

Sistema de codificación espacial de documentos geocientíficos

D. Hernández Sanz⁽¹⁾, J.A. Carroza García⁽¹⁾, E. Manchado Macías⁽²⁾,
M. Bernat Rebolal⁽¹⁾ y R. Hernández Manchado⁽¹⁾

(1) Instituto Geológico y Minero de España. Área de Tecnologías y Sistemas de la Información, Secretaría General
c/ Ríos Rosas, 23. 28003 Madrid

E-mails: d.hernandez@igme.es - ja.carroza@igme.es - m.bernat@igme.es - r.hernandez@igme.es

(2) Universidad de Salamanca. Facultad de Ciencias. Plaza de los Caídos, s/n. 37008 Salamanca
E-mail: a87578@usal.es

RESUMEN

Actualmente muchos organismos, bibliotecas y fondos documentales permiten a los usuarios realizar búsquedas entre sus documentos a través de Internet. Las consultas son temáticas, se realizan a partir de diversos campos, como pueden ser título, autor, descripción, tema, país de edición, idioma, etc. Si analizamos los documentos existentes en bibliotecas y fondos documentales sobre las Ciencias de la Tierra, podemos observar que la mayoría de ellos, dado el carácter espacial de dichas ciencias, se refieren a espacios concretos de la superficie terrestre. Esta relación entre espacios y documentos, es la que permite a estos ser georreferenciados. La georreferenciación permite realizar consultas espaciales, es decir, buscar documentos que traten acerca de una parte de la superficie terrestre elegida por el usuario. Una variante de la georreferenciación, la geocodificación, permite que las consultas espaciales resulten eficaces en un medio como Internet en que la rapidez de respuesta es fundamental.

Palabras clave: base de datos, consulta SQL, documentos, geocodificación, georreferenciación

Spatial codification system of geoscientific documents

ABSTRACT

Nowadays many organizations and libraries allow the users to make searches among their documents through Internet. The searches are thematic, made from diverse fields, such as title, author, description, subject, country of edition, language, etc. If existing documents in libraries about the so-called Earth Sciences are analyzed, most of them, given the spatial character of these sciences, talk about specific areas of the terrestrial surface. This relation between areas and documents, is what allows these to be georeferenced. The georeferencing of documents allows to make spatial searches, what means looking for documents that affect to a specific location on the Earth chosen by the user. Geocodification, as a variation of georeferencing, allows the spatial statements to be useful in an environment as Internet, where the speed to answer is essential.

Key words: database, documents, geocodification, georeferencing, SQL statement

Introducción: las búsquedas en el fondo documental del IGME

Actualmente muchos organismos, bibliotecas y fondos documentales permiten a los usuarios realizar búsquedas entre sus documentos a través de Internet. Este es el caso del Sistema de Información Documental (SID) del Instituto Geológico y Minero de España (IGME). Las búsquedas se pueden realizar a partir de diversos campos, que en el SID del IGME son título, autor, descriptor, año de edición, código, país, idioma y área. Todo ello se completa con un

Tesoro de Ciencias de la Tierra, compendio en varios idiomas de términos relacionados y jerarquizados.

Pero si analizamos los documentos existentes en el fondo documental del IGME, podemos ver que la gran mayoría, más de un 80%, están referidos a un espacio definido, a un área concreta. Dado el carácter espacial de éstos y las ciencias que son objeto de estudio en el IGME, podemos deducir que muchas de las búsquedas que los usuarios realizan en el SID, podrían tener carácter espacial. Esto en la actualidad no es posible, ya que la opción de consulta bajo criterios espaciales no existe. Esta falta es suplida por

los usuarios buscando documentos cuyo título o palabras clave contengan el nombre del área que les interesa.

Objetivos

Poder realizar consultas con criterios *realmente* espaciales parece por lo tanto un método innovador, que añadido a las consultas temáticas tradicionales, permitirá aumentar la precisión de las búsquedas.

