

“La Web Semántica y el lenguaje RDF”

Jesús Cáceres Tello
Servicios Informáticos
Profesor Asociado del Dpto. Ciencias de la Computación
Escuela Técnica Superior de Informática
Universidad de Alcalá
jesus.caceres@uah.es

Resumen. Este artículo empieza por definir qué es la Web Semántica y cuáles son sus características, para después definir cada uno de los elementos de los que se compone, entrando más en profundidad en uno de ellos como es el lenguaje RDF. Además define los URI's como elementos necesarios en la definición de un RDF Schema.

1 Introducción

Actualmente la información presente en la WWW se encuentra estructurada mediante lenguajes de etiquetado que únicamente describen la forma en que dicha información debe ser presentada al usuario por el navegador (colores, maquetación, interacción...), pero no expresan nada sobre su significado, es decir, su semántica.

2 La Web Semántica

El objetivo de la Web semántica es que toda la información presente en la WWW sea comprensible no sólo por los humanos sino también por las propias máquinas, produciéndose una conversión de la estructura actual de la web en una gran estructura de almacenamiento de la información, es decir, catalogar la información de los recursos web, páginas HTML, documentos PDF, videos, archivos de sonido, etc., mediante lo que se denomina ontologías, esto es, mediante el significado de las palabras, no mediante palabras claves.

Nace por el objetivo de un ingeniero software, Tim Berners-Lee, miembro fundador del W3C, de ampliar la interoperabilidad entre sistemas y reducir la intervención de operadores en procesos de flujo de información.

La Web Semántica constituye una extensión de la actual Web en el cual se ha añadido soporte para bases de datos en formato legible por la máquina. La actual Web soporta documentos, páginas de texto e imágenes que están diseñadas únicamente

para el entendimiento humano. La Web Semántica añade soporte para bases de datos, grandes colecciones de información organizadas para ser procesadas por las máquinas.

3 URI's: Identificadores Uniformes de Recursos

Como hemos visto en el punto anterior, el objetivo de la Web Semántica es el de identificar ítems, nombres de documentos, contenidos de documentos; si queremos buscar algo sobre “un tipo de golosinas”, “el tráfico en Madrid”, “Adolfo Suárez”, todos estos conceptos son ítems en la Web, bien, para identificarlos en la Web, necesitamos usar identificadores, pero eso sí usaremos un sistema uniforme de identificadores y cada ítem identificará a un único recurso. A estos identificadores los llamaremos URI “Uniform Resource Identifiers”. Los URIs pueden identificar cualquier cosa y para la WWW tener una URI significa “estar en la Web”.

Existen diferentes tipos de identificadores URI's, entre otros están los siguientes:

- Identificador Personal : Es una URI que identifica a una persona., <http://www2.uah.es/jcaceres>
- Identificador de Concepto : Identifica el concepto de la última modificación que tuvo lugar, <http://www.uah.es/modificado>.
- Identificador Descentralizado: Llamado también UUID. Se crea por la combinación de tiempo y dirección de la tarjeta de red Ethernet. Es un número aleatorio si no se dispone de uno fijo en la tarjeta, uuid:045455gh-3fgf-5ghg-565f-4343343cgcvc. Este identificador es único, lo que permite crear identificadores sin la utilización de sistemas centralizados como los dominios de nombres.
- Identificador Persistente: También llamado TAG, es una combinación de dirección de mail o nombre de dominio, la fecha en la que el propietario de esa dirección creó el identificador y el nombre del identificador, tag:jesus.caceres@uah.es,22-01-2006:Semantica.
- Identificador Criptográfico: Se denomina también ESL, es un URI que utiliza criptografía en una firma digital para asegurarse que nadie pueda crear identificadores URIS con el nombre de otro. También se utiliza para verificar informaciones, asegurándose de que la clave de descryptación es la misma para la firma de la información como para el URI, esl:SHA2:aFDFdfdfsxEEEWSFDSGFhggfhgfr54546546546443DFDSFDSFddf dsfsf3435435ds45er45fggfdFDG4354gfdG6545FDG546554:Semantica

Esto es una pequeña muestra de los tipos de URIs que se pueden encontrar. De cualquier manera se puede registrar un nuevo esquema URI para un propósito determinado

4 Componentes de la Web Semántica

Podemos decir que la Web Semántica se basa dos conceptos fundamentales:

1. La descripción del significado que tiene los contenidos en la Web
2. La manipulación automática de estos significados.

