

WEB SEMÁNTICA APLICADO A LA CARRERA DE INGENIERÍA EN SISTEMAS

León Richar
UNL CIS-loja

Elivar Largo
UNL CIS-loja
Affiliation

Edgar Tuza
UNL CIS-loja

En el presente artículo realizaremos la Web Semántica aplicada a la carrera de Ingeniería en Sistemas de la Universidad Nacional de Loja, primero haremos un introducción a la Web Semántica sus componentes, características para posteriormente con la ayuda de la herramienta Protégé hacer el diseño clases, relaciones, restricciones de la ontología y por ultimo la conexión con Java utilizando Jena.

Palabras Clave: Web Semántica, ontología, clases, relaciones, restricciones.

WEB SEMÁNTICA

Introducción

La Web Semántica es una web extendida, la misma está dotada de un mayor significado. Se desarrolla con lenguajes universales que permitirán a los usuarios encontrar respuestas a sus preguntas de una forma más rápida y sencilla gracias a la mejor estructuración de la información. La misma permitirá a los usuarios delegar tareas a las herramientas de la Web Semánticas, las cuales podrán ser capaces de procesar la información. El término de Web Semántica es algo desconocido para muchos usuarios, varias personas afirman que es la sucesora de la Web actual. Sabemos que la comprensión del concepto de la misma puede resultar ser algo confuso. La Web Semántica es una extensión de la World Wide Web en la que los contenidos de la Web pueden ser expresado mucho mas que en un lenguaje natural, y también en un formato que pueda ser entendido, interpretado y usado por diferentes software, permitiéndoles buscar, compartir e integrar información más fácil.

Definición

Es una Web que contiene mucha información donde podemos relacionar diferentes recursos. Con la misma podemos solucionar problemas habituales como son los de realizar búsquedas en Internet. Los buscadores nos arrojan resultados gracias a palabras claves, están nos muestra informaciones que no se relacionan con lo que deseamos encontrar, y que podamos relacionar recursos en la Web; supongamos que contamos con un artículo publicado, podemos preguntar cuáles artículos ha publicado ese autor mediante las relaciones que existen(McGuinness, 2004).

Principales Componentes:

Entre los principales componentes de la Web Semántica podemos encontrar XML, XML Schema, RDF, RDF Schema y OWL. La descripción de OWL en inglés “Ontology Web Language” describe las funciones y relaciones de cada componente de la Web Semántica:

1. **XML:** Provee una sintaxis elemental para las estructuras de contenidos dentro de documentos.
2. **XML Schema:** Es un lenguaje para proporcionar y restringir la estructura y el contenido de los elementos contenidos dentro de documentos XML.
3. **RDF:** Es un lenguaje simple para expresar modelos de los datos, que refieren a los objetos “recursos” y a sus relaciones. Un modelo de RDF-based se puede representar en sintaxis de XML.
4. **RDF Schema:** Es un vocabulario para describir propiedades y clases de recursos RDF-based, con semántica para generalizar-jerarquías de las propiedades y clases.
5. **OWL:** Es un mecanismo para desarrollar temas o vocabularios específicos en los que podamos asociar esos recursos.

¿Cómo funciona?

Supongamos que la Web tiene la capacidad de construir una base de conocimiento sobre las preferencias de los usuarios y que, a través de una combinación entre su capacidad de conocimiento y la información disponible en Internet, sea capaz de atender de forma exacta las demandas de información por parte de los usuarios en relación, por ejemplo, a reserva de hoteles,

vuelos, médicos, libros, etc.

Si esto ocurriese así en la vida real, el usuario, en su intento, por ejemplo, por encontrar todos los vuelos a Praga para mañana por la mañana, obtendría unos resultados exactos sobre su búsqueda. Sin embargo la realidad es otra. La figura 1 muestra los resultados inexactos que se obtendrían con el uso de cualquier buscador actual, el cual ofrecería información variada sobre Praga pero que no tiene nada que ver con lo que realmente el usuario buscaba. El paso siguiente por parte del usuario es realizar una búsqueda manual entre esas opciones que aparecen, con la consiguiente dificultad y pérdida de tiempo. Con la incorporación de semántica a la Web los resultados de la búsqueda serían exactos. La figura 2 muestra los resultados obtenidos a través de un buscador semántico. Estos resultados ofrecen al usuario la información exacta que estaba buscando. La ubicación geográfica desde la que el usuario envía su pregunta es detectada de forma automática sin necesidad de especificar el punto de partida, elementos de la oración como "mañana" adquirirían significado, convirtiéndose en un día concreto calculado en función de un "hoy". Algo semejante ocurriría con el segundo "mañana", que sería interpretado como un momento determinado del día. Todo ello a través de una Web en la que los datos pasan a ser información llena de significado. El resultado final sería la obtención de forma rápida y sencilla de todos los vuelos a Praga para mañana por la mañana (van Harmelen, 2008).



