

Conflación Semántica: Análisis de fenómenos en el contexto de los WFS

J. A. Ramos, L. M. Vilches

Ontology Engineering Group – Facultad de Informática
Universidad Politécnica de Madrid
Avda. Montepríncipe sn, 28660 Boadilla del Monte (Madrid)
{jarg, lmvilches}@fi.upm.es

Resumen

Este artículo pretende hacer un análisis crítico de los servicios de fenómenos (WFS), en concreto a su parte relativa a los fenómenos que contemplan estos servicios. Así, se repasará el estándar OGC para estos servicios y se analizará la accesibilidad a esta información, la lengua en la que se encuentran los nombres de los fenómenos y la descriptividad de estos nombres. Todo ello de cara a conseguir unos servicios verdaderamente útiles, de los que se pueda sacar el mayor beneficio posible tanto por otras aplicaciones como por otros usuarios. Este análisis se aplicará a un conjunto de servicios de libre acceso disponibles en la Web y para cada caso se irán viendo las características y su adecuación al estándar o a las recomendaciones clave para conseguir la mayor usabilidad de cara a la explotación.

Palabras clave: fenómeno, conflación semántica, WFS, lexicón.

1 Introducción

Como se afirma en [1], España es un ejemplo en el campo de las IDE (Infraestructuras de Datos Espaciales) tanto en el ámbito europeo como globalmente. La abundancia de proyectos, servicios, geoportales y, en general, la existencia de recursos (conformes con las especificaciones del *Open Geospatial Consortium* (OGC)) públicos, libres y accesibles vía Internet implementados nos coloca en una situación privilegiada. Esta situación se ha producido por la confluencia de un conjunto de circunstancias y factores descritos en detalle en [1].

No obstante, la integración de la Información Geográfica (IG) proveniente de múltiples fuentes pone de manifiesto diversas problemáticas tanto en el contexto de la representación gráfica como en las clasificaciones de conocimiento asociadas. Estas problemáticas están vinculadas al término confluencia. Este concepto está adquiriendo una doble perspectiva; por un lado, la visión tradicional del término confluencia [2], vinculada a la representación gráfica de la información geo-espacial mediante información vectorial o ráster. Por tanto, el núcleo del proceso de confluencia se centra en identificar y asociar los elementos geométricos comunes entre las múltiples fuentes, reconciliar las diferencias entre ellos y construir un resultado integrado. Por otro lado, aparece un nuevo punto de vista que relaciona la perspectiva semántica con el concepto confluencia -vinculada con las clasificaciones de conocimiento-. Este hecho es resultado de la alta variabilidad existente en los tipos de fenómenos presentes en la descripción de los distintos servicios que soportan IG en el contexto de las IDE. Esta variabilidad repercute en diferencias y variaciones notables a nivel de conceptual, como manifiesta el uso de diversas terminologías para representar una misma información. Este hecho impide que el proceso de búsqueda de servicios sea rápido y eficaz, y consecuentemente, la superposición y mezcla de datos de diferentes fuentes en un entorno interactivo. Desde la perspectiva de la confluencia semántica se presenta la posibilidad de integrar información acudiendo a la descripción de los fenómenos presentes en los servicios *Web Feature Services* (WFS).

En este trabajo nos centraremos en la perspectiva semántica del término confluencia con objeto de analizar la descripción de los fenómenos geográficos (punto de partida para el modelado de la información geo-espacial según OGC [3]), presentes en diferentes WFS estudiados.

2 Web Feature Services

El *Web Feature Service* (WFS) [4] del OGC es un servicio estándar que ofrece un interfaz de comunicación que permite interactuar con los mapas servidos por el estándar *Web Map Service* (WMS). Para la realización de las diferentes operaciones permitidas por este servicio se utiliza el lenguaje *Geographic Markup Language* (GML) [5].

Dentro de la especificación WFS las principales operaciones vinculadas con los procesos de descripción, recuperación y visualización de los fenómenos geográficos son: *GetCapabilities* (indica qué tipos de fenómenos pueden servirse y

qué operaciones son soportadas sobre cada tipo fenómeno), **DescribeFeatureType** (describe la estructura de cualquier tipo de fenómeno descrito en el servicio) y **GetFeature** (recupera las instancias pertenecientes a los fenómenos).

