



Web Semántica y Representación del Conocimiento con Ontologías

German E. Yahuita Quisbert
Postgrado en Informática
Universidad Mayor de San Andrés - UMSA
La Paz, Bolivia
german@umsa.bo

Resumen— Este artículo presenta a las web semánticas como una propuesta de arquitectura para la próxima generación de aplicaciones en la WWW¹⁸ basadas en: ontologías para la estructuración de la representación del conocimiento y sus relaciones. Y al lenguaje OWL¹⁹ para la implementación de inferencias de un aserto de interés para el usuario de la www. La descripción de los “recursos” (información de un dominio determinado) se realiza basada en el estándar propuesta por RDF²⁰ y RDF Schema²¹

Palabras Claves - ontología, semántica, web, información, conocimiento, clases

Abstract – This paper presents a web semantic as a proposed architecture for the next generation of applications in WWW, based structuring of ontologies for knowledge representation and relationships. And the OWL language for implementing inference of a claim of interest to the user of the www. The description of the "resources" (information of a certain domain) is performed based on the proposed standard for RDF and RDF Schema

Palabras Claves— *ontology, semantic, web, information, knowledge, Class.*

I. INTRODUCCIÓN

La WWW surgió a consecuencia de la necesidad de compartir información en los años 80, y por la gran difusión que la misma sufrió, se introduce en los aspectos comunicacionales de la vida de las personas en las familias de las sociedades avanzadas. Ofreciendo diferentes tipos de información y servicio; consultas a las cuentas bancarias, reservas de boletos para viajes, Acceso a las últimas

noticias, comunicación con otras personas casi instantáneamente, culminado con las redes sociales.

Este conjunto de servicios actualmente se conoce bajo el denominativo de Web 2.0

La dinámica que generó la Web 2.0 permitió la participación de la sociedad de forma activa subiendo información a través de blogs, wikis, repositorio, etc.

La información almacenada creció enormemente en todas las áreas del quehacer humano; diseminado en los miles de servidores públicos interconectados en la WWW.

La localización de la información en este contexto se volvió sumamente ineficiente en cuanto a la pertinencia de los resultados obtenidos, debido principalmente a la gran cantidad de referencias (enlaces) que se obtiene al realizar una consulta en los motores de búsqueda como: yahoo, google, etc. A esta situación también se debe sumar los cambios y actualizaciones permanentes que se suceden. De hecho, se estima que para el año 2015 se tendrá aproximadamente 5.7 Zetabites de información²².

El estado de situación existente, llevaron al surgimiento de muchas propuestas e iniciativas que se enfocaron a paliar estos problemas, y actualmente a las mismas se las conoce como la Web 3.0 que buscan incorporar nuevas técnicas de indexación y almacenamiento de la información. Lo que se espera mejore las estrategias de búsqueda, ya que incorporan nuevos elementos semánticos que cooperan a los motores de búsqueda,

¹⁸ WWW World Wide Web

¹⁹ OWL Ontology Web Language, un lenguaje de marcado para publicar y compartir datos usando ontologías en la WWW.

²⁰ RDF Resource Description Framework, Marco de Descripción de Recursos, un lenguaje de descripción del W3C

²¹ RDF Schema es una extensión semántica de RDF. Un lenguaje primitivo de ontologías que proporciona los elementos básicos para la descripción de vocabularios

²² Inf., IDC International Data Group (IDG)

Los trabajos de investigación realizados en el marco de la Web 3.0, mostraron a la web semántica y el uso de las ontologías, como una propuesta viable a ser mejorada y establecerse como un standard para las futuras aplicaciones a ser utilizadas en la www.

Web Semántica [1]

Tim Berners-Lee, planteo una arquitectura para el desarrollo de aplicaciones web y la denominó Web Semántica, esta web es; donde las aplicaciones son capaces de efectuar un procesamiento de la información mucho más “profunda”. La web propuesta está caracterizada por programas capaces de “comprender” el contenido de las páginas web, y por tanto, de procesar y relacionar la información contenidas en estas.

