

**Web Semántica e Información Geográfica:
Una interrelación necesaria ante las problemáticas actuales**

**Semantic Web and Geographical Information:
A necessary inter-relationship faced with current problems**

VILCHES BLÁZQUEZ, LUIS MANUEL ¹
CORCHO, OSCAR ²
RODRÍGUEZ PASCUAL, ANTONIO ³
BERNABÉ POVEDA, MIGUEL ÁNGEL ⁴

^{1,3} Instituto Geográfico Nacional
C/ General Ibáñez de Íbero, 3. 28003. Madrid. España.
Tlf.: [+34] 915979660 / 61

¹ lmvilches@fomento.es

³ afrodriguez@fomento.es

² Departamento de Inteligencia Artificial. Facultad de Informática. Universidad Politécnica de Madrid. Campus de Montegancedo S/N. Boadilla del Monte. 28660 Madrid. España

² ocorcho@fi.upm.es

⁴ E.T.S.I. en Topografía, Geodesia y Cartografía. Universidad Politécnica de Madrid. Km 7.5 de la Autovía de Valencia. 28031 Madrid. España.

Tlf.: [+34] 913366487

⁴ ma.bernabe@upm.es

Resumen

Internet nos pone al alcance una cantidad ingente de documentación de cualquier tipo, sobre cualquier tema y en cualquier formato, consecuencia de que la Web actual constituye el mayor almacén de información recopilado por personas. Resultado de este gran volumen de información y de su crecimiento constante surgen diversos problemas que repercuten en la obtención de resultados. Con el objetivo de solventar la problemática actual e incrementar las potencialidades de las Web surge la Web Semántica, cuyo principal instrumento para la mejora de la organización y gestión de la información son las ontologías. La unión entre ontologías e información geográfica se presenta como la alternativa necesaria para poder solventar los principales problemas que atañen a esta información.

Palabras clave: Web Semántica, Ontologías, Información Geográfica, Sistemas de Información Geográfica, Infraestructuras de Datos Espaciales, Catálogo de fenómenos y Tesauros.

Abstract

There is a huge quantity of information at the Internet about any issue and format, due to the current Web is the biggest information store which is compiled by people. Different problems appear as result of this huge quantity of information and its rise up. It has an effect on information search and retrieve. Semantic Web appears to solve these problems and increasing power of current Web, and ontologies are main tool to improve information structuring and management. The relationship between ontologies and geographic information is necessary task to solve current problems of this information.

Keywords: Semantic Web, Ontologies, Geographic Information, Geographic Information Systems, Spatial Data Infrastructures, Feature Catalogues and Thesaurus.

1. INTRODUCCIÓN

Las ventajas que ofrece Internet para la búsqueda y recopilación de cualquier tipo de información en general y geográfica en particular son innegables, consecuencia de que la Web actual constituye el mayor almacén de información recopilado por personas. Debido a esto, se puede acceder a una cantidad ingente de documentación de cualquier tipo sobre cualquier tema y en cualquier formato. Pero el crecimiento progresivo de la WWW está provocando que los buscadores tengan dificultades para mantener la calidad de recuperación de los resultados. Además, los buscadores sólo acceden a lo que se denomina contenido estático de la Web e ignoran la parte dinámica de la misma, es decir, aquellas páginas generadas a partir de bases de datos. Es por esto, por lo que se produce el surgimiento de la Web Semántica, cuyo principal instrumento para la mejora de la organización e interrelación de los recursos presentes en la Web son las ontologías.

Estos problemas tampoco resultan ajenos a los diferentes servicios web que presentan Sistemas de Información Geográfica (SIG) e Infraestructuras de Datos Espaciales (IDEs). Es por ello que se hace necesario una estrecha ligazón entre Web Semántica, con las ontologías como principal instrumento, e Información Geográfica (IG).

Este artículo está dirigido a la comunidad geográfica que está interesada en conocer los avances tecnológicos con repercusión directa sobre la IG. Para ello, se presenta la situación de contexto de la Web actual versus Web Semántica en la sección 2. En la sección 3 se realiza una breve introducción al concepto de ontología. A continuación, se describen algunos problemas de los geodatos y se describen brevemente las principales formas de organización de dicha información para poner de manifiesto la necesidad de llevar el desarrollo de ontologías en el dominio de la IG en la sección 4. Finalmente, en la sección 5 se proporcionan algunas conclusiones al trabajo presentado.

