

Georreferenciación de documentación ferroviaria mediante indexación semántica y su publicación, usando software libre

José Gómez Castaño^{(1)/(2)}

⁽¹⁾Grupo de Astronomía Extragaláctica e Instrumentación Astronómica
Dept. Astrofísica y CC. De la Atmósfera – UMC, jgomez03@pdi.ucm.es

⁽²⁾Subdirección de Sistemas Operacionales, ADIF, Edificio 23, Calle Hiedra S/N, Estación Madrid Chamartín, Madrid

RESUMEN

Para la gestión y explotación del ferrocarril, es necesario manejar gran cantidad de documentación, la mayoría de ella de carácter oficial, vinculante y reglamentaria. Actualmente se usan gestores documentales que permiten la indexación de estos usando metadatos cumplimentados manualmente. Su publicación se lleva a cabo mediante aplicaciones que la ponen a disposición del usuario. Respecto a la documentación reglamentaria, existe una aplicación desarrollada a medida, que permite la consulta, reparto y control de distribución de los mismos. Los metadatos están constituidos por atributos alfanuméricos que identifican las características de los documentos, siendo la única referencia a la ubicación el código de la estación, o tramo contable.

En este trabajo se desarrolla una metodología para llevar a cabo la indexación y georreferenciación de todo tipo de documentación, haciendo hincapié en la reglamentaria, utilizando algoritmos semánticos y herramientas basadas en software libre. Así mismo la publicación añadiendo este atributo adicional, permite el cruce de la misma con cualquier otro elemento espacial.

El objetivo es llevar a cabo análisis geoespaciales en los que los documentos se utilizan asociados a entidades espaciales, de esta forma es posible llevar a cabo análisis junto a la posición de los trenes, su recorrido o cualquier información geográfica externa ligada al terreno, como información catastral, estructuras, archivos SHP, etc,

Palabras clave: Ferrocarril, SIG, Documentación, Indexación Semántica, Software Libre

INTRODUCCIÓN

La documentación es un elemento fundamental a la hora de desarrollar la actividad ferroviaria. Cuando se trata de normalizar las actuaciones y procedimientos, es necesario tener un soporte documental al día y a disposición de todos los usuarios y empresas que intervienen en la operación. Hoy día no solo se trata de una labor interna de una empresa, sino que, desde la liberalización del ferrocarril, y la

participación de gran cantidad de empresas auxiliares, esta normativa debe ser accesible a través de diferentes canales hasta hacerla llegar al mayor número de interesados posible, con la mayor celeridad. Actualmente se gestionan más de 6000 documentos reglamentarios, que van desde simples noticias hasta modificaciones al Reglamento General de Circulación, y que hay que poner a disposición de todos los implicados, muchos de ellos implicados en regular la seguridad de las operaciones ferroviarias.

Esto se ha conseguido hasta ahora, con el uso de aplicaciones basadas en gestores documentales. En ADIF existe desde 2009 la aplicación de Registro de General de Documentación Reglamentaria [ADIF 2009]. Esta permite la distribución y trazabilidad del reparto de estos documentos. Este software está basado en búsquedas e índices alfanuméricos sobre bases de datos relacionales. La asociación de la documentación a un lugar en concreto se reduce a los metadatos alfanuméricos que hacen referencia a la estación o línea en la que se encuentra el elemento objeto del documento

En este trabajo se persiguen dos objetivos.

- Llevar a cabo un análisis semántico del contenido de los documentos, y a partir de él, asociarle una entidad geoespacial.
- Realizar búsquedas usando índices espaciales que puedan utilizarse junto a motores de búsqueda tradicionales

Estado actual

En la actualidad, los gestores documentales son básicamente repositorios de documentos. Estos se almacenan en sistemas de ficheros, y se les asocia una serie de metadatos para describir su contenido, autoría y toda la información que se considere relevante. Estos metadatos sirven para llevar a cabo búsquedas usando diferentes criterios de acuerdo con las necesidades del usuario.