El trabajo nace con la pretensión de que cualquier usuario que acceda al fondo documental del IGME, a través de intranet o Internet, pueda realizar, además de las consultas temáticas tradicionales, consultas espaciales, buscando por el nombre de alguna unidad espacial, o incluso por unas coordenadas geográficas concretas (Fig. 1).

Para ello es necesario que los documentos estén asociados a su área de estudio, es decir, los documentos deben estar georreferenciados.



Fig. 1. Ejemplo de búsqueda: actual y espacial
Fig. 1. Example of search: now and spatial

Asociación de información espacial a los documentos

La asociación de información espacial a los documentos se puede llevar a cabo mediante dos sistemas (Fig. 2): la georreferenciación propiamente dicha, y la geocodificación. Y ambos han sido analizados, concluyendo ventajas e inconvenientes para cada uno de ellos.

Georreferenciación

La georreferenciación consiste en asociar coordenadas geográficas al documento, ubicarlo en la superficie terrestre a partir de un par de coordenadas (para un documento que trata de un elemento puntual), o a partir de múltiples coordenadas (documento acerca de un elemento lineal o de un área).

La característica fundamental de la georreferenciación y también su principal ventaja frente al otro método es la precisión con que se ubica el documento sobre el terreno. La resolución espacial del sistema es perfecta, al poder relacionar un documento con un punto exacto definido por su coordenada X y su coordenada Y.

Pero este método presenta inconvenientes en dos fases, en la geocodificación y en la consulta. Durante la geocodificación resulta difícil pensar que el operador que da de alta el documento en el fondo documental (que se guía habitualmente por el título) sea capaz de ubicarlo perfectamente, ya que no conocerá las coordenadas precisas de su área de estudio.

Durante la consulta nos encontramos con dos situaciones problemáticas, la necesidad de implementar un SIG en el servidor que realice intersecciones de coberturas buscando documentos en cada una de ellas, y la integración de las consultas temáticas tradicionales (actualmente en lenguaje SQL contra la base de datos del fondo documental en SQL Server) con la nueva consulta espacial (requeriría una herramienta del tipo a ArcSDE que transmitiera la información espacial entre la aplicación SIG y la base de datos SQL Server). Ambas situaciones redundan negativamente en un aspecto fundamental para los usuarios, la velocidad de respuesta ante las búsquedas.

Geocodificación

La geocodificación consiste en asociar códigos a los documentos, y cada uno de esos códigos representa el nombre de un espacio concreto y reconocido existente sobre la superficie terrestre.

