

Hacia la integración de gestión documental e información espacial

Raúl Moreno¹, Antonio Quintanilla¹, Javier Sánchez¹.

¹Instituto de Desarrollo Regional de la Universidad de Castilla - La Mancha

Antonio.Quintanilla@uclm.es

Resumen

Integrar información espacial en un gestor de documentación permite a los usuarios de éste interactuar de forma más eficiente, al poder acceder a los documentos a partir de su posición en un mapa. En este artículo se abordan los problemas subyacentes de tal integración, proponiendo una infraestructura de bajo coste para efectuarla. En adición, presentamos un módulo desarrollado para el gestor Alfresco y cuya misión es habilitarlo para trabajar con capas en formato vectorial.

Palabras clave: SIG, documento, ECM, IDE, Alfresco, espacial.

1 Introducción

De manera similar a cómo una Infraestructura de Datos Espaciales (IDE) proporciona un paradigma de organización y acceso respecto de la información geográfica, la Gestión de Contenido Empresarial (ECM, por su sigla en inglés) proporciona lo propio respecto de documentos e información general. La ECM es un conjunto formado tanto por prácticas metodológicas (e. g. la digitalización de documentos en papel) como por herramientas (eventualmente agrupadas en un único paquete software). De entre estas últimas una herramienta esencial se corresponde con el gestor de documentación (DMS, por su sigla en inglés), i. e. con el software que se encarga de controlar el almacenamiento en repositorio, el acceso, el flujo de trabajo, y el versionado asociados a cualquiera de los documentos de interés para la entidad administrativa o empresarial involucrada.

En la actualidad existen varias opciones DMS, ya sea en forma de productos exclusivamente dedicados a ese rol, como por ejemplo DSpace, o en forma de productos ECM que lo incluyen, como por ejemplo OpenText Suite. La mayoría de

tales opciones conllevan licencias económicamente costosas u ofrecen licencias gratuitas pero no las funcionalidades requeridas para un DMS de garantías. Sin embargo, una excepción la constituye Alfresco Community Edition [1], que supone una licencia gratuita sin perjuicio alguno de las funcionalidades necesarias; este producto es comúnmente aceptado como uno de los mejores software ECM de código abierto y ha sido evaluado favorablemente por varios autores, siendo Nijman [2] una muestra.

La consideración de datos espaciales dentro de la ECM, en general, y dentro de la gestión documental, en particular, significa una mejora ostensible de los beneficios aportados por éstas a la entidad sobre la que se instauran. Nótese que implica posibilitar enlazar con localizaciones geográficas (con geometrías, en la práctica) los documentos gestionados, consiguiendo de ese modo abrir un nuevo abanico de interesantes aplicaciones para los usuarios finales. Así, un usuario puede consultar las localizaciones asociadas a un documento y análogamente los documentos asociados a una localización. Ejemplos de áreas de aplicación son el turismo (e. g. acceder a la carta de un restaurante al seleccionarlo sobre el mapa), la seguridad pública (e. g. ver en el mapa el lugar de los hechos al seleccionar un documento relativo a una denuncia) o el mundo empresarial (e. g. obtener documentos pertenecientes a las sedes corporativas que estén situadas en una región seleccionada sobre el mapa), existiendo otras muchas como señala Sherman [3].

No obstante, la integración ECM/SIG no queda exenta de desafíos; entre ellos destaca el criterio a emplear para la indexación espacial de los documentos, la comunicación entre los distintos componentes software implicados, la interacción de los usuarios con el sistema final resultante y la adecuada modificación/extensión del DMS empleado.

Proponemos una infraestructura que integre sistema de información geográfica (SIG) y DMS con un coste económico de implantación significativamente bajo. Hemos elegido Alfresco como DMS a emplear y adicionalmente hemos implementado un módulo de extensión para incluir como documentos gestionados en este DMS capas en formato vectorial; a pesar de que en el tratamiento interno existan diferencias respecto de los documentos convencionales, el usuario puede acceder a ellas e incluso modificarlas mediante programas SIG sin percibir estas diferencias y abstrayéndose de la complejidad inherente a ellas.

