
Agregador automático de servicios web geoespaciales

Automatic aggregator of geo-spatial web services

Javier LACASTA, Francisco J. LOPEZ-PELLICER, Walter RENTERIA-AGUALIMPIA, Javier NOGUERAS-ISO

Universidad de Zaragoza, María de Luna 1, 50018, Zaragoza, España, {jlacasta|fjlopez|walterra|jnog}@unizar.es>

Resumen

La información geoespacial es comúnmente proporcionada por servicios web estándar. Sin embargo, en muchos casos es difícil encontrar qué servicios contienen la información deseada. Este artículo propone un proceso para construir un modelo geoespacial de datos enlazados que facilite la búsqueda de información geoespacial en la web.

Palabras clave: Análisis de contenido. Indización. Recuperación de información. Información geoespacial. Servicios web OGC.

Abstract

Geospatial information is commonly provided by standardized web services. However, in many cases, it is difficult to find the services containing the desired resources. This paper proposes a process to construct a Linked Data model of geospatial resources that facilitates the search of geospatial information in the web.

Keywords: Content analysis. Indexing. Information retrieval. Geospatial information. OGC web services.

1. Introducción

La información geoespacial es un recurso común para la toma de decisiones a nivel corporativo y personal (European Parliament and Council, 2007). Los ciudadanos la usan para tareas como planear rutas o seleccionar restaurantes. Los gobiernos la aplican en áreas como el catastro o la gestión del medio-ambiente.

En las últimas dos décadas, se ha estandarizado la forma en la que la información espacial es creada y proporcionada. Un ejemplo de esto son los estándares del Open Geospatial Consortium (OGC) y la International Organization for Standardization (ISO). Su trabajo ha permitido armonizar la forma en que los datos geoespaciales y sus metadatos son almacenados en catálogos (ISO, 2003a) y como se proporcionan a través de servicios web (ISO, 2005). Siguiendo estas normas, se han desarrollado catálogos comunitarios que describen el contenido de los servicios geoespaciales públicos (Nogueras-Iso, 2009).

Sin embargo, estos catálogos no solucionan todos los problemas de acceso a la información. Por una parte, requieren el registro voluntario y manual de los geo-servicios. Por otra parte, la falta de experiencia en la descripción de servicios web (Olfat, 2012) ha generado una abundancia de recursos geoespaciales pobremente descritos.

El uso de tecnologías semánticas para anotar la web (schema.org, micro formatos, datos enla-

zados...) han servido para mejorar la forma en la que los recursos no textuales son descritos y publicados. Estas mismas tecnologías pueden usarse en el campo geoespacial para crear un catálogo semántico de servicios (y datos) geoespaciales.

Como avance en esta dirección, este artículo propone un proceso para generar un modelo de datos enlazados (*Linked Data*) a partir de descripciones de servicios OGC existentes. El proceso usa un *crawler* especializado para identificar servicios web OGC y obtener descripciones acerca de su contenido. Estas descripciones son emparejadas con conceptos de modelos de organización del conocimiento comunes para generar una colección semánticamente anotada que pueda ser publicada como datos enlazados.

2. Trabajo relacionado

Hay varios trabajos en la literatura que describen el uso de modelos semánticos en infraestructuras espaciales. Janowicz (2010) hace una completa revisión sobre cómo usarlos para mejorar el descubrimiento, acceso, actualización, procesamiento y visualización de los datos espaciales. Para el descubrimiento sugiere el uso de ontologías para expandir consultas. Adicionalmente, recomienda el uso de lenguajes formales para describir semánticamente los recursos. En esta línea, Lutz (2006) propone un modelo para formular consultas intuitivamente que está basado en el uso de vocabularios de dominio conocidos. Latre (2009) se centra en la in-

teroperabilidad de modelos. Se propone la integración de datos hidrológicos usando una ontología multilingüe que facilita el alineamiento entre modelos de datos locales de fuentes diferentes. Otro ejemplo es el buscador SPIRIT (Jones, 2004) que incluye una ontología geográfica que juega un papel vital en la formulación e interpretación de las consultas, generación de índices espaciales y extracción de metadatos. Finalmente, Lacasta (2007) se enfoca en definir la especificación de un servicio de ontologías para proporcionar modelos de clasificación usados en el campo geoespacial. Estos modelos son convenientemente descritos por el interfaz del servicio para facilitar su uso e integración.

