

Ingeniería ontológica: El camino hacia la mejora del acceso a la información geográfica en el entorno web

Luis Manuel Vilches Blázquez[†], Antonio F. Rodríguez Pascual[†], Miguel Ángel Bernabé Poveda[‡]

[†] Subdirección General de Aplicaciones Geográficas.
Instituto Geográfico Nacional.

C/ General Ibáñez de Íbero, 3. 28003. Madrid.

Tlf: 91 597 9660 Fax: 91 597 9646 e-mail: {lmvilches|afrodriguez}@fomento.es

[‡] ETSI Topografía, Geodesia y Cartografía.
Universidad Politécnica de Madrid.

Km. 75 Autovía de Valencia, 28031. Madrid.

Tlf: 91 336 6487. Fax: 91 336 7932. e-mail: ma.bernabe@upm.es

Resumen

La consideración actual de interoperabilidad se centra en el establecimiento de un lenguaje común que permite difundir y compartir cualquier tipo de información geográfica. Esto es lo que se ha dado en llamar interoperabilidad sintáctica. Tal concepción obvia los problemas derivados de la heterogeneidad semántica de la geo-información.

Las formas actuales de organización de la geo-información (catálogos de fenómenos y tesauros) no son eficaces para solucionar los problemas derivados de la integración semántica de conjuntos de geodatos. Por ello, el desarrollo de ontologías se presenta como el instrumento adecuado para alcanzar interoperabilidad semántica en el entorno de las Infraestructuras de Datos Espaciales.

Esta comunicación, también, presenta un caso de uso en el que se describen las principales características de una ontología-marco de fenómenos hidrográficos con vistas a ser implementada en la IDEE. Dicha ontología tiene como objetivo definir un modelo de explotación más eficaz en la consulta, acceso y recuperación de la geo-información en el entorno web.

Palabras clave: Ontología, Web Semántica, Ingeniería Ontológica, Información Geográfica, Catálogo de Fenómenos, Tesouro, Infraestructura de Datos Espaciales, IDEE.

1 Introducción

Una de las principales motivaciones para el desarrollo de las Infraestructuras de Datos Espaciales (IDEs) es hacer más eficiente el trabajo con los geodatos [1, 2]. Para ello, se crean los estándares del Comité Técnico 211 del International Organization for Standardization (ISO/TC 211)¹, World Wide Web Consortium (W3C)² y del Open Geospatial Consortium (OGC)³ con los que se resuelven problemas derivados de la diferencia de formatos de información, consiguiendo interoperabilidad, es decir, *“la capacidad de comunicarse, para ejecutar programas, o para transferir datos entre varias unidades funcionales de una manera que requiera al usuario tener poco o nada de conocimiento de las características únicas de esas unidades”*[3]. Con esto se alcanza una sintaxis homogénea para datos y servicios geográficos, es decir, interoperabilidad sintáctica.

La información ofrecida a través de los diferentes servicios [Web Map Service (WMS), Web Feature Service (WFS), Web Coverage Service (WCS), etc.] debe ser almacenada, ofertada y mantenida al nivel más adecuado [4]. Pero es generada, mantenida y actualizada por diversos productores, originando gran disparidad de fuentes y terminología. Adicionalmente, y debido a lo anterior, se encuentran diferencias de escala, tratamiento e interés. Esta situación no facilita los procesos de búsqueda, acceso e interpretación a la Información Geográfica (IG), resultado de la ausencia de una interoperabilidad semántica.

Este artículo está dirigido a la comunidad geográfica que está interesada en conocer los avances tecnológicos en las IDEs con repercusión directa sobre la IG. A continuación, se describe el estado de estructuración actual de la geo-información. En la sección 3 se realiza una introducción al concepto de ontología, citando algunas de sus ventajas y posibilidades. Un breve comentario sobre las características de una ontología marco de fenómenos hidrográficos se realiza en la sección 4. Finalmente, en la sección 5 se proporcionan algunas conclusiones y líneas de trabajo futuro.

