

Cuaderno Activa, ISSN 2027-8101. N°5,  
Enero-Diciembre de 2013, pp. 51-56.  
Tecnológico de Antioquia, Colombia.



## Comparación de dos métodos de representación semántica en interoperabilidad a la luz de la lingüística computacional

María Isabel Marín Morales  
Magíster en Ingeniería  
Docente asistente  
Universidad de San Buenaventura  
maria.marin@usbmed.edu.co



UNIVERSIDAD DE  
SAN BUENAVENTURA

Recibido: 5 de mayo 2013  
Aceptado: 10 de junio 2013

### Resumen

Interoperabilidad es la capacidad que exhiben las aplicaciones informáticas para compartir datos y procesos. Los diferentes sistemas de información subyacentes en la web y al interior de las empresas requieren interoperar, ya que contienen grandes cantidades de información y procesos que se deben complementar para realizar un eficiente procesamiento y análisis de la información. La interoperabilidad en el nivel de datos se aborda desde diferentes niveles, entre ellos el nivel semántico. Existen varias aproximaciones de representación semántica en interoperabilidad, de donde nace la necesidad de elegir la más adecuada para realizar aplicaciones. Por ello, en este artículo se analizan y comparan dos métodos de representación semántica de la interoperabilidad, a la luz de la lingüística computacional. El primer método se basa en la representación de la información sobre esquemas XML, y el segundo consiste en la utilización de ontologías como base de conocimiento de los diferentes sistemas por interoperar.

**Palabras clave:** Interoperabilidad, Representación semántica, Lingüística computacional, Ontología, XML, OWL, Protégé.

### Comparison of two methods of semantic representation in interoperability in light of computational linguistics

### Abstract

Interoperability is software ability to share data and processes. Web-based and organizational information systems must “interoperate”, since they carry big amounts of data and processes, which must complement each other so that an efficient information processing and analysis is made. Interoperability on the data-level is addressed on various levels; among them, the semantic level. Some projects explore the semantic representation of interoperability —the need to choose the more pertinent representation arises from there. Consequently, this paper

analyzes and compares two methods of semantic representation of interoperability in the light of computational linguistics. The first method is based on the representation of information in XML schemas and the second one is based on the usage of ontologies as a knowledge base of several systems to be interoperated.

**Keywords:** Interoperability, Semantic representation, Computational linguistics, Ontology, XML, OWL, Protégé.

## 1. Introducción

Hoy en día es común encontrar que las empresas se encuentran soportadas por sistemas en la web y aplicaciones de escritorio ubicadas en forma distribuida, los cuales requieren compartir datos y procesos. Para hacer posible esta tarea, en las ciencias de la computación surge el concepto de interoperabilidad, definido, precisamente, como la capacidad que exhiben los programas informáticos para compartir datos y procesos (Zhang et al., 2007). Este requisito se fundamenta en la necesidad de procesamiento y análisis de la información desde diferentes fuentes de datos y con herramientas distribuidas espacialmente.

La interoperabilidad se puede dar entre procesos y datos, y esta última, a su vez, se aborda desde varios niveles. Gómez (2007) realiza una clasificación de la interoperabilidad en el nivel de datos, en Sintaxis, Estructura, Infraestructura y Semántica. Es así que, en interoperabilidad, se definen métodos para facilitarla en cada uno de los niveles mencionados (Shanzhen, 1999).

Por otro lado, la interoperabilidad a nivel semántico se está convirtiendo en una iniciativa dirigida a la representación de la información disponible en la web y en las fuentes de datos, de una manera comprensible no solo por los seres humanos, sino por los computadores. Esto permitirá a las máquinas mostrar la información de forma análoga a como lo hacen hoy y además “comprenderla” con fines de apoyo al intercambio, al análisis y a la reutilización de datos a través de diferentes aplicaciones o comunidades (Navarrete et al., 2004). Por lo anterior, existen varias aproximaciones de representación basadas en almacenamiento de la información en esquemas XML (Abdalla, 2003) y en la utilización de ontologías (Xianming et al., 2009). De aquí nace la necesidad de elegir la metodología más adecuada para llevar a cabo el desarrollo de aplicaciones.