Centrado en la generación de descripciones semánticas de datos geoespaciales, Silva (2006) describe un proceso para identificar automáticamente el ámbito de documentos web. Para ello usan un proceso de reconocimiento y desambiguación de referencias geográficas en texto y asignan el ámbito espacial de los documentos con una función de ranking. Un componente central del proceso es una ontología usada como fuente de nombres y relaciones entre conceptos geográficos. Otro ejemplo es Stadler (2011), que presenta la transformación del modelo de OpenStreetMap usado para recopilar entidades espaciales de forma colaborativa a RDF usando como base vocabularios ya existentes.

Fuera del campo geoespacial hay muchos trabajos que describen procesos (semi-) automáticos para generar descripciones semánticas de recursos textuales. Por ejemplo, Wu (2007) propone un proceso para transformar los info-boxes de la Wikipedia en descripciones semánticas. Esto se hace procesando el contenido de los artículos para identificar categorías que puedan ser añadidas a los info-boxes. Bizer (2009) va más lejos describiendo un proceso para transformar la Wikipedia en su versión semántica (DBpedia) mediante el uso de diferentes técnicas de extracción de conocimiento. Thesen (2014) realiza un proceso equivalente para extraer conocimiento y anotar semánticamente la Encyclopedia of Life, un repositorio sobre las diferentes formas de vida de la tierra. Para ello usa como modelo base la DBpedia y herramientas de procesamiento de lenguaje natural.

Nuestra aproximación va en la línea de Bizer (2009), pero con diferencias estructurales causadas por el tipo de recurso procesado. Las descripciones de los servicios geoespaciales son generalmente escasas, y los datos binarios asociados (formas geométricas y valores numéricos) no pueden ser usados de forma sencilla para enriquecer las descripciones. Por tanto,

muchas aproximaciones de generación basadas en el procesado de texto libre no son apropiadas.

3. Generación de un modelo geoespacial de datos enlazados

Esta sección presenta nuestra propuesta de generación de un modelo geoespacial de datos enlazados a partir de descripciones de servicios web OGC. El diagrama del proceso se muestra en la Figura 1.

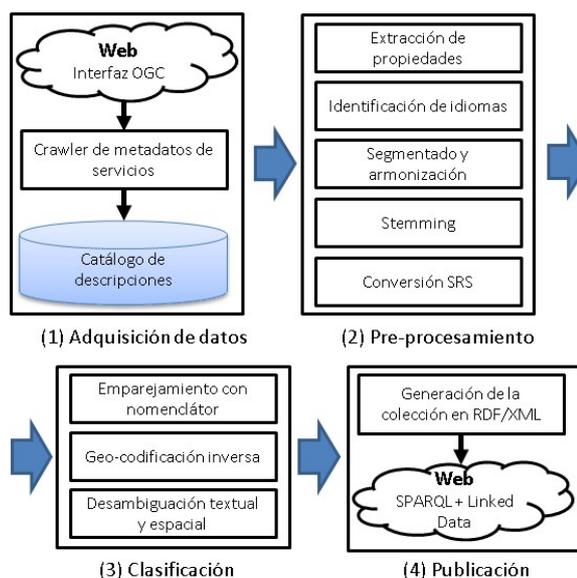


Figura 1. Proceso de anotación

El primer paso es la adquisición de las descripciones proporcionadas por los servicios web OGC a través de un *crawler* especializado. Dado que la estructura de estos metadatos varía dependiendo del tipo de servicio, la fase de pre-procesamiento extrae su contenido y lo armoniza separando la descripción de cada servicio de la descripción de los recursos contenidos. Estas descripciones son enriquecidas en la etapa de clasificación con conceptos de modelos de organización del conocimiento (temáticos, de lugar, organizaciones, y temporales). Finalmente, se generan metadatos RDF (Brickley, 2014) usando un modelo de Dublin Core extendido (ISO, 2003b) para su publicación como datos enlazados.