Figura 1: Aplicación actual RGD

En algunos casos, los metadatos incluyen información sobre la ubicación del objeto al que hacen referencia. Así por ejemplo la Consigna de Instalaciones de una estación, lleva asociado el código de esa estación. En muchos casos esta no se encuentra ni siquiera en forma de metadato, sino que solo aparece descrita en el propio texto del documento. La búsqueda de un documento que afecta a una estación se hace introduciendo el código de la misma y buscando las coincidencias de este dentro de la base de datos. Para tener un acceso más rápido, se crean índices sobre los atributos más habituales, como el código de la estación o el código de la línea a que afecta

En este trabajo se aborda la indexación de los documentos desde el punto de vista geográfico. En Gómez Castaño, J, 2013 ya se han analizado ampliamente las posibilidades de convertir elementos ferroviarios, como el recorrido de un tren, en un elemento espacial, y se ha desarrollado una aplicación para su representación y análisis Gómez Castaño, J 2013. En otros trabajos, como Cedeira-Pena, 2012, y Luaces 2008, también se ha analizado el uso de gestores documentales basados en atributos geográficos.

La diferencia de este trabajo con otros anteriores es que se lleva a cabo un rastreo de todo el documento, no se ciñéndose exclusivamente a los metadatos. El texto de todo el documento se rastrea en busca de referencias conocidas y el mismo queda asociado a todas ellas

Metodología

Podemos distinguir dos partes en este trabajo. La primera se encarga de la indexación y georreferenciación de los documentos y la segunda del proceso de búsquedas

El modulo de georreferenciación rastrea todo el documento. En primer lugar convierte en texto plano el fichero pdf, y sobre este se hace la comparación con la base ontológica creada. Los metadatos siguen el mismo proceso.

Para cada topónimo encontrado se almacena la referencia del índice textual y una lista de las geometrías asociadas al topónimo.

Las relaciones del tipo "estar contenido en", se implementan como una consulta Postgis. De esta forma, se comprueba si una estación, St_Point, pertenece a una geometría de tipo ST_Line o ST_Polygon, las cuales definen a una Jefatura, delegación al estar dentro del polígono delimitado para cada una.

Indexación

El proceso de indexación se puede dividir en dos partes. La primera llevada a cabo por el gestor documental tradicional, se encarga del tratamiento de los metadatos

alfanuméricos. Es variado el software disponible para esta tarea, tanto de uso libre como privativo. En la aplicación actual en producción se usa el producto Content Manager [IBM] de IBM, a través de un API disponible en Java. Como soluciones libres se pueden encontrar Alfresco [ALFRESCO], por ejemplo.

El proceso de indexación espacial, se basa en reconocer patrones semánticos que hagan referencia a localizaciones. Para ello se ha creado un diccionario de topónimos que recogen los nombres geográficos relacionados con elementos ferroviarios.

De esta forma se ha construido una tabla donde figuran los nombres junto a su posición geográfica. Este tesoro consiste en un dominio restringido y especializado fácilmente manejable, Los nombres se han extraído de la literatura relacionada con geografía ferroviaria. Por otra parte, se han tomado sus correspondientes nombres del Nomenclátor [NEM]. En la tabla aparecen un ejemplo de algunos de estos nombres

Tabla 1: Ejemplo de nombres geográficos ferroviarios

Id
Madrid Chamartín
Bifurcación La Negrilla
Subestación Linares
PK 15,500
Línea Hendaya - Madrid
Viaducto de Cáceres
Paso a Nivel PK 50+541

El flujo de trabajo que se sigue consiste en la introducción del documento en el sistema. Indexación por metadatos alfanuméricos. Reconocimiento de texto. Comparación de las palabras contenidas en los metadatos *Título, Resumen, Palabras Clave, Líneas, Estaciones Afectadas*, con la base de topónimos. Búsqueda de palabras internas en el texto y comparación con la base de topónimos.

Cada topónimo encontrado genera una entrada en una tabla, que relaciona el identificador del documento, con el objeto espacial al que representa el topónimo. De esta manera podemos encontrar los siguientes tipos de relaciones:

- ST_Point: Para localizaciones puntuales. Estaciones, bifurcaciones, subestaciones eléctricas, señales...
- ST_Line: Para trayectos lineales. Líneas, trayectos, tramos,
- ST_Polygon: Para recintos. Zonas de seguridad,

Como se expuso en Gómez Castaño, J, 2011, INSPIRE, y concretamente el ANEXO I, permite disponer de un marco en el que tener una nomenclatura común para diferentes elementos geográficos relacionados con el ferrocarril que permita la interoperabilidad. Sin embargo, la nomenclatura utilizada en cada país en su documentación, no está adaptada a la misma. Para poder hacer esta correlación, se ha creado una tabla que relaciona algunos de los tipos de elementos ferroviarios nacionales, con su equivalente en INSPIRE.

