

Diseño de una ontología de dominio para la interoperabilidad de sistemas de información en el ámbito del medio natural: “Aplicación al proyecto HITA”

E. Saiz Riquelme¹ y M. Vicente Albaladejo²

¹Planisferio Digital S.L.
eduardo.saiz@planisferiodigital.com

²SIGA Servicio de Información e Integración Ambiental
Consejería de Agricultura y Agua de la Región de Murcia
mariano.vicente@carm.es

Resumen

Del análisis histórico de la interoperabilidad de la información geográfica se desprende el alto grado de profundización y estandarización en los aspectos sistémicos. No obstante, la multiplicación de las fuentes de información geográfica, plantea además la necesidad de resolver la interoperabilidad semántica; cuestión que puede pasar por la definición de ontologías de dominio.

En este trabajo presentamos la arquitectura de un sistema que, haciendo uso de una ontología de dominio, consigue que los distintos subsistemas que lo componen logren interoperar entre ellos.

Palabras clave: alfresco, ontología, medio natural, proyecto HITA

1. Introducción

De las diversas facetas que reviste el trabajo con tecnologías de información geográfica y explotación y difusión de datos espaciales, el de la puesta en marcha de servicios para la publicación de los mismos (geoportales, visores, servicios OGC,...), es la que mayor auge ha experimentado en los últimos años.

Sin embargo, en el ámbito de los organismos que dedican estas tecnologías a una lógica distinta, los avances en la interoperabilidad deben implicar también a otros subsistemas de información involucrados en la gestión del negocio.

Con ese objetivo, lograr esta interoperabilidad, se presenta en este trabajo el diseño de una arquitectura de sistema que, apoyada en una ontología de dominio, pretende integrar la gestión de información geográfica con las herramientas administrativas y documentales utilizadas en el seno de la Dirección General del Patrimonio Natural y Biodiversidad (en adelante, DGPNB).

Georreferenciar automáticamente contenidos administrativos y documentales y establecer funcionalidades de consulta híbrida (basadas en texto y localización), son algunos de los retos a los cuales se pretende encontrar soluciones adecuadas y efectivas.

Se presenta pues un avance de los trabajos desarrollados con este planteamiento en torno al Proyecto HITA, en el que sobre la “figura de gestión” de los Montes de Utilidad Pública se desempeñan numerosas labores interdisciplinarias, tales como los levantamientos topográficos,

elaboración de una base de datos espacial, documentación fotográfica *in situ*, digitalización y organización de archivos, desarrollo de aplicaciones de gestión, etc.

2. Trabajo relacionado

En cuanto a la georreferenciación de contenidos documentales, en [1] se presenta una arquitectura de sistema que permite acceder a los contenidos de un Sistema de Gestión Documental (SGD) tanto a través del ámbito textual como geográfico de los mismos. Este sistema incorpora un módulo de georreferenciación automático, que se basa en un servicio de *gazetteer* y en un servicio proveedor de geometrías, basándose éste último en una ontología espacial.

Se menciona también la problemática que puede suponer un mecanismo de georreferenciación de contenidos de forma automática. Por último, la explotación del sistema se realiza a través de un visor cartográfico basado en un servidor de mapas.

En [2], los mismos autores rediseñan el sistema dotándolo de más flexibilidad, para lo cual definen un flujo de trabajo para la gestión dinámica de contenidos. El objetivo es no limitar el sistema a colecciones de documentos existentes, sino que éste se pueda adaptar a organismos públicos que manejan grandes cantidades de información.

3. Arquitectura del sistema

La arquitectura del sistema desarrollado atiende a una estructura modular basada en la naturaleza de la información gestionada por cada parte del mismo.

Los tipos de información manejados son los siguientes:

- *Información de negocio*: consistente en los atributos (datos alfanuméricos) de los objetos de negocio. Este tipo de información es gestionada por los aplicativos de negocio, que implementan la lógica del mismo. En este caso, están basados en el Sistema Gestor de Bases de Datos Oracle, y en la tecnología Java EE.
- *Información geográfica*: consistente en la geometría de los objetos de negocio. Este tipo de información es gestionada por bases de datos geográficas (GeoBD). La DGPNB cuenta actualmente con el gestor SDE 9.2 sobre Oracle.
- *Información documental*: consistente en los contenidos documentales (carpetas y ficheros) relacionados con los objetos de negocio o que son resultado de su gestión, tales como los expedientes administrativos, la documentación técnica, las colecciones temáticas digitalizadas, etc. Este tipo de información es gestionado por el SGD Alfresco [3], disponiendo actualmente en producción de la versión 2.1 Community.