La arquitectura propuesta, tuvo buena acogida en la comunidad de desarrolladores; que se creó un equipo en el World Wide Web Consortium (W3C) que se dedican a mejorar, extender y estandarizar el sistema y muchos lenguajes, publicaciones y herramientas han sido ya desarrollados.

¿Cuál es su definición?

Es la Web que contiene mucha información donde identificamos las relaciones y los recursos en un contexto determinado. Las mismas permiten sostener estrategias que se orientan a resolver los problemas que actualmente aquejan al rendimiento de las diferentes aplicaciones que se encuentran en la WWW.

¿Cuál es su objetivo?

El objetivo es crear un medio universal que permita el intercambio de datos y brindar un mayor significado a la misma para que puedan ser interpretadas por las máquinas.

II. PRINCIPALES COMPONENTES

Los principales componentes de la Web Semántica son: XML, XML Schema, RDF, RDF Schema y OWL.

OWL “Ontology Web Language” describe las funciones y relaciones de cada componente de la Web Semántica:

XML: le provee a la web semántica una sintaxis elemental para las estructuras de contenidos dentro de documentos.

XML Schema: Es un lenguaje para proporcionar y restringir la estructura y el contenido de los elementos contenidos dentro de documentos XML.

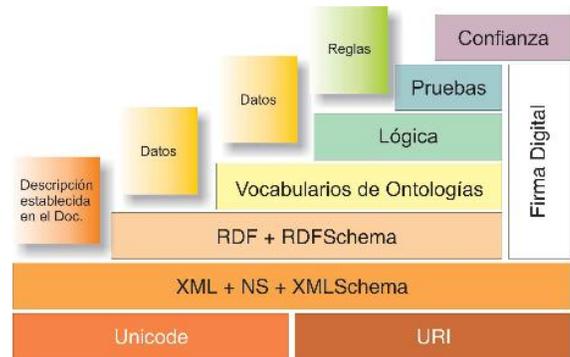


Fig. 1. Arquitectura Web Semántica.

Unicode:

Es un estándar cuyo objetivo es proporcionar el medio por el cual un texto en cualquier forma e idioma pueda ser codificado para el uso informático. El mismo nos permite mostrar información en cualquier idioma y con la certeza de que no aparezcan símbolos extraños.

URI:

Son cadenas que permiten acceder a cualquier recurso de la Web. En la Web Semántica las URIs son las encargadas de identificar objetos. Todos los objetos pueden ser identificados mediante una URI. Si dos objetos cuentan con la misma URI pueden existir colisiones. El grupo de trabajo del W3C está intentando resolver este problema.

XML+NS+xmlschema:

Esta es la capa más técnica de la Web Semántica. En la que se encuentran agrupadas las diferentes tecnologías que posibilitan la comunicación entre las diferentes aplicaciones.

El estándar Extensible Markup Language (XML) ofrece un formato común para el intercambio de documentos,

Namespaces (NS) proporciona el método para cualificar los elementos y atributos de nombres usados en los documentos XML asociándolos con los espacios de nombre identificados por referencias URIs.

XML Schema es un lenguaje que permite describir la estructura y restringir el contenido de los documentos XML.



RDF: es un lenguaje simple que define un modelo de datos para describir a los objetos “recursos” y a sus relaciones, mediante tripletas sujeto-predicado-objeto. Los dos primeros serán URIs y el tercero puede ser URI o un valor literal. Este modelo de RDF se puede representar en sintaxis de XML

RDF Schema: Es un vocabulario RDF para describir propiedades y clases de recursos RDF, con semántica para generalizar-jerarquías de las propiedades y clases.

RDF+rdfschema: Está basada en la capa anterior (XML NL, XML-SCHEMA), define el lenguaje universal con el que podemos expresar diferentes ideas en la Web Semántica.