Figure 1. Resultados obtenidos con un buscador normal (McGuinness, 2004)

La forma en la que se procesará esta información no sólo será en términos de entrada y salida de parámetros sino en términos de su SEMÁNTICA. La Web Semántica como infraestructura basada en metadatos aporta un camino para razonar en la Web, extendiendo así sus capacidades.

No se trata de una inteligencia artificial mágica que permita a las máquinas entender las palabras de los usuarios, es sólo la habilidad de una máquina para resolver problemas bien

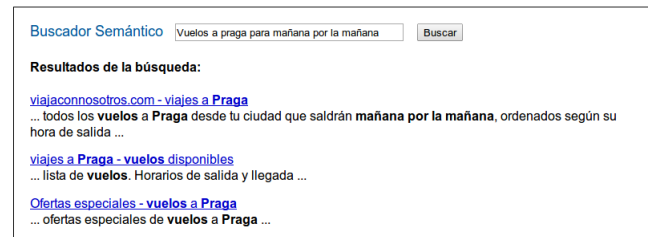


Figure 2. Resultados obtenidos con un buscador semántico (McGuinness, 2004)

definidos, a través de operaciones bien definidas que se llevarán a cabo sobre datos existentes bien definidos.

Para obtener esa adecuada definición de los datos, la Web Semántica utiliza esencialmente RDF, SPARQL, y OWL, mecanismos que ayudan a convertir la Web en una infraestructura global en la que es posible compartir, y reutilizar datos y documentos entre diferentes tipos de usuarios.

- RDF proporciona información descriptiva simple sobre los recursos que se encuentran en la Web y que se utiliza, por ejemplo, en catálogos de libros, directorios, colecciones personales de música, fotos, eventos, etc.
- SPARQL es lenguaje de consulta sobre RDF, que permite hacer búsquedas sobre los recursos de la Web Semántica utilizando distintas fuentes de datos.
- OWL es un mecanismo para desarrollar temas o vocabularios específicos en los que asociar esos recursos. Lo que hace OWL es proporcionar un lenguaje para definir ontologías estructuradas que pueden ser utilizadas a través de diferentes sistemas. Las ontologías, que se encargan de definir los términos utilizados para describir y representar un área de conocimiento, son utilizadas por los usuarios, las bases de datos y las aplicaciones que necesitan compartir información específica, es decir, en un campo determinado como puede ser el de las finanzas, medicina, deporte, etc. Las ontologías incluyen definiciones de conceptos básicos en un campo determinado y la relación entre ellos.

Otra tecnología que ofrece la Web Semántica para enriquecer los contenidos de la Web tradicional es RDFa. Mediante RDFa se pueden representar los datos estructurados visibles en las páginas Web (eventos en calendarios, información de contacto personal, información sobre derechos de autor, etc.), a través de unas anotaciones semánticas incluidas en el código e invisibles para el usuario, lo que permitirá a las aplicaciones interpretar esta información y utilizarla de forma eficaz. Por ejemplo, una aplicación de calendario podría importar directamente los eventos que encuentra al navegar por cierta página Web, o se podrían especificar los datos

del autor de cualquier foto publicada, así como la licencia de cualquier documento que se encuentre. Para extraer el RDF se podría utilizar GRDDL, una técnica estándar para extraer la información expresada en RDF desde documentos XML, y en particular, de las páginas XHTML.(van Harmelen, 2008)

Capas de la Web Semántica

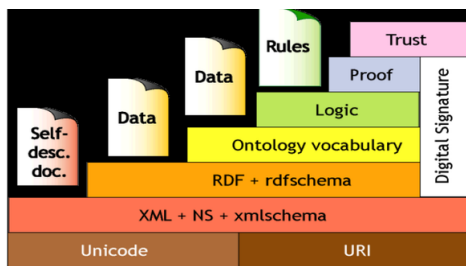


Figure 3. Capas Web Semántica

1. **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.
2. **URI, Las referencias:** 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).
3. **XML + NS + xmlschema:** 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.
4. **RDF + rdfschema:** Es un lenguaje simple mediante el cual definimos 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).
5. **Lenguajes de Ontologías:** 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.

6. **Lógica:** En esta capa se establecen reglas de inferencia, es decir, una ontología puede expresar la regla por ejemplo: "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 Central de Vzla, al estar en la ciudad de Caracas, debe estar situada en Venezuela, y debería por lo tanto estar formateado según los estándares Venezolanos.
7. **Pruebas (Proof):** Se intercambiarán "pruebas" escritas en el lenguaje unificador de la Web Semántica. Este lenguaje posibilita las inferencias lógicas realizadas a través del uso de reglas de inferencia.
8. **Confianza (Trust):** Hasta que no se haya comprobado de forma exhaustiva las fuentes de información, los agentes deberían ser muy escépticos acerca de lo que leen en la Web Semántica. Digital Signature (Firma digital): Utilizada por los ordenadores y agentes para verificar que la información ha sido ofrecida por una fuente de confianza.(Doldan, 2007)

ONTOLOGÍA

Una ontología define los términos y las relaciones básicas para la comprensión de un área del conocimiento, así como las reglas para poder combinar los términos para definir las extensiones de este tipo de vocabulario controlado.

