

LA ONTOLOGÍA Y LA WEB SEMÁNTICA: RECOMENDACIONES DEL W3C.

I. INTRODUCCIÓN

❖ LA WEB HOY

En poco más de una década, la **World Wide Web o Red Global Mundial**, conocida comúnmente como la Web, se ha impuesto como la alternativa más potente y rápida para la difusión internacional de información, accesible toda ella a través de Internet. Fue concebida por **Tim Berners-Lee**, quien, en 1989, presentó su ambicioso proyecto en el CERN (Suiza).

En la actualidad, el **volumen documental** que alberga es asombroso. Se dice que todo está en la Red, y lo que no está es porque no existe. A la **Web superficial** o visible hay que sumarle la **Web profunda** o invisible, compuesta por todos los contenidos que están ahí, pero que no aparecen al realizar consultas en los buscadores, debido normalmente a que son recursos dinámicos pertenecientes a bases de datos.

Sin embargo, a pesar de contar con este ingente almacén documental, es bien sabido que la capacidad Internet para satisfacer necesidades concretas de información es limitada. Por un lado, la **inestabilidad, obsolescencia** y, en ocasiones, **falta de calidad** de muchos contenidos suponen aún grandes inconvenientes. Por otro, la abrumadora obtención de resultados, pertinentes y no pertinentes, ante una búsqueda genérica denota una importante **falta de precisión** en la Web. Este habitual problema se debe principalmente a que **la Web actual carece de capacidad para expresar significados**. Únicamente se limita a recoger cadenas de caracteres indizadas en grandes bases de datos, a buscar palabras clave, planas e inconexas, cuando son requeridas y a presentarlas en la pantalla del ordenador para su visualización por parte de un ser inteligente. Pero **la máquina no sabe realmente lo que la información significa**.

De aquí arranca la apuesta por construir de una **Web Semántica (WS)**, también llamada Web Inteligente.

❖ LA WEB SEMÁNTICA: WEB DEL FUTURO

La WS es una corriente promovida por el propio **Tim Berners-Lee**, cuyo fin es **lograr que las máquinas puedan entender** y, por tanto, **utilizar lo que la Web contiene**. De ahí que se haya tomado el concepto de **Semántica**, que desde el punto de vista lingüístico es la disciplina que **estudia el significado** de los términos.

Tal y como expresó Berners-Lee en un artículo publicado por la revista *Scientific American* en 2001, la **WS es una extensión de la Web actual**, no una sustitución, ya que mantiene las características esenciales que han hecho de ésta un éxito: descentralización, máxima facilidad de acceso, apertura al crecimiento, etc.

Se pretende desarrollar **una Web más cohesionada**, donde sea más fácil **localizar, compartir e integrar información y servicios**, para sacar un mayor partido a los recursos disponibles. La WS persigue el **establecimiento de una forma universal de representar las relaciones entre los datos**, y entre éstos y sus significados, para que un sistema automático (agente) sea capaz de sacar sus propias conclusiones respecto a las búsquedas realizadas. Así, esta nueva Web estaría poblada de agentes inteligentes de software capaces de navegar y realizar operaciones por nosotros, optimizando resultados y ayudando o, incluso sustituyendo, a las personas en tareas rutinarias o inabarcables.

Pero, para que las máquinas empiecen a **procesar “conocimiento”** y **no únicamente “datos”**, es necesario que la información esté convenientemente estructurada y clasificada. Esto conlleva un mayor **desarrollo de estándares** y una utilización consensuada de los llamados **“lenguajes para la representación del conocimiento”**, con los que las personas vinculadas al mundo de la Documentación estamos bien familiarizados (ej. DC).

Las **tecnologías actuales más importantes** para construir la WS son: el metalenguaje **XML**, la arquitectura **RDF**, y el lenguaje **OWL**. Todas ellas están respaldadas por el **W3C**, que paso a tratar a continuación:

II. EL WORLD WIDE WEB CONSORTIUM (W3C)

❖ CONCEPTO

El **World Wide Web Consortium (W3C)** es un organismo encargado de **velar por la normalización en Internet**. Está presidido desde su creación en **1994** por **Tim Berners-Lee**, quien, como se ha mencionado, es a su vez el máximo impulsor de la WS. En la actualidad, sus miembros ascienden a más de 400 organizaciones.