El resto de este artículo se estructura como sigue. La sección 2 pone de manifiesto una revisión de la literatura referente a la integración DMS/SIG. En la sección 3 se describe la arquitectura general de la infraestructura que nos planteamos materializar, así como se analiza someramente cada uno de los componentes implicados. La sección 4 describe el mecanismo que hemos llevado a cabo para

conseguir la adecuada inclusión de capas vectoriales en el repositorio del gestor. Finalmente, en la sección 5 quedan indicadas las conclusiones derivadas.

2 Trabajos relacionados

En los últimos años se han llevado a cabo varios esfuerzos relacionados con crear arquitecturas que incorporen información espacial en la gestión documental y/o en la ECM. Asimismo, existe constancia de la implementación de varios sistemas al uso.

Hasta donde conocemos, el trabajo de Swenson [4] es la primera aproximación importante al problema. En ella se describe la creación de un sistema para facilitar el trabajo de la División de Permisos de la ciudad de Indianápolis (EE.UU.). Plantea emplear la localización geográfica de los documentos relativos a permisos de construcción, en la indexación de éstos y por ende posibilita su acceso vía mapa. Deja patente el beneficio que esto supone gracias al análisis de las situaciones más usuales a las que se enfrentan los funcionarios en relación con dichos permisos. Gracias a esta temprana experiencia se pueden vislumbrar algunas vicisitudes que pueden surgir en la integración que nos ocupa; no obstante, la solución de Swenson está sesgada hacia el escenario de aplicación y el gestor de documentación considerado resulta harto primitivo.

Algo más sofisticado que el anterior es el tipo de DMS considerado por Crosswell en [5], donde se exponen teóricamente los desafíos a los que la instauración de un DMS se puede enfrentar, inclusive la integración con un SIG. En la misma línea se sitúa el trabajo de Nordheim et al. [6], en el cual toda una serie de datos empíricos relativos a distintas soluciones ECM son empleados para extraer (entre otros hechos) que la integración con SIG es requerida a menudo.

Algunas experiencias más actuales de implantación de sistemas reales que integren gestión documental e información espacial son revisadas a continuación. Para facilitar la revisión hemos optado por agruparlas según el tipo de aplicación al que queden orientadas, considerando los siguientes: aplicaciones de información patrimonial, aplicaciones de turismo, aplicaciones de gestión urbana, y aplicaciones genéricas.

Con respecto a la información patrimonial, el sistema de Hosse et al. [7] es probablemente uno de los acercamientos pioneros. Fue desarrollado por la Universidad Tecnológica de Múnich para el estado federado de Baviera (Alemania) y no emplea un DMS sino que maneja independientemente los documentos georeferenciados. La descripción de un sistema similar es realizada por Gaio et al. [8];

tampoco se emplea un DMS, en su lugar un corpus sin operaciones (i. e. un repositorio estático) es utilizado. Por último, en [9] se puede acceder al Sistema de Información Patrimonial creado por el Consorcio de Santiago de Compostela. Tal sistema emplea OpenText Content Server para la gestión documental y ArcGIS para la parte SIG.

Con respecto a los sistemas orientados a turismo, hasta donde conocemos todos efectúan la integración con un Sistema de Gestión de Contenidos (CMS, por su sigla en inglés); ninguno la efectúa con un DMS. Nótese que la utilización de CMS impide, entre otras funcionalidades, mantener un control de versiones de los documentos geo-referenciados. A modo ilustrativo de trabajos que realizan integraciones en ese sentido valga comentar los llevados a cabo por Ahlers et al. [10], por Rioja et al. [11] y por Rojas-Sola et al. [12]. El primero está enfocado hacia dispositivos móviles y su objetivo es proporcionar al usuario información geo-localizada relativa a distintos lugares de interés turístico de la región de Montafon (Austria), pudiendo tratarse tanto de información en texto como de información multimedia (narraciones sobre el lugar). El segundo se refiere a GeoSistur, que consiste en una plataforma web de información turística; una de sus características más destacables queda ligada a la capacidad del usuario de consultar contenidos cercanos a una localización ordenados por distancias. El tercero concierne a la obtención de una herramienta para museos virtuales; utiliza Joomla como CMS y resulta destacable que permite al usuario interactuar con realidad virtual aumentada.