3.1. Adquisición de los datos

OGC ha liderado desde 1994 el desarrollo de especificaciones abiertas para acceder a la información geográfica. Ha generado un conjunto de especificaciones de servicios web que comparten un formato de petición común para

devolver información técnica y funcional (Get-Capabilities).

Para descubrir servicios OGC publicados en la web, hemos usado un *crawler* geoespacial especializado (Lopez-Pellicer, 2012b). Este *crawler* realiza consultas con términos relacionados con los estándares OGC (por ejemplo, request, getCapabilities, service, endpoint, WSDL) a motores de búsqueda tradicionales (*Bing* y *Google*) y analiza los resultados para identificar servicios. Además, explora los enlaces de cada página obtenida teniendo en cuenta la información cercana a cada enlace, lo relevante que es la página original y un valor de decaimiento que limita el número de enlaces a seguir.

3.2. Pre-procesamiento

Las descripciones de los servicios obtenidas por el *crawler* son procesadas para extraer términos útiles para su clasificación. La extracción se realiza usando hojas de estilo para cada tipo de servicio OGC. Este proceso genera una colección de registros de metadatos RDF describiendo cada recurso proporcionado por un servicio y otro registro para el servicio en sí mismo (enlazados entre ellos). Esta separación permite identificar los recursos geoespaciales independientemente del servicio web que los contiene. Es más, con la desambiguación apropiada puede ser posible identificar que dos recursos en servicios diferentes son equivalentes.

El contenido extraído es armonizado para simplificar las tareas de clasificación. Se realizan tareas como reemplazar guiones por espacios, partir palabras conteniendo mayúsculas o números y transformar las coordenadas espaciales de los registros a WGS 84 —latitud, longitud— (NIMA, 2000). También se identifica el idioma del texto, ya que en la etapa de clasificación se usan técnicas de procesamiento de lenguaje natural que son dependientes del idioma. Todo esto se hace usando expresiones regulares y librerías de software especializadas tales como GeoTools (Turton, 2008) y LingPipe (Carpenter, 2011).

3.3. Clasificación

Los metadatos RDF son enriquecidos con conceptos de modelos de organización del conocimiento. Se mantiene la información textual original (título, resumen, coordenadas) y se añaden relaciones a modelos espaciales, temáticos, de autoridades y temporales.

La información espacial es descrita principalmente en los servicios OGC mediante referencias espaciales directas (bounding box), pero

también se pueden encontrar referencias indirectas (topónimos) como texto libre en las descripciones del servicio. Para extraer estas referencias espaciales clasificamos los recursos de acuerdo a una ontología de divisiones administrativas (Lopez-Pellicer, 2012a). Primero se transforman los *bounding boxes* en topónimos usando un *geocoder* inverso creado a partir de la ontología (Renteria-Agualimpia, 2013). Después, se procesa el texto de las descripciones para identificar topónimos adicionales usando el software de procesamiento de lenguaje natural GATE (Cunningham, 2011). Se ha realizado un proceso básico de identificación de nombres basado en comparación de las palabras del texto (*tokens*) con un *gazetteer* (ignorando acentos y mayúsculas). Dicho *gazetteer* se ha generado a partir de las etiquetas de la ontología con los nombres oficiales y alternativos de los topónimos en el idioma del documento. Para reducir los falsos positivos se han filtrado aquellos que se corresponden con palabras del idioma del documento comúnmente usados en la descripción de recursos geoespaciales —por ejemplo, Bosque (Cádiz)—.

La descripción temática de un servicio suele estar explícitamente indicada en ciertos campos de los metadatos OGC. Sin embargo, no se indican los vocabularios de los que se han sacado los diferentes términos. Esto dificulta proporcionar una búsqueda homogénea. Para solventar este problema, los metadatos se han procesado, al igual que en el caso anterior, para identificar términos de diversos modelos de organización del conocimiento (GEMET, 2010; AGROVOC, 2009; EUROVOC, 2008; UNESCO, 2008). Las definiciones de conceptos, sinónimos y relaciones proporcionadas por estos modelos pueden ser usadas para mejorar los resultados de las consultas. En este caso, para tratar el problema de la existencia de plurales y aumentar los conceptos identificados, se ha realizado un proceso de obtención de raíces (*stemming*). Este proceso no se ha aplicado de forma general, ya que suele introducir falsos positivos; solo se ha usado para aquellos casos en los que no se ha identificado ningún emparejamiento con los modelos de organización del conocimiento seleccionados.