Por ello en este artículo se analizan y comparan dos métodos que facilitan la interoperabilidad en el nivel semántico, a la luz de la representación semántica definida en la lingüística computacional (LC). En este campo interdisciplinar, la semántica se considera como

el estudio del significado del lenguaje. La LC define los criterios de comparación, dada la naturaleza computacional que la representación semántica requiere para ser legible por la máquina (Mitkov, 2003).

Este artículo se estructura así: en la sección 2 se expone el marco teórico que acerca al lector a los conceptos básicos del tema, en la sección 3 se hace una recopilación de los avances en interoperabilidad, con énfasis en el nivel semántico, en la sección 4 se explica el proceso y los resultados obtenidos en la comparación de los métodos a la luz de la lingüística computacional y, finalmente, en la sección 5 se sintetizan las conclusiones y se plantea el trabajo futuro.

## 2. Marco teórico y trabajos previos

Una ontología describe los conceptos y relaciones que son importantes en un dominio particular, y proporciona un vocabulario resultante de un consenso así como una especificación computarizada del significado de los términos utilizados en el vocabulario. En los últimos años, las ontologías se adoptaron en muchas empresas y en la comunidad científica como una forma de compartir y reutilizar el conocimiento del dominio. Las ontologías son fundamentales para aplicaciones como los sistemas de gestión de la información y los servicios de la web semántica. Los lenguajes más usados para su elaboración son el Ontology Web Language (OWL) y el Resource Description Framework (RDF) (Xianming et al., 2009).

El RDF es un framework para metadatos en la World Wide Web (WWW), desarrollado por el World Wide Web Consortium (W3C). Este modelo se basa en la idea de convertir las declaraciones de los recursos en expresiones con la forma sujeto-predicado-objeto (conocidas en términos RDF como tripletas). El sujeto es el recurso, es decir, aquello que se está describiendo. El predicado es la propiedad o relación que se desea establecer acerca del recurso. Por último, el objeto es el valor de la propiedad o el otro recurso con el que se establece la relación. La combinación de RDF con otras herramientas como RDF Schema permite añadir significado a las páginas (W3C, 2004).

OWL es un lenguaje de marcado para publicar y compartir datos usando ontologías en la web. Este lenguaje tiene como objetivo facilitar un modelo de marcado construido sobre RDF y codificado en el Extensible Markup Language (XML). Junto al entorno RDF y otros componentes, estas herramientas hacen posible el proyecto de web semántica. OWL tiene tres variantes: Lite, DL y Full, que incorporan diferentes funcionalidades. OWL Lite se construyó de tal forma que toda sentencia se pueda resolver en tiempo finito, en tanto que la versión más completa de OWL, DL, puede contener 'bucles' infinitos (Xianming et al., 2009).

XML es un lenguaje de etiquetado de documentos (se introducen etiquetas entre el texto), publicado por el W3C como una recomendación en 1998. Se trata de una simplificación de SGML (Standard Generalised Markup Language), cuya aplicación más conocida es HTML. XML modela los documentos como conjuntos de elementos que contienen cadenas de caracteres. El principio y el final de cada elemento se delimitan con etiquetas. La idea fundamental detrás de XML consiste en marcar o etiquetar la información, de modo que a cada porción (elemento) del documento la delimita una etiqueta de comienzo seguida de la correspondiente etiqueta de cierre (Martínez, 2000).

Por otra parte, la lingüística computacional es un campo interdisciplinar orientado al procesamiento del lenguaje por computador. Desde hace unos cincuenta años viene creciendo y desarrollándose en forma exponencial. Se expandió teóricamente por el desarrollo de modelos formales y computacionales del lenguaje. En este periodo de maduración, su campo de aplicación creció considerablemente y ahora cubre el análisis fonológico, morfológico, lexicográfico, sintáctico y semántico del lenguaje (Mitkov, 2003).