4.1 La descripción del significado

La descripción del significado requiere la intervención de otros conceptos como:

1. La Semántica, como significado procesable por máquinas
2. Los Metadatos, como contenedores de información semántica sobre los datos. Dentro de este concepto podrían estar los siguientes lenguajes:
 - a. XML: Aporta la sintaxis superficial para los documentos estructurados, pero sin dotarles de ninguna restricción sobre el significado.
 - b. XML Schema: Es un lenguaje para definir la estructura de los documentos XML.
 - c. RDF: Es un modelo de datos para los recursos y las relaciones que se puedan establecer entre ellos. Aporta una semántica básica para este modelo de datos.
 - d. RDF Schema: Es un vocabulario para describir las propiedades y las clases de los recursos RDF, con una semántica para establecer jerarquías de generalización entre dichas propiedades y clases.
 - e. OWL: Añade más vocabulario para describir propiedades y clases tales como relaciones entre clases, cardinalidad (únicamente uno), igualdad, tipologías de propiedades más complejas, caracterización de propiedades (simetría) o clases enumeradas.
3. Las Ontologías: Proviene del campo de la Inteligencia Artificial y son vocabularios comunes para las personas y aplicaciones que trabajan en un dominio, según el W3C, una ontología define los términos que se usan para describir y representar un cierto dominio (cierta área de interés)

4.2 La manipulación automática de los significados

La manipulación automática de los significados se hace a través de dos factores importantes como son:

1. La lógica matemática, que permite establecer reglas para tratar el contenido semántico.
2. Los motores de inferencia, que permite combinar conocimientos conocidos para elaborar otros nuevos conocimientos.

5 El lenguaje RDF

El lenguaje RDF es un lenguaje para representar información sobre recursos en la World Wide Web. Está recogido en 6 recomendaciones del W3C: Primer, Concepts, Syntax, Semantics, Vocabulary (Schema) y Test Cases.

Está particularmente indicado para representar Metadatos sobre recursos web, tales como títulos, autor, modificaciones de los datos de la página web, copyright y otras licencias de información sobre documentos web, así como la disponibilidad para algunos recursos compartidos. Se trata de un modelo de datos para objetos (recursos).

Puede también utilizarse para representar información sobre ciertas cosas que pueden ser identificadas en la Web, aunque no puedan ser directamente recuperadas en la misma, por ejemplo, información sobre artículos disponibles desde servicios on-line (como información sobre especificaciones, precios, disponibilidad, etc..)

RDF está basado en la idea de que los objetos a describir poseen propiedades que a su vez tienen valores. Estos objetos pueden ser descritos formulando ‘declaraciones’ que especifican estas propiedades y valores.

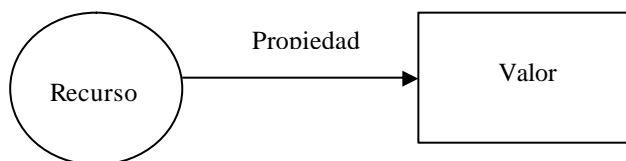


Figura 1 – Para el lenguaje RDF, todos los objetos poseen propiedades que a su vez tienen valores.

RDF fue desarrollado gracias a los auspicios de la W3C (*World Wide Consortium*) y fue diseñado como fundamento para el procesamiento de los metadatos y permitir el intercambio de éstos. Se trata de un modelo que permite definir relaciones semánticas entre distintas URIs asociándoles un conjunto de propiedades y valores.