Estas operaciones están conformadas por diversos elementos encargados de definir las características de los servicios mediante diversos parámetros. En este trabajo se pone especial atención en aquellos elementos cuyo objetivo es describir los tipos de fenómenos geográficos. Así, dentro de la operación *GetCapabilities* se define el elemento **FeatureType** (recoge una lista de los tipos de fenómenos y operaciones de cada tipo de fenómeno disponibles en un determinado servicio) y **FeatureTypeList** (su propósito es contener una lista de tipos de fenómenos derivada de gml:AbstractFeatureType. Dentro del FeatureTypeList pueden ser utilizados elementos como “nombre”, “título”, “resumen”, “palabras clave”, etc. para describir cada tipo de fenómeno). Por otro lado, dentro de la operación *DescribeFeatureType* aparece definido el elemento **TypeName**, encargado de codificar los nombres de los tipos de fenómenos descritos.

Considerando estas operaciones y elementos, a partir de una URL (por ejemplo, <http://ideg.xunta.es/wfs/request.aspx>) donde reside un servidor web, se puede invocar un servicio web añadiendo a la URL una serie de parámetros sobre el servicio que se invoca, la funcionalidad del servicio que se invoca y los valores de los parámetros de entrada de esa funcionalidad. Un ejemplo completo podría ser la invocación del WFS de CORINE, de la operación DescribeFeatureType y con valores para los parámetros de entrada NAMESPACE y TypeName:

```
http://www.idee.es/wfs/IGN-Corine/wfs?  
SERVICE=WFS&VERSION=1.1.0&REQUEST= DescribeFeatureType  
&NAMESPACE=xmlns(clc=http://www.idee.es/corine_land_cover)  
&TypeName=clc:Change_90_00_Area
```

Tras el análisis anterior, se ve que es invocando la petición GetCapabilities de un WFS se obtiene un XML como resultado en el que aparece, entre otras cosas, la lista de tipos de fenómenos (etiqueta <FeatureTypeList>). Esta lista está compuesta por tipos de fenómenos (etiqueta <FeatureType>) y dentro de cada uno de ellos aparece un nombre (etiqueta <Name>). Los WFS que siguen la estandarización propuesta por Open Geospatial Consortium tienen en las etiquetas más información, como <wfs:FeatureTypeList>, pero no todos los servicios WFS analizados en este trabajo siguen esta norma.

3 Contexto. WFS a considerar

Como se describe en [1] desde su aparición hace unos seis años, las primeras implementaciones de IDE en España, como la IDE de Cataluña (IDEC) y poco después la IDE de España (IDEE), han proliferado las implementaciones tanto de recursos IDE, es decir, geoportales, servicios, documentación, utilidades, aplicaciones, etc., como de auténticas IDE, si se considera una verdadera IDE como un SIG distribuido completo en sí mismo. Este hecho ha derivado en una renovación y revolución del sector español de la producción y gestión, entendidas ambas en un sentido amplio, de la IG y se ha realizado un proceso de reingeniería de los procesos de producción muy notable y en un tiempo record.

Desde la perspectiva de los servicios OGC, en una visión global puede apreciarse la proliferación de los servicios WMS. Esta proliferación es patente si se recurre a la búsqueda en catálogos de este servicio. A nivel global esto se puede comprobar en páginas como WMS-Sites¹ o Skylab², mientras, a nivel español un buen referente de esta información es el directorio de servicios de la IDEE, dónde aparecen más de 150 servicios de este tipo. Por el contrario, no resulta tan prolífico el número de servicios conformes a la especificación WFS. En el contexto global algunos de estos servicios pueden ser encontrados en Fmepedia³ o GeoServer⁴, mientras en el ámbito español el listado de servicios WFS presentes en el directorio de la IDEE no superan la veintena.

En este trabajo se ha intentado cubrir la mayor parte de los WFS existentes en la península. La restricción geográfica al territorio nacional peninsular está respaldada por dos motivos: similares características geográficas y similares características sociopolíticas. Aun conociendo la diversidad geográfica peninsular, la comparación de conjuntos de fenómenos y sus especificaciones de un territorio más extenso conllevaría mucha más diversidad de fenómenos y además supondría fronteras idiomáticas (que aun así ya las hay como se verá más adelante) y políticas (en cuando al apoyo o promoción de la gestión y explotación de información geográfica).