Algunas de las herramientas usadas con mayor frecuencia para facilitar la interoperabilidad entre sistemas son: CORBA, DCOM, EJB, P2P, Mobile Agent, etc. Estas herramientas presentan dificultades en su extensión ya que, entre ellas, manejan modelos de datos diferentes y sus protocolos son propietarios (Young et al., 2003). Para enfrentar este problema, son muchas las iniciativas propuestas y en ocasiones detalladas para un dominio de conocimiento específico. Entre ellas, Shanzhen (1999) define un modelo de interoperabilidad llamado InterModel5, basado en Infraestructura de información espacial (Spatial Information Infrastructure —SII—),

el cual conecta cinco niveles de interoperabilidad: 1) El nivel institucional, teniendo en cuenta las políticas y la cultura de las organizaciones, 2) nivel semántico, 3) nivel del servicio de aplicación, 4) nivel de la transformación de los recursos y 5) nivel del descubrimiento de los recursos. La definición se hace en un nivel de descripción, no se llega ni a prototipos ni a definición de lenguajes.

Por otro lado, con respecto a la interoperabilidad semántica, Abdalla (2003) identifica, en el intercambio de información, conflictos estructurales y semánticos, como el uso de conceptos en diferentes contextos, unidades, entre otros, y propone una solución utilizando las especificaciones de XML para representar la información tanto a nivel de su estructura como a nivel semántico. Además, realiza previamente una única definición de tipo de documento (Document type definition —DTD—). Finalmente, propone como trabajo futuro la interoperabilidad semántica basada en este formato con DTD diferentes.

Xianming et al. (2009) muestran las desventajas de las especificaciones JSR168 y WSRP para propiciar la interoperabilidad semántica entre portlets y propone el uso de ontologías. Para el diseño de estas utiliza la herramienta Protégé, desarrollada por el Centro de Stanford para la Investigación en Informática Biomédica de la Escuela de Medicina de la Universidad de Stanford, que permite especificar bases de conocimiento en el lenguaje basado en XML y OWL.

Shanzhen et al. (2002) expresan la necesidad de la representación semántica al identificar estándares para la interconexión entre el dominio específico de la información geográfica (IG), como GML y G-XML del Open Geospatial Consortium (OGC), que no permiten una especificación del conocimiento. Así, proponen solucionar el problema de la interoperabilidad semántica, cooperación, fusión e integración entre IG con la utilización de ontologías desarrolladas sobre el lenguaje RDF. Finalmente, concluyen sobre la falta de lenguajes de desarrollos de ontologías maduros para dominios específicos, sugiriendo la extensión de los ya existentes. En su trabajo identifican una alta demanda de reglas para inferir sobre diferentes ontologías y la falta de herramientas para asociar reglas automáticamente de varias de estas herramientas.

Cui & Wu (2005) exponen las diferentes maneras de

trabajar con ontologías. Inicialmente, el enfoque de ontologías individuales, en donde varias fuentes de información tienen una misma visión del dominio, es decir, utiliza una misma ontología. Luego, el enfoque de múltiples ontologías, en el que la propia ontología describe cada fuente de información, identificando como problema la falta de vocabulario común que hace difíciles las comparaciones. Finalmente, el enfoque híbrido, en donde una ontología con el vocabulario compartido recoge las definiciones comunes de las ontologías específicas de cada fuente de información. Los autores concluyen que la mejor manera de representar el conocimiento es a través de la construcción de ontologías híbridas.

### 3. Comparación de dos métodos de interoperabilidad semántica

La comparación realizada en este trabajo se hace entre el método de interoperabilidad semántica basado en el almacenamiento de la información en estructuras XML (Abdalla, 2003) y el método de interoperabilidad semántica basado en la utilización de ontologías (Xianming et al., 2009), esto, a la luz de la definición de representación semántica dada en Mitkov (2003) y de algunas métricas de interoperabilidad encontradas en la literatura.

#### 3.1 Interoperabilidad semántica basada en el almacenamiento de la información en estructuras XML

Una de las finalidades del metalenguaje de marcado XML es la de compartir información entre varias aplicaciones. Por ser metalenguaje, posibilita personalizar la estructura en una deseada, haciendo que los parsers, o analizadores sintácticos, de las aplicaciones puedan entender e interpretar fácilmente los datos.