Con respecto a la gestión urbana, han sido realizados bastantes sistemas similares al de la primigenia experiencia de Indianapolis. Por ejemplo, Gharaibeh et al. presentan en [13] un sistema de apoyo a la gestión del drenaje de aguas pluviales de la ciudad de El Paso (EE.UU.). El mismo aborda entre otras cuestiones la geo-localización de documentos asociados a las diferentes estructuras de drenaje. Está centrado en la parte SIG y no utiliza DMS alguno sino que los documentos son gestionados en un repositorio propio creado a propósito. Otro ejemplo en esta línea de aplicación es el realizado por Hale et al. [14] y queda destinado a la gestión de peticiones/quejas al Buro de Aguas y Aguas Residuales del Departamento de Trabajos Públicos de la ciudad de Baltimore (EE.UU.). En tal trabajo se contempla la integración SIG con ECM al completo (i. e. incluyendo la gestión documental, el control de flujos de trabajo, etc.); para cada parte se utiliza respectivamente ArcGIS Server y el gestor OnBase. Sin tener en cuenta los casos de uso y el ámbito de aplicación y considerando los aspectos meramente tecnológicos, la anterior es posiblemente la propuesta más pintiparada a la nuestra.

Con respecto a los sistemas genéricos no orientados a aplicaciones específicas, uno de los primeros se corresponde con desarrollado por Bain [15]. Utiliza ArcIMS

para la parte SIG y por ende permite la interacción vía web. Un importante aspecto positivo es que soporta varios DMSs, en concreto Sharepoint Portal Server, Documentum 5 y FileNET. Un importante aspecto negativo es que no permite realizar búsquedas espaciales de documentos; en otras palabras, el usuario sólo puede obtener los documentos asociados explícitamente a una geometría, e. g. no puede indicar un *bounding box* y obtener los documentos contenidos en éste. El tener que asociar explícitamente las geometrías implica un problema al actualizar las referencias cruzadas tal y como su propio autor reconoce. Por otro lado, las compañías ESRI, Documentum, EMC, Nirvana, MetaCarta y GTI ofrecen de forma conjunta una solución [16] para integrar SIG/ECM, pero sólo puede ser utilizada si la organización destinataria dispone del respectivo producto de cada compañía implicada. Del mismo modo, la compañía docSTAR ofrece MapConnect [17], un módulo adicional para el DMS comercializado por la misma. Finalmente, en Strötgen et al. proponen TimeTrails [18], cuya característica más definitoria consiste en que no sólo considera la localización de los documentos en el espacio sino también en el tiempo. Emplea el paquete PostGIS y no contempla DMS interno; utiliza fuentes de documentos externas.

3 Infraestructura propuesta

En aras de acometer los desafíos señalados en la sección 1, resulta necesario identificar el conjunto de aquellas entidades tales que la unión de sus funciones a través de unos determinados protocolos de comunicación pueda dar como resultado una infraestructura que cumpla con la funcionalidad global esperada. Como respuesta a la posibilidad de implantar la propuesta en una IDE ya existente y/o en una organización con una ECM ya implantada, se ha diseñado un sistema completamente modular que permite tanto un despliegue conjunto de sus componentes desde cero como el acoplamiento individual de cualquiera de éstos en una infraestructura desplegada a priori.

3.1 Visión general

Establecemos que son necesarias las siguientes entidades debidamente enlazadas: un DMS, un módulo espacial para éste, un servidor de mapas, y una interfaz de usuario. Nuestro cometido consiste en enlazar adecuadamente tales entidades y en desarrollar una interfaz web de demostración y el módulo espacial. En adición nos proponemos incluir un geo-localizador para asociar geometrías a los documentos y un buscador semántico espacial. La figura 1 muestra la arquitectura presumible

para el sistema final, incluyendo los protocolos y/o librerías necesarios para garantizar la comunicación.

