

WEB SEMÁNTICA, MÁS DE UNA DÉCADA DE SU APARICIÓN

D. Bernal¹, A. Castro², J. González³, Grupo de Investigación en el Manejo de la Información – GIMI Universidad Pedagógica y Tecnológica de Colombia,

Recibido enero 7 de 2014 – Aceptado abril 22 de 2014

<http://dx.doi.org/10.18566/puente.v8n1.a07>

Resumen – Mucho se habla de Web Semántica, sin embargo su implementación ha sido escasa debido a la falta de conocimiento de las tecnologías que la componen. En el artículo se describen la Web Semántica y sus antecedentes, posteriormente se presentan las tecnologías que la soportan de una forma sencilla y gráfica, para finalmente reflexionar sobre el papel del usuario final en la Web Semántica y sobre el futuro de esta.

Palabras clave – Web Semántica, XML, RFD, OWL, SPARQL.

Abstract – Much is talk about Semantic Web, however its appliance has been poor due to the lack of knowing about technologies composing it. In this article there are described Semantic Web and its background; then are shown its backing technologies in a simple and graphic way and finally there is a reflection about final user role on Semantic Web and its future.

Key words – Semantic Web, XML, RFD, OWL, SPARQL.

I. INTRODUCCIÓN

La importancia que ha tomado el proceso de captura y manejo de la información, junto a las diversas necesidades organizacionales y personales de los usuarios de Internet, han motivado el desarrollo de tecnologías que pretenden ser un medio para garantizar un acceso más simple y natural a esta información, que arrojen resultados de búsqueda que correspondan con lo que el usuario necesita y permitan un mejor acceso al conocimiento.

Expertos en el tema del manejo de la información, describen la sociedad actual como una “Sociedad de la Información” [1], y resaltan la importancia que las empresas otorgan a la monitorización del conocimiento; “las principales estrategias de negocios sostienen que el conocimiento está remplazando al capital financiero como el recurso clave para las empresas”.

¹ D. Bernal, Universidad Pedagógica y Tecnológica de Colombia, email: diego.bernal@uptc.edu.co

² A. Castro, Universidad Pedagógica y Tecnológica de Colombia, email: alexander.castro01@uptc.edu.co

³ J. González, Universidad Pedagógica y Tecnológica de Colombia, email: juansebastian.gonzalez@uptc.edu.co

Este hecho demuestra que la información se ha convertido en el activo esencial para la toma de decisiones en cualquier ámbito dentro de las organizaciones.

En este sentido, y a pesar de los esfuerzos que se llevan a cabo para optimizar la gestión de la información, persisten diversos inconvenientes en aspectos aparentemente sencillos como la recolección, almacenamiento y representación de los datos que se manejan, pues aunque “es cierto que muchas de las organizaciones confían en las tecnologías de la información (TI) para apoyar muchos de sus procesos de información, también hay una gran cantidad de información y conocimiento que no es capturado o representado por estos sistemas basados en computadores” [2].

Por otro lado, cabe destacar que es en Internet donde se presenta un mayor problema con el manejo de la información, esto se debe principalmente a dos factores: la cantidad de datos que se maneja, el tráfico IP global se ha multiplicado por ocho en los últimos cinco años y se triplicará en los próximos cinco años, alcanzando en 2016 un tráfico IP mundial de cerca de 1,3 Zettabytes (1.073.741.824 Terabytes) por año [3]; y la complejidad de la información, debido a que, cada día, los elementos tienen más características y relaciones lo que aumenta dicha complejidad.

Adicionalmente “la mayoría de las páginas Web aún están diseñadas para el consumo por parte de los humanos y no para ser procesadas por máquinas, los computadores solo están siendo usados para mostrar la información” [4], se podría agregar a esta idea que los computadores no están entendiendo lo que muestran y por eso la recuperación de la información resulta tan difícil y tan deficiente, este último planteamiento se viene dando desde hace mucho tiempo, “el mayor problema que enfrentan los usuarios de motores de búsqueda en la Web hoy en día es la calidad de los resultados que obtienen. Si bien los resultados son a menudo entretenidos y amplían los horizontes de los usuarios, suelen ser frustrantes y consume un tiempo precioso” [5].