RDF y el lenguaje de Esquema RDF (*RDFS*) se basó en investigaciones sobre metadatos realizadas por comunidades de Bibliotecas digitales, pudiendo considerarse RDF como una implementación del WF (*Warwick Framework*) donde RDF es una evolución de este último, que permite que cada vocabulario de metadatos posea una sintaxis distinta.

Cabe hacer notar que el lenguaje RDF mismo no es independiente de la sintaxis, ya que está implementado sobre XML¹ (*eXtensible Markup Language*).

¹ Dado que RDF está implementado sobre XML, Los esquemas XML pueden utilizarse para validar la sintaxis de las expresiones RDF/XML.

A diferencia de las Definiciones del Tipo de Documentos (DTDs) de XML y los Esquemas XML (XMLSCHEMA), Un Esquema RDF entrega información sobre la interpretación de las sentencias del modelo de datos RDF y especifica restricciones que deben seguirse por éstos.

Veamos un ejemplo concreto donde se muestran una serie de declaraciones o sentencias: "hay una persona identificada por <http://www.antonio.edu/contact#me>, cuyo nombre es "Antonio López", cuya dirección de correo electrónico es "antonio@uah.es", y cuya Universidad es "Universidad de Alcalá" que podría representarse como el grafo RDF de la siguiente figura:

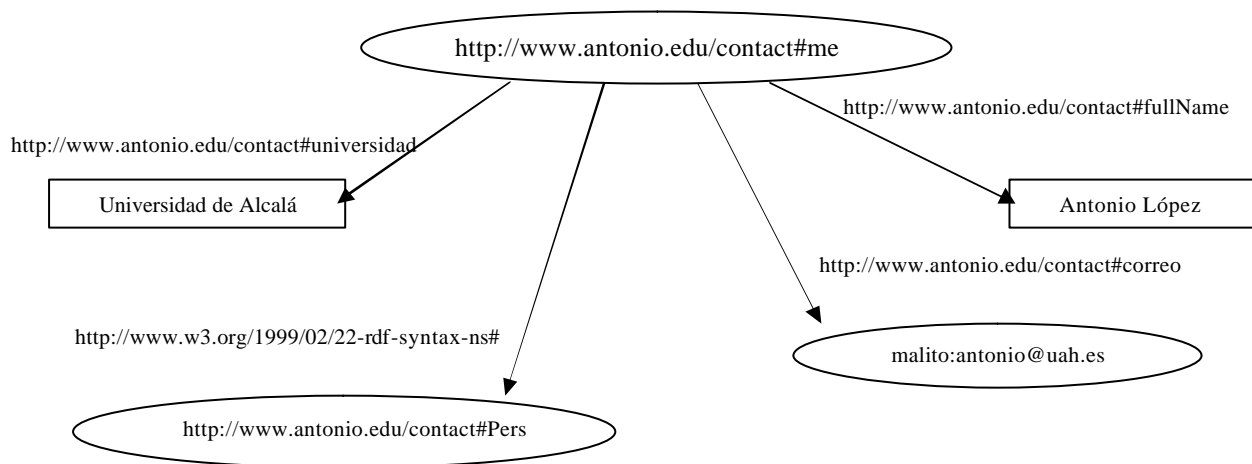


Figura 2 – Grafo RDF que describe los datos de una persona

RDF también provee una sintaxis basada en XML (llamada *RDF/XML*) para guardar e intercambiar estos grafos. Este ejemplo es una pequeña muestra de RDF en RDF/XML correspondiente al grafo de la ilustración anterior.

```

<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:contact="http://www.antonio.edu/contact#">

  <contact:Person rdf:about="http://www.antonio.edu/contact#me">
    <contact:fullName>Antonio López</contact:fullName>
    <contact:correo rdf:resource="mailto:antonio@uah.es"/>
    <contact:universidad>Universidad de Alcalá</contact:universidad>
  </contact:Person>

</rdf:RDF>