El objetivo es conseguir una sinergia (basada en la interoperabilidad) entre las fuentes de información de la organización, de forma que el sistema ofrezca un acceso unificado al conjunto de ésta.

Para que los tres subsistemas tengan la misma visión del conocimiento de la organización, y por tanto, puedan interoperar entre ellos, el desarrollo se centra en el diseño de una ontología de dominio que será implementada en cada uno de ellos (como se explica en el siguiente apartado, la implementación consiste en replicar el esquema conceptual de la ontología).

Realmente el sistema no define una verdadera arquitectura, formada por distintos componentes, sino que se basa en la interconexión de los ya existentes, objetivo que se logra con la citada ontología.

Otro aspecto importante del sistema es la inexistencia de información duplicada; esto es, cada subsistema sólo almacenará la información de su propio tipo. A veces ocurre que la información se encuentra duplicada o incluso dispersa por varios sitios, lo que puede suponer un problema de integridad de la misma.

4. Ontología

4.1. Introducción

Las ontologías son un intento de definir el mundo de tal forma que todos los subsistemas que trabajan sobre él tengan una visión común del mismo.

Dado que la razón de ser de una ontología es la de poder comunicarse, se hace necesario emplear lenguajes que permitan representar el conocimiento bajo una base de definición común.

No obstante, en el presente caso, el interés por el uso de ontologías radica exclusivamente en el aspecto conceptual de las mismas y en los beneficios que se obtienen de su uso; es decir, el objetivo no pasa por una comunicación explícita de conocimiento entre los distintos subsistemas a través de un lenguaje de definición de ontología, sino que lo que se persigue es que éstos implementen dicha ontología (el esquema conceptual de la misma), y así puedan interoperar entre ellos.

Esa interoperabilidad se basa en la aplicación del mismo esquema conceptual (adaptado a la lógica de cada subsistema de información: tablas de GeoBD, tipos de contenidos del SGD...) y en el establecimiento de un sistema de *identificadores únicos de instancia* para las diferentes clases de la ontología: cada identificador (código) lleva implícita la información de clase e instancia.

4.2. Diseño de la ontología

Habitualmente ocurre que un mismo objeto de la realidad posee distintas interpretaciones dependiendo de la semántica o de los intereses del observador. En [4] se discute este tema, así como el concepto de rol, que se puede definir como un punto de vista concreto para observar un mismo fenómeno.

Guiados por esta idea, en el diseño de esta ontología identificamos las dos siguientes clases:

- **Objeto de referencia:** se caracteriza por tener existencia propia independientemente de la actuación de nuestra organización. No requiere interpretación. Ejemplos de objetos de referencia son las especies de flora y fauna, lugares o hábitats.
- **Objeto de gestión:** a diferencia del anterior, es el resultado de la acción y gestión de la DGPNB. Sería el equivalente al rol que un determinado objeto de referencia puede desempeñar en un determinado momento. Ejemplos de objetos de gestión son los montes, las vías pecuarias, espacios naturales protegidos, las especies protegidas, etc.

A modo de ejemplo, el conjunto montañoso conocido como “Sierra Espuña” se define en el modelo como un objeto de referencia (perteneciente a la clase *lugar*), pero cada conjunto de funciones que sobre él se desempeñan, determinadas por diferentes instrumentos de regulación, conlleva la definición en el modelo de distintos objetos de gestión (Monte del Catálogo de Utilidad Pública, Reserva Nacional de Caza, Lugar de Importancia Comunitaria, Parque Regional, Zona de Especial Protección para las Aves...).

En ambos casos se puede distinguir entre objetos referidos a un territorio determinado o no, situación que originará la aparición de nuevas subclases determinadas por esta cuestión. Un *objeto de gestión* cuya definición conlleva la determinación de un área geográfica concreta da

lugar a la subclase *objeto de gestión territorial*, en tanto que aquellos que son independientes de un lugar específico serán los *objetos de gestión no territorial*. El mismo razonamiento es válido para la clase *objeto de referencia*.