II. WEB SEMÁNTICA

Partiendo de los principios presentados en el aparte anterior, surgió la necesidad de buscar una estrategia que contribuyera a resolver algunos de los problemas relativos al manejo de la información como son, gestión de grandes cantidades de datos, recuperación de la información en datos no estructurados y la falta de relaciones entre los datos. Posiblemente, el esfuerzo más grande en este campo se dio en 2001, cuando Tim Berners Lee, considerado el padre de la Web y miembro del World Wide Web Consortium (W3C), propone el concepto de “Web Semántica”, lo que originó un movimiento que lleva aproximadamente doce años intentando colaborar con la situación mencionada; Lee junto con unos compañeros, proponen dos ideas fundamentales sobre la Web Semántica, la primera plantea esta como “una extensión de la Web actual en la cual la información tiene un significado bien definido, propiciando el trabajo cooperativo entre ordenadores y personas” y la segunda menciona que “la Web Semántica aporta estructura al contenido significativo de las páginas Web, creando un entorno donde los agentes de software se moverán de una página a otra fácilmente llevando a cabo tareas sofisticadas para los usuarios” [6], proporcionando las bases para lo que hoy se conoce como “Web Semántica”.

La W3C aporta una definición más técnica de Web Semántica, “la visión de la Web Semántica es ampliar los principios de la Web, de los documentos a los datos. Los datos deben ser accedidos utilizando la arquitectura Web general, por ejemplo, URI (Uniform Resource Identifier), los datos deben estar relacionados entre sí de la misma forma que están relacionados los documentos (o partes de documentos). Esto también significa la creación de un marco común que permite que la información sea compartida y reutilizada a través de la aplicación, para ser procesada automáticamente por herramientas, así como de forma manual, incluso revelando posibles relaciones nuevas entre los pedazos de datos” [7].

En la actualidad se maneja un concepto más amplio que define la Web Semántica como un movimiento que busca darle a la Web una estructura definida y cambiar el manejo actual de la información (dando una representación más formal a esta), apoyándose en el uso de tecnologías que permitan pasar de estructuras desorganizadas y aisladas a estructuras que permitan el manejo de datos (y metadatos) fuertemente relacionados, logrando a su vez una mejor (más subjetiva y/o más rápida) y más fácil automatización de los procesos de recuperación y manejo de la información.

Si bien, el concepto no ha cambiado y la misión de la Web Semántica sigue intacta, son evidentes dos cosas: sus postulados no han llegado a aplicarse [8] y no existe un camino claro para su implementación [9]; una de las causas de esto puede ser el desconocimiento que se tiene al iniciar el tema y la dificultad de acceder a información clara y precisa sobre el tema.

III. TECNOLOGÍAS DE LA WEB SEMÁNTICA

Como se mencionó inicialmente, la Web Semántica se vale de un conjunto de tecnologías para cumplir sus fines, sin embargo para alguien que inicia en el tema se hace difícil distinguir las funciones y características de cada tecnología, cuáles son las más relevantes y en que radica la diferencia entre estas. Es por ello que a continuación se presentan las tecnologías más relevantes de la Web Semántica de una manera sencilla y se hace una introducción adecuada a cada una:

A. XML(*Extensible Markup Language*)

En un esfuerzo por dar una estructura a los documentos de la Web, en 1996 nace el lenguaje de marcado extensible (XML, por sus siglas en inglés), que es un formato de texto simple y muy flexible derivado de SGML (ISO 8879); si se quiere hablar de Web semántica es obligatorio hablar de XML, debido a que este lenguaje es el soporte de muchas de sus tecnologías [10].

XML describe la sintaxis para marcar un documento al tiempo que se crea una sencilla relación entre las palabras que estructuran el mismo. De igual modo, este lenguaje proporciona un mecanismo para imponer restricciones en el diseño de almacenamiento y una estructura lógica [11].

XML pretende darle una estructura a un documento de texto sin afectar la simplicidad de este, usando etiquetas para hacerlo más entendible a las máquinas y más adecuado para su manejo en Internet.

Este lenguaje ha sido acogido ampliamente, convirtiéndolo en la referencia para todos los lenguajes del mismo tipo, y adaptado a diversas circunstancias debido a que se plantea como un estándar para el intercambio de información estructurada que parte del postulado “los computadores son estúpidos por eso hay que especificarles todo”, por ejemplo, si a una persona se le dan indicaciones de la ubicación de un lugar es muy probable que llegue, sin embargo, si se trata de una máquina, será necesario especificar la dirección exacta del sitio de una manera que la pueda comprender.