Tabla 2: Relación entre designación nacional y fenómenos INSPIRE

España	INSPIRE
Estación	RailwayNode
	RailwaySatationNode
Bifurcación	RailwayNode
Línea	RailwayLine

Area Estacion	RailwayYardArea
Tramo	RailwayLink
Zona Seguridad	RailwayArea

Además de estos elementos, se han tipificado otros no teniendo una transcripción en INSPIRE, dentro de la ontología. Por ejemplo, Delegaciones Territoriales, Jefaturas de distinto tipo, Municipios, Provincias y Comunidades Autónomas

El proceso de reconocimiento de los topónimos se lleva a cabo a partir del documento PDF, permitiendo el sistema introducir la documentación de forma tradicional, no impactando en los procesos actuales. El software Lucene, permite recorrer el contenido del documento, y realizar la comparación con la base de topónimos. Cada uno de ellos se almacena en la base de datos

Indexación de Puntos Kilométricos

Un capítulo a parte requiere la indexación de Puntos Kilométricos. Como se ha demostrado en Gómez Castaño, y otros 2014, existen elementos en la infraestructura que se encuentran en plena vía. Estos pueden ser bifurcaciones, agujas, apartaderos, subestaciones, pasos a nivel, e incluso un tramo de obras. los cuales no tienen un nombre definido geográficamente en ninguna toponimia. La forma de localizarlos es por medio del punto kilométrico de la línea en la que se encuentran. Así, por ejemplo, se designa la aguja KM 32,500.

En los casos en los que si existe una geolocalización previa del elemento, esta se ha incorporado a la ontología ferroviaria, utilizando el código correspondiente al elemento. En caso contrario, se aplica el procedimiento descrito y el PK nominal del elemento es el que ese se geolocaliza, siendo esta la posición que queda asociada al mismo. De esta forma se obtiene un ST_Point, para un PK aislado, o un ST_Line, para el caso de un tramo entre dos Pks, de bloqueo, por ejemplo.



Figura 2: Geolocalización de Puntos Kilométricos

Búsquedas

Para las búsquedas se ha aplicado la técnica geo-first o text-first, según el caso. Si la búsqueda implica entidades geográficas, se implementa la primera solución. En primer lugar se hace un recorrido por el índice espacial, y devolviendo el conjunto de documentos seleccionado. A este se le aplica el índice textual. En el caso de búsquedas meramente textuales, se aplica el segundo.

Desde el interfaz de búsquedas se rellenan los campos por los que se necesita filtrar el resultado. Este interfaz está compuesto por un formulario. Los campos los

componen parte metadatos alfanuméricos y parte geográficos. En el caso de que solo se filtre por el primer tipo, la búsqueda se hace solo en el gestor documental. Si la búsqueda implica localizaciones, se incluye la búsqueda sobre la base de datos espacial. Los atributos alfanuméricos se utilizan de forma tradicional y el campo espacial sirve como índice cuando entran en juego localizaciones.

Para hacer esta doble búsqueda en dos gestores diferentes, es necesario que el motor de búsqueda, desarrollado en java, se conecte al motor tradicional y a la base de datos espacial. Los gestores documentales de código abierto disponen de una API abierta. Esta permite integrar otro software de externos que haga uso de sus capacidades de búsqueda. De esta manera se ha utilizado el gestor Alfresco como motor textual.

Herramientas utilizadas

Las herramientas utilizadas han sido las siguientes:

- Base de datos PostGIS
- Servidor de aplicaciones Tomcat 7
- Lenguaje de scripting Python
- Lenguaje Java para la interacción con el motor de indexación
- Motor de indexación Lucene [Lucene 2007]
- Gestor documental Alfresco

Condición de pertenencia

Una vez asociado un objeto geoespacial a los documentos, estos pueden ser tratados con herramientas que impliquen geoprosos. Básicamente se busca la condición de inclusión. Un ST_Point está incluido en una ST_Line o en un $ST_Polygon$. A su vez una ST_Line está incluida en un $ST_Polygon$. Aplicando esta regla meramente espacial, se pueden conocer los documentos asociados a su geometría que están dentro de otra que los contiene.

$$ST_{Point} \subset ST_{Line} \subset ST_{Polygon}$$

Las búsquedas sobre atributos espaciales se han implementado directamente sobre la base de datos, usando el estándar SFSQL. Esto permite extender la funcionalidad a otros gestores de bases de datos espaciales que cumplan el estándar.