2. WEB ACTUAL VERSUS WEB SEMÁNTICA: CONTEXTO

Uno de los mayores problemas a los que nos enfrentamos hoy día en la Sociedad de la Información es la sobrecarga de información, un problema que se potencia por el constante crecimiento de

recursos disponibles en la Web. Este hecho repercute de forma directa en la navegación a través de la red para la búsqueda de una determinada información útil o necesaria, incrementando la cantidad de tiempo invertido a pesar de la potencia y sagacidad crecientes de los buscadores disponibles.

La Web actual, conocida como Web Sintáctica, se caracteriza por enlazar gran cantidad de recursos entre sí mediante hipervínculos, esto se conoce como grafo dirigido. Los ordenadores realizan la presentación visual (tarea fácil) y las personas navegan, seleccionan e interpretan el contenido (tarea difícil). Esto es resultado del lenguaje utilizado en el desarrollo de páginas web (HTML - *HyperText Markup Language*), únicamente incluye información sobre características de visualización de la información (tipo de letra, color, etc.), por lo que el significado sólo es accesible a las personas y no para los ordenadores. Por este motivo, mientras el usuario visualiza e interpreta el contenido de una página web los ordenadores ven algo similar al contenido que se muestra en la figura 1.

FIGURA 1

Esto pone de manifiesto la falta de capacidad de las representaciones en que se basa la Web actual para expresar significados, ya que mientras la presentación de los datos en el navegador es inmediatamente comprendida por una persona, para el ordenador es una tarea imposible poder entender cuál es la semántica del documento, ya que aparece entremezclada con las etiquetas del formato de representación elegido. Esta misma limitación expresiva hace que la noción de semántica que manejan los buscadores web se limite a palabras clave con pesos, pero planas e inconexas, lo que no permite reconocer ni solicitar significados más elaborados (Castells, 2003).

Como respuesta ante la problemática derivada de las dificultades que conllevan las posibilidades de acceso masivo a los datos, la complejidad de su interpretación posterior y la dificultad que supone la búsqueda de una determinada información de forma precisa y rápida surge la Web Semántica¹, como vía para ofrecer soluciones más efectivas a los usuarios de Internet.

¹ <http://www.w3.org/2001/sw/>

Esta “nueva” Web, según (Berners-Lee y otros 2001), vendría a ser una extensión de la Web actual dotada de significado, esto es, un espacio donde la información tendría un significado bien definido, de manera que pudiera ser interpretada tanto por agentes humanos como por agentes computerizados para poder trabajar en cooperación. La Web Semántica (Berners-Lee y otros 2001) propone superar las limitaciones de la Web actual mediante la introducción de descripciones explícitas del significado, de la estructura interna y la estructura global de los contenidos y servicios disponibles en la WWW. Frente a la semántica implícita, el crecimiento caótico de recursos y la ausencia de una organización clara de la actual Web, la Web Semántica aboga por clasificar, dotar de estructura y anotar los recursos con semántica explícita procesable por máquinas.

Como se comentó con anterioridad, actualmente la Web se asemeja a un grafo dirigido, formado por nodos (recursos) del mismo tipo y arcos (hiperenlaces) que conectan los diferentes recursos igualmente definidos. Por ejemplo, no se hace distinción entre una página personal y el portal de una tienda on-line, como tampoco se distingue explícitamente los enlaces a las asignaturas que imparte un profesor de los enlaces a sus publicaciones. Por el contrario, en la Web Semántica cada nodo se clasifica mediante un tipo (profesor, tienda, pintor, libro) y los arcos representan relaciones explícitamente diferenciadas (pintor-obra, profesor-departamento, libro-editorial). Estas diferencias entre la Web Sintáctica y la Web Semántica se pueden apreciar gráficamente en la figura 2.

FIGURA 2

En la Web Semántica, las páginas no almacenan únicamente el contenido como un conjunto de palabras sin relación, tal y como sucede en la Web actual, sino que también incluyen su significado y su estructura. Además, los lenguajes de la Web Semántica (RDF - Resource Description Framework, RDF-S - Resource Description Framework-Schema, y OWL - Web Ontology Language) están basados en XML (eXtensible Markup Language), muchos más ricos que HTML y permiten representar el significado y la estructura del contenido a través de interrelaciones entre

conceptos. Esto contribuye a convertir el contenido de la Web en entendible y procesable por agentes software y servicios, por lo que contribuye de manera importante a un cambio en la óptica del trabajo, ya que se tiende hacia un cambio gradual desde la recuperación “manual” de información hacia la delegación y consecución de tareas.