Para entender el funcionamiento de XML, se puede recurrir al caso de una biblioteca, donde un libro tiene diversos atributos como un título y autor, de modo tal que es posible vincularlos en este lenguaje tal como se evidencia en la Fig. 1.

```
<?xml version="1.0" encoding="UTF-8"?>
<libro titulo="E-data" id="1">
  <autor>Jill Dyché</autor>
</libro>
```

Fig. 1. Ejemplo XML

En este sentido, se dice que el “libro” tiene características como un “título” y un “autor”, de modo tal que un intérprete podría relacionarlos. Pero es aquí donde se encuentra el primer inconveniente, XML fue diseñado como un lenguaje para marcar documentos y no como un modelo que sirva para dar significado a un conjunto de datos, es decir, para una persona es claro que “título” y “autor” corresponden a características propias de un “libro”, pero una máquina no sabe esto, pues la relación no está expresada explícitamente. La relación entre libro y autor podría significar cualquier cosa como una dependencia, asociación de

características o relación padre-hijo, por citar algunos ejemplos.

Y es la flexibilidad de XML la que impide crear una estructura lógica robusta, eficiente y totalmente descriptiva ajustada a las necesidades de la Web Semántica, además su estructura (de árbol) impide modelar muchas situaciones del mundo real. Se puede concluir que XML se queda como una sintaxis (que cumple muy bien sus funciones de transporte y almacenamiento de la información) pero no alcanza a llegar a la complejidad de un modelo de datos.

En un esfuerzo por hacer más descriptivo a XML nace XMLS (XML Schema), el cual expresa vocabularios compartidos y permite a las máquinas verificar las normas hechas por las personas [12]. Este proporciona un medio para definir la estructura, el contenido y la semántica de los documentos XML; en la Fig. 2. se hace uso de XSD, en este se aprecia una “introducción” a lenguajes más robustos (pero no necesariamente más complicados) como RDF y RDFs los cuales serán abordados más adelante.

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="libro">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="autor" type="xs:string" />
      </xs:sequence>
      <xs:attribute name="titulo" type="xs:string" />
      <xs:attribute name="id" type="xs:string" />
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Fig. 2. Ejemplo de XMLS (XSD).

Y aunque con el uso de XMLS los documentos se vuelven más descriptivos y formales, como se puede notar en el ejemplo, no se contribuye a dotar a la Web de significado, pues aún quedan serios vacíos en el manejo de relaciones y metadatos.

En este punto se puede concluir dos cosas, la primera, XML está diseñado para trabajar con la información, pero se queda corto cuando se hace necesario trabajar con los metadatos, “desde el principio, parte del proyecto Web Semántica ha sido la de crear un estándar hermano para los metadatos”, en este caso se habla de RDF (Resource Description Framework), tecnología que se profundizara en el siguiente bloque [13], y la segunda, XML cumple las funciones para las cuales fue diseñado y es una tecnología que provee una buena estructura, base de otras tecnologías de la Web Semántica.

Para alguien que desee dotar una página Web con algo de sentido más no de semántica, XML se puede adaptar a sus necesidades y además es un buen punto de partida para entrar a trabajar con otras tecnologías de la Web Semántica, sin embargo llegara un punto donde necesite relacionar los datos de una forma más compleja para ello tendrá que usar otras tecnologías como las que se explicaran más adelante.

B. RDF (Resource Description Framework)

Los metadatos son la clave para la recuperación de la información, un metadato es la información acerca de la información, son datos que describen otros datos, por ejemplo una canción puede tener un metadato que indique a que género musical pertenece y la canción puede existir sin esta

información, pero si se conoce, da una pista del tipo de canción y por lo tanto facilita su recuperación.

RDF es “un lenguaje para representar información sobre recursos en la World Wide Web” [14], el cual está especialmente diseñado para representar metadatos sobre recursos web, como el título, autor y fecha de modificación de una página Web, los derechos de autor y licencias sobre un documento Web, o el calendario de disponibilidad de algún recurso compartido”, una de las principales características es que RDF está diseñado para trabajar con información que puede ser intercambiada entre aplicaciones sin pérdida de significado y a su vez se basa en la idea de relacionar objetos que utilizan identificadores Web (llamados Identificadores Uniformes de Recursos, o URIs) y la descripción de los recursos en términos de propiedades simples y valores de propiedad.

Adicionalmente, recordando que “gran parte de la motivación para la Web Semántica viene del valor encerrado en bases de datos relacionales. Para liberar este valor, los objetos de base de datos deben ser exportados a la Web como objetos de primera clase, por lo que se deben asignar a un sistema de URIs” [15], en este punto se resalta la importancia de las URIs como un sistema, no solo para enlazar sino para identificar los recursos, permitiendo el paso de los documentos a los datos.

