dispositivo o de un ser humano, y estos pueden ser combinados con otros datos para crear diferentes abstracciones del medio ambiente, o pueden integrarse a la cadena de procesamiento de datos en una aplicación existente para apoyar contexto y conocimiento de la situación. En todos estos casos, es importante que los datos heterogéneos puedan ser perfectamente integrados a un tipo de dato combinable con otros, o con los datos del mundo físico.

La Web semántica puede apoyar esta integración al permitir la interoperabilidad entre diferentes fuentes; sin embargo, todavía se requiere el análisis y la correlación entre los diferentes modelos de descripción semántica para facilitar la integración de datos (Rovira, 2001).

- **Abstracción de datos**

La abstracción de datos se ocupa de las formas en que los datos de mundo físico están representados y gestionados. Con las descripciones semánticas, los datos pueden también caracterizarse en diferentes niveles de abstracción. Esto se logra con el razonamiento semántico ofrecido por los lenguajes de consultas semánticas.

El acceso a los datos se puede implementar en bajos niveles mediante el uso de lenguajes de programación de bajo nivel y los sistemas operativos. Obviamente, la heterogeneidad de los dispositivos y redes hace que el acceso a datos a través de las redes sea una tarea difícil.

- **Servicios de búsqueda y descubrimiento**

Son dispositivos o entidades que pueden proporcionar datos. Los mecanismos de búsqueda y descubrimiento permiten la localización de recursos o servicios que proporcionan los datos relativos a una entidad de interés en el mundo físico.

- **Razonamiento semántico e interpretación**

El formalismo de representación del conocimiento mediante un lenguaje académico utilizado por los profesores e investigadores

permite inferir nueva información o conocimiento de las afirmaciones y los paradigmas vigentes.

El razonamiento semántico es un instrumento importante en el dominio de esta nueva forma de investigar, tales como la búsqueda de recursos, la abstracción de datos y extracción de conocimiento (Tramullas y Olvera (2001).

Los algoritmos de inferencia reales se aplican por lo general para que los investigadores no tengan que preocuparse por las complejidades del proceso de razonamiento en sí.

## **Algunas herramientas semánticas para la generación de conocimiento**

- **Bases de conocimiento:** El objetivo de una base de conocimientos es el de modelar y almacenar bajo forma digital un conjunto de conocimiento, ideas, conceptos o datos que permitan ser consultados o utilizados (Abadal, 2005).

Existen varios métodos y programas para crear bases de conocimientos:

**Wikis:** Los wiki son ideales para enlazar entre ellos una gran cantidad de pequeños contenidos de información poco estructurados (por lo general de un párrafo a una página).

Un elemento primordial en los wikis son los hipervínculos. El hecho de que los usuarios participen en la creación y edición del contenido permite que estos evolucionen rápidamente.

una de las ventajas de estos es que es posible añadir rápidamente contenido y vincular a otro contenido después. Entre los wikis más conocidos, se encuentran MediaWiki (el wiki de Wikipedia).

**CMS:** El CMS es un programa que está orientado a la publicación de documentos de gran tamaño, por lo que no es muy flexible para realizar cambios, pero permite que haya una validación de los documentos antes de la publicación por lo tanto existe más control.

Por lo general es posible otorgar permisos a los usuarios - permisos de acceso solo a ciertos documentos, permisos de edición, entre otros.-

Una de las desventajas radica en que los enlaces entre documentos no es fácil de hacer.

**Mindmap y concept map:** Los Mindmaps y los concept maps son muy similares. Aquí no se ingresan frases largas, sino por lo general solo algunas palabras o expresiones (conceptos, verbos, elementos, etc.). Después se crean conexiones entre todos los elementos ingresados, lo que permite que aparezcan visualmente las relaciones, jerarquías y agrupamientos.

Esto permite una organización rápida de la información, y poder memorizarla fácilmente.

Para citar un ejemplo, FreeMind es un programa de mapa mental gratuito.

**Sistema experto:** El sistema experto permite modelar un conjunto de conocimientos con el propósito de que un usuario pueda explotarlos sin tener que ser un experto: en éste se limita a ingresar la información que dispone y el sistema experto le da una respuesta. Ejemplo de uso: Diagnostico de problemas.