En definitiva, la Web Semántica mantiene los principios que han hecho un éxito la Web actual, como son los principios de descentralización, compartición, compatibilidad, máxima facilidad de acceso y contribución o la apertura al crecimiento y uso no previsto de antemano. En este contexto un problema clave es alcanzar un entendimiento entre las partes que han de intervenir en la construcción y explotación de la Web, es decir, entre usuarios, desarrolladores y programas de diverso perfil. Esta visión rescata la noción de *ontología* del campo de la Inteligencia Artificial como vehículo para cumplir este objetivo (Castells, 2003)

3. UNA BREVE INMERSIÓN EN EL UNIVERSO ONTOLÓGICO

El término Ontología procede del mundo de la filosofía, concretamente, según el Diccionario de la Real Academia Española², es la parte de la metafísica que trata del ser en general y de sus propiedades transcendentales, o propiedades que traspasan los límites de lo meramente experimental. Este concepto ha sido adoptado por la informática, específicamente en el dominio de la Inteligencia Artificial. Su adopción ha dado origen a una gran variedad de definiciones de este “nuevo” concepto, producto de los puntos de vista de diversos autores y, fundamentalmente, de las diferentes formas de construir y utilizar las ontologías como sistemas informáticos.

Una de las definiciones más divulgadas es la aportada por (Gruber, 1993). Afirma que una ontología constituye una especificación explícita y formal de una conceptualización compartida. Esta definición tiene una serie de conceptos claves que ayudan a comprenderla, así por “conceptualización” se entiende un modelo abstracto de la realidad, de tal manera, que mediante esta se identifican los conceptos relevantes de un área. Por “explícita” se entiende que todos sus

² <http://www.rae.es/>

componentes deben estar definidos explícitamente. Respecto a “formal”, se refiere al hecho de que la ontología debe ser entendible por las máquinas. Por último, “compartida” refleja el hecho de que una ontología debe capturar conocimiento consensuado/aceptado por un grupo o comunidad de expertos, esto es, no debe ser privado o algo individual (Studer y otros, 1998).

Desde la óptica de la Información Geográfica (IG) esta definición nos hace entender que las ontologías no están formadas únicamente por meros conceptos con una cierta organización, lo que no distaría de las formas más habituales de organización de la IG (catálogos de fenómenos, diccionarios de datos y/o tesauros), sino que también se van a definir relaciones, atributos, reglas y axiomas entre conceptos que enriquecen y contribuyen a ampliar el vocabulario del dominio o área de trabajo. En definitiva, las ontologías van a ayudar al mundo geográfico a definir los significados de los fenómenos contenidos en los geodatos, por lo que estas pueden proporcionar la base del entendimiento en el dominio de la IG. Aunque, a menudo desarrollar una ontología de un dominio no es la meta en sí, sólo es un proceso de definición de un conjunto de datos y sus estructuras para que otros programas los usen, no obstante, algunas de las principales razones que pueden conducir a la construcción de una ontología, son (Noy y otros, 2001);

- Compartir el entendimiento común sobre un área de conocimiento entre personas y máquinas.

- Permitir la reutilización de conocimientos de un dominio o área de conocimiento.

- Permite cambiar las especificaciones de conocimiento de un dominio si se producen cambios en el mismo. Además, las especificaciones explícitas del dominio de conocimiento son útiles para nuevos usuarios que deben aprender el significado de los términos del área.

- Analizar el conocimiento de un dominio es posible una vez realizada una especificación completa de los términos que componen al dominio. El análisis formal de los términos es extremadamente valioso al intentar reutilizar ontologías existentes y pretender extenderlas.

En definitiva, en esta sección se pueden extraer varias conclusiones. Por un lado, en las ontologías el conocimiento se especifica a través de conceptos, es decir, estas están formadas por una agrupación de conceptos que dan como resultado una forma de ver el mundo, común y compartida, acorde con cierta perspectiva ante un dominio de conocimiento. Por otro lado, el hecho de que entre los conceptos existen relaciones, reglas de inferencia, axiomas, etc. que contribuyen a enriquecer el conocimiento del dominio en cuestión. Por último, hay que destacar, que esas conceptualizaciones formalizadas permiten la comunicación entre expertos y sistemas informáticos, consecuencia directa de que la información no sólo es entendible por las personas, como sucedía hasta este momento, sino que también es entendida por las máquinas, lo que repercutirá en una constante reutilización del conocimiento. Por tanto, desde la óptica geoespacial, la construcción de ontologías dará origen a una importante mejora en la representación de la IG, repercutiendo de forma directa en los sistemas de recuperación, consulta y análisis de la misma.