XML permite representar de una forma deseada los datos de manera estructural más no semánticamente. En Abdalla (2003) se plantea un mecanismo para representar semánticamente la información para compartir entre aplicaciones. Esto se logra con la inserción de etiquetas que embeben los valores semánticos posibles de una propiedad.

Por ejemplo, al obtener dentro de un documento XML la expresión `<weight> 30 </weight>`, la máquina puede identificar que se trata de una propiedad de peso=30

del elemento del cual weight es hijo. Sin embargo, no queda claro si se refiere a 30 kilos o a 30 libras. Con el método propuesto por Abdalla (2003) es posible resolver esta duda definiendo las propiedades semánticas posibles en ese contexto, como se muestra en la Tabla 1. Así, cuando la máquina se encuentra con una expresión como la mostrada en la Tabla 2, identificará inmediatamente que se trata de 30 kilogramos.

Una vez es posible representar semánticamente el conocimiento mediante XML, es preciso identificar el modelo de interoperabilidad. Abdalla (2003) propone un agente de mapeo (Semantic Mapping Agent—SMA—); este componente de la interfaz entre una y otra representación del conocimiento, homologa las propiedades semánticas de uno y otro XML.

**Tabla1.** Representación semántica de la propiedad Peso

Representación semántica de la propiedad Peso
<pre> &lt;SemanticProperty&gt;   &lt;PropertyName&gt;weightUnit&lt;/PropertyName&gt;   &lt;PropertyValue&gt;Kgs&lt;/PropertyValue&gt;   &lt;PropertyValue&gt;Lbs&lt;/PropertyValue&gt; &lt;/SemanticProperty&gt; </pre>

Fuente: Elaboración propia (2013)

**Tabla2.** Representación semántica a través de una estructura XML

La propiedad "Peso" dentro de la información
<pre> &lt;weight semanticProperty="weightUnit,Kgs"&gt;   30 &lt;/weight&gt; </pre>

Fuente: Elaboración propia (2013)

#### 3.2 Interoperabilidad semántica basada en el uso de ontologías

Las ontologías, como resultados de consensos, proporcionan el vocabulario correspondiente a un dominio. En esta medida, cumplen el papel de clarificadoras entre dos aplicaciones a interoperar.

Xianming et al. (2009) proponen un esquema de ontologías como un servicio de portlet remoto para ilustrar la interoperabilidad semántica de estos. Este método en particular se compara con el mencionado en el numeral anterior.

Los portlets son la tecnología núcleo de los portales web. Estos últimos, a su vez, pueden poseer muchos portlets que requieren acceder a diferentes fuentes de información y aplicaciones. Si un portlet no soporta interoperabilidad semántica, hace que el trabajo cooperativo en los portales de las empresas no sea del todo exitoso, pues, en situaciones, como al referirse a la palabra “notebook”, por ejemplo, no estaría claro si se puede relacionar con una papelería o con el computador.

Xianming et al. (2009) construyeron las ontologías en el editor de OWL, Protégé-OWL, y las administraron como un servicio web en un servidor de portlets remoto.

El proceso de interoperabilidad se da como sigue: un portlet local puede requerir un servicio de ontología a través de WSRP. Cuando dos portlets tienen conflictos semánticos al transmitir un mensaje, un portlet debe enviar un requerimiento de servicio de ontología. Para esto, el portlet debe crear un objeto SOAP y crear en este un requerimiento SOAP con la solicitud hacia el servicio web de la ontología. El servicio web de la ontología abre el requerimiento SOAP y busca la equivalencia semántica, de acuerdo con el parámetro que le enviaron. El resultado se envía de nuevo como una respuesta SOAP hacia el objeto SOAP. Finalmente, el objeto SOAP abre el resultado y lo envía al portlet que lo creó. Así, el portlet entiende semánticamente el mensaje y continúa cooperando con otros portlets.

#### 4. Elementos en una representación semántica

Mitkov (2003) plantea, como uno de los elementos más importantes a la hora de verificar un modelo de representación semántica, considerar la arquitectura del interfaz sintáctico-semántico en este.