Se trata de convertir la información en conocimiento mediante unas estructuras de conocimiento formalizadas (las ontologías) que referencien los datos, por medio metadatos, bajo un esquema común normalizado sobre algún dominio del conocimiento. Los metadatos no sólo especificarán el esquema de datos que debe aparecer en cada instancia, sino que también podrán contener información adicional de cómo hacer deducciones sobre ellos, es decir, cómo establecer axiomas que podrán, a su vez, aplicarse en los diferentes dominios que trate el conocimiento almacenado.(Lapiente, 2010)

Beneficios de utilizar ontologías se pueden resumir de la siguiente forma:

- Proporcionan una forma de representar y compartir el conocimiento utilizando un vocabulario común
- Permiten usar un formato de intercambio de conocimiento proporcionan un protocolo específico de comunicación
- Permiten una reutilización del conocimiento

Componentes

1. **Conceptos:** son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etc.
2. **Relaciones:** representan la interacción y enlace entre los conceptos de un dominio. Suelen formar la taxonomía del dominio. Por ejemplo: subclase-de, parte-de, parte-exhaustiva-de, conectado-a, etc.
3. **Funciones:** son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología. Por ejemplo, pueden aparecer funciones como: asignar-fecha, categorizar-clase, etc.
4. **Instancias:** se utilizan para representar objetos determinados de un concepto.
5. **Reglas de restricción o axiomas:** son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Por ejemplo: "Si A y B son de la clase C, entonces A no es subclase de B", "Para todo A que cumpla la condición B1, A es C", etc. Los axiomas, junto con la herencia de conceptos, permiten inferir conocimiento que no esté indicado explícitamente en la taxonomía de conceptos.

PROTÉGÉ

Protégé es una herramienta para el desarrollo de Ontologías y Sistemas basados en el conocimiento creada en la Universidad de Stanford. Protégé está desarrollada en JAVA y puede funcionar perfectamente bajo WINDOWS.

Las aplicaciones desarrolladas con Protégé son empleadas en resolución de problemas y toma de decisiones en dominios particulares. La herramienta Protégé emplea una interfaz de usuario que facilita la creación de una estructura de frames con clases, slots e instancias de una forma integrada.

Lo podemos descargar de la siguiente dirección <http://protege.stanford.edu/>, basta con descomprimirlo y ejecutar el archivo run.sh

APACHE JENA

Apache Jena es un framework Java para construir aplicaciones basadas en ontologías.

Jena se desarrolló en HP Labs en el 2000, en 2009 HP cedió el proyecto a la fundación Apache que decidió adoptarlo en noviembre de 2010.

Su Arquitectura incluye:

- API para trabajar (leer, procesar, escribir) ontologías RDF y OWL
- Motor de inferencia para razonar sobre ontologías RDF y OWL
- Estrategias de almacenamiento flexible para almacenar triplets RDF en memoria o fichero
- Motor de queries compatible con especificación SPARQL

DESCRIPCIÓN DE LA PRÁCTICA

Como ejemplo vamos a modelar la ontología aplicada a la Carrera de Ingeniería en Sistemas de acuerdo a los siguientes datos:

- El área a la que pertenece la carrera
- Las cuatro carreras que tiene el área en las cuales está incluido Sistemas
- Los diez módulos que se requiere para aprobar la carrera
- La malla actual y la anterior de la carrera
- Las materias de cada módulo
- Los docentes y estudiantes de la carrera, centrándose más en el décimo módulo

RESULTADOS

Una vez realizado el análisis de acuerdo a los datos mencionados anteriormente se ha obtenido el siguiente modelo:

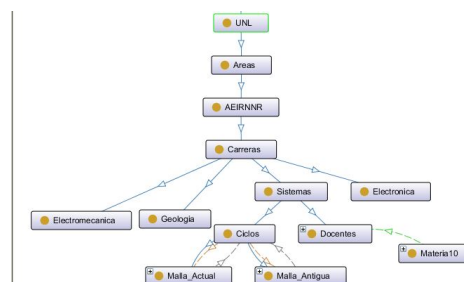


Figure 4. Gráfico de la Ontología

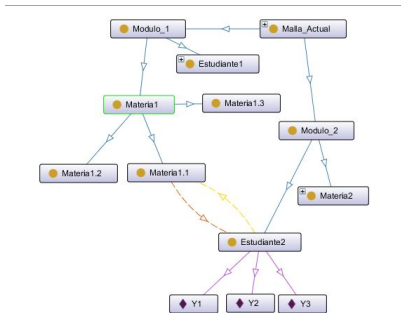


Figure 5. Esquema de malla actual

Descripción

Como se puede observar la ontología cuenta (ver Fig. 5) con dos mallas: malla actual y malla antigua. En el caso de la malla actual tenemos dos ciclos llamados Módulo 1 y Módulo 2 los mismos que cuentan con estudiantes y materias respectivas, para el manejo del arrastre de materias existe una restricción en la que se especifica que los estudiantes del módulo dos puede tomar clases de las materias del módulo uno siempre y cuando estas no sean troncales.