A continuación se muestran dos diagramas de clases¹ con la representación tanto de los objetos de referencia como de los de gestión:

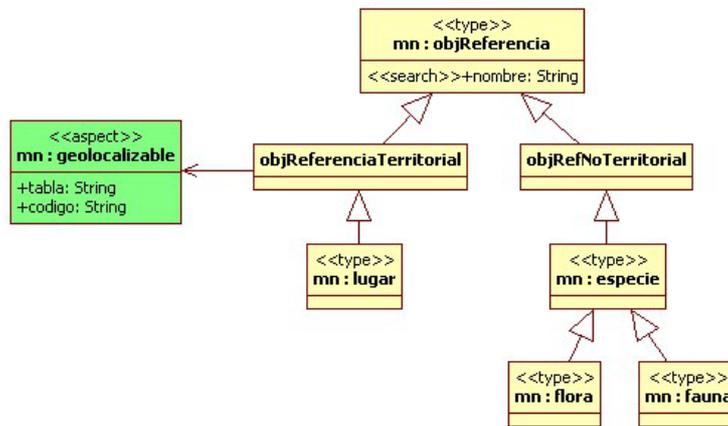


Figura 1. Objetos de referencia

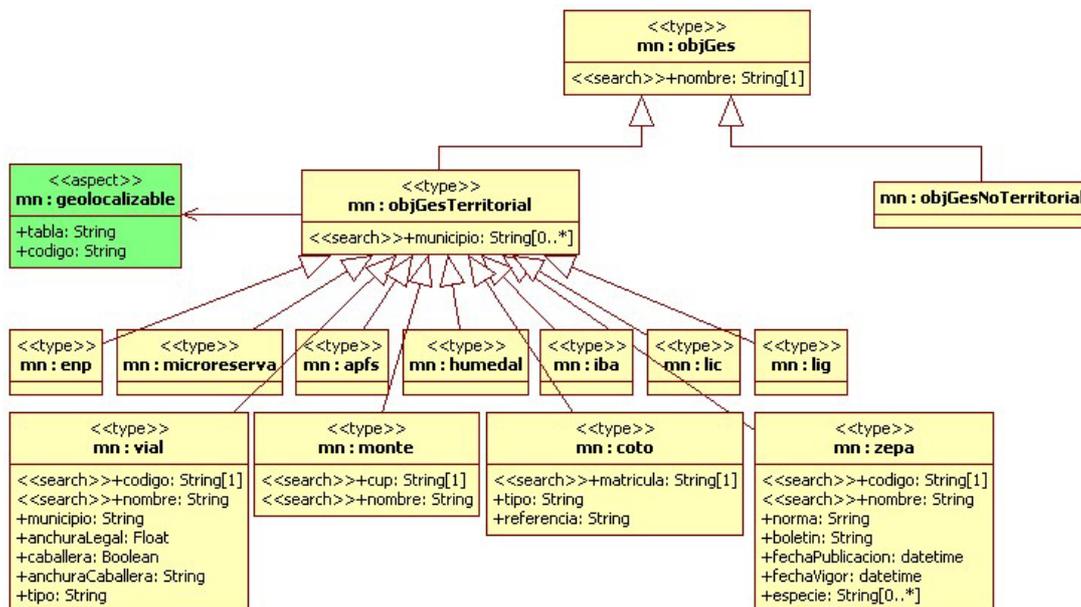


Figura 2. Objetos de gestión

Para el caso de los *objetos de gestión territorial*, que tienen especial importancia en la gestión interna de la DGPNB, la ontología propuesta se extiende incluyendo nuevas clases con el objetivo de representar el conocimiento relativo a los límites de los mismos.

¹ Los diagramas están diseñados en lenguaje UML.

Así, todo *objeto de gestión territorial* tendrá un límite de gestión, que no es más que la geometría que delimita su ámbito espacial de aplicación.

En la siguiente figura se muestra el diagrama de clases con la representación de los límites de gestión:

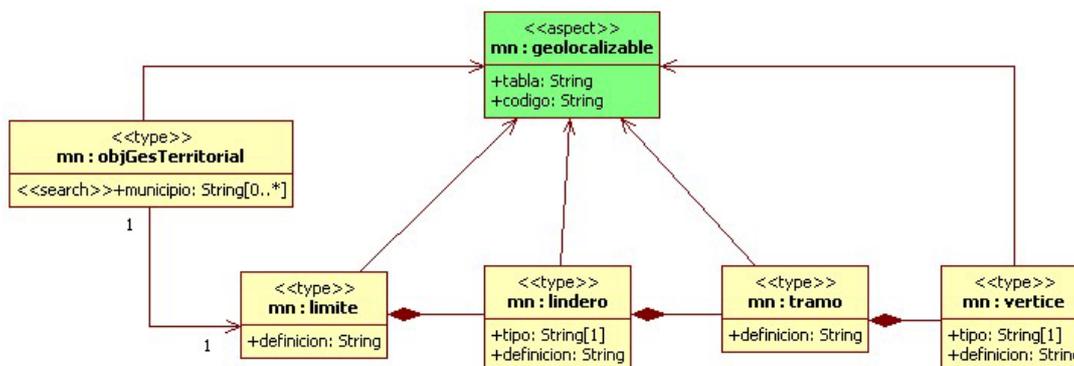


Figura 3. Límites de objetos de gestión

4.3. Implementación de la ontología en el gestor documental

Normalmente los SGD, combinados con los Sistemas de Recuperación de la Información (SRI), se engloban en los llamados Servicios Basados en el Conocimiento y en la Información No Estructurada (SBC-INE).

No obstante, mediante el diseño de un modelo de contenidos junto con una estructura de carpetas adecuada, se puede lograr que el SGD almacene información que atienda a la lógica y estructura del negocio.

La ontología mostrada puede ser implementada fácilmente en el SGD de la DGPNB. En el caso particular de Alfresco, éste permite la creación de tipos y aspectos. Un tipo es la explicitación de una entidad o concepto que aparece en el análisis de nuestro universo en discurso. Un aspecto es una cualidad de un tipo, que tiene un carácter transversal, a diferencia de los tipos, que siguen una estructura jerárquica.

En los diagramas mostrados anteriormente se han utilizado dos estereotipos para indicar si una clase será implementada mediante un tipo o un aspecto en Alfresco. Así, se puede observar la existencia del aspecto *geolocalizable*, que se emplea para dotar a un contenido (instancia de la ontología) de información geográfica como cualidad intrínseca. Como atributos de este aspecto se encuentran la tabla y el código del registro que se emplea para relacionar la instancia de Alfresco con su correspondiente en la GeoBD.

La existencia del aspecto *geolocalizable* se explica al observar que en el repositorio documental habrá contenidos que no representen instancias de la ontología, y por tanto, no tendrán asociado ninguna información geográfica (su equivalente en la GeoBD).

5. Funcionalidad

5.1. Mecanismos de explotación

El sistema no define ningún mecanismo de explotación específico, sino que trabaja con los propios de cada tecnología utilizada en el sistema y lo que ésta permita. Con ello entendemos que se maximiza la reutilización de funcionalidades ya desarrolladas de forma nativa en las respectivas herramientas.

En la actualidad, se ha implementado un servicio Web de generación de ficheros KML que permite integrar los tres tipos de información en un único punto de acceso. Este servicio Web ha sido implementado utilizando una funcionalidad propia de Alfresco llamada Web Scripts [5]. Ésta, que se basa en la técnica REST, permite obtener el estado de uno o varios contenidos del repositorio de Alfresco.

Para cada clase de la ontología existe una plantilla FreeMarker [6] (motor de plantillas utilizado por Alfresco) que encapsula la lógica de negocio de la misma. Así, por ejemplo, en la plantilla se especifica el subsistema de donde obtener la información de negocio (no siempre se dispone de un aplicativo de gestión, por lo que esta información se almacena bien en la GeoBD, bien en el SGD), así como los atributos de interés para el usuario en cada caso.

Así, el funcionamiento del servicio diseñado es como sigue:

- Para uno o varios contenidos del repositorio (que son instancias de clases de nuestra ontología) se desea obtener un fichero KML que aglutine los distintos tipos de información que se dispone de ellos.
- Para cada uno de esos contenidos, el servicio genera un objeto del tipo Placemark [7] (propio de KML) cuya descripción (información asociada a éste y que el usuario podrá visualizar en cualquier aplicación cliente que soporte el estándar KML) es el resultado de ejecutar la correspondiente plantilla FreeMarker asociada a la clase del contenido.
- La geometría del objeto Placemark es recuperada de la GeoBD. Para ello se utilizan los valores del aspecto geolocalizable que tiene asociado cada una de las instancias de la ontología.

Para seleccionar el conjunto de contenidos sobre los que generar el KML se ha utilizado la sintaxis de consulta propia de Lucene [8], que es el sistema de recuperación de información utilizado por Alfresco, y que de alguna forma se puede ver como una especie de lenguaje de consulta sobre la ontología.