La información de quien ha creado un recurso o servicio también ha sido procesada para mejorar su calidad. Los proveedores de los datos son puntos de conexión naturales entre recursos y ayudan a extender los resultados con elementos relevantes. En este caso, no se ha seguido la misma aproximación que para los campos temático y espacial, ya que no se encontró ninguna ontología de autoridades que cubriera las orga-

nizaciones que crean información geoespacial. En su lugar hemos usado un transductor de estado finito llamado JAPE (Thakker, 2009). JAPE es un lenguaje de expresiones regulares integrado en GATE que permite identificar patrones genéricos en base a las anotaciones de los *tokens*. Usando este lenguaje se ha definido una plantilla que identifica organizaciones a partir de un conjunto de patrones que contienen las estructuras de nombrado más comunes de organizaciones (por ejemplo, Instituto <adjetivo> nacional). En este caso se ha hecho un análisis del texto con el etiquetador posicional de GATE para identificar nombres, adjetivos, verbos y adverbios. Sobre este texto etiquetado se ha usado el procesador de JAPE con la plantilla previamente indicada.

El último tipo de información procesada es la fecha a la que hace referencia la información de cada recurso. Los metadatos de los servicios OGC no proporcionan explícitamente esta información, pero algunas veces se puede encontrar como parte de alguno de los campos descriptivos (título, resumen). Para identificarlas también se ha usado JAPE usando patrones de fechas y representando los resultados de acuerdo a la ontología de tiempo del W3C (Hobbs, 2006).

Una descripción más detallada de todos estos procesos puede encontrarse en Lacasta (2014).

4. Validación del proceso

Para comprobar la calidad de los resultados obtenidos a través de este proceso, se ha usado para recopilar los Servicios OGC WMS de mapas (Beaujardiere, 2006) disponibles en España y se han analizado los resultados obtenidos.

Un WMS produce mapas de datos geo-referenciados en formato de imagen como PNG, GIF o JPEG. Estas imágenes son construidas superponiendo las capas de datos espaciales contenidas en el servicio (*layers*). Se ha seleccionado este tipo de servicio al ser uno de los más comunes. Sin embargo, se puede usar con otros tipos servicios OGC espaciales descritos de forma similar al WMS. Por ejemplo, servicios de fenómenos (*features*) (WFS) (Vretanos, 2005), o de coberturas (WCS) (Baumann, 2012). Un WFS permite describir operaciones de manipulación de datos geográficos (obtener y modificarlos). A diferencia de la imagen proporcionada por un WMS, que es estática, el resultado devuelto por un WFS son datos numéricos que se pueden procesar. El WCS es similar al WFS pero para colecciones de datos *raster*. A diferencia de los WFS, que permiten la descarga de información sobre elementos vectoriales (carre-

teras, ríos), los WCS proporcionan coberturas con los valores de cada variable del modelo usado para cada uno de los píxeles de una cuadrícula.

Las descripciones de los WMS de España se han obtenido restringiendo el *crawler* a servicios OGC de tipo WMS bajo el dominio .es. Se han identificado 1075 servicios con un total de 15651 capas. Dentro de las descripciones se han procesado los campos: Title (título), Name (nombre), Abstract (resumen), KeywordList (palabras clave), ContactOrganization (contacto) y EX GeographicBoundingBox (coordenadas). La tabla I indica el porcentaje de servicios y capas que tienen dichas propiedades. Puede observarse que la mayor parte de los servicios tienen título y nombre, pero el resto de propiedades están muchas veces incompletas. Solo la falta de contactos y *bounding box* en capas y servicios es una característica de las descripciones; el resto es debido a una mala descripción.