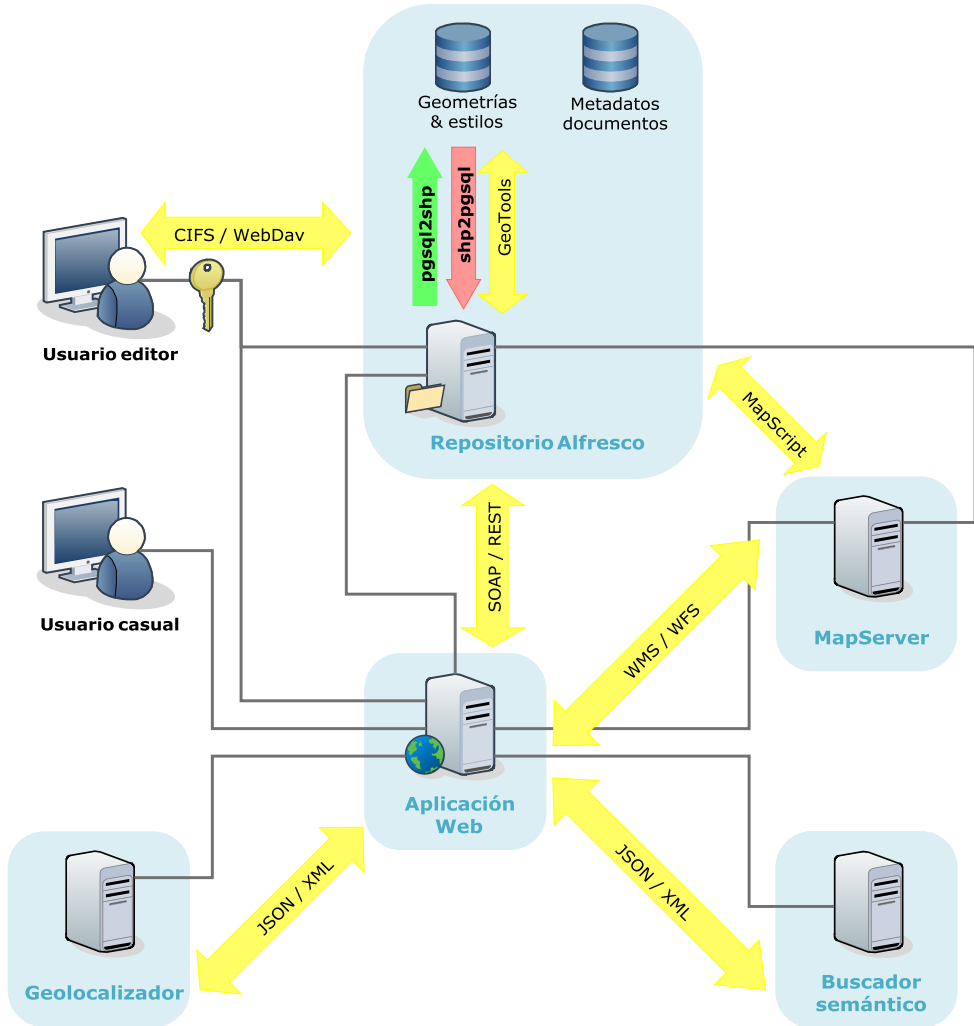


Figura 1. Arquitectura del sistema propuesto

3.2 Aplicación web

La interfaz de usuario ha de posibilitar tanto la interacción habitual con el DMS, i.e. exploración del repositorio, búsqueda de documentos según propiedades y/o

contenido, etc., como la interacción con un visor SIG. Esta última consiste principalmente en visualizar las geometrías asociadas a un documento seleccionado y en indicar como criterio de búsqueda de documentos un elemento de la capa mostrada, un conjunto de elementos, o un bounding box. Nótese que a diferencia de otras propuestas, e. g. la de Bain, en nuestro caso se pueden realizar búsquedas por regiones no enlazadas directamente a los documentos.