Teniendo la definición planteada es destacable una diferencia entre XML y RDF, la primera tiene una estructura de árbol rígido, sin la posibilidad de relaciones mientras que la segunda incorpora simples declaraciones acerca de los recursos en forma de grafo donde los nodos representan los recursos y los arcos sus propiedades y valores. Además XML no tiene una semántica inherente, es decir es demasiado flexible para ser considerado un lenguaje de representación de datos formal, por otro lado RDF tiene semántica declarativa específica y esta es independiente del contexto en que se encuentre, lo que le da mayor robustez como lenguaje de representación de datos [16].

Como se aprecia, la naturaleza de RDF está basada en etiquetas a las cuales básicamente se le asigna la estructura de la Fig. 3.

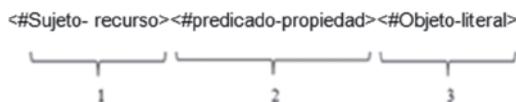


Fig. 3. Estructura RDF.

A esto, se le conoce como tripleta o declaración, una visión sencilla pero muy funcional de la estructura básica que sirve como lineamiento para describir un dato. El modelo de datos de RDF está basado en URI, un identificador único que permite relacionar los datos, estos son el recurso base para las tripletas. Así, la *propiedad* 2 es una característica utilizada para describir y enlazar un *recurso* 1 con su *objeto* 3. El *objeto* puede ser otro *recurso* u otro *literal*. De este modo, es posible crear un conjunto de tripletas RDF comúnmente conocidas como Grafos RDF, que brindan versatilidad y consistencia haciendo de estas una excelente forma para sentar las bases de la Web Semántica. Es importante destacar que los nombres de recursos en la web semántica deben cumplir dos requisitos: en primer lugar, una descripción del recurso identificado debe ser recuperable con tecnologías Web estándar, en segundo lugar, un esquema de nombres no debe confundir los documentos y las cosas descritas en los documentos [17].

De esta forma y retomando el ejemplo planteado en la Fig. 1., es posible escribir una sentencia RDF que describa las propiedades de un libro, Fig. 4.

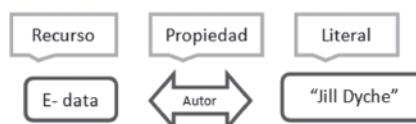


Fig. 4. Sentencia RDF para descripción de un libro.

A su vez, este sencillo grafo puede ser extendido de modo tal que la última parte del mismo, no sea un literal sino otro recurso, como se presenta en la Fig. 5.

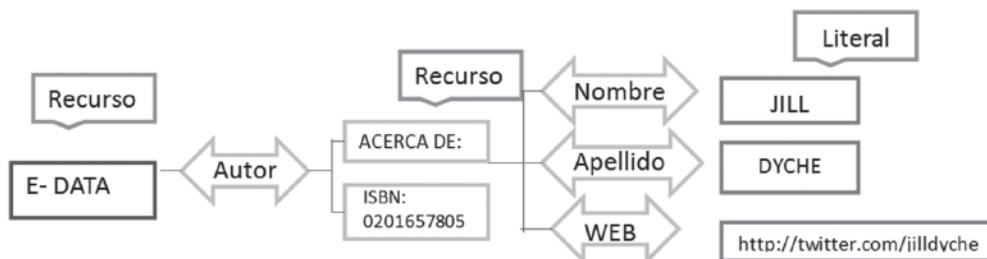


Fig. 5. Extensión de Sentencia RDF.

Consecuentemente, pueden irse agregando recursos en la medida que se requiera.

El código RDF para el grafo de la Fig. 4 se vería de la manera expresada en la Fig. 6.

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="E-data">
    <dc:creator>Jill Dyche</dc:creator>
  </rdf:Description>
</rdf:RDF>
```

Fig. 6. Código RDF para el grafo de la Fig. 4.

Un término que vale la pena aclarar es “XML Namespace”, este es un contenedor abstracto que congrega distintos identificadores únicos, cada uno de los cuales puede tener su propia descripción, se podría hacer un símil con un diccionario en donde cada palabra tiene su significado, y todas las palabras se encuentran organizadas de una forma estructurada. En el ejemplo anterior, se definen dos namespaces, uno para RDF y otro para DC (Dublin Core Metadata Initiative) los cuales tienen identificadores como “Description”, “About” o “Creator”, este se denota con la sigla “xmlns”, lo que indica que se asocian una serie de etiquetas (como rdf o dc, Ejemplo 3) a un namespace en específico, de modo que se tengan acceso a sus identificadores. En palabras simples un namespace es la dirección a un catálogo de etiquetas que permiten describir recursos en RDF.

Es posible concluir que RDF es una tecnología esencial para la Web Semántica, se basa en XML y permite describir recursos de una forma estándar teniendo en cuenta sus metadatos y el uso de URIs y Namespaces. Además, con esta tecnología se añade significado a las páginas web y su uso no es complejo. Así como para XML existe XMLs, en RDF hay **RDFS** (RDF Schema) que pretende hacer más descriptivo a RFD, este permite clasificar los recursos, permitiendo al software entender de una mejor forma el objeto que se pretende describir. La ventaja de usar RDFS es que al agregar etiquetas como “type” se permiten establecer relaciones jerárquicas y de herencia para los recursos. Siguiendo con el ejemplo, si se quiere describir un objeto que es una instancia de una clase, se usa la etiqueta “Class” y se hace la descripción del grafo como de costumbre, con la ventaja de que se pueden crear relaciones más complejas entre los recursos.

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://librosEjemplo.com/libros/E-data">
    <rdf:type rdf:resource="http://librosEjemplo.com/libros#Libro"/>
    <dc:title>E-data</dc:title>
    <dc:creator>Jill Dyche</dc:creator>
  </rdf:Description>
  <rdfs:Class rdf:about="http://librosEjemplo.com/libros#Libro">
    <rdfs:label>Libro</rdfs:label>
  </rdfs:Class>
</rdf:RDF>
```