**El formato “foro”:** El formato “foro” es una herramienta de uso típico: en este hay una lista de temas y haciendo clic en el titulo de un tema se puede leer el artículo. Normalmente los artículos son clasificados en categorías. Dependiendo del foro, es más o menos fácil hacer enlaces de un artículo a otro o incluir imágenes. La adición de comentarios depende del programa.

- **Buscadores semánticos:**

**Ngine:** Funciona mediante la introducción de conceptos. Lo primero que muestra al buscar una palabra es la información que tiene sobre esta. Por ejemplo: si se busca lámpara, mostrará sus tipos, medidas, etcétera, si se buscas un medicamento, dirá los datos que tiene sobre esta. Tiene en su base de datos más de 8 millones de conceptos.

**Hakia:** Coloca los resultados de una búsqueda en Webs, News, Blogs, Twitter, Image y Video. Regularmente ofrece una definición de la búsqueda extraída de Wikipedia.



**Kosmix:** Los resultados de una búsqueda los presenta directamente con artículos de blogs o imágenes.

**DuckDuckGo:** Su inteligencia es tal que solicitará al usuario más información en caso de que ésta no haya sido suficiente para realizar la búsqueda.

**Evri:** Filtra los resultados por artículos, citas, imágenes y tweets. Además está ligada a redes sociales como Facebook o MySpace.

- **Servicios Web Semánticos:**

**Freebase:** El concepto es similar a Wikipedia, contenidos generados por los mismos usuarios, pero todo queda centralizado en documentos estructurados.

**GetGlue:** Los usuarios pueden obtener recomendaciones para ver programas de televisión, leer libros, escuchar música.

**Flipboard:** Permite obtener una revista de perfiles de Facebook Twitter. Realiza búsquedas en tiempo real y tendencias en buscadores personalizados, lo que le confiere un gran valor en semántica tecnológica.

**Hunch:** Muestra las películas, libros, destinos, etcétera de interés del usuario, basándose en búsquedas anteriores.

**Apture:** Es un servicio de búsqueda semántica contextual.

**Google Squared:** Mediante una pregunta simple, en lenguaje natural, se recibe respuesta simple.

- **Links contextuales:**

Los links (o links buildings) son uno de los factores más importantes para posicionar una página web en la red de internet y se utilizan para buscar información. Existen más de 150 estrategias de link building y entre los que más se destacan encontramos los links contextuales.

Los links editoriales (o contextuales), son aquellos links que alguien recibe naturalmente de menciones de terceros en sus respectivos blogs o páginas.

Estos links se obtienen gracias a la calidad del contenido de determinadas páginas web y esto se debe a que el contenido es bueno que y por lo tanto se va posicionando en la web

- **Funciones que aportan las distintas capas de la Web Semántica.**

**Unicode:** el alfabeto. Se trata de una codificación del texto que permite utilizar los símbolos de diferentes idiomas sin que aparezcan caracteres extraños. De esta forma, se puede expresar información en la Web Semántica en cualquier idioma.

**URI:** las referencias. URI es el acrónimo de "Uniform Resource Identifier" o Identificador Uniforme de Recursos, identificador único que permite la localización de un recurso que puede ser

accedido vía Internet. Se trata del URL (descripción de la ubicación) más el URN (descripción del espacio de nombre).

**XML + NS + xmlschema:** se trata de la capa más técnica de la Web Semántica. En esta capa se agrupan las diferentes tecnologías que hacen posible que los agentes puedan entenderse entre ellos. XML ofrece un formato común para intercambio de documentos, NL (namespaces) sirve para cualificar elementos y atributos de nombres usados en XML asociándolos con los espacios de nombre identificados por referencias URI y XML Schema ofrece una plantilla para elaborar documentos estándar. De esta forma, aunque se utilicen diferentes fuentes, se crean documentos uniformes en un formato común y no propietario.

**RDF + rdfschema:** basada y apoyada en la capa anterior, esta capa define el *lenguaje universal* con el cual se pueden expresar diferentes ideas en la Web Semántica. RDF es un lenguaje simple mediante el cual se definen sentencias en el formato de una 3-upla o triple (sujeto: el recurso al que nos referimos; predicado: el recurso que indica qué es lo que estamos definiendo; y objeto: puede ser el recurso o un literal que podría considerarse el valor de lo que acabamos de definir).