Campo	Servicios	Capas
Título	97,39%	96,09%
Nombre	97,67%	91,28%
Resumen	92,65%	52,54%
Palabras Clave	61,20%	37,51%
Contacto	71,62%	N.A.
Bounding Box	N. A.	100,00%

Tabla I. Disponibilidad de las propiedades

La tabla II muestra el nivel de completitud de los metadatos de servicio y de capa obtenidos después de aplicar el proceso a las descripciones obtenidas por el *crawler*.

Campo	Servicios	Capas
Bounding Box	N.A.	100%
Topónimo	96,93%	89,31%
Palabra Clave	91,62%	73,54%
Creador	64,83%	15,82%
Fecha	26,04%	9,12%

Tabla II. Disponibilidad de propiedades procesadas

Puede observarse que el nivel de éxito es heterogéneo. Un caso llamativo es la información de topónimo. A pesar de no haber *bounding boxes* en los servicios, ha sido posible extraer dicha

información de las descripciones textuales. Como problemas hay que destacar, la información de creador y fecha. La cobertura del creador es menor que los datos originales debido a que en muchos casos la información original describe personas no organizaciones, y nuestra aproximación actual no trata este caso. El bajo porcentaje de fechas es debido a la falta de referencias temporales en la mayoría de descripciones.

Para determinar la calidad de los registros obtenidos, hemos revisado un conjunto aleatorio de 200 servicios y capas con valores en las propiedades generadas. La tabla III muestra la precisión de los elementos (% de valores correctos). El valor N. A. indica que el análisis no se ha realizado debido a la falta de datos.

<i>Campo</i>	<i>Servicios</i>	<i>Capas</i>
Bounding Box	N. A.	99,00%
Topónimo	92,43%	93,75%
Palabra Clave	79,46%	84,13%
Creador	97,95%	82,67%
Fecha	94,54%	99,00%

Tabla III. Calidad de los registros en términos de precisión

En la tabla III puede observarse que unos pocos *bounding box* son incorrectos. Esto es un error crítico ya que impide el uso de la capa (refiere a un área errónea). Estos errores pueden estar causados por problemas como el uso inadecuado de los sistemas de referencia espacial o el intercambio de las coordenadas X e Y (Hill, 2006). Respecto a los errores en las palabras clave se ha detectado que son debidos a la existencia de conceptos no temáticos en los tesauros seleccionados. Por ejemplo, el concepto "Aprobación" está en GEMET pero no es útil para la clasificación. Eliminar estos conceptos de los tesauros mejoraría la calidad de los resultados.

La publicación del modelo generado como datos enlazados simplemente requiere la creación de un punto de acceso SPARQL (W3C, 2013) y un servicio de transformación en datos enlazados. Como ventajas de esta forma de publicación hay que destacar la posibilidad de definir filtros de búsqueda complejos y permitir la construcción de sistemas de inferencia basados en los modelos de organización del conocimiento usados. Para realizar esta tarea, en nuestro sistema se ha utilizado el servicio de consulta SPARQL

Fuseki (Apache, 2014) y el servicio de publicación como datos enlazados Pubby (Cygniak, 2011).

5. Conclusiones

Este artículo ha descrito un proceso para crear un modelo de datos enlazados de descripciones de servicios web geoespaciales. Este proceso identifica servicios web OGC públicos, y enriquece sus metadatos haciendo uso de modelos de organización del conocimiento (nomenclátors, tesauros, ontologías). El artículo ha mostrado su uso para reclasificar servicios de mapas OGC en España, generando un modelo de datos enlazados con calidad aceptable.

Como trabajo futuro se quiere mejorar el proceso de clasificación a través del uso de las relaciones entre los servicios y los recursos contenidos para enriquecer las descripciones. Por otra parte, se quiere avanzar en el soporte de inferencia, para que el usuario del sistema pueda decidir sobre cómo utilizarla en el sistema.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el Gobierno de España a través del proyecto TIN2012-37826-C02-01; el Instituto Geográfico Nacional (IGN); GeoSpatiumLab S. L. y Zeta Amaltea S. L. El trabajo de Walter Renteria Agualimpia ha sido cofinanciado por el Gobierno de Aragón a través de la beca B181/11.