A continuación, se detalla un ejemplo de invocación al servicio para generar el KML del monte con código M0146 y de sus contenidos relacionados que se encuentren dentro del espacio de trabajo LIMITES/LINDEROS (que es un subespacio de trabajo del propio monte).

```
http://localhost:8080/alfresco/wcs/geolocalizable/kml.kml?create&nodes=PATH:"/app:company_home/cm:OBJETOS_DE_GESTION/cm:MONTES/cm:M0146/.";monte;false|PATH:"/app:company_home/cm:OBJETOS_DE_GESTION/cm:MONTES/cm:M0146/cm:LIMITES/cm:LINDEROS//*";mojon;false&description=Monte%20M0146
```

Como se puede apreciar, se puede incluir en la llamada al servicio algunos parámetros adicionales, como puede ser el estilo KML que se aplicará al Placemark, así como un booleano para indicar si se desea mostrar el centroide de un polígono en lugar de éste.

En la siguiente captura de pantalla se puede observar el fichero KML resultante de la invocación anterior:

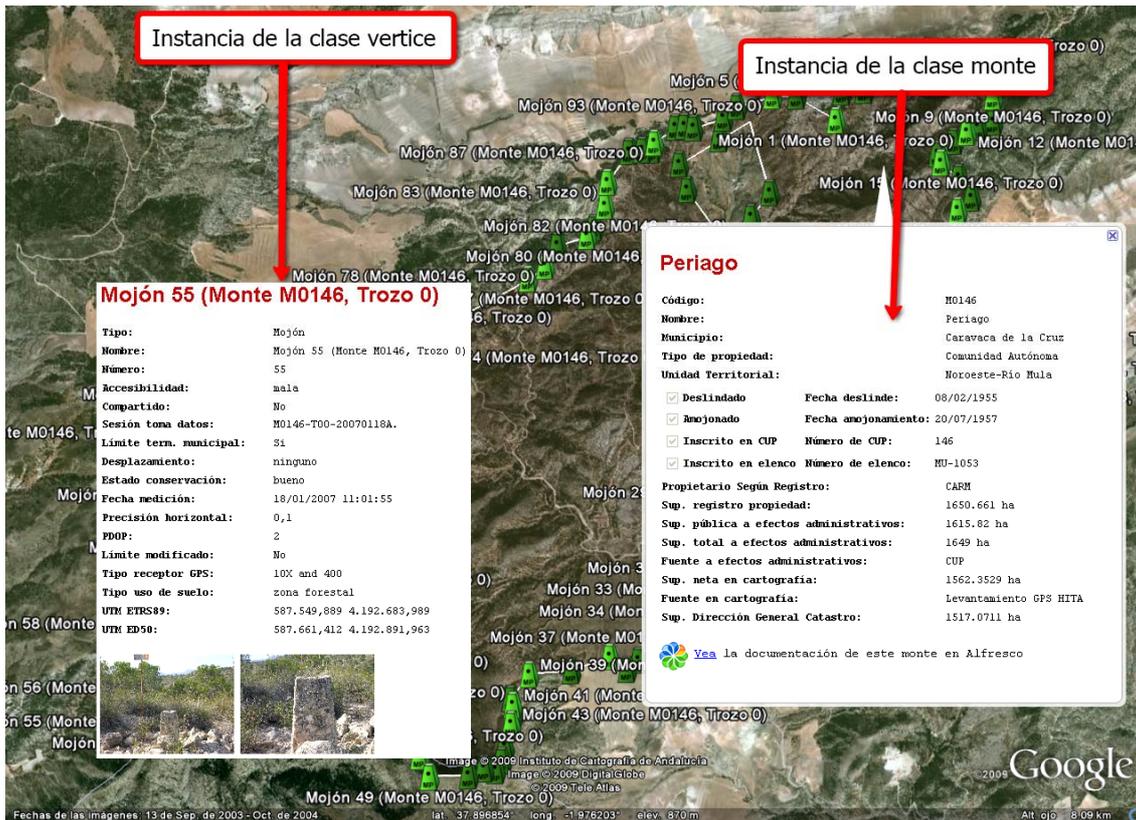


Figura 4. KML visualizado en Google Earth

El KML generado incluye instancias de dos clases distintas, *monte* y *vertice*, cada una de las cuales tiene su propia lógica de negocio. En el caso de los vértices, la información de negocio se obtiene de la GeoBD, mientras que en el caso de los montes éste se recupera de un aplicativo de gestión ad hoc.