Aplicaciones

La principal aplicación de esta metodología es la de agilizar las búsquedas de todo tipo de documentación asociada a las líneas e instalaciones ferroviarias. Tanto las Empresas Ferroviarias como los Administradores de la Infraestructura se pueden beneficiar de la implementación de estas soluciones en sus respectivos gestores documentales. Así mismo es posible la aplicación a otros ámbitos fuera del ferrocarril.

A continuación se detallan algunos ejemplos implementados:

Autorización de acceso

La documentación tiene diferentes niveles de acceso para que solo los usuarios con ciertos privilegios en el sistema, puedan visualizarlos, editarlos, etc. En la actualidad, la asociación de un usuario a un área se llevaba a cabo mediante atributos alfanuméricos. Teniendo en cuenta la localización del usuario, previamente almacenada en un campo GEOM de la base de datos PostGIS, es posible comprobar

si este, tiene autorización, definiendo esta como correcta, si se cumple que la localización del usuario está dentro del fenómeno al que está asociado el documento.

Ejecución de trabajos

A la hora de llevar a cabo trabajos ya sea en plena vía o en instalaciones, es posible conocer la documentación que le afecta. Al conectarse al sistema se identifica la posición del usuario y se envía al buscador como referencia de su posición.

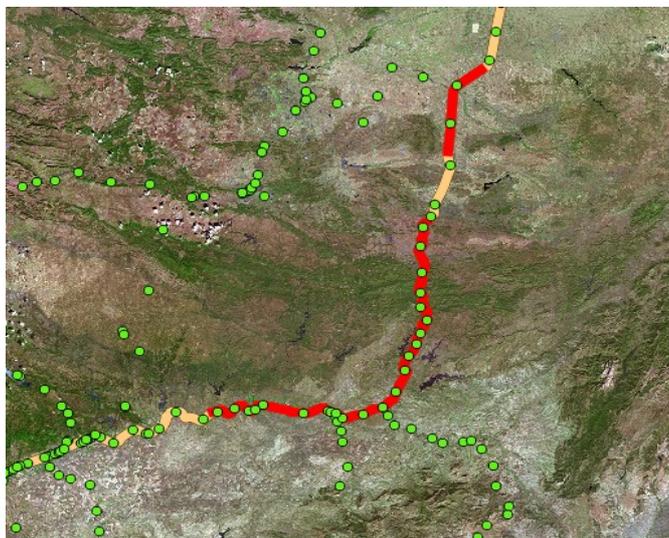


Figura 3: Definición de tramos con trabajos sobre una línea

Documentación en ruta

Los trabajadores de vía y los trenes, necesitan tener acceso a la documentación relacionada con la instalación en la que se encuentran. Así por ejemplo, las brigadas de instalaciones necesitan tener acceso a las Consignas de subestaciones eléctricas, bloqueos, señales, etc. A partir de la posición del trabajador, se interroga al sistema y este devuelve aquella relacionada con la misma. Si no se está in-situ en el lugar de trabajo, se puede consultar la documentación usando un mapa o el código de la instalación.

Documentación no ferroviaria

Existe gran cantidad de documentación no ferroviaria, pero que tiene relación con esta actividad. Esta es creada por administraciones locales, Autonómicas o Nacionales. Estos documentos se incorporan al sistema permitiendo su búsqueda mediante el mismo procedimiento. Así la información sobre edificaciones, carreteras etc, se referencia a las líneas ferroviarias sin tener que introducir ningún otro metadato.

Conclusión

Se ha mostrado una metodología para la indexación, recuperación y publicación de información documental aplicadas al ferrocarril, en la que se utilizan índices espaciales para relacionar cada documento con las localizaciones a las que afecta el mismo.