¹ <http://www.wms-sites.com/catalog>

² http://www.skylab-mobilesystems.com/en/wms_serverlist.html

³ [http://www.fmepedia.com/index.php/WFS_\(Web_Feature_Service\)_Servers](http://www.fmepedia.com/index.php/WFS_(Web_Feature_Service)_Servers)

⁴ <http://geoserver.org/display/GEOS/Available+WMS+and+WFS+servers>

Los servicios analizados en este trabajo son los siguientes:

- CORINE

<http://www.ideo.es/wfs/IGN-CORINE/wfs>

- Cartociudad

<http://www.cartociudad.es/wfs-vial/services>

- Xunta de Galicia

<http://ideg.xunta.es/wfs/request.aspx>

- Deputación da Coruña

<http://www.dicoruna.es/webeiel/ideac/wfs>

- IDEE

<http://www.ideo.es/IDEE-WFS/ogcwebservice>

- Catastro

<http://ovc.catastro.meh.es/Cartografia/WFS/ServidorWFS.aspx>

- Conf. Hidrográfica del Ebro.

<http://80.255.113.15/DMA-OWS/ogcwebservice>

- Nomenclátor Geográfico Conciso de España

<http://www.ideo.es/IDEE-WF-Nomenclator-NGC/services>

- Nomenclátor NOMGEO

<http://www.ideo.es/IDEE-WFS-Nomenclator-NG/services>

- Confederación Hidrográfica del Duero

<http://www.imrame.chduero.es/duerowfd>

- Concejos de Asturias

<http://www.cartografia.asturias.es/wfsconcejos/request.asp>

- Coruña servicio de nomenclátor

<http://www.dicoruna.es/webeiel/ideac/snm>

- Atlas Virtual de Aves Terrestres de España

<http://161.111.161.171/cgi-bin/AtlasAves.exe>

- Instituto de Desarrollo Económico del Principado de Asturias

<http://www2.idepa.es/wfspolygonos/request.asp>

Para el caso de Cartociudad se pueden encontrar seis servicios diferentes: código, vial, portal, código, sección, distrito y domicilio. En este trabajo se va a trabajar con el WFS de vial, puesto que los seis tienen idénticas características y los fenómenos contemplados son los mismos.

4 Proceso de análisis

El primer parámetro a analizar es la accesibilidad a los datos de los WFS. Para una mayor calidad del servicio, se ha de seguir el estándar y que el acceso a los datos sea uniforme. Así, se ha desarrollado una aplicación Java que crea un lexicón con los fenómenos contemplados en el WFS a partir de la URL del mismo.

A partir de los lexicones con los fenómenos contemplados en cada WFS, se va a realizar un estudio manual de cada uno para más tarde formular unas directrices generales a seguir para evitar los errores que actualmente se detecten.

Dentro del análisis manual de los lexicones de fenómenos se mirará la cantidad de fenómenos, la lengua en que están, la descriptividad de los nombres, la aparición de partículas léxicas (afijos) no necesarias, etc.

5 featureLexicon

Se ha desarrollado una sencilla aplicación Java, llamada *featureLexicon*, que obtiene los nombres de los fenómenos (*FeatureType*) de un WFS y genera un fichero de texto plano con todos los nombres, que se corresponde con un lexicón. Esta aplicación tiene como entradas la URL donde se encuentra el servidor y el fichero de salida.

6 Análisis de resultados

De manera general, los ficheros y el número de términos obtenidos como nombre de los fenómenos gracias a la ejecución de la aplicación *featureLexicon* son los siguientes: lexiconcorine(5), lexiconcartovial (3), lexicongalicia (22), lexiconcoruna (395), lexiconavate (3), lexiconchd (48), lexiconconchedma (19), lexiconidee (21), lexiconidepa (17), lexiconnngc (2), lexiconnomgeo (2).