4. ONTOLOGÍAS: UNA NECESIDAD DERIVADA DE LA PROBLEMÁTICA ACTUAL DE LA INFORMACIÓN GEOGRÁFICA

Las ciudades y el territorio se transforman y se adaptan a nuevas necesidades y nuevos roles, a los que no son ajenos los fenómenos que conlleva el nuevo paradigma de la Sociedad de la Información. Gestionar la complejidad de dicha transformación requiere Información Global (Guimet, 2003) común y compartida. Este hecho motiva que los recursos de información disponibles sobre el territorio sufran un crecimiento progresivo, aunque las dificultades de acceso a la información impiden una visión homogénea por parte de los diferentes actores presentes en el territorio.

Hoy día, los servicios Web de los diferentes SIG e IDEs presentan una mera integración de información a modo de superposición de capas con gran diversidad temática, en ocasiones, provenientes de diferentes productores. Este hecho, pone de manifiesto contenidos (conceptos) y estructuras (organización de la información) heterogéneas derivados de la falta de consenso, de

diferencias de intereses y necesidades y de las inercias de los procesos de producción. Todos estos factores generan dificultades en las tareas de consulta, recuperación, explotación, actualización y visualización de la geo-información, ante las que todo usuario demanda sencillez, eficacia y seguridad.

Gran parte de esta problemática es derivada del empleo de las formas más comunes de estructuración de fenómenos geográficos, tales como catálogos de fenómenos y tesauros. Estos son gestionados por la mayoría de los sistemas de información en el entorno geográfico, ya sean SIG o IDEs, pero no solucionan las dificultades comentadas con anterioridad, consecuencia de la pobre y rudimentaria modelización de la IG, es decir, la estructuración de los nombres (conceptos), códigos, atributos y otras características asociadas a la geometría. Las definiciones usualmente más extendidas de Catálogo y Tesoro, son las que se exponen a continuación:

1. Catálogo de fenómenos (*feature catalogue*); define los tipos de elementos (*features*) sus operaciones, atributos y asociaciones representadas en los datos geográficos. Estos son indispensables para convertir datos en información utilizable (ISO, 2005). En la práctica este tipo de catálogos poseen importantes limitaciones, tales como la ausencia de cualquier tipo de estructuración y de relación entre elementos de manera explícita. Lo único que puede encontrarse, en ocasiones, es una jerarquía entre clases de fenómenos, determinada por los códigos asociados a las mismas.

2. Tesauros, conforme al International Standard Organization (ISO), es un vocabulario de un lenguaje de indización controlado (conjunto controlado de términos extraídos del lenguaje natural y utilizados para representar, de forma breve, los temas de los documentos), organizado formalmente con objeto de hacer explícitas las relaciones, a priori, entre conceptos (por ejemplo “más genérico” o “más específico que”) (ISO, 1985 – 1986).

La construcción de tesauros supone una considerable mejora en la estructuración de la información respecto a los catálogos de fenómenos, consecuencia de la desaparición de la imprecisión y

ambigüedad en el uso del lenguaje; motivada por la existencia de sinónimos y polisemias; y del establecimiento de relaciones (ej.: Término Genérico, Término Específico, Use, etc.) entre los conceptos.

Las limitaciones estructurales, comentadas anteriormente, y la utilización de diversos vocabularios para describir la información presente en los servicios Web de SIG e IDEs, evidencian diversos problemas que se manifiestan al preguntar e interpretar resultados producidos por la búsqueda sobre diferentes catálogos distribuidos (Bernad y otros, 2003). Esto refleja la necesidad de un cambio en dichas herramientas (SIG e IDEs) para dejar de ser utilizadas como “simples” almacenes de datos con representación espacial y de escasa utilización como instrumentos en la planificación territorial, toma de decisiones y/o gestión de recursos, para dar un salto cualitativo en funcionalidad y posibilidades.

La Ingeniería Ontológica (Gómez-Pérez, 2003), surgida de la Web Semántica, proporciona soluciones a los problemas actuales relacionados con la accesibilidad y las búsquedas distribuidas de IG. Fundamentalmente, esta mejora se basa en el cambio de perspectiva en los procesos de búsqueda, ya que se pasa de utilizar palabras clave, es decir, consideración únicamente de aspectos sintácticos, a centrarse en los significantes de los conceptos, es decir, en la semántica de la información. De esta manera, se obvia la asunción de que los datos deben ser entendidos, exclusivamente, por los usuarios y se pasa a un proceso de entendimiento recíproco entre hombre y máquina, en el que las máquinas pasan a “comprender” los datos que procesan. Por estas razones, la definición de modelos ontológicos globales que logren fácil accesibilidad y común estructura de la información geoespacial se hace más que necesaria, más aún si la información compartida por todos los actores del territorio puede originar flujos de información y de conocimiento.