El modelo RDF o Resource Description Framework es un modelo común (Framework) que permite hacer afirmaciones sobre los recursos (Description) y que hace posible que estos recursos pueden ser nombrados por URIs (Resource). Por su parte RDF Schema provee un vocabulario definido sobre RDF que

permite el modelo de objetos con una semántica claramente definida. Esta capa no sólo ofrece descripción de los datos, sino también cierta información semántica. Tanto esta capa como la anterior corresponden a las anotaciones de la información (metadatos).

**Lenguaje de Ontologías:** ofrece un criterio para catalogar y clasificar la información. El uso de ontologías permite describir objetos y sus relaciones con otros objetos ya que una ontología es la especificación formal de una conceptualización de un dominio concreto del conocimiento. Esta capa permite extender la funcionalidad de la Web Semántica, agregando nuevas clases y propiedades para describir los recursos.

**Lógica:** además de ontologías se precisan también reglas de inferencia. Una ontología puede expresar la regla "Si un código de ciudad está asociado a un código de estado, y si una dirección es el código de ciudad, entonces esa dirección tiene el código de estado asociado". De esta forma, un programa podría deducir que una dirección de la Universidad Complutense, al estar en la ciudad de Madrid, debe estar situada en España, y debería por lo tanto estar formateado según los estándares españoles. El ordenador no "entiende" nada de lo que está procesando, pero puede manipular los términos de modo mucho más eficiente beneficiando la inteligibilidad humana (Tramullas, 2001).

**Pruebas:** resulta necesario el intercambio de "pruebas" escritas en el lenguaje unificador (se trata del lenguaje que hace posible las inferencias lógicas hecha posibles a través del uso

de reglas de inferencia tal como es especificado por las ontologías) de la Web Semántica.

**Confianza:** los agentes deberían ser muy escépticos acerca de lo que leen en la Web Semántica hasta que hayan podido comprobar de forma exhaustiva las fuentes de información.

**Firma digital:** bloque encriptado de datos que serán utilizados por las computadoras y los agentes para verificar que la información adjunta ha sido ofrecida por una fuente específica confiable.

## CONCLUSIÓN

Existe una amplia gama de herramientas de la Web Semántica que facilitan el trabajo de la gestión y generación de nuevo conocimiento y que son necesarias conocer para un buen uso de la misma. Sin embargo, hay varios desafíos de la investigación en la aplicación de tecnologías semánticas.

La mayoría de estos desafíos están estrechamente relacionados con el dinamismo y la capacidad de penetración de quién busca información.

Será necesario que la comunidad de investigadores continúe explorando el nuevo uso de las tecnologías semánticas, junto con la aplicación de otros campos como la informática de servicios, minería de datos y de las ciencias sociales para mejorar el tratamiento y la utilización de los datos semánticos. En suma, el objetivo de la Web Semántica es que la Web pase de ser una colección de documentos a convertirse en una base de conocimiento.

## FUENTES DE INFORMACIÓN

1. AGUILLO, Isidro (2001). "Información científica en la web: retos y tareas para los documentalistas del siglo XXI". En: Fuentes, M.E. (dir.). Anuario de biblioteconomía, documentación e información. Barcelona: COBDC, 2001, p. 33-50
2. HÍPOLA, P.; ÉITO, R. (2000). "Edición digital: formatos y alternativas". El profesional de la información, v. 9, n. 10, octubre 2000, p. 4-15
3. NUNBERG, G. (comp.) (1998). El futuro del libro: ¿esto matará eso?. Barcelona: Paidós, 1998, 314 p.
4. ROVIRA, Cristòfol (2001). "Herramientas de ayuda a la navegación". Temes de Disseny, n. 18, abril 2001, , p. 66-73
5. TRAMULLAS, Jesús; OLVERA, M. Dolores (2001). Recuperación de la información en Internet. Madrid: Ra-Ma, 232 p.
6. ABADAL, E.; Codina L. (2005). Bases de datos documentales: características, funciones y método. Madrid: Síntesis.
7. PENROSE, R. (1991). La nueva mente del emperador. Madrid: Mondadori
8. TRAMULLAS, J. (2001). "La recuperación de información en el World Wide Web". En J. López Yepes (coord.) Manual de Ciencias de la Documentación. Madrid: Síntesis