6.1 Accesibilidad

La aplicación *featureLexicon* ha generado lexicones vacíos para tres de los servicios: Catastro, Concejos de Asturias y nomenclátor de Coruña. Para el caso de los demás servicios ha creado ficheros con contenido.

De aquí ya se puede tener una primera medida sobre la accesibilidad de ciertos servicios. La invocación de estos servicios o la salida que se obtiene de las consultas no sigue las reglas generales y por tanto un software que los trate, como es el caso de *featureLexicon*, tendrá dificultades y el acceso a estos recursos tendrá que ser tratado de manera especial. La otra opción sobre la causa que puede generar un lexicón de fenómenos vacío es que, en verdad, el servicio siga los formatos estándares pero no tenga ningún fenómeno o no tenga ningún *FeatureType* con los campos de nombre rellenos.

En cualquiera de los casos, hay una deficiencia en estos tres servicios, a nivel de estandarización o a nivel de completitud de los campos del estándar; lo que supone que un problema de accesibilidad.

6.2 Lengua

Los lexicones obtenidos están en tres lenguas diferentes:

- Castellano: cartovial, coruna, chd, idee, depa, ngc y nomgeo.
- Inglés: corine, chedma.
- Gallego: galicia.

En esta lista faltaría el lexicón avate, que por los términos (3) con que cuenta el lexicón, no se puede determinar el lenguaje en que están. Para ser exactos, los términos de este lexicón son: *Dlyr*, *Blyr* y *cuad50*.

6.3 Descriptividad

La descriptividad de los nombres se va a analizar para este trabajo de forma independiente a la cantidad de fenómenos contemplados por el WFS. Por ejemplo, el fenómeno “río” es descriptivo en un lexicón tanto de 20 fenómenos como en uno de 200 fenómenos. También lo sería “río 1ª categoría” en ambos casos. Sin embargo, el fenómeno “entidad” no es descriptivo como fenómeno, puesto que no delimita los topónimos que bajo ese fenómeno se clasifican.

Otro aspecto a tener en cuenta en la descriptividad de los fenómenos es la aparición de afijos no descriptivos de los fenómenos. Estos afijos pueden contener información sobre las características gráficas de la representación asociada al fenómeno, pero no sobre el fenómeno en sí. Por ejemplo, contienen afijos no descriptivos fenómenos como “río_contorno”, “río_superficie” y “río_línea”. En estos casos, los afijos sobre características gráficas suponen la aparición de tres fenómenos en el lexicón cuando tan sólo es un fenómeno geográfico con tres posibles representaciones.

Con estas dos consideraciones previas, se va a analizar a continuación la descriptividad de los nombres de los lexicones no vacíos obtenidos de los WFS mencionados en la sección 3.

- **CORINE.** Los nombres no son significativos y además contienen afijos y partículas seguramente con semántica sobre el origen de los datos. Ejemplos: *Class90*, *Class00*, *CORINE90_Area*.
- **Cartociudad-vial.** Los tres nombres de fenómenos no son descriptivos ni significativos. Ejemplos: *Entidad*, *entidadlocal_*.
- **Galicia.** Los nombres son descriptivos, pero contienen afijos relativos a la escala de las entidades reflejadas. Ejemplos: *CONCELLO_25*, *CONCELLO_1000*, *COMARCA_25*, *PRESA_200*.
- **Coruña.** Los nombres son descriptivos, pero aparecen muchas repeticiones debido a que aparecen en multitud de ocasiones afijos sobre la representación gráfica asociada, como “_superficie”, “_geometria”, “_linea_co” o “vista_”. Ejemplos: *vista_depuradoras_geometria*, *vista_carreteras_geometria_linea_co*.
- **Avate.** Los nombres no son descriptivos. En este caso, los nombres son incluso crípticos, ya que no tienen significado conocido por la comunidad. Fenómenos: *Dlyr*, *Blyr* y *cuad50*.
- **CHD.** Los nombres son descriptivos. Tan sólo en unos pocos casos aparecen nombres con afijos sobre la situación subterránea o superficial del fenómeno. Ejemplos: *Abastecimientos_Superficiales*, *Abastecimientos_Subterráneos*, *Piscifactorias*, *Presas*, *Municipios*, *Campos_Golf*.
- **CHEDMA.** Los nombres son descriptivos y no tienen afijos. Ejemplos: *LakeWaterBody*, *Fish*, *Nitrates*, *ProtectedArea*, *CoastalwaterBody*.
- **IDEE.** Aunque los nombres son descriptivos, todos contienen afijos con otra información. Ejemplos: *BDLL25Provincia*, *BDLL200Provincia*, *EGMConstruccionArea*, *EGMViaComunicacionLinea*, *VerticeRedRegente*, *VerticeRedOrdenInferior*.
- **IDEPA.** Los nombres son descriptivos. Tan sólo aparecen 4 de los 17 fenómenos con un afijo relativo a la condición gráfica (vista). Ejemplos: *SEGURIDAD*, *PARKING*, *ESTACION_SERVICIO*, *VistaGParcelasWMS*, *VistaGFasesWMS*.
- **NGC.** Los nombres no contienen afijos y tampoco son descriptivos. Los dos fenómenos son: *Entidad*, *TipoEntidad*.
- **NOMGEO.** Los nombres no contiene afijos y tampoco son descriptivos. Los dos fenómenos son: *Entidad*, *TipoEntidad*.