¹ <http://www.isotc211.org/>

² <http://www.w3.org/>

³ <http://www.opengeospatial.org/>

2 Organización actual de la Información Geográfica

La IG alojada en los servicios web de las IDEs presenta contenidos y estructuras heterogéneas, derivados de la falta de consenso y de las inercias de los procesos de producción. Estos factores obstaculizan la consecución de una interoperabilidad *sensu stricta* (sintáctica y semántica) generando dificultades en las tareas de consulta, recuperación, explotación, actualización y visualización de la geo-información, en las que el usuario demanda sencillez, eficacia y seguridad.

Actualmente, la modelización semántica de la IG, es decir, la estructuración de los nombres, códigos, atributos y otras características asociadas a la geometría, es pobre y rudimentaria en la mayoría de los sistemas informáticos que la gestionan, ya sean Sistemas de Información Geográfica (SIG) o Infraestructuras de Datos Espaciales (IDEs). Esto dificulta la posibilidad de conseguir eficientes búsquedas, accesos y recuperaciones de dicha información, consecuencia de las formas de estructuración de fenómenos geográficos empleadas por las organizaciones productoras de cartografía en nuestro país, entre las que destacan:

1.- El catálogo de fenómenos (*feature catalogue*) es la forma más sencilla y utilizada para la organización de los conceptos asociados a la pura geometría (puntos, líneas, polígonos y textos) que describen la IG. Se corresponde con un simple listado de fenómenos geográficos agrupados en clases homogéneas, por ejemplo: municipio, montaña, acequia, iglesia, carretera, etc. Normalmente cada clase viene identificada por un código único que sirve para marcar las instancias (el Municipio de Madrid, el Teide o la Catedral de Burgos) pertenecientes a cada clase y que, eventualmente, pueden ir acompañadas de una definición y un conjunto de atributos.

Este listado posee 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. La norma internacional ISO19110⁴ “Geographic Information – Methodology for feature cataloguing” establece cómo crear y describir de manera normalizada un catálogo de fenómenos.

2.- Por otro lado, los tesauros, referencia básica de multitud de trabajos de todas las áreas de conocimiento, son vocabularios controlados que manifiestan relaciones, tanto semánticas como genéricas, entre conceptos que reflejan una parte del

⁴ <http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=39965>

conocimiento. Con su uso se describe, básicamente, las relaciones entre los campos semánticos de cada término, considerando la inclusión de unos en otros, la similitud o la existencia de cierto solape, sin describir más matices.

En la actualidad, existen dos normas internacionales para la construcción de tesauros, concretamente, la ISO 2788-1986⁵, orientada a la creación y desarrollo de tesauros monolingües y la ISO 5964-1985⁶, dirigida a los tesauros multilingües. Estas normas han sido adoptadas y traducidas por la Asociación Española de Normalización y Certificación (AENOR)⁷.

La construcción de tesauros supone una considerable mejora en la estructuración de la información respecto a los catálogos de fenómenos. Este hecho deriva 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. Estas mejoras provocan un leve acercamiento hacia la consecución de cierta interoperabilidad semántica, pero aún con muchas limitaciones, consecuencia de la simplicidad de las relaciones y de su falta de entendimiento con las máquinas.

En definitiva, las formas de organización actuales no facilitan la integración semántica del conjunto de la información, debido a sus carencias estructurales y a su ausencia de legibilidad por las máquinas, ya que están pensadas y realizadas para ser entendidas por el hombre. Por estos motivos, la gestión más sencilla y rutinaria de los datos se hace complicada y, en ocasiones, harto difícil. Además, esta situación provoca el incremento del coste en tiempo de cualquier información conseguida en la red, a pesar de la potencia y sagacidad crecientes de los buscadores actuales disponibles.

3 Ontologías

Las limitaciones estructurales, comentadas anteriormente, y la utilización de diversos vocabularios para describir la información presente en los servicios IDEs, evidencian diversos problemas que se manifiestan al preguntar e interpretar resultados producidos por la búsqueda sobre diferentes catálogos distribuidos [5]. Esto refleja la necesidad de un cambio que tenga como objetivo proporcionar un salto cualitativo, en funcionalidad y posibilidades, a la IG presente en la web.