Referencias

- AGROVOC (2009). Agriculture Vocabulary. // Food and Agriculture Organization of the United Nations. <http://aims.fao.org/standards/agrovoc> (28-05-2014).
- Apache (2014). Fuseki: serving RDF data over HTTP. // Apache Software Foundation. http://jena.apache.org/documentation/serving_data (28-05-2014).
- Baumann, P. (ed.) (2012). OGC WCS 2.0 Interface Standard - Core. Version 2.0.1. Open Geospatial Consortium, 2012.
- Beaujardiere, J. (ed.) (2006). OpenGIS Web Map Server Implementation Specification. Version 1.3.0. Open Geospatial Consortium, 2006.
- Bizer, C.; Lehmann, J.; Kobilarov, G.; Auer, S.; Becker, C.; Cygniak, R.; Hellmann, S. (2009). Dbpedia - a crystallization point for the web of data. // Web Semantics: Science, Services and Agents on the World Wide Web. 7:3 (Septiembre 2009) 154-165.
- Brickley, D.; Guha, R. V. (2014). RDF Schema 1.1. W3C Recommendation 25 February 2014. // World Wide Web Consortium, 2014.
- Carpenter, B.; Baldwin, B. (2011). Natural language processing with LingPipe 4. LingPipe Publishing, 2011.
- Cunningham, H.; Maynard D.; Bontcheva, K. (2011). Text processing with GATE (Version 6). Sheffield, UK: University of Sheffield Department of Computer Science, 2011.
- Cygniak, R.; Bizer, C. (2011). Pubby A Linked Data Frontend for SPARQL Endpoints. // Freie Universität Berlin.