Las ventajas del uso de esta solución conjuntamente a los índices alfanuméricos tradicionales son:

- Los elementos se crean sin tener que preocuparse por relacionarlos con el resto, no haciendo falta mantener tablas de relación. Por el hecho de estar referidos a una localización, es esta la que crea la relación con el resto
- Es más sencillo crear ámbitos. Si se necesita agrupar los documentos que afectan a una área de trabajos, basta definirla usando las coordenadas que abarca, bien manualmente o relacionándolas con una geometría ya existente. Los documentos relacionados con la misma son los que sus geometrías se encuentran dentro de ese polígono
- Es posible integrar documentación no ferroviaria. Existe gran cantidad de documentos procedente de administraciones locales, autonómicas o privadas, emergencias, etc. En ellas se hace referencia a localizaciones que pueden afectar a las instalaciones o a la zona de seguridad de la vía

En la figura 4 se muestra un mapa de calor indicando la distribución de documentos

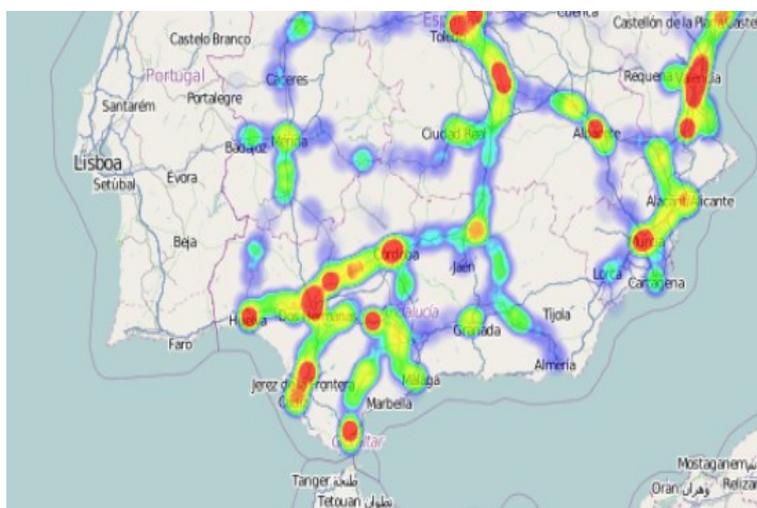


Figura 4: Mapa de calor con la distribución de la documentación

Agradecimientos

El autor desearía agradecer al personal de IDEAdif por proporcionar la geometría de vía

REFERENCIAS

- ◆ ADIF, 2009 “Registro General de Documentación Reglamentaria”
- ◆ ALFRESCO <http://www.alfresco.com/es>
- ◆ Cedeira-Pena, Ana, Luaces, Miguel, Pedreira, Oscar, Seco, Diego “Un Sistema de Gestión Documental y Workflow con Indexación Temática y Geográfica de los Documentos”, JIIDE 2012
- ◆ DIRECTIVA 2007/2/CE DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 14 de marzo de 2007 por la que se establece una infraestructura de información espacial en la Comunidad Europea (Inspire)
- ◆ GÓMEZ CASTAÑO, J, 2014. “Análisis de la correlación existente entre los Puntos Kilométricos Nominales y Reales en la geometría de la línea ferroviaria”. Vía Libre Técnica, Fundación de los Ferrocarriles Españoles, 2014

- ◆ GÓMEZ CASTAÑO, J, (2012) “INSPIRE: Base de la implementación de una Infraestructura de Datos Espaciales para la interoperabilidad de datos geográficos del ferrocarril en Europa”, VII Congreso de Innovación Ferroviaria, Zaragoza, 2012
- ◆ GÓMEZ CASTAÑO, J, 2013 “Tratamiento geoespacial del recorrido de trenes y de tramos ferroviarios”, IV Jornadas Ibéricas de infraestructuras de Datos Espaciales, Toledo, 2013
- ◆ GÓMEZ CASTAÑO, J, 2013 <http://www.youtube.com/watch?v=63qbSx4hoxU>
- ◆ IBM Content Manager
<http://www-03.ibm.com/software/products/es/conman/>
- ◆ Luaces, M. R., Paramá, J. R., Pedreira, O., Seco, D. (2008) LBD LOCAL: Un Sistema para la Recuperación de Documentos con Referencias Geográficas. En Actas de las II Jornadas de SIG Libre, Girona, España
- ◆ LUCENE (2007) Apache Lucene, <http://lucene.apache.org>
- ◆ NEM, Nomenclator Español