⁵ <http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=7776>

⁶ <http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=12159>

⁷ <http://www.aenor.es>

La Ingeniería Ontológica, surgida de la Web Semántica⁸, proporciona los requisitos necesarios para mejorar las búsquedas de IG. Esta mejora se debe a que en lugar de utilizar palabras clave en los procesos de búsqueda, se centra en los significantes de los conceptos, es decir, en la semántica de la información. De esta manera, se obviará la asunción de que los datos deben ser entendidos exclusivamente por los usuarios y se pasará a un proceso de entendimiento recíproco entre hombre y máquina, en el que las máquinas pasarán a “*comprender*” los datos que procesan, actuando sin la necesaria y continuada supervisión actual.

Circunscribiéndose a las ontologías, una de las definiciones más divulgadas es la aportada por [6]. Afirma que una ontología constituye *una especificación explícita y formal de abstracciones mentales, las cuales se conforman mediante un acuerdo de la comunidad experta en un dominio y en un diseño para un propósito específico*. Las *abstracciones mentales* establecen vocabularios que reflejan los fenómenos de la realidad a través de conceptos consensuados por una comunidad experta [7, 8, 9]. Por otro lado, la *especificación formal* de un vocabulario se puede dar en forma de lista plana de palabras, diccionario, taxonomía, diagrama Entidad – Relación, modelo UML (Unified Markup Language), esquema de XML (eXtensible Markup Language) y muchos otros posibles [10]. Además, [11] sostiene que las ontologías definen un *conjunto de relaciones entre sus términos, así como las reglas que combinan términos y relaciones que amplían las definiciones dadas en el vocabulario*.

Entre las ventajas proporcionadas por las ontologías destacan i) La disminución de la confusión semántica. Reduce la ambigüedad terminológica al considerar sinónimos y polisemias, repercutiendo sobre la comunicación. 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 *mappings* (cotejos) empleados para describir correspondencias entre fenómenos (ej.: río, river, rivière y fleuve) y entre diferentes ontologías (ej.: ontología de fenómenos hidrográficos y ontología de las ciudades)

De esto se deduce que las ontologías constituyen el complemento ideal para las IDEs, más aún una vez que éstas comienzan a extenderse concediendo acceso público y abierto a la geo-información mediante múltiples servidores y servicios.

⁸ <http://www.w3.org/2001/sw/>

A diferencia de las formas de organización actuales, la construcción de ontologías se presenta como una buena solución para alcanzar la interoperabilidad semántica entre los diferentes catálogos de IG. De esta manera se originará una importante mejora en la representación de esta información, que repercutirá de forma directa en los sistemas de minería de datos.

En definitiva, las ontologías deben permitir conceptualizar el conocimiento, generando formas comunes y compartidas de ver el mundo, y ayudar a que los sistemas gestionen más información que datos, consecuencia de que el experto ha cedido parte de su conocimiento a los sistemas.

4 Un caso de uso: Ontología de fenómenos hidrográficos

El IGN ha comenzado a desarrollar una ontología de fenómenos hidrográficos, con la pretensión de establecerla como un marco semántico genérico y de uso para todas las organizaciones productoras. Su objetivo es optimizar la búsqueda y recuperación de la IG soportada por la Infraestructura de Datos Espaciales de España (IDEE).

4.1. Modelado de la ontología hidrográfica

El modelado de esta ontología está basado, fundamentalmente, en cuatro criterios:

- La Directiva Europea Marco del Agua⁹.
- Criterios semánticos, es decir, clasificación acorde a los significados.
- La clasificación y estructura reflejada en el Proyecto SDIGER¹⁰ (IDE para la gestión de las cuencas de los ríos Adour-Garonne y Ebro) [12].
- La herencia de las fuentes de estudio, para facilitar posibles mappings y para ser consecuente con las jerarquías de fenómenos creadas por los expertos.

4.2. Fuentes

Debido a la pretensión de establecerse como modelo de convergencia se han tenido en cuenta multitud de fuentes de información a diferentes escalas (EuroGlobalMap, EuroRegionalMap, Bases Cartográficas Numéricas del IGN, Diccionarios de datos de productores autonómicos, Tesauro de la UNESCO, GEMET, etc).

⁹ http://europa.eu.int/eurlex/pri/en/oj/dat/2000/l_327/l_32720001222en00010072.pdf

¹⁰ <http://www.idee.es/sdiger/>

En esta ontología aparecen alrededor de 100 conceptos relevantes pertenecientes al dominio de la hidrografía, incluyendo conceptos como río, embalse, mar, laguna, estuario, rápido, etc.