Entre las ventajas que las ontologías pueden proporcionar a la IG destacan (Vilches y otros, 2006);

i) La disminución de la confusión semántica. Reduce la ambigüedad terminológica al considerar sinónimos y polisemias, repercutiendo sobre la comunicación y gestión de la información. ii) La

posibilidad de reutilización de conocimientos. Esto permite el aprovechamiento de ontologías realizadas sobre cualquier área de la IG, consecuencia de que el desarrollo de ontologías refleja formas concretas de ver el mundo. iii) La traducción e intersección semántica a través de correspondencia (mappings) empleados para describir semejanzas entre fenómenos (ej.: río, river, rivière y fleuve) y entre diferentes ontologías (ej.: ontología de fenómenos hidrográficos y de fenómenos urbanos). iv) Integración de información de diferentes fuentes o bases de datos, lo que contribuye a crear la necesaria armonización de la IG.

De todo esto se deduce que las ontologías constituyen el complemento ideal para los SIG Web y las IDEs, más aún una vez que éstos comienzan a extenderse, concediendo acceso público y abierto a la geo-información mediante múltiples servidores y servicios y, en la medida en que pueden contribuir, de forma efectiva y práctica, a mejorar la gestión de la información (consulta, obtención y recuperación) y, por tanto, ayudar en los procesos de toma de decisiones en la planificación y gestión territorial.

En definitiva, las ontologías van a aportar muchas utilidades al mundo de la información geográfica, entre las que destacan, según Torres (2003), la mejora en la comunicación, habida cuenta de su dedicación a reducir la confusión terminológica y conceptual en el ámbito del dominio geográfico y la interoperabilidad semántica, ya que las ontologías potenciarán el intercambio de datos geográficos gracias a la semántica que se encuentra en ellas.

5. CONCLUSIONES

La abundante información que genera la Sociedad de la Información en la que vivimos inmersos provoca que el funcionamiento de la Web actual no sea adecuado para la gestión de tales volúmenes de información. Consecuencia de esta situación se produce el origen de la Web Semántica, cuyas aplicaciones y herramientas tienen como finalidad proporcionar un enorme salto cualitativo en funcionalidad y posibilidades para la red y sus usuarios. En suma, el objetivo de la Web Semántica

es que la Web pase de ser una colección de documentos a convertirse en una base de conocimiento y para esto las *ontologías* son el instrumento más eficaz.

La imbricación de las ontologías con la información geográfica parece ser más que necesaria, dadas las enormes ventajas que proporcionan a los usuarios de esta información. La implementación de las ontologías en el campo de la IG hará posible mejorar las búsquedas de información, el procesamiento de los datos en la red, una explotación de resultados más potente, se reducirá el tiempo empleado, se abrirán nuevos horizontes y posibilidades que potenciarán multitud de usos y, todo ello, revertirá, de forma directa, en un aumento de la confianza sobre la red, su utilidad y su uso, mientras, por otro lado, esto repercutirá en un aumento cualitativo de la potencia de explotación de los datos vía SIG o vía IDE.

6. BIBLIOGRAFÍA

Berners-Lee, T.; Hendler, J.; Lassila, O. (2001) "The Semantic Web", *Scientific American*.

Thomas R. Gruber. (1993) "A Translation Approach to Portable Ontology Specifications". *Knowledge Acquisition*, 5(2), pp.199-220.

Torres Rodríguez, N. (2003) *Imágenes en la web semántica: estándares, aplicaciones y organización de sitios en la red*. Universidad Carlos III de Madrid. (Tesina).

Gómez-Pérez, A.; Fernández-López, M.; Corcho, O. (2003). *Ontological Engineering: with examples from the areas of knowledge management, e-commerce and the Semantic Web*. London. Springer-Verlag.

Castells, P. La Web Semántica (2003) C. Bravo, M. A. Redondo (Eds.), *Sistemas Interactivos y Colaborativos en la Web*. Ediciones de la Universidad de Castilla - La Mancha, ISBN: 84-8427-352-0, pp. 195-212.