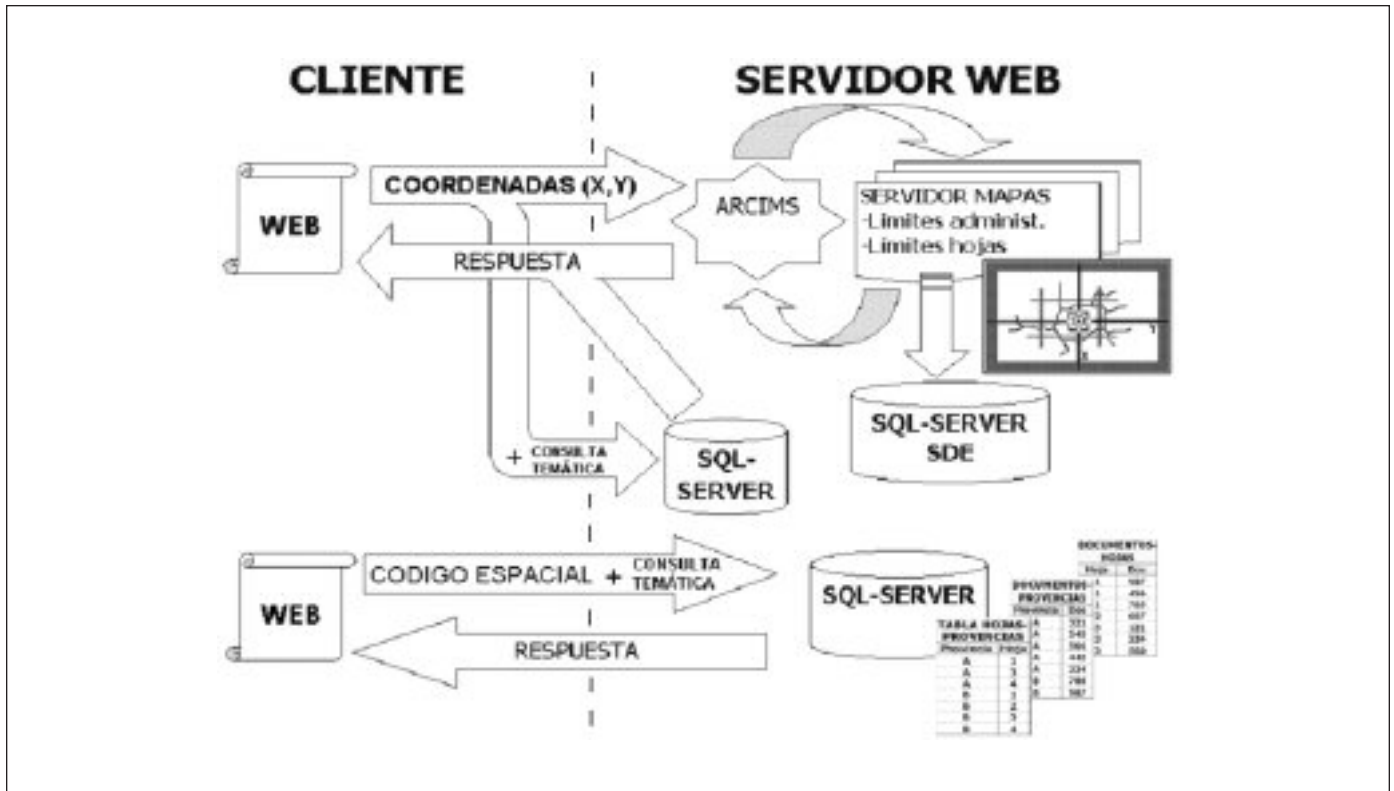


Fig. 2. Georreferenciación VS Geocodificación
 Fig. 2. Georeferencing VS Geocodification

Este sistema tiene un inconveniente destacado, la pérdida de resolución espacial. Al asociar un espacio a un código que representa un área en vez de a unas coordenadas estamos limitando la precisión del sistema a las dimensiones de los espacios representados por los códigos.

Este sistema tiene por el contrario varias ventajas. Una es la facilidad con la que el operador al dar de alta un documento decide que código asociarle, ya que estos por lo general representan espacios conocidos, "familiares".

Una segunda es que no necesita implementar un SIG en servidor, ya que las relaciones entre coberturas, van a ser en realidad relaciones entre tablas alfanuméricas. Esto es debido a un proceso previo en el que mediante un software de análisis espacial (para nuestro caso se usa ArcGis 9.0) realizamos intersecciones entre las coberturas de las unidades espaciales que queremos asociar a los documentos, y el resultado son tablas cuyos registros relacionan unidades espaciales de diferentes tipos cuando éstas se superponen sobre la superficie terrestre (Fig. 3).

Y una tercera es la facilidad de integración de la consulta temática con la espacial, ya que ambas son lanzadas contra una misma base de datos (el SID del

IGME en SQL Server). Ello permite que la búsqueda dé lugar a una consulta única en lenguaje SQL.

Sin olvidar la primera ventaja, las dos últimas son fundamentales para la velocidad de respuesta. Y la velocidad de respuesta en un medio como Internet a la hora de difundir la información hace que optemos para el IGME por este segundo sistema, la codificación espacial o geocodificación, cuyo desarrollo explicaremos en adelante.