```

Este RDF/XML también contiene URIs, y propiedades como `mailbox` y `fullName` (en forma abreviada), y sus respectivos valores `antonio@uah.es`, y Antonio López. Como HTML, este RDF/XML es procesable por la máquina y, usando URIs, se pueden enlazar las piezas de información a lo largo de la Web. Sin embargo, al contrario que el hipertexto convencional, los URIs de RDF pueden hacer referencia a cualquier cosa identificable, incluyendo cosas que pueden no ser directamente recuperables en la Web (tales como la persona Antonio López). El resultado es que, en adición para describir tales cosas como páginas web, RDF puede también describir coches, negocios, gente, noticias, eventos etc. E incluso, las propiedades RDF que tienen los URIs, para precisamente identificar las relaciones que existen entre los términos enlazados. En RDF tanto los sujetos, como las propiedades y los objetos, son recursos.

5.1 Propiedades del lenguaje RDF

Las propiedades básicas de RDF (`'rdfs'` y `'rdf'`) pueden dividirse en distintas categorías: fundamentales, para la definición de esquemas y utilitarios (que en rigor no son absolutamente necesarios, pero son utilizados ampliamente en un gran dominio de aplicaciones).

5.1 Propiedades fundamentales

rdf:Resource En RDF se denomina 'recurso', al igual que en la mayoría de sistemas de metadatos, a cualquier cosa que pueda poseer una URI, por lo que el potencial descriptivo de RDF está sujeto entre otros, a los objetos que puedan poseer una.

rdf:Property Son utilizados como predicados de los triples, la semántica de un triple depende de la 'property' utilizada como predicado. RDF considera las propiedades como clases de primer orden (a diferencia de la mayoría de los lenguajes de modelación donde las propiedades son atributos de una clase).

rdf:Value Es una propiedad que provee RDF para denotar valores primordiales.

rdf:Statement Si se necesita declarar quien es el autor de una declaración en particular, es necesario que la declaración sea un recurso. *rdf:Statement* sirve para tornar una declaración formulada anteriormente en un recurso. Estos recursos debe tener por lo menos 3 propiedades: *rdf:subject*, *rdf:object*, *rdf:Predicate*.

5.2 Propiedades para la definición de esquemas

Permiten la definición de esquemas, es decir, vocabularios de recursos para utilizar con RDF. En los esquemas, nuevos recursos pueden definirse como una 'especialización' de los anteriores. Esta es una poderosa característica de `'rdfs'` dado que en ella radica la extensibilidad en cuanto a elaboración de nuevos esquemas.

rdfs:subPropertyOf Se aplica a las propiedades que pueden ser interpretadas como un subconjunto de otras propiedades. Esta propiedad es importante en los esquemas para efectos de inferencia, por ejemplo si 'violín' es una sub-propiedad de 'instrumento'.

to musical' y algún triple que tenga 'violín' como predicado también debe ser considerado el hecho de tener 'instrumento musical' como predicado. Esta regla denota una relación de subconjunto, por lo que se aplica transitividad.

rdfs:Class Las Clases son recursos que denotan conjuntos de recursos.

rdf:type Denota las instancias que tienen una propiedad valorada por una clase. Las clases son estructuradas al igual que las propiedades, en un subconjunto jerárquico denotado por la propiedad

rdfs:subClassOf. Se aplica a las instancias que tienen una propiedad valorada por una clase. Las clases poseen una clase-universal llamada *rdf:Resource* a la que se aplican las mismas reglas que *rdfs:subPropertyOf*.

rdfs:domain y *rdfs:range* Estas se aplican a propiedades que son valoradas por clases. Ellas son utilizadas para restringir el conjunto de recursos que pueden tener una determinada propiedad (el dominio de la propiedad) y un conjunto de valores válidos para la propiedad (el rango de la propiedad). Una propiedad no puede tener mas de un valor para *rdfs:range*, no así para

rdfs:domain. Estas propiedades son utilizadas entre otras cosas para efectuar chequeos (automáticos) de verificación

5.3 Propiedades para contenedores

Estos podrían ser definidas en esquemas externos, pero como son de uso muy común se definen en el esquema núcleo. Un ejemplo de estas propiedades son *rdfs:label* y *rdfs:comment* que son útiles a la hora de describir un recurso mediante un texto en lenguaje natural. El primero es para dar un nombre y el segundo es para proveer de una descripción más larga.