- <http://wifo5-03.informatik.uni-mannheim.de/pubby/> (28-05-2014).
- European Parliament and Council (2007). Directive 2007/2/ec of the european parliament and of the council of 14 march 2007 establishing an infrastructure for spatial information in the european community (inspire). // Official Journal of the European Union- 50 (Abril 2007) 1-14.
- EUROVOC (2008). Multilingual Thesaurus of the European Union. // European Parliament. <http://eurovoc.europa.eu/> (28-05-2014).
- GEMET (2010). GEneral Multilingual Environmental Thesaurus. // European Environment Agency. <http://www.eionet.eu.int/GEMET> (28-05-2014).
- Hill, L. (2006). Georeferencing Elements in Metadata Standards. // Georeferencing: The Geographic Associations of Information. Digital Libraries and Electronic Publishing, 2006. 155-184.
- Hobbs, J.R.; Pan, F. (2006). Time Ontology in OWL. W3C Working Draft 27 September 2006. // World Wide Web Consortium, 2006.
- ISO (2003a). Geographic information - Metadata. ISO 19115:2003. // International Organization for Standardization, 2003.
- ISO (2003b). Information and documentation - The Dublin Core metadata element set. ISO 15836:2003, International Organization for Standardization, 2003.
- ISO (2005). Geographic information - Services. ISO/DIS 19119. // International Organization for Standardization, 2005.
- Janowicz, K.; Schade, S.; Bröring, A.; Kebler, C.; Maue, P.; Stasch, C. (2010). Semantic enablement for spatial data infrastructures. // Transactions in GIS. 14:2 (Abril 2010) 111-129.
- Jones, C. B.; Abdelmoty, A. I.; Finch, D.; Fu, G. (2004). The spirit spatial search engine: Architecture, ontologies and spatial indexing. // Proc. 3rd Int. Conf. On Geographic Information Science (2004) 125-139.
- Lacasta, J.; Nogueras-Iso, J.; Béjar, R.; Muro-Medrano, P. R.; Zarazaga-Soria, F. J. (2007). A web ontology service to facilitate interoperability within a spatial data infrastructure: Applicability to discovery. // Data & Knowledge Engineering. 63:3 (Diciembre 2007) 947-971.
- Lacasta, J.; Lopez-Pellicer, F. J.; Renteria-Agualimpia, W.; Nogueras-Iso, J. (2014). Improving the visibility of geospatial data on the Web. // Aceptado en Digital Libraries. ACM/IEEE Joint Conference on Digital Libraries. International Conference on Theory and Practice of Digital Libraries. (Septiembre 2014).
- Latre, M.A.; Lacasta, J.; Mojica-Abrego, E.; Nogueras-Iso, J.; Zarazaga-Soria, F. J. (2009). An approach to facilitate the integration of hydrological data by means of ontologies and multilingual thesauri. // Lecture Notes in Geoinformation and Cartography. Advances in GIScience. (2009) 155-171.
- Lopez-Pellicer, F. J.; Lacasta, J.; Florczyk, A.; Nogueras-Iso, J.; Zarazaga-Soria, F. J. (2012a). An ontology for the representation of spatio-temporal jurisdictional domains in information retrieval systems. // International Journal of Geographical Information Science. 26:4 (2012) 579-597.
- Lopez-Pellicer, F. J.; Renteria-Agualimpia, W.; Nogueras-Iso, J.; Zarazaga-soria, F. J.; Muro-Medrano, P.R. (2012b). Towards an active directory of geospatial web services. Bridging the Geographic Information Sciences. // Lecture Notes in Geoinformation and Cartography. Bridging the Geographic Information Sciences. (2012) 63-79.
- Lutz, M.; Klien, E. (2006). Ontology-based retrieval of geographic information. // International Journal of Geographical Information Science. 20:3 (March 2006) 233-260.
- NIMA (2000). Nima technical report tr8350.2. World geodetic system 1984, its definition and relationships with local geode-tic systems. Department of defense, 2000.
- Nogueras-Iso, J.; Barrera, J.; Rodríguez-Pascual, A.; Recio, R.; Laborda, C.; Zarazaga-Soria, F. J. (2009). SDI Convergence: Research, Emerging Trends, and Critical Assessment, chapter Development and deployment of a services catalog in compliance with the INSPIRE metadata implementing rules. // The Netherlands Geodetic Commission. (2009) 21-34.
- Olfat, H.; Kalantari, M.; Rajabifard, A.; Senot, H.; Williamson, I. P. (2012). Spatial metadata automation: A key to spatially enabling platform. // International Journal of Spatial Data Infrastructures Research. 7 (2012) 173-195.
- Renteria-Agualimpia, W.; Lopez-Pellicer, F. J.; Lacasta, J.; Zarazaga-Soria, F. J.; Muro-Medrano, P. R. (2013). Identifying hidden geospatial resources in catalogues. // Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics, 2013.
- Silva, M.J.; Martins, B.; Chaves, M.; Afonso, A. P.; Cardoso, N. (2006). Adding geographic scopes to web resources. // Computers, Environment and Urban Systems. 30 (2006) 378-399.
- Stadler, C.; Lehmann, J.; Höffner, K.; Auer, S. (2011). LinkedGeoData: A core for a web of spatial open data. // Semantic Web. 3:4 (2011) 333-354.
- Thakker, D; Osman, T.; Lakin, P. (2009). GATE JAPE Grammar Tutorial. <http://gate.ac.uk/sale/thakker-jape-tutorial/GATE%20JAPE%20manual.pdf> (28-05-2014).
- Thessen A.E.; Sims Parr, C. (2014). Knowledge Extraction and Semantic Annotation of Text from the Encyclopedia of Life. // PLoS ONE. 9:3 (Marzo 2013).
- Turton, I. (2008) Geo tolos. // Open Source Approaches in Spatial Data Handling. Springer, 2008. 153-169.
- UNESCO (2008). UNESCO Thesaurus. UNESCO. <http://databases.unesco.org/thesaurus/> (28-05-2014).
- Vretanos, P. A. (ed.) (2005). OpenGIS Web Feature Service Implementation Specification. Version 1.1.0. Open Geospatial Consortium, 2005.
- Wu, F.; Weld, D. S. (2007). Autonomously semantifying wikipedia. // Proceedings of the sixteenth ACM conference on Conference on information and knowledge management. (2007) 41-50.
- W3C (2013). SPARQL 1.1 Overview. W3C Recommendation 21 March 2013. // World Wide Web Consortium. 2013.

Enviado: 2014-04-09. Segunda versión: 2014-06-02.
Aceptado: 2014-07-10.