OWL: Es un mecanismo para desarrollar temas o vocabularios específicos en los que podamos asociar esos recursos.

Ontología [2]

La ontología es una antigua disciplina que en sentido filosófico, se define como un esquema específico de categorías que refleja una visión específica del mundo. Desde el punto de vista informático ontologías son teorías que especifican un vocabulario relativo a un cierto dominio. Este vocabulario define entidades, clases, propiedades, predicados y funciones, además de las relaciones entre estos componentes. Las ontologías toman un papel clave en la resolución de interoperabilidad semántica entre sistemas de información y su uso.

Con el fin de comprender la definición de ontología desde el campo de la informática, citaremos algunas definiciones que por su contenido se consideran valiosas dentro del estado del arte.

La primera, fue dada por Neches [3] y colegas, quienes propusieron que: “Una ontología define las condiciones básicas y relaciones que comprenden el vocabulario de un área del tema así como las reglas para combinar condiciones y las relaciones para definir extensiones del vocabulario ”. Esta definición descriptiva dice qué hacer para construir una ontología, y nos da algunas pautas vagas: la definición identifica condiciones básicas y relaciones entre las condiciones, identifica las reglas para combinar las condiciones, y proporciona las definiciones de tales condiciones y relaciones. Note que, según la definición de Neches [3] , una ontología incluye no sólo las condiciones

que se definen explícitamente en ella, sino que también incluye el conocimiento que puede ser inferido.

La definición más popular de Ontología en la Informática es la dada por Gruber [3], quien define una ontología como “Una especificación explícita de una conceptualización, es decir, que proporciona una estructura y contenidos de forma explícita que codifica las reglas implícitas de una parte de la realidad; estas declaraciones explícitas son independientes del fin y del dominio de la aplicación en el que se usarán o reutilizarán sus definiciones”.

Para Guarino [4] “La ontología describe una cierta realidad con un vocabulario específico, usando un conjunto de premisas de acuerdo con un sentido intencional de palabras del vocabulario”.

La definición más formal la encontramos en el documento de Web Ontology Language de la W3C²³ que indica que “Una ontología define los términos a utilizar para describir y representar un área de conocimiento. Las ontologías son utilizadas por las personas y las aplicaciones que necesitan compartir información en un dominio determinado (área de conocimiento o temática específica, tales como el área de medicina, fabricación de herramientas, bienes inmuebles, reparación automovilística, gestión financiera, etc).

III. REPRESENTACIÓN DEL CONOCIMIENTO EN LAS WEB SEMÁNTICAS

La representación del conocimiento es un elemento fundamental en la implementación de las web semánticas, ya que permite a las aplicaciones implementar funciones que incorporan estrategias de localización de información que sean óptimas y eficientes.

La ontología apoyada en otras tecnologías de la rama de la Inteligencia Artificial y de la Lingüística, generaron propuestas para las consultas y representaciones del conocimiento, que se las conoce como OWL, RDF y RDF Scheme

Ontology Web Language(OWL) surgió de ese conjunto de propuestas para convertirse en un estándar recomendado por el W3C (World Wide Web Consortium

²³ El Consorcio World Wide Web (W3C), es una comunidad internacional, que trabajan conjuntamente para desarrollar estándares Web. Liderado por el inventor de la Web Tim Berners-Lee

2010) [5] En esencia es un lenguaje que permite describir semántica procesable de forma automatizada utilizando motores de inferencia para responder a cualquier tipo de consultas.

El lenguaje está sustentada en el modelo de datos que proporciona el vocabulario RDF (Resource Description Framework) [6] y sobre las proposiciones lógicas de RDF Schema [7], Un vocabulario que permite definir la semántica de los asertos definidos en RDF. En otras palabras, OWL extiende la semántica que es posible definir con RDF Schema y adopta la manera que tiene RDF para estructurar la información para describir en la web los recursos, objetos físicos, conceptos abstractos o cualquier otra cosa con identidad propia.