Se provee de propiedades que materializan las funcionalidades de *Warwick framework*, tales como las colecciones de recursos, obtenidas mediante *rdfs:Container*, propiedad modelada por alguna instancia de estas tres subclases:

<i>rdfs:Bag</i>	colección desordenada
<i>rdfs:Seq</i>	colección ordenada
<i>rdfs:Alt</i>	un alternativo

5.2 Características del lenguaje RDF

RDF posee un modelo de datos simple para su procesamiento por parte de las aplicaciones.

El modelo RDF extendido con el RDF Schema es poderoso, dado que puede utilizarse como un modelo general para expresar metadatos sobre recursos Web. RDF puede ser expresado en una sintaxis XML que permite utilizarlo en muchos ambientes y plataformas.

Es posible conjugar múltiples grafos RDF en uno solo y utilizar los vocabularios definidos en múltiples esquemas.

RDF posee semánticas que generan una base para razonar sobre el significado de una expresión RDF.

RDF posee un vocabulario extensible, basado en URIs.

RDF provee un marco en el cual cualquier persona puede realizar declaraciones simples sobre cualquier cosa, inevitablemente se realizarán declaraciones inconsistentes o sin sentido, dado que todas las personas ven el mundo en forma distinta.

6. Conclusiones.

Gracias a la Web Semántica y a la utilización de lenguajes de Metadatos como el RDF la usabilidad y aprovechamiento de la Web y sus recursos interconectados aumentará gracias a una serie de factores:

- Los documentos etiquetados con información semántica (estilo la etiqueta <meta> de HTML utilizada para facilitar el trabajo a los robots). Se pretende que esta información sea interpretada por el ordenador con una capacidad comparable a la del lector humano. El etiquetado puede incluir Metadatos descriptivos de otros aspectos documentales o protocolarios.
- Los vocabularios comunes de Metadatos (ontologías) y mapas entre vocabularios permiten a quienes elaboran los documentos disponer de nociones claras de cómo deben etiquetarlos para que los agentes automáticos puedan usar la información contenida en los Metadatos (por ejemplo el metadato autor debería significar “autor de la página” y no “autor del objeto descrito en la página”)
- Los agentes automáticos que realizan tareas para los usuarios de estos metadatos de la Web Semántica.
- Los servicios web que proporcionan información a los agentes (por ejemplo un servicio de garantías a quien un agente pudiera consultar sobre si un comercio electrónico tiene un historial de mal servicio o de generar correo basura).

Los proveedores primarios de esta tecnología son las URIs que identifican los recursos junto con XML y los namespace. Si a esto se añade un poco de lógica mediante RDF, la Web estará lista para alcanzar las aspiraciones de su inventor Tim Berners-Lee.

Referencias

- Juan Carlos Camara. Posibilidades de Websemantico http://www.iua.upf.es/~jblat/material/doctorat/students/jccbis/Web_Semantico.htm
- MÉNDEZ RODRÍGUEZ, Eva María. RDF: Un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio. <http://www.cobdc.org/jornades/7JCD/1.pdf>
- RDF Model Theory.W3C Working Draft 29 April 2002. <http://www.w3.org/TR/rdf-mt/#graphsyntax>
- RDF Tutorial. Pierre-Antoine Champin. <http://www710.univ-lyon1.fr/~champin/rdf-tutorial/rdf-tutorial.html>
- Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Working Draft 08 November 2002. <http://www.w3.org/TR/2002/WD-rdf-concepts-20021108/>
- W3C. Resource Description Framework (RDF). <http://www.w3.org/RDF/>