Para materializar una interfaz de las características indicadas hemos elegido implementar una aplicación web que se comunique mediante protocolos estándar con el resto de componentes del sistema; por tal razón otras aplicaciones web desarrolladas por terceras partes pueden ser fácilmente integradas. Tenemos en cuenta la existencia de editores así como de usuarios que sólo puedan leer.

3.3 Gestor de documentación

Como componente DMS hemos elegido Alfresco Community Edition. Al ser éste un producto de código abierto, nuestra propuesta goza de una doble ventaja respecto de otras como por ejemplo la de Hale et al. o la de Bain. Por una parte, la organización destinataria se ahorra coste de licencia. Por la otra, existe la posibilidad de extender el código del gestor para hacerlo más eficiente a propósito de la gestión espacial.

De hecho, nuestra propuesta incluye dos módulos de extensión para Alfresco. Uno de ellos queda destinado a las cuestiones relativas a la comunicación con el servidor de mapas, así como a las relativas al almacenamiento de capas, de sus estilos de visualización y de geo-referencias de documentos; en el caso de las capas se contempla aplicar el algoritmo de simplificación Ramer-Douglas-Peucker [19] antes del almacenamiento. El otro queda destinado a conseguir que Alfresco sea capaz de gestionar adecuadamente capas en formato vectorial (véase sección 4).

3.4 Servidor de mapas

El servidor de mapas escogido ha sido MapServer, que es comúnmente reconocido como la mejor opción gratuita para servir a dicho fin. Su misión en nuestro sistema consiste en atender peticiones realizadas tanto por Alfresco como por la propia aplicación web.

3.5 Módulos adicionales

Consideramos integrar un geolocalizador externo para permitir al usuario conocer la geometría correspondiente a una determinada dirección postal; de este modo se

facilita al editor el geo-referenciar los documentos del repositorio. Asimismo, consideramos integrar un buscador semántico externo para facilitar al usuario las búsquedas espaciales.

4 Gestión de capas en repositorio

Si bien en el sistema que proponemos las capas vectoriales son almacenadas en una base de datos espacial (concretamente PostGIS) y esto habilita el correcto funcionamiento de toda la infraestructura, también resulta posible mantenerlas almacenadas en el repositorio nativo de Alfresco. Si se opta por la segunda opción entonces es necesario que Alfresco sea capaz de trabajar con ellas como si fueran documentos típicos (aplicables versionado, control de flujo, etc.) y que, al mismo tiempo, mantenga al usuario abstraído de características tales como la naturaleza multi-archivo de una capa Shapefile. Por esta causa, hemos desarrollado una extensión modular que habilite Alfresco para dicho cometido. También hemos creado una interfaz web para interactuar con los documentos del repositorio, inclusive las capas; la comunicación entra la una y el otro se lleva a cabo vía REST. Una captura de la esta interfaz puede verse en la figura 2.

Nótese que el usuario sólo vislumbra un archivo (.shp) por capa a pesar de que el formato Shapefile implica como mínimo tres; el resto de archivos de cada capa permanecen en el repositorio asociados en todo momento al mostrado pero salvaguardados de la manipulación directa.

Alfresco posibilita interactuar con sus documentos creando una unidad de red enlazada al repositorio vía protocolo CIFS. Es especialmente interesante en nuestro caso pues los usuarios pueden utilizar programas SIG de escritorio, e. g. ArcView, para editar capas del repositorio como si éstas estuvieran en disco local. La figura 3 muestra una captura de una unidad de red enlazada a un repositorio Alfresco que contiene varias capas Shapefile. En este tipo de acceso la abstracción de ficheros auxiliares (.dbf, .shx, etc.) y los eventuales permisos de usuario se mantienen.

Por su parte, para que Alfresco sea capaz de gestionar capas vectoriales son necesarias toda una suerte de modificaciones que extiendan su proceso habitual de gestión de documentos, el cual queda destinado a documentos lógicos asociados a un único archivo físico. La base del nuevo proceso que hemos desarrollado queda expuesta en la figura 4.