Como se puede ver en el análisis anterior, la calidad de los fenómenos contemplados por los servicios no depende de la cantidad de fenómenos, ya que

servicios con muchos fenómenos contienen multitud de reiteraciones de fenómenos en nombres con variaciones en los afijos geométricos.

De los servicios analizados, es el servicio asturiano IDEPA y el de la Confederación Hidrográfica del Duero los que tienen una mejor calidad en cuanto a los fenómenos contemplados. En el extremo opuesto, se sitúan los servicios de los nomenclátors (NG, NOMGEO y Cartociudad) que contienen un número mínimo de fenómenos y no descriptivos, y el servicio del Atlas Virtual de Aves (AVATE) que comparte errores e incluso tampoco tienen una lengua reconocible.

7 Conclusiones y líneas futuras

De este trabajo se pueden sacar como conclusiones generales que hay cierta dejadez por parte de los proveedores de servicios a la hora de atender este aspecto de los fenómenos dentro de los WFS, ya que se han centrado más en hacer accesibles los datos que en que los datos sean accesibles de una manera óptima y sencilla. Hay que destacar que una buena parte de los servicios analizados tiene nombres de fenómenos descriptivos pero desvirtuados por afijos con información no relativa a los fenómenos en sí, sino a su representación u otros.

El siguiente paso de este análisis será, basado en este análisis, establecer unas recomendaciones sobre el estándar para que, a la hora de la construcción y puesta en funcionamiento de servicios web de fenómenos, éstos sean lo más fácilmente explotables y el trabajo invertido sea verdaderamente rentable debido a la utilidad y disponibilidad que a las fuentes de información geográfica se les está otorgando.

Agradecimientos. Este trabajo ha sido parcialmente subvencionado por Centro Nacional de Información Geográfica (CNIG) en el contexto del proyecto CENIT “España Virtual”, financiado por el CDTI bajo el programa INGENIO 2010.

Bibliografía

- [1] González Gonzalez, C., Rodríguez Pascual, A., Mas Mayoral, S., Abad Power, P., Alonso Jimenez, J.A., Sánchez Maganto, A., Soteres Dominguez, C., Potti Manjavacas, H. (2009) Neocartography: use of protocols and standard interfaces in the construction of the geographic mashups. In GEOMATICA’09. La Habana (Cuba).

- [2] OGC 2008, 'OWS-5 Conflation Engineering Report. Discussion Paper', Pete Brennan (ed.). Version: 0.1.0. Open Geospatial Consortium
- [3] Buehler, K. (2003) OpenGIS Reference Model. Open Geospatial Consortium (OGC).
- [4] Open Geospatial Consortium – OGC (2005) Web Feature Service Implementation Specification Version: 1.1.0
- [5] Cox, S., Daisay, P., Lake, R., Portele, C., & Whiteside, A. (Eds.) (2002) OpenGIS Geography Markup Language (GML) Version 3.0. Open Geospatial Consortium. Document available at <http://www.opengeospatial.org/standards/gml>