4.3. Relaciones ontológicas

Los diferentes conceptos que componen esta ontología aparecen estructurados en una taxonomía y relacionados mediante cuatro relaciones taxonómicas: Subclase-de, Descomposición-Disjunta, Descomposición-Exhaustiva y Partición.

Un concepto “ C_1 ” es *Subclase-de* otro concepto “ C_2 ” si y sólo si todas las instancias de “ C_1 ” son también instancias de “ C_2 ” [13].

Una *Descomposición-Disjunta* de un concepto C es un conjunto de subconceptos de C que no tienen instancias comunes y que no cubren C , es decir, puede haber instancias del concepto C que no son instancias de ninguno de los conceptos que forman la descomposición. [13] Un ejemplo de esta relación aparece en la Figura 2.

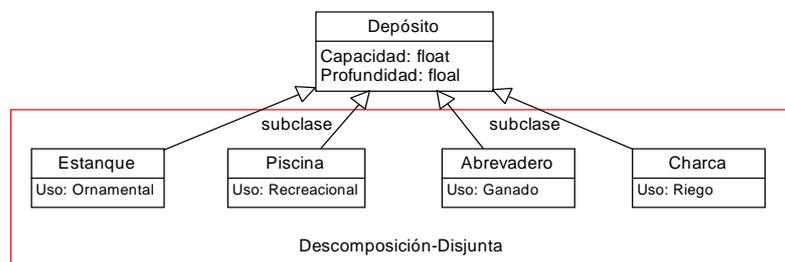


Fig. 1. Ejemplo de Descomposición-Disjunta

Una *Descomposición-Exhaustiva* de un concepto C es un conjunto de subconceptos de “ c ” que lo cubren, es decir, tal que no existe ninguna instancia de C que no sea instancia de al menos uno de los conceptos de la descomposición. Los conceptos que pertenecen a este conjunto pueden tener instancias y subconceptos comunes [13]. La Figura 3 muestra un ejemplo de este tipo de relación.

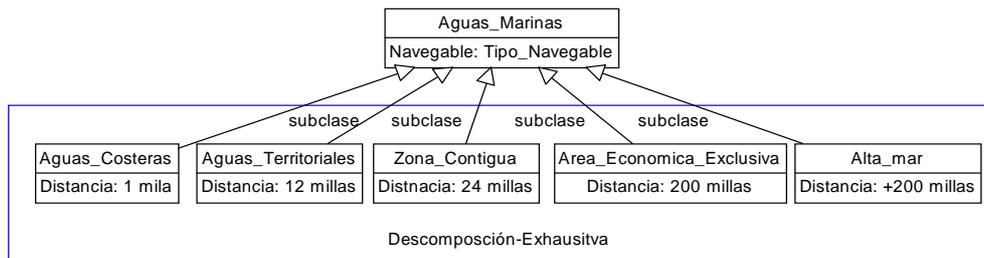


Fig. 2. Ejemplo de Descomposición-Exhaustiva

Una *Partición* de un concepto C es un conjunto de subconceptos de C que no tienen instancias ni subconceptos comunes y que cubren C [13]. Un ejemplo de partición es mostrado en la Figura 4.

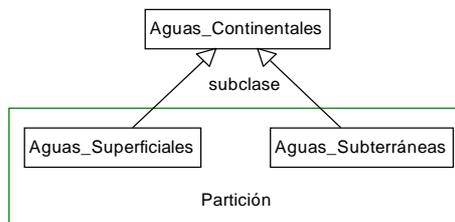


Fig.3. Ejemplo Partición

El siguiente paso del proceso de construcción de esta ontología fue la construcción de un diccionario de conceptos hidrográficos, consecuencia de la ausencia generalizada de definiciones en las fuentes consultadas. De esta manera, se contribuye al respaldo semántico de la IG. A continuación, se establecieron relaciones *ad hoc* entre los diferentes conceptos. Este tipo de relaciones, contribuyen de forma explícita al enriquecimiento de la ontología, logrando un sistema mucho más potente y eficaz. Un ejemplo de esta relación en la Figura 1.

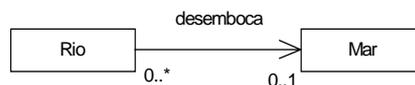


Fig.4. Ejemplo de relación ad hoc.