Fig. 7. Sentencia RDF con uso de Class.

En la Fig. 7. se define la clase libro, y a la instancia se le indica que es una subclase de libro y se le asignan propiedades, este planteamiento se acerca más a los principios de la programación orientada a objetos, lo que hace que la descripción de recursos sea más natural y completa.

C. OWL (Web Ontology Language)

Es una extensión del modelo de datos de RDF el cual busca describir el significado implícito y el por qué se da determinada conceptualización al contexto de cierto elemento. OWL provee un conjunto de

semánticas para la construcción de modelos de datos complejos, vocabularios y lógica del software; es decir, se puede definir como una especie de conjunto de herramientas orientado a objetos, que relaciona las tripletas de RDF a clases, asociaciones y otras relaciones complejas [18]. Con OWL el usuario define clases, las cuales pueden ser sub-clases e instanciadas, al tiempo que las propiedades permite definir los atributos de cada clase, además permite definir

relaciones entre las clases, cardinalidad de atributos, equivalencia ya sea entre clases o entre propiedades, clases enumeradas, propiedades de los atributos entre otras operaciones que se le adicionan a RDF [19]. Como se aprecia en la Fig. 8., se definen propiedades del objeto y de los atributos, lo que permite relacionar las instancias de recursos autor con la instancia del recurso libro específico dándole un nombre a esta relación, por lo tanto añadiéndole significado.

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:Libro="http://librosEjemplo.com#Libro"
  xmlns:Persona="http://librosEjemplo.com#Persona"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

  <!-- Object Properties -->
  <owl:ObjectProperty rdf:about="http://librosEjemplo.com#esAutorDe"/>

  <!-- Data properties -->
  <owl:DatatypeProperty rdf:about="http://librosEjemplo.com#tieneTitulo"/>