La información se representa mediante sentencias que toman la forma de tripletas sujeto-propiedad-objeto, donde el sujeto indica a qué o quién se refiere la sentencia (es decir, el concepto o recurso que queremos describir), la propiedad indica una característica determinada del sujeto, y el objeto indica el valor que toma la propiedad (que puede ser literal o un nuevo concepto). Por lo tanto, se pueden distinguir dos tipos de tripletas atendiendo a los elementos que las conforman: recursos-propiedad-recursos y recursos-propiedad-valor, esta descripción se realiza en función de las propiedades que caracterizan a los recursos y los valores que estas toman, de forma que es posible enriquecer la definición de un recurso simplemente añadiéndole nuevas propiedades.

Entendiendo, como Recursos: Cualquier cosa que pueda nombrarse mediante una URI.

Propiedad; Característica o atributo de un recurso (Tiene asociado una URI y un significado concreto, además que puede relacionarse con otras propiedades)

Aunque existen varias formas de representar las proposiciones RDF la sintaxis más extendida es la basada en el metalenguaje XML (eXtensible Markup Language) [8], que en los últimos años se ha convertido en un estándar para el intercambio de información en la Web.

Las tripletas RDF pueden ser representado gráficamente a través de grafos conexos; donde los conceptos se representan con elipses y las propiedades con flechas direccionadas que conectan conceptos con otros conceptos o literales.

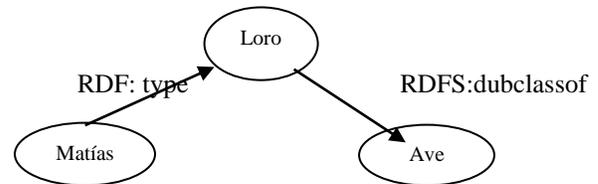


Fig. 2. Representación gráfica de un aserto en RDF.

Y OWL [5] define la semántica y reglas básicas de razonamiento basadas en la lógica de predicados y la lógica descriptiva, ejemplo:

1. Pertenencia a clases: si x es una instancia de la clase C , y C es una subclase de D , entonces x es una instancia de la clase D
2. Equivalencia de clase: Si A es una clase equivalente a la clase B , y B es equivalente a la clase C , entonces A es equivalente a C .
3. Clasificación de Instancias: La imposición de determinado pares propiedad/valor como condición suficiente para ser miembro de una clase, nos permite definir tantas instancias como sujetos cumplan esta condición.

Sin embargo a pesar de las muchas capacidades de OWL no es posible definir inferencias complejas, la expresividad y eficiencia en el procesamiento son también otro problema que presenta este lenguaje.

Para conciliar capacidad expresiva con eficiencia de procesamiento el grupo de trabajo del World Wide Web Consortium que se encarga del desarrollo normativo de OWL [5] se vio en la necesidad de definir tres sub-lenguajes (o especies) para OWL con el fin de poder ofrecer a los desarrolladores diferentes combinaciones de expresividad-capacidad de razonamiento que permitieran adaptar las ontologías a la complejidad y necesidades específicas de un sistema determinado:

OWL Full es la especie de mayor capacidad expresiva ya que integra todas las primitivas definidas en el lenguaje (para ser rigurosos ni siquiera debería considerarse como un sub-lenguaje) y permite combinarlas de forma arbitraria con RDF y RDF Schema.

OWL DL (Description Logics) por su parte restringe el modo en que se pueden utilizar los constructores de OWL y RDF de manera que las



ontologías se ajustan a los requerimientos de los motores de inferencia basados en lógica descriptiva.

OWL Lite impone condiciones más estrictas sobre los elementos de OWL que se pueden utilizar, lo cual disminuye su capacidad expresiva pero facilita el razonamiento con las ontologías diseñadas usando este sub-lenguaje.