El último paso conducirá a la implementación de esta ontología en la IDEE. Contribuirá a gestionar esta información de manera más eficiente y autónoma, consiguiendo resolver las dificultades estructurales de la IG y armonizar un marco común con el que caminar hacia la interoperabilidad semántica.

5 Conclusiones y trabajo futuro

La consideración actual de interoperabilidad se circunscribe a la homogeneización de la sintaxis de comunicación e intercambio, obviando el contenido de la información (semántica). La implementación de ontologías en las IDEs actuales puede conducir a importantes mejoras en los procesos de búsqueda, acceso, procesamiento y explotación de la IG. Este hecho dará lugar a la apertura de nuevos horizontes y posibilidades en los usos y utilidades de los servicios IDE, derivando en un aumento de la confianza sobre la red, su utilidad y su uso. Por tanto, las ontologías se presentan como el instrumento más eficaz para alcanzar la interoperabilidad semántica que la IG necesita.

La siguiente fase del trabajo irá encaminada al enriquecimiento de esta ontología marco mediante la definición de constantes, axiomas formales y reglas. Además, se procederá al poblamiento de instancias obtenidas de diferentes fuentes. De esta manera, se pretende contribuir a la formación de una ontología más completa y potente. No obstante, se prevé la ampliación de la construcción de ontologías a otros dominios del catálogo de fenómenos del IGN.

6 Referencias

- [1] L. McKee, Who wants a GDI?, In Geospatial Data Infrastructure – Concepts, cases and good practice, New York, Oxford University Press, 2000, pag. 13-24.
- [2] D. Nerbert, Developing Spatial Data Infrastructures: The SDI Cookbook, Version 1.1, Global Spatial Data Infrastructure, Technical Committee. 2001
- [3] ISO/TC-211 & OGC, Geographic information Services Draft ISO/DIS 19119. OpenGis Service Architecture. vs.4.3. Draft Version, ISO & OGC. 2002.
- [4] Commission of the European Communities, Proposal for a Directive of the European Parliament and of the Council establishing an infrastructure for spatial information in the Community (INSPIRE), COM(2004) 516 final, 2004/0175 (COD), 2004.
- [5] L. Bernad, U. Einspanier, S. Haubrock, S. Hübner, W. Kuhn, R. Lessing, M. Lutz, U. Visser, Ontologies for intelligent search and semantic translation in Spatial Data Infrastructures, Photogrammetrie - Fernerkundung - Geoinformation (6). 2003.
- [6] T. R. Gruber, “A Translation Approach to Portable Ontology Specifications”. *Knowledge Acquisition*, 1993, 5(2), pag.199-220.
- [7] Y. Bishr, “Overcoming the semantic and other barriers to GIS interoperability” *Geographic Information Science*, 1998, 12(4), pag.199-314.

- [8] F. Harvey, W. Kuhn, H. Pundt, Y. Bishr, "Semantic interoperability: A central issue for sharing geographic information" *The Annals of Regional Science*, 1999, 33(2), pag. 213-232.
- [9] A. P. Sheth, "Changing focus on interoperability in information systems from system, syntax, structures to semantics". *Interoperating geographic information systems*. 1999. Boston, Kluwer Academic Publishers.
- [10] C. Batini, S. Ceri, B. Navathe, *Conceptual Database Design*. 1992. Redwood City, California, The Benjamin/Cummings publishing Company.
- [11] Neches *et al.*, "Enabling technology for knowledge sharing". *AI Magazine*, 1991, 12(3), pag.16-56.
- [12] M.A.Latre, F.J.Zarazaga-Soria, J.Nogueras-Iso, R. Béjar, P.R.Muro-Medrano. SDIGER: A cross-border inter-administration SDI to support WFD information access for Adour-Garonne and Ebro River Basins. *Proceedings of the 11th EC GI & GIS Workshop, ESDI Setting the Framework*. 2005. pag 5-7. Alghero, Sardinia (Italia).
- [13] Corcho O, Fernández-López M, Gómez-Pérez A, López-Cima. *A Building legal ontologies with METHONTOLOGY and WebODE. Law and the Semantic Web*. Springer-Verlag. 2005, Editor: Bran Selic. ISBN: 3540250638.