Se detalla por último un pequeño fragmento de código de la plantilla FreeMarker asociada a la clase *monte*:

```
<#ftl encoding="ISO-8859-1">
<#assign data=montesDB.getData("select MO.* from monte MO where
MO.id_monte = '${CODIGO}')")

<table width="500px">
<#if data["ID_MONTE"]??>
  <tr>
    <td width="200px"><strong>Código: </strong></td>
    <td>${data["ID_MONTE"]}</td>
  </tr>
</#if>
<#if data["NOMBRE"]??>
  <tr>
    <td><strong>Nombre: </strong></td>
    <td>${data["NOMBRE"]}</td>
  </tr>
</#if>
```

El objeto *montesDB* permite acceder al subsistema que almacena la información de negocio para el caso de la clase *monte*.

5.2. Georreferenciación automática de contenidos

La georreferenciación de un contenido se puede realizar de forma explícita, asignándole la información geográfica manualmente, o de forma automática, en el que se utilizan técnicas de identificación de información geográfica, como pueden ser los servicios de gazetteer, o los servicios proveedores de geometría [1].

El primer caso tiene como principal inconveniente la eficiencia del proceso en sí, que dependiendo del volumen de contenidos a tratar puede llevarnos a desestimar el mismo.

El segundo caso, por contra, adolece de cierta inexactitud, en tanto que la información geográfica extraída no siempre es la correcta, o incluso la deseable o necesaria.

En tanto que el SGD de la DGPNB se ajusta a la ontología diseñada (mediante el diseño de un modelo de contenidos y de una estructura de carpetas específica), el proceso de georreferenciación se reduce a ubicar los contenidos en el lugar adecuado, esto es, la carpeta correspondiente al objeto al que está asociado, lo que constituye un método ágil y preciso.

6. Conclusiones y trabajo futuro

Esta estrategia para integrar distintos sistemas viene originada en la necesidad de adaptar sistemas heterogéneos cuyos componentes evolucionan de forma separada e independiente, y que a veces incluso son gestionados por departamentos igualmente distintos. La imposición de un marco de trabajo único y perdurable para todo el conjunto de herramientas implicadas, es en la práctica imposible. Definir una ontología resulta una opción transversal, y por tanto más robusta en un contexto administrativo cambiante, como el que sufre la administración ambiental en la Región de Murcia.

La ontología mostrada se completa con otras clases tales como tipos de expedientes administrativos o documentos de diversa índole, que se definen en el modelo de contenidos del SGD, pero que no tienen su equivalente en la GeoBD. Normalmente, estos tipos de contenidos están asociados a algún objeto de gestión, por lo que pueden ser recuperados a partir de la información geográfica de éste. En caso contrario, la ubicación de estos contenidos dentro del repositorio documental no permitirá dotarlos de una georreferenciación explícita (la facilitada por el objeto de gestión asociado), por lo que se hace necesario habilitar algún mecanismo que permita asociar a los mismos de información geográfica para su posterior recuperación (en caso de que la dispongan).

Con esta pretensión, actualmente se está estudiando la posibilidad de utilizar Local Lucene [9], que es una extensión de Lucene que añade a éste funciones de búsqueda geográfica.

En lo que se refiere a los mecanismos de explotación, se encuentra actualmente en desarrollo la construcción de un visor de mapas que, apoyándose en un servicio WMS, ofrezca la misma funcionalidad que el servicio de generación de ficheros KML, incluyendo además la posibilidad de realizar consultas híbridas, en las que interviene de forma conjunta tanto la componente textual como la geográfica.

Referencias

- [1] Luaces, M. R., Paramá, J. R., Pedreira, O., Seco, D.: LBD LOCAL: Un Sistema para la Recuperación de Documentos con Referencias Geográficas (2008)
- [2] Cerdeira-Pena A., Luaces M. R., Pedreira O., Seco D.: Un Sistema de Gestión Documental y Workflow con Indexación Temática y Geográfica de los Documentos. (2008).
- [3] Alfresco. <http://www.alfresco.com/>

- [4] Torres Fonseca, F.: Ontology-Driven geographic information systems (2001)
- [5] Web Scripts. http://wiki.alfresco.com/wiki/Web_Scripts
- [6] FreeMarker. <http://freemarker.org/>
- [7] KML. <http://code.google.com/apis/kml/documentation/kmlreference.html#placemark>
- [8] Lucene. <http://lucene.apache.org/>
- [9] Local Lucene.
http://www.nsshutdown.com/projects/lucene/whitepaper/locallucene_v2.html