La **misión** del W3C es «guiar a la Web hacia su máximo potencial mediante el desarrollo de protocolos y pautas comunes que promuevan su evolución y garanticen su interactividad». Sus tres **objetivos** a más largo plazo son crear una web:

- **De acceso universal**, para todos y desde cualquier sitio.
- **Semántica**, gracias al desarrollo de tecnologías adecuadas para optimizar su uso.
- **De confianza**, teniendo en cuenta los nuevos aspectos legales, comerciales y sociales.

❖ ACTIVIDAD NORMATIVA

El W3C lleva a cabo su actividad normativa a través de **RECOMENDACIONES**, **aceptadas como verdaderos estándares** tanto por la industria como por los internautas. Cada recomendación es una especificación estable desarrollada por un Grupo de Trabajo del W3C y revisada por los miembros del consorcio. A través de ellas se favorece la **“interoperabilidad”** de las tecnologías web, es decir, la capacidad de sistemas heterogéneos para interactuar entre sí e intercambiar información.

❖ RECOMENDACIONES PARA LA WS: RDF y OWL

En febrero de 2004 el consorcio W3C anunció la aprobación de **dos recomendaciones** fundamentales para la construcción de la WS: **RDF** (*Resource Description Framework*) y **OWL** (*Ontology Web Language*). Ambas, elaboradas por sendos **Grupos de Trabajo** y compuestas por **6 documentos** respectivamente, cuentan con una **guía de uso**, con **especificaciones técnicas e informáticas** así como con **casos y conjuntos de pruebas** para ayudar a aprender, utilizar y aplicar estas tecnologías.

Veamos cada una de ellas por separado:

III. RDF (MARCO DE DESCRIPCIÓN DE RECURSOS)

RDF ofrece una **superestructura** para la **descripción de recursos web** de forma sencilla. Es un lenguaje capaz de **representar metadatos**, es decir, datos sobre datos, que ayudan a la identificación, clasificación o recuperación de documentos en línea. Cuenta con un **conjunto claro de reglas**, que permite la **codificación, intercambio y procesamiento** automático de los metadatos normalizados así como la **construcción de una estructura semántica no ambigua**, gracias a la utilización de identificadores permanentes de recursos URI. Todo ello lo convierte en **el estándar más popular y extendido** en la comunidad web hoy en día.

El elemento de construcción básica de RDF es la **sentencia** o “tripleto”, que consta de:

- **Sujeto.** Es el recurso, es decir, todo aquello que puede ser descrito.
- **Predicado.** Introduce la propiedad o atributo que va a detallarse sobre el recurso.
- **Objeto.** Es el valor de dicha propiedad.