Código para importar datos de la ontología (OWL)

```
public String leerClasesySub(String ruta) {
    String xys = "";
    OntModel model = ModelFactory.createOntologyModel();
    OntModel a = ModelFactory.createOntologyModel();
    OntModel b = ModelFactory.createOntologyModel();
    try {
        InputStream file = new FileInputStream(ruta);
        b.read(file, "");
    } catch (Exception ex) {
        JOptionPane.showMessageDialog(null, "Archivo no compatible...\n"
            + "los archivos deben ser de extensión wxml");
    }
    Iterator<DatatypeProperty> propiedades = a.listDatatypeProperties();
    Iterator<OntClass> clases = b.listClasses();
    while (clases.hasNext()) {
        if (clases.next().getLocalName() != null) {
            xys = xys.concat(String.valueOf(clases.next().getLocalName()).concat("\n"));
        }
    }
    return xys;
}
```

Figure 6. Método de importación de el archivo OWL

Descripción

A través método obtenemos las clases de la ontología, para esto utilizamos la libreria Jena: primero establecemos la ruta del archivo, luego obtenemos la lista de clases de la ontología utilizando el método listClasses() que pertenece a la clase OntModel, de ahí se recorre la lista para su posterior presentación.

```
public String leerClasesySub(String ruta) {
    String xys = "";
    OntModel model = ModelFactory.createOntologyModel();
    OntModel a = ModelFactory.createOntologyModel();
    OntModel b = ModelFactory.createOntologyModel();
    try {
        InputStream file = new FileInputStream(ruta);
        b.read(file, "");
    } catch (Exception ex) {
        JOptionPane.showMessageDialog(null, "Archivo no compatible...\n"
            + "los archivos deben ser de extensión wxml");
    }
    Iterator<DatatypeProperty> propiedades = a.listDatatypeProperties();
    Iterator<OntClass> clases = b.listClasses();
    while (clases.hasNext()) {
        if (clases.next().getLocalName() != null) {
            xys = xys.concat(String.valueOf(clases.next().getLocalName()).concat("\n"));
        }
    }
    return xys;
}
```

Figure 7. Método de obtención de Importación

Descripción

A través método obtenemos las clases de la ontología, para esto utilizamos la libreria Jena: primero establecemos la ruta del archivo, luego obtenemos la lista de clases de la ontología utilizando el método listClasses() que pertenece a la clase OntModel, de ahí se recorre la lista para su posterior presentación



Figure 8. Interfaz de la aplicación

Descripción

En esta figura tenemos la pantalla principal del programa, la cual nos permite seleccionar un archivo .owl de los cuales se extrae: clases, individuos y propiedades.

CONCLUSIONES

- Las ontologías permiten organizar y describir el conocimiento de manera que las búsquedas

en internet son mas eficientes y se obtienen en menor tiempo

- Se puede obtener las clases, relaciones, individuos utilizando la libreria Jena
- A través de la práctica se obtuvieron conocimientos acerca de la estructura y funcionamiento de una ontología aplicado a la Web Semántica.

RECOMENDACIONES

- Es muy importante tomar en cuenta tener una buena lógica ya que esto tambien influye en las restricciones y las relaciones que tendrá nuestra ontología
- Los nombres de las relaciones deben estar acordes con lo que se desea expresar para darle un mejor sentido

References

- Doldan, F. (2007). REDES BAYESIANAS Y RIESGO OPERACIONAL. (German) [On universidad de a coruña]. *Revista Galega de Economía*, 16(núm. extraord). doi: http://www.usc.es/econo/RGE/Vol16_ex/Castelan/art1c.pdf
- Lapiente, M. J. L. (2010). Hipertexto: El nuevo concepto de documento en la cultura de la imagen. doi: <http://www.hipertexto.info/documentos/ontologias.htm>
- McGuinness, D. L. (2004). Lenguaje de Ontologías Web (OWL) Vista General. doi: <http://www.w3.org/2007/09/OWL-Overview-es.html>
- van Harmelen, F. (2008). Guía Breve de Web Semántica. doi: <http://www.w3c.es/Divulgacion/GuiasBreves/>