La elección de una u otra especie depende del problema que queramos resolver y de cuáles son las necesidades específicas del sistema en el que queremos implantar nuestra ontología (o lo que es lo mismo, hay que tener claro si la prioridad de nuestro sistema es la interoperabilidad o la facilidad de procesamiento). Por esta razón, hoy las implementaciones prácticas de ontologías se hacen utilizando OWL DL u OWL Lite, ya que con OWL Full son inviables.

IV. DISEÑO DE LA ONTOLOGÍA

La implementación de la web semántica, requiere de la identificación de los Conceptos, Relaciones, Instancias, Atributos, Axiomas y reglas de un dominio determinado. Una vez identificado estos elementos, se debe elaborar:

- i) Glosario de términos importantes
- ii) Taxonomía de conceptos
- iii) Descripción de las Relaciones Binarias y de Conceptos
- iv) Definir Reglas
- v) Crear Instancias
- vi) Identificar los Axiomas

A la conclusión de esta actividad se puede, ya plantear un prototipo de ontología, con la herramienta “protege”²⁴.

La implementación de la ontología en un almacenamiento persistente se la realiza a través de la herramienta Jena²⁵, posteriormente, se implementan la interfaces para la aplicación web utilizando, Json²⁶ que es una alternativa más óptima a XML.

V. CONCLUSIONES

²⁴ Protégé es una herramienta para el desarrollo de ontologías y Sistemas basados en el conocimiento creada en la Universidad de Stanford

²⁵ Jena es un framework construido por Apache con el que se permite construir aplicaciones web semánticas y aplicaciones Linked Data

²⁶ Json. JavaScript Object Notation, es un formato ligero para el intercambio de datos. Json es un subconjunto de la notación literal de objetos de JavaScript que no requiere el uso de XML

La web semántica es la respuesta a las necesidades de almacenamiento, relacionamiento e intercambio de información, que permite realizar búsquedas, extraer, interpretar y procesar conocimiento que podrán ser realizado por las maquinas.

Pese a que existe mucha investigación en relación a las web semánticas, podemos ver que la misma es muy prometedora y actualmente se está desarrollando la infraestructura básica y se está experimentando con la creación de componentes básicos que puedan explotar esta infraestructura.

Lo que se muestra en este artículo es solo una visión general del trabajo que se está realizando, y seguramente en los próximos años esta tecnología se consolidara para apuntalar a una WWW mucho más inteligente.

Finalmente, las web semánticas (Web3.0) basadas en Ontologías se fortalece como una alternativa para la publicación de páginas web que en conjunto se las que se conoce como la World Wide Web versión 3.0

Y queda pues continuar profundizando la investigación de representación de la información utilizando las nuevas tecnologías que surgen en el ámbito del desarrollo de aplicaciones web.

REFERENCIAS

- [1] Bernes Lee Tim, Hendeler James , Lassila Ora Scientific American.
- [2] Contreras Jesús y Martínez Comche Juan Antonio. Tutorial de Ontologías. Universidad Complutense de Madrid.
- [3] Neches, Robert; Gruber, Thomas; Fikes, Richard; Finin, Timothy; Patil, Ramesh; Senator, Edward & Swartout, William. “Enabling Technology for Knowledge Sharing”. En: Artificial Intelligence Magazine. Vol. 12, num. 3, American Association for Artificial Intelligence, 1991, Press. Menlo Park, CA, USA, pp. 36-56.
- [4] Guarino N., "Formal Ontology, Conceptual Analysis and Knowledge Representation" en International Journal of Human and Computer Studies, 1995, pp. 625-640
- [5] OWL web ontology language overview DL McGuinness, F Van Harmelen W3C recommendation, 2004.
- [6] RDF/XML Syntax Specification (Revised), David Beckett (ed.), W3C Recommendation 10 February 2004.
- [7] RDF vocabulary description language 1.0: RDF Schema. W3C Recommendation, 2004.
- [8] RDF/XML Syntax Specification (Revised), Brian McBride, W3C Recommendation, 10 February 2004