Ejemplo: Descripción de un recurso musical en RDF

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:cd="http://www.recshop.fake/cd#">
  <rdf:Description
    rdf:about="http://www.recshop.fake/cd/Empire
    Burlesque">
    <cd:artist>Bob Dylan</cd:artist>
    <cd:country>USA</cd:country>
    <cd:company>Columbia</cd:company>
    <cd:price>10.90</cd:price>
    <cd:year>1985</cd:year>
  </rdf:Description>
</rdf:RDF>
```

El elemento <rdf:Description> contiene la descripción del **recurso** <http://www.recshop.fake/cd/EmpireBurlesque>. Los elementos <cd:artist>, <cd:country>, <cd:company>, etc. son las **propiedades** del recurso. Y Bob Dylan, USA, Columbia, etc. son los **valores** de esas propiedades.

Es una manera de **darle la información desmenuzada** al ordenador para que la “entienda” e identifique cada parte de la sentencia esté en el orden que esté.

El **Esquema RDF o RDFS** es una extensión semántica que enriquece el modelo básico. Es un vocabulario que permite definir un primer sistema de **jerarquías** entre las clases de recursos, especificando las propiedades y relaciones admitidas entre ellas.

Pero **RDF** no deja de ser un **marco abstracto** para describir recursos que **requiere de una sintaxis concreta** para expresar el conocimiento y transferirlo. De todas las formas sintácticas definidas para la formulación escrita de RDF, la más extendida es, sin lugar a dudas, la basada en **XML** (*Extensible Markup Language*), un metalenguaje surgido en 1998 bajo los auspicios del W3C. Se trata de un lenguaje de marcas, derivado de SGML, **específicamente pensado para ser utilizado en el entorno web**.

XML es la herramienta utilizada para **estructurar y presentar los contenidos web**. Sin embargo, **ofrece una capacidad limitada para expresar semántica**. Por eso, se entiende que **RDF** es a la **Semántica** lo que **XML** es a la **Sintaxis**. Gracias a **RDF** se expresan **afirmaciones** y mediante su lenguaje de base **XML** se define la **estructura de tales afirmaciones**. Por tanto, XML responde a la necesidad de **representar sintácticamente el modelo planteado por RDF** para archivos legibles por ordenador.

La otra recomendación del W3C relacionada con la WS e íntimamente ligada a RDF es OWL.

IV. OWL (LENGUAJE DE ONTOLOGÍA WEB)

❖ CONCEPTO ONTOLOGÍA

La Ontología es un **concepto filosófico** (parte de la metafísica que trata del ser en general y de sus propiedades trascendentales), adoptado por la **Informática**. En su aplicación a la WS, alude al intento de formular un **exhaustivo y riguroso mapa conceptual dentro de uno o varios dominios dados**, con la finalidad de facilitar la comunicación e intercambio de información (**interoperabilidad**) entre distintos sistemas.

Las ontologías definen de forma estándar y consensuada **un vocabulario de conceptos** así como las **relaciones entre ellos** dentro de área concreta del conocimiento, formando **redes jerárquicas semánticas**. Actúan a modo de tesauros, pero en lenguaje informático. Así, recogen **reglas lógicas y restricciones** para hacer “comprender” a las máquinas los conceptos que manejan dentro de un determinado campo. Ej. Una ontología de arte establece que todos los escultores son artistas pero no todos los artistas son escultores.

Por supuesto, la adopción de **ontologías comunes** por parte de todos los participantes en la WS es clave para que puedan trabajar de forma autónoma y con la garantía de “hablar el mismo idioma”.

Aunque son compatibles con cualquiera de los lenguajes de representación del conocimiento existentes (basados normalmente en **XML**), el más común es **RDF**.

Una ontología contiene:

- **Conceptos**: Son las clases o ideas básicas que se intentan formalizar dentro de un dominio determinado. Ej. En una ontología de deportes, cada clase sería un deporte: fútbol, baloncesto, tenis, etc.
- **Instancias**. Representan objetos determinados de un concepto (subclases).
- **Relaciones**. Representan la interacción y enlace entre conceptos del dominio (jerarquías). Ej. Subclase-de, parte-de, conectado-a, etc.
- **Funciones**. Tipo concreto de relación. Ej. Asignar-fecha.
- **Axiomas**. Son teoremas que aplican la lógica en las relaciones. Ej. Si A tiene relación con B, B tiene relación con A.