  <!-- Instancias -->
  <rdf:Description rdf:about="http://librosEjemplo.com#Dyche-Jill">
    <rdf:type rdf:resource="http://librosEjemplo.com#Persona"/>
    <Persona:esAutorDe rdf:resource="http://librosEjemplo.com#E-data"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://librosEjemplo.com#E-data">
    <rdf:type rdf:resource="http://librosEjemplo.com#Libro"/>
    <Libro:tieneTitulo>E-data</Libro:tieneTitulo>
  </rdf:Description>
</rdf:RDF>
```

Fig. 8. Ejemplo uso OWL.

En este punto se hace necesario definir una ontología como “vocabulario de términos que se formalizan, a menudo cubriendo un dominio específico y que es compartido por una comunidad de usuarios” [20], este es un acercamiento a lo que es la semántica en el mundo cotidiano, no es lo mismo la jerga de un policía que de un profesor, así usen el mismo término en el mismo contexto se pueden referir a cosas completamente diferentes, pues se manejan relaciones, términos y significados diferentes en cada contexto, por ello se hace necesario agrupar términos y sus significados en vocabularios formales u ontologías.

La importancia de OWL radica en el uso de ontologías, pues “uno de los objetivos más importantes de la Web Semántica es facilitar el compartir datos y su integración a través del uso de ontologías. La Web Semántica no se basa en una sola enorme ontología, que lo abarca todo ya que alcanzar un consenso mundial en el vocabulario no es factible. Más bien se alienta al desarrollo de los pequeñas ontologías que se

puedan compartir y su reutilizar para la interoperabilidad de los datos” [21]. Un ejemplo de una ontología es la clasificación de productos y sus funciones realizada por una tienda virtual como Amazon.

Definir una ontología puede ser complejo pero partiendo de un modelo de datos y OWL es posible, sin embargo “no existe una sola ontología correcta para cualquier dominio” [22], el diseño de ontologías es un proceso creativo y por lo tanto puede ser subjetivo.

En este punto se puede concluir que OWL permite agregarle valor (mediante la implementación de nuevas etiquetas) a las relaciones y descripciones de RDF [23], pues se amplía su modelo de datos y se lleva más hacia un paradigma orientado a objetos, convirtiéndose en un lenguaje de modelado ideal para la creación de ontologías web, que permite mejor integración con las tecnologías de manejo de la información actuales y una descripción de recursos más natural y completa.

D. SPARQL (Protocol and RDF Query Language)

Hasta el momento se han mencionado lenguajes para el modelo de información, pero estos serían poco útiles sin el respaldo de un lenguaje que permita la recuperación de dicha información.

SPARQL es un lenguaje de consultas para RDF y también un protocolo para transmitir estas consultas, “se puede utilizar para expresar consultas a través de diversas fuentes de datos, almacenados de forma nativa como RDF o visto como RDF a través de middleware. SPARQL contiene funciones para realizar consultas de patrones requeridos y opcionales en grafos, junto con sus conjunciones y disyunciones. SPARQL también soporta la agregación, subconsultas, la negación, la creación de los valores de las expresiones, las pruebas de valor extensible y consultas restrictivas de grafo RDF fuente. Los resultados de consultas SPARQL pueden ser conjuntos de resultados o gráficos RDF” [24], en resumen este es un lenguaje de consultas en crecimiento, pero sin la posibilidad de operaciones CUD (Creación, Actualización y Borrado) sin embargo a diferencia de otros lenguajes de consultas, permite combinar varias fuentes de información de forma dinámica, según sea necesario para crear nuevas fuentes de información. Una consulta en SPARQL para obtener los libros que tenga como autor a “Jill”, Fig. 9.

```
PREFIX dc:<http://purl.org/dc/elements/1.1/>
SELECT ?title
WHERE
{ ?title dc:creator ?Creator}
```

Fig. 9. Estructura consulta SPARQL.

Un ejemplo de aplicación SPARQL se da con Watson (computador con inteligencia artificial) de la empresa IBM (International Business Machines), “Watson proporciona diferentes ‘perspectivas’, desde la búsqueda más simple por palabra clave, hasta consultas complejas y estructuradas con SPARQL” [25]. En conclusión SPARQL “permite acercarse a la naturaleza semi-estructurada de los grafos RDF en el mundo real, con la ayuda de patrones opcionales y alternativos” [26], lo que significa que brinda la posibilidad de hacer consultas sencillas con una sintaxis similar a la de SQL (Structured Query Language) sobre recursos RDF.

E. HTML5 (HyperText Markup Language)

Hasta el momento se han mencionado varias tecnologías que ayudan a cumplir los principios de la Web Semántica, sin embargo para el usuario común,

quien es el que mayor cantidad de contenido genera, también existen tecnologías más simples para que contribuya a dotar de sentido la Web.

HTML es un lenguaje para definir la estructura gráfica de una página Web, a pesar de que se podría pensar que su función se limita a la parte estética, la W3C ha fortalecido el estándar con la incorporación de nuevos elementos como etiquetas semánticas que permiten dotar a estos componentes gráficos de significado e importancia dentro de la estructura de la página [27]. Cada una de estas etiquetas indica el contenido que deberían poseer, de este modo, aplicaciones web como buscadores pueden clasificar de mejor forma la información que posee una página.