##### 4.1 Métricas de comparación

Para comparar estos dos métodos de interoperabilidad semántica y caracterizar las faltas o ventajas de uno u otro se establecieron métricas con base en Latorres (2004) y Mitkov (2003):

**Adaptabilidad al cambio:** Se define como la capacidad de cada método para adaptarse a actualizaciones del dominio.

**Visualización gráfica:** Define las facilidades en el medio para editar el dominio a través de una interfaz de usuario gráfica y amigable.

A partir de esta comparación, y analizando puntualmente la información recopilada, es posible concluir que el método de representación del conocimiento mejor definido es el que para su fin hace uso de las ontologías.

#### 5. Conclusiones

En este trabajo se compararon dos métodos de interoperabilidad semántica. El primero consiste en almacenar la información en estructuras XML. Este método permite interoperar a través de interfaces de apareamiento de una estructura semántica hacia otra y con la utilización de DTD. El segundo método propone la utilización de ontologías. En particular, se comparó el método propuesto en Xianming et al. (2009), que consiste en un esquema de ontologías como un servidor remoto portlet. Con este método se tiene un vocabulario común entre aplicaciones. Se comparó con base en las métricas propuestas en Latorres (2004) y Mitkov (2003). Finalmente, se obtuvo que es más recomendable usar las ontologías como método de representación del conocimiento.

#### Referencias

Abdalla, K. (2003). A Model for Semantic Interoperability Using XML. Proceedings of the 2003 Systems and Information Engineering Design Symposium. pp. 107-111. IEEE Press.

Cui, W. & Wu, H. (2005). Using Ontology to Achieve the Semantic Integration and Interoperation of GIS. Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. IEEE International, Vol. 2, pp. 836-838.

Gómez, L. (2007). Interoperabilidad en los sistemas de información documental (SID): la información debe fluir. Vol. 3, N° 01, pp. 23-39.

- Latorres, E. (2004). Modelo de Semántica y Representación del Conocimiento. II Congreso de Enseñanza. Montevideo.
- Martínez, M. (2000). Extensible Markup Language (XML): Una solución para modelar documentos y sus interrelaciones basadas en la semántica de la información. Universidad de Zaragoza: Prensas Universitarias de Zaragoza, Vol. 6, N° 2, pp. 121-151
- Mitkov, R. (2003). Semantics. In: S. Lappin (editor). The Oxford Handbook of Computational Linguistics. Oxford University Press.
- Navarrete, T., Blat, J. & Ruiz, M. (2004). Semantic Interoperability of Field-based Thematic Geographic Information. Proceedings of the 15th International Workshop on Database and Expert Systems Applications, pp. 792-796.
- Shanzhen, Y. (1999). An interoperability GIS model based on the spatial information infrastructure. Geoinformatics and Socioinformatics. The Proceedings of Geoinformatics'99 Conference. Ann Arbor, N° 1, pp. 1-5.
- Shanzhen, Y., Lizhu, Z., Chunxiao, X., Qilun, L. & Yong, Y. (2002). Semantic and Interoperable WebGIS. Proceedings of the Second International Conference on Web Information Systems Engineering, Vol. 2, pp. 42-47.
- W3C (World Wide Web Consortium). (2004). Disponible en: <http://www.w3.org/RDF/> [Consultado, septiembre de 2013].
- Xianming, L., Weihua, L. & Shixian, L. (2009). Using Ontology to Support Portlet Semantic Interoperability. Education Technology and Computer Science, 2009. ETCS '09. First International Workshop on, N° 2, 534-537.
- Young, P., Chaki, N., Berzins, V. & Luqi. (2003). Evaluation of Middleware Architectures in Achieving System Interoperability. Proceedings of the 14th IEEE International Workshop on Rapid Systems Prototyping (RSP'03), pp. 108-116.
- Zhang, S., Gan, J., Miao, L., Lv, G. & Huang, J. (2007). Study on GML Spatial Interoperability based on Web Service. 31st Annual International Computer Software and Applications Conference. Vol. 1, pp. 649-656.