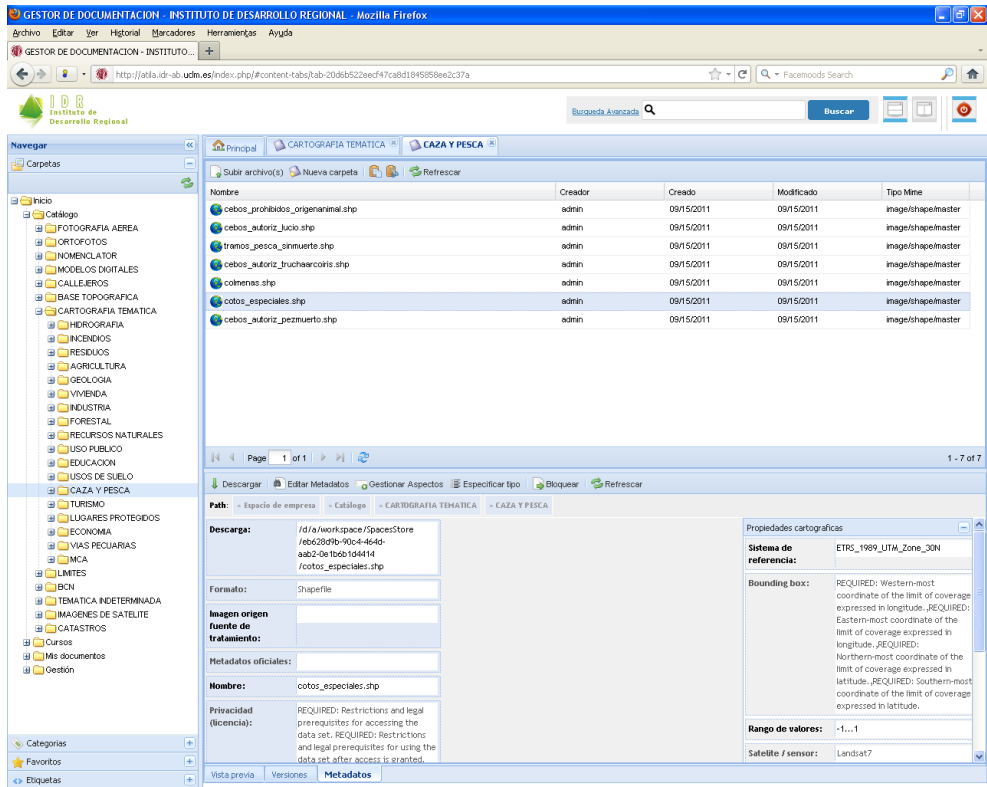


Figura 2. Interfaz web creada para el gestor documental

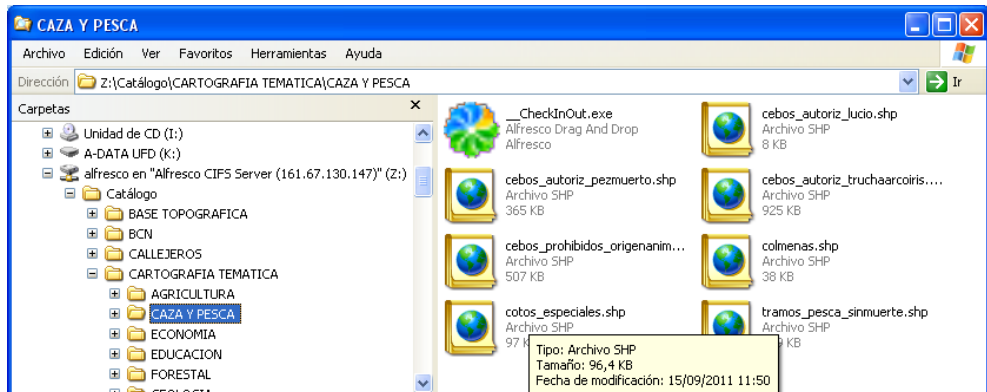


Figura 3. Interacción con shapefiles del repositorio vía unidad de red

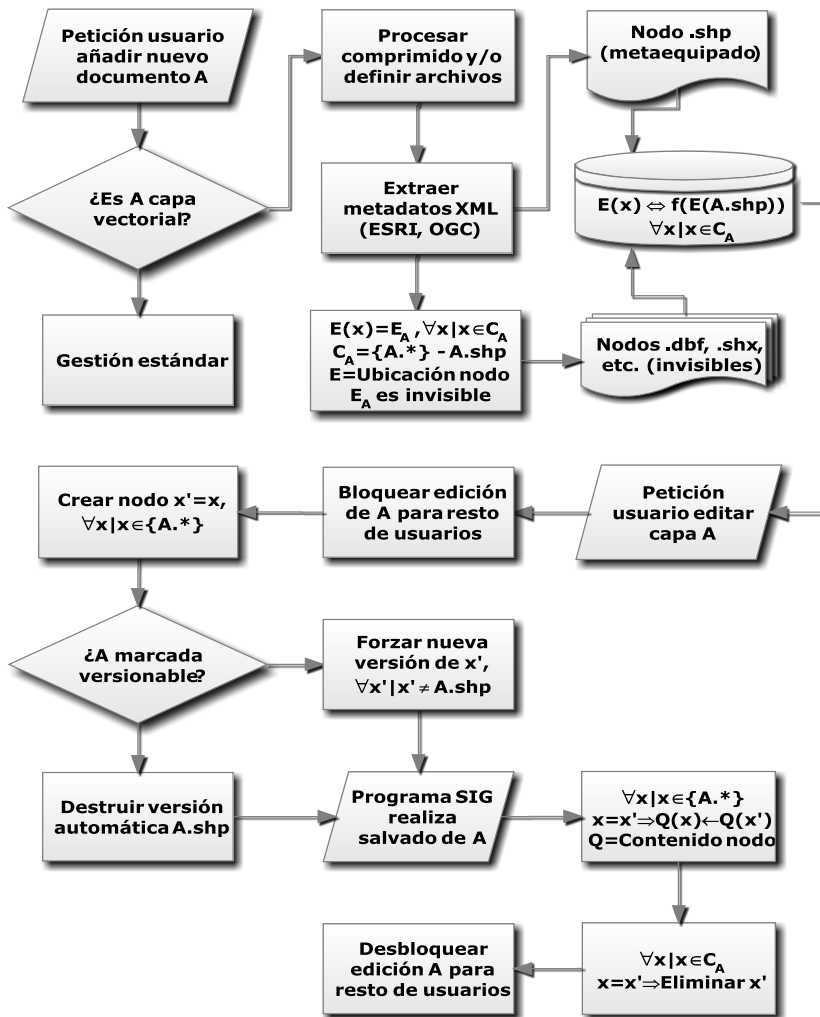


Figura 4. Mecanismo base empleado en la gestión de capas vectoriales

5 Conclusiones

Bajo determinados contextos, poder acceder a un documento a partir de una localización geográfica brinda a los usuarios de un gestor documental con una considerable mejora respecto de la eficiencia en la exploración. Análogamente, poder recuperar las localizaciones vinculadas a un documento aumenta la

capacidad de trabajo de estos mismos usuarios. Para lograr que un gestor de documentación sea capaz de otorgar las capacidades anteriores, es necesario integrarlo con un sistema de información geográfica y en similar medida modificar/extender el modelo de datos de su repositorio.

Si bien en la literatura se pueden encontrar varias propuestas al respecto, la mayoría están basadas en productos software propietarios y por ende suponen un coste elevado para la organización destino. En este artículo se ha abordado el problema proponiendo una infraestructura que en su conjunto cumpla las funcionalidades indicadas pero con un coste económico mínimo. Adicionalmente, se ha presentado un módulo desarrollado a fin de proporcionar al gestor Alfresco la habilidad de trabajar con capas en formato vectorial, de manera análoga a documentos de cualquier otro tipo. Como cuestión futura queda completar la implementación de la infraestructura de tal modo que se pueda realizar una demostración funcional.