Visto el concepto de ontología y su posible aplicación a la WS, paso a explicar el OWL.

❖ OWL

El Lenguaje de Ontologías Web (OWL) sirve para **definir ontologías web estructuradas**, mediante la descripción refinada de las diferentes clases, propiedades y relación entre las mismas (cardinalidad, disyunción, igualdad, simetría, etc.) El Grupo de Trabajo de Ontología Web del W3C afirma que «supone **un gran paso adelante** en la representación y organización del conocimiento en la World Wide Web». Está basado **en el lenguaje previo DAML+OIL**, derivado, a su vez, de la fusión entre el lenguaje **DAML** (de la Agencia de defensa DARPA de los EE.UU.) y **OIL** (desarrollado en Europa). OWL trata de unificar todas las ventajas de DAML+OIL como lenguaje estándar para la definición de ontologías, pero resolviendo sus carencias, especialmente como formato de intercambio.

DARPA Agent
Markup Language
Ontology Inference
Language

OWL puede formularse en RDF, que ofrece la base adecuada para desarrollar ontologías, y mejora la capacidad expresiva de RDFS. Al construirse sobre RDF, las ontologías OWL podrán ser distribuidas en diferentes sistemas y ser compatibles con otros estándares web.

No obstante, escribir en lenguajes como RDF y OWL resulta sumamente difícil y propenso a errores. Afortunadamente, existen en el mercado **entornos gráficos para visualizar y construir ontologías** de forma más razonable, como **Protégé**. Este software, desarrollado por la Universidad de Stanford, es el que más usuarios atrae actualmente. Permite editar ontologías con una interfaz sencilla y en un entorno de menús, botones, cuadros de diálogo o representaciones gráficas fáciles de usar, tan importantes para una tarea tal de abstracción y síntesis. Aunque Protégé posee su propio lenguaje interno para definir ontologías, admite trabajar con RDF y OWL.

V. CONCLUSIÓN: EL CAMINO HACIA LA WS

La WS es aún una visión, un proyecto de futuro muy ambicioso, que permitirá, con ayuda de la **Inteligencia Artificial**, realizar un sinnúmero de operaciones en la Web, mucho más amplias que las ofertadas hoy en día. El tener **toda la información etiquetada sintácticamente y semánticamente** facilitará la implementación eficaz de los llamados **agentes inteligentes**, capaces de ofrecer información Web pertinente, en función de los intereses y circunstancias personales de cada usuario (personalización máxima).

Esta situación imaginaria tiene ya su base real, materializada en los proyectos piloto realizados y en los grandes avances logrados para su creación en cuanto a estándares e infraestructura. Las principales empresas, como IBM, Microsoft, etc. participan activamente en su desarrollo, así como la comunidad investigadora, especialmente la universitaria. Por supuesto, el proyecto no hubiera sido posible sin el apoyo e impulso de la W3C, que junto con el sitio oficial www.semanticweb.org, se encarga de ofrecer toda la información disponible sobre los progresos en este ámbito. El interés por la WS se refleja en la celebración anual del Congreso internacional de la WWW, que en 2009 ha tenido lugar en la Universidad Politécnica de Madrid. También queda patente con la publicación de la revista *Journal of Web Semantics*.

En el terreno de las bibliotecas, la WS podrá ser decisiva de cara a la construcción de una **Biblioteca Digital Universal**, donde todo sea accesible de forma rápida y precisa, se encuentre donde se encuentre. Por supuesto, aún queda mucho camino por recorrer y **la transición de la Web actual a la WS** puede implicar un **coste altísimo** (en tiempo, dinero y esfuerzo), ya que no sólo se trata de estructurar la información web venidera, sino también la ya existente, labor que se prevé irrealizable.

RECAPITULACIÓN DE LOS PUNTOS TRATADOS

LA ONTOLOGÍA Y LA WEB SEMÁNTICA: RECOMENDACIONES DEL W3C.

I. INTRODUCCIÓN

- ❖ LA WEB HOY
- ❖ LA WEB SEMÁNTICA: WEB DEL FUTURO

II. EL WORLD WIDE WEB CONSORTIUM (W3C)

- ❖ CONCEPTO
- ❖ ACTIVIDAD NORMATIVA
- ❖ RECOMENDACIONES PARA LA WS: RDF y OWL

III. RDF (MARCO DE DESCRIPCIÓN DE RECURSOS)

IV. OWL (LENGUAJE DE ONTOLOGÍA WEB)

- ❖ CONCEPTO ONTOLOGÍA
- ❖ OWL

V. CONCLUSIÓN: EL CAMINO HACIA LA WS