```
<!DOCTYPE html>
<html Lang="es">
  <head></head>
  <body>
    <nav></nav>
    <section></section>
    <aside></aside>
    <footer></footer>
  </body>
</html>
```

Fig. 10. Etiquetas en HTML5

Por ejemplo, se puede pensar en “<nav>” como un conjunto donde se encuentran varios enlaces de navegación ósea categorías de gran orden jerárquico en que se divide la página. Estas nuevas etiquetas responden a las necesidades de los usuarios de la Web, las cuales radican en contar con “mecanismos en HTML que claramente y sin ambigüedades permitan a los desarrolladores agregar semántica más rica y significativa, no de pseudo-semántica a su margen de beneficio. Este es quizás el objetivo más urgente para el proyecto HTML5” [28], esto tiene una sencilla explicación, en un afán de tener un mejor posicionamiento en los buscadores, los creadores de páginas web no usan adecuadamente las etiquetas a su disposición, esto presenta un reto tanto para los motores de búsqueda como para los creadores de HTML5, no basta con crear etiquetas semánticas sino que hay que crear mecanismos que obliguen a un uso adecuado de éstas, por ello uno de los temas más nuevo pero importantes en Web Semántica son las “redes de confianza” [29], puedes de que sirve tener tecnologías que ayuden a darle sentido a la información si esta no es confiable.

Por otra parte es interesante ver como la integración de HTML5 con tecnologías más propias de Web semántica como RDF se está dando a un nivel transparente para el usuario final, destacando

que la combinación de estas tecnologías ya ha sido usada para mejorar la búsqueda de videos en la Web [30].

Aunque este es un pequeño esfuerzo para organizar la estructura de una página Web y es un punto de inicio interesante para alguien que quiera aplicar los principios de la Web Semántica, con HTML5 solo se puede dar significado a la estructura de la página pero no a su contenido.

Para futuras versiones de HTML se debería considerar más etiquetas semánticas que se orienten a la organización del contenido en sí, a los metadatos y a la inclusión de tecnologías como RDF al alcance de las personas del común.

IV. CONCLUSIONES

La Web Semántica es un movimiento que busca darle a la Web una estructura definida y cambiar el manejo actual de la información (dándole una representación más formal a esta), se apoya en el uso de tecnologías que permitan pasar de estructuras desorganizadas y aisladas a estructuras que permitan el manejo de datos (y metadatos) fuertemente relacionados. Esta tiene como objetivo lograr una mejor (más subjetiva y/o más rápida) y más fácil automatización de los procesos de recuperación y manejo de la información.

En la Web Semántica existen muchas tecnologías, se toma como base XML, para construir RDF que tiene una estructura bien definida, luego con OWL se puede agregar complejidad en las relaciones y atributos de RDF, con SPARQL se puede consultar documentos RDF y por ultimo aunque HTML5 aunque por el momento solo define la estructura visual de las pagina Web en un futuro la dará significado al contenido por medio de etiquetas que sean fáciles de usar para el usuario final. Aunque las tecnologías de Web Semántica parecen complejas, con un juicioso estudio de ellas, es posible dotar a la Web de significado, sin embargo es importante tener en cuenta dos cosas importantes, la primera que los que más generan contenido en la web y quienes la van a dotar de significado son los usuarios finales y hay que concentrarse en ellos y hacer que las tecnologías les lleguen de una manera amable y en lo posible transparente, y la segunda cuestión de importancia es que se va a ser necesario que la gente aprenda a usar estas tecnologías con información verídica ósea hay que contribuir a crear redes de confianza.

REFERENCIAS

- [1] M. Powell, "Information, development and global change" in Information Management: For Development Organizations, 2nd ed., Oxfam, pp. 5-7, 2003.
- [2] M. Hinton, "Introducing Information Management: the business approach" in Introducing Information Management, Routledge, pp. 1-7, 2006.
- [3] Cisco, *Cisco Visual Networking Index: Forecast and Methodology*, 2011-2016, Cisco, 2012, Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf.
- [4] K. Breitman, et. al., "Semantic Web: Concepts, Technologies and Applications", Springer, United States, 2007.
- [5] S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine" in: Seventh International World-Wide Web Conference, April 14-18, 1998.
- [6] T. Berners-Lee et. al., "The Semantic Web", Scientific American, vol. 284, no. 5, pp. 2837, May 2001, Available: <http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html>.
- [7] Editores W3C, "W3C Semantic Web Frequently Asked Questions", W3C, 2009, Available: <http://www.w3.org/RDF/FAQ>.