Agradecimientos. Este trabajo es fruto de una de las líneas de I+D que viene realizando el IDR dentro del proyecto “RED DE INTELIGENCIA E INNOVACIÓN DE TURISMO”, financiado por el Ministerio de Ciencia e Innovación, Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica. Subprograma IMPACTO.

Referencias

- [1] Alfresco Community Edition, <http://www.alfresco.com/community/>
- [2] Nijman, A.S.: A Proposal and Selection of a Document Management System for Staatsolie Suriname. M.S. thesis, Delft University of Technology, Netherlands (2009)
- [3] Sherman, S.: GIS and Document Management Integration. In: Proceedings of the 27th Annual Esri International User Conference. ESRI, San Diego (2007)
- [4] Swenson, A.D.: Mapping your documents: issues integrating GIS with document management. In: Sailing, M.J. (ed.) URISA Annual Conference Proceedings, pp. 210-218. Orlando (2000)
- [5] Crosswell, P.: Concepts and Project Design Considerations for Electronic Document Management System Implementation. In: Sailing, M.J. (ed.) URISA Annual Conference Proceedings, pp. 556-568. Long Beach (2001)
- [6] Nordheim, S., Päivärinta, T.: Customization of Enterprise Content Management Systems: An Exploratory Case Study. In: Proceedings of the Proceedings of the 37th Annual Hawaii International Conference on System

- Sciences. HICSS '04, vol. 4, pp. 1-9. IEEE Computer Society, Washington (2004)
- [7] Hosse, K., Schilcher, M.: Temporal GIS for Analysis and Visualization of Cultural Heritage. In: Proceedings of CIPA XIX international Symposium. Antalya (2003)
- [8] Gaio, M., Sallaberry, C., Etcheverry, P., Marquesuzaa, C., Lesbegueries, J.: A global process to access documents' contents from a geographical point of view. *J. Vis. Lang. Comput.* 19, 3-23 (2008)
- [9] Sistema de Información Patrimonial de Santiago de Compostela, <http://sip.consorciodesantiago.org/SIPWeb/>
- [10] Ahlers, D., Boll, S., Wichmann, D.: Virtual Signposts for Location-based Story-telling. In: Proceedings of the International GI Days Conference. GI Days, Münster (2008)
- [11] Rioja, R. et al.: GeoSistur: integración de componentes tecnológicos IDE en un CMS para la construcción de geoportales turísticos. In: Actas de las I Jornadas Ibéricas de Infraestructuras de Datos Espaciales. JIIDE'2010, Lisboa (2010)
- [12] Rojas-Sola, J.I., Castro-García, M., Carranza-Cañadas, M.P.: Content management system incorporated in a virtual museum hosting. *J. Cult. Herit.* 12, 74-81 (2011)
- [13] Gharaibeh, N., Camacho, G., Elgendy, M., Ramirez, I.: Geographic Information System Framework for Stormwater Drainage Asset Management. *J. Transp. Resrch. Brd.* 2121, 1-12 (2009)
- [14] Hale, H.E., DeShields, J., Reichert, C., Beck, S.: Implementing a GIS/ECM Based Public Utility CSR Tracking System. In: Proceedings of the 30th Annual Esri International User Conference. ESRI, San Diego (2010)
- [15] Bain, A.: Documents and GIS: A Marriage Built to Last. In: Proceedings of the 23th Annual Esri International User Conference. ESRI, San Diego (2003)
- [16] White paper: Integration and management of geospatial data and related documents in collaborative environments. EMC® & Nirvana®.
- [17] docSTAR™ MapConnect, <http://www.docstar.com/gis-document-management>
- [18] Strötgen, J., Gertz, M.: TimeTrails: A System for Exploring Spatio-Temporal Information in Documents. In: Bertino, E., Atzeni, P., Lee, K., Chen, Y., Tay, Y. C. (eds.) VLDB 2010. The Proceedings of the VLDB, vol. 3, pp. 1569-1572. VLDB Endowment (2010)
- [19] Douglas, D. H., Peucker, T.K.: Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *J. Canad. Cart.* 10, 112-122 (1973)