Studer, Benjamins, Fensel. Knowledge Engineering: Principles and Methods. *Data and Knowledge Engineering*. 25 (1998) 161-197

Noy, N. F.; McGuinness, D. L. (2001) *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.

GUIMET, J.: (2003) *Internet, información y territorio*. Proceedings del 1er Congreso Internacional sobre Territorio y Ciudad. La metrópolis presente y futura. Barcelona.

International Standards Organization, ISO 19110 (2005) *Geographic information -- Methodology for feature cataloguing*.

International Standards Organization, ISO 5964:1985 (1985) *Documentation - Guidelines for the establishment and development of multilingual thesauri*.

International Standards Organization, ISO 2788:1986 (1986) *Guidelines for the establishment and development of monolingual thesauri*.

Bernad, L.; Einspanier, U.; Haubrock, S.; Hübner, S.; Kuhn, W.; Lessing, R.; Lutz, M.; Visser, U. (2003) *Ontologies for intelligent search and semantic translation in Spatial Data Infrastructures, Photogrammetrie - Fernerkundung - Geoinformation* (6).

Vilches Blázquez L.M., Rodríguez Pascual, A.F., Bernabé Poveda, M.A. (2006). *Ingeniería ontológica: El camino hacia la mejora del acceso a la información geográfica en el entorno web*. Avances en las Infraestructuras de Datos Espaciales. Granell, Carlos & Gould, Michael (editores). Col·lecció "Treballs d'Informàtica i Tecnologia". Núm. 26. Universitat Jaume I. pág 95 – 103. ISBN: 84-8021-590-9. Castellón de la Plana.

ÍNDICE DE FIGURAS

```
<html><head><title>⌘ □ Ⓜ Ⓜ </title></head>
<body>
<h1>✂ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ </h1>
<table>
<caption>✂ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ </caption>
<tr>
<td>Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ </td><td>Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ </td><td>Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ Ⓜ </td>
</tr> . .
</body>
</html>
```

Figura 1. Ejemplo de visualización y entendimiento del código HTML por los ordenadores.

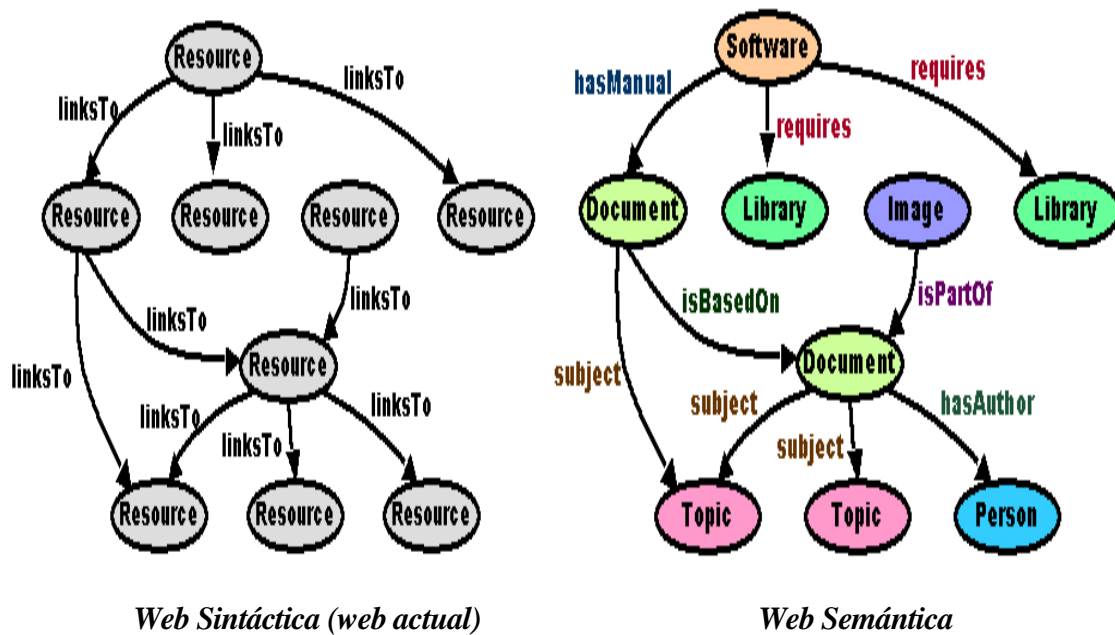


Figura 2. Ejemplo de relaciones entre recursos de la Web Sintáctica frente a la Web Semántica.