- [8] Editores W3C, "Extensible Markup Language (XML)", W3C, 2012, Available: <http://www.w3.org/XML/>.
- [9] S. Dobson, "The semantic web: good ideas poorly supported?", Simondobson, Feb 2012, Available: <http://www.simondobson.org/2012/02/semantic-web/>.
- [10] D. Zambonini, "The 7 (f) laws of the Semantic Web", Oreillynet, Jun 2006, Available: http://www.oreillynet.com/xml/blog/2006/06/the_7_flaws_of_the_semantic_we.html.
- [11] T. Bray, et. al., "Extensible Markup Language (XML) 1.1 (Second Edition)", W3C, 2006, Available: <http://www.w3.org/TR/xml11/>.
- [12] D. Fallside and P. Walmsley, "XML Schema Part 0: Primer Second Edition", W3C, 2004, Available: <http://www.w3.org/TR/xmlschema-0/>.
- [13] J. Bosak and T. Bray, "XML and the Second-Generation Web", Scientific American, vol. 280, pp. 89-93, 1999.
- [14] E. Miller et. al., "RDF Primer", W3C, 2004, Available: <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>.
- [15] J. Pollock, "Semantic Web for Dummies", Wiley, pp. 84, 2009.
- [16] S. Decker, et. al., "The Semantic Web: The roles of XML and RDF", IEEE Internet Computing, vol. 15, pp. 63-74, Oct 2000.
- [17] L. Sauermaun, et. al, "Cool URIs for the semantic web", German Research Center for Artificial Intelligence, 2007, Available: http://scidok.sulb.unisaarland.de/volltexte/2011/3944/pdf/TM_07_01.pdf.
- [18] N. Shadbolt, et. al., *The Semantic Web Revisited*, Intelligent Systems, IEEE, vol. 21, pp. 96-101, Jan 2006.
- [19] J. Golbeck et. al., "Organization and Structure of Information using Semantic Web Technologies", University of Maryland, 2009, Available: <http://www.mindswap.org/papers/Handbook.pdf>.
- [20] Editores W3C, "OWL 2 Web Ontology Language Document Overview (Second Edition)", W3C, 2012, Available: <http://www.w3.org/TR/owl2-overview/>.
- [21] H. Lushan, et. al., "Finding Semantic Web Ontology Terms from Words", Proceedings of the Eighth International Semantic Web Conference, Springer, 2009.
- [22] N. Noy, et. al., "Ontology Development 101: A Guide to Creating Your First Ontology". Stanford Knowledge Systems Laboratory Technical Report, Mar 2001, Available: http://protege.stanford.edu/publications/ontology_development/ontology101-noymcguinness.html.

- [23] D. Tester, “*RDFS vs. OWL*”, Cambridge Semantics Inc, 2013, Available: <http://www.cambridgesemantics.com/semantic-university/rdfs-vs.-owl>.
- [24] Editores W3C, “*SPARQL 1.1 Query Language*”, W3C, 2013, Available: <http://www.w3.org/TR/sparql11-query/>.
- [25] E. Motta and M. d’Aquin, “*Watson, more than a Semantic Web search engine*”, Semant Web, vol. 2, pp. 55-63, Jan 2011.
- [26] P. McCarthy, “*Search RDF data with SPARQL*”, IBM Developers Works, May 2005, Available: <http://www.ibm.com/developerworks/xml/library/j-sparql/>.
- [27] E. Lewis, “*Using HTML5’s New Semantic Tags Today*”, MSDN Magazine - Script Junkie, Nov 2010, Available: <http://msdn.microsoft.com/en-us/magazine/gg454786.aspx>.
- [28] J. Allsopp, “*Semantics in HTML 5*”, A list Apart, Jan 2009, Available: <http://alistapart.com/article/semanticsinhtml5>.
- [29] J. Golbeck, et. al., “*Trust Networks on the Semantic Web*”, Proceedings of Cooperative Intelligent Agents, Springer, pp. 238-249, 2003.
- [30] L. Sikos, “*Advanced (X) HTML5 Metadata and Semantics for Web 3.0 Videos*”, DESIDOC Journal of Library & Information Technology, vol. 31, pp. 247-252, Jul 2011.

Enviado para postulación de publicación el día 07 de abril de 2014. Afiliación de los autores: Universidad Pedagógica y Tecnológica de Colombia.

BIOGRAFÍA



Diego Mauricio Bernal Torres, finalizó los estudios de Ingeniería de Sistemas y Computación en la Universidad Pedagógica y Tecnológica de Colombia – UPTC en el año 2013. Actualmente cursa la Especialización en Bases de Datos en la misma Institución y ha sido participe de investigaciones relacionadas con NFC, Inteligencia de Negocios y Web Semántica.



Alexander Castro Romero, Ingeniero de Sistemas y Computación de la Universidad Pedagógica y Tecnológica de Colombia en el año 2013. Actualmente ejerce como docente de la Institución donde se graduó. Es estudiante de la Maestría en Tecnología Informática y su área de investigación principal es la de “nuevas tecnologías para el manejo de la información”.



Juan Sebastián González Sanabria, actualmente se desempeña como docente de tiempo completo en la Universidad Pedagógica y Tecnológica de Colombia UPTC. Finalizó sus estudios de Ingeniería de Sistemas y Computación en el año 2011, y posterior a esto curso estudios de Especialización en Bases de Datos, ambas en la UPTC. Desde el año 2009 ingresó como semillero investigador al Grupo de Investigación en el Manejo de la Información – GIMI –, donde actualmente realiza la labor de investigador en diferentes proyectos.