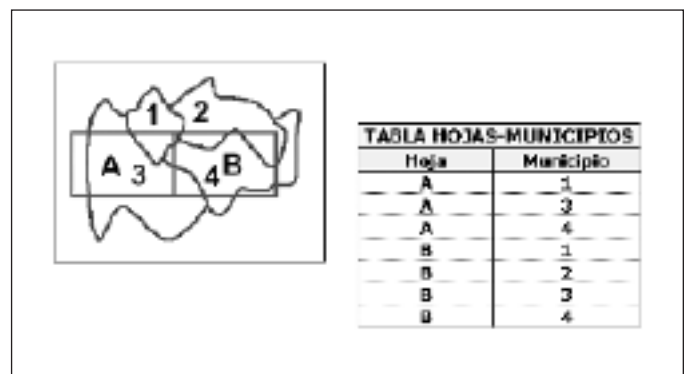


Fig. 3. Creación de tablas que relacionan unidades espaciales
 Fig. 3. Creation of tables that relate spatial units

La pérdida de resolución espacial puede ser corregida buscando algún tipo de unidad cuyo tamaño sea reducido.

Diseño de una base de datos para la geocodificación de documentos

Elección de unidades espaciales a asociar

Como primer paso a la hora de diseñar una base de datos se deben estudiar las posibles unidades geográficas que se pueden asociar a los documentos. Se han seleccionado para el SID del IGME siete tipos, englobados en tres clases: unidades administrativas (países, comunidades, provincias y municipios), límites de hojas del Mapa Geológico Nacional (serie 1:50.000 y serie 1:200.000), y un grupo heterogéneo compuesto por unidades geológicas, geográficas, etc (a las que denominaremos geounidades). Estos siete tipos de unidades son los que relacionaremos con los documentos.

Establecimiento de la unidad básica

Las relaciones entre cada clase y entre cada unidad se establecen a partir de la unidad o unidades mínimas (básicas), que pueden ser la unidad menor de cada clase, o la unidad menor de entre todas las clases. Esta unidad o unidades básicas nos darán la resolución espacial de nuestro sistema de geocodificación. Para el caso del IGME la unidad básica seleccionada es el municipio (Fig. 4), la menor de las unidades administrativas, y relaciona todas las demás.

Rediseño de la base de datos existente

Para implementar la geocodificación se debe rediseñar la base de datos del SID, y se crean nuevas tablas de tres tipos (Fig. 5):

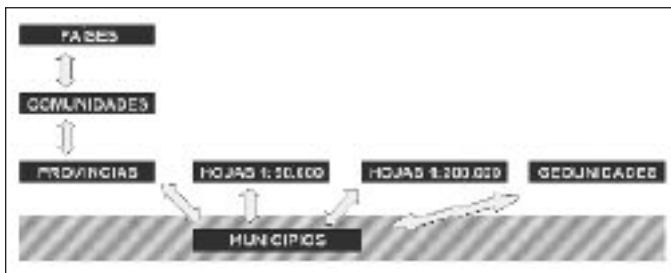


Fig. 4. Relaciones entre unidades espaciales: unidad básica
Fig. 4. Relations among spatial units: basic unit

- Tablas de unidades: siete tablas, una por tipo de unidad: países, comunidades, provincias, municipios, hojas 1:50.000, hojas 1:200.000 y geounidades, con dos campos: código de unidad y nombre de unidad.
- Tablas que relacionan tipos de unidades: siendo la unidad básica el municipio, a través de intersecciones de coberturas del resto de unidades con la cobertura de municipios, se obtienen tablas que relacionan el municipio con el resto de unidades. Este tipo de tablas tienen dos campos: código de municipio (unidad básica) y código de unidad (Fig. 3).
Para relacionar unidades administrativas entre sí no es necesario realizar este proceso y crear nuevas tablas, la relación se establece añadiendo en las "tablas de unidades" de unidades administrativas un campo con el código de la unidad administrativa inmediatamente superior que incluye a cada una.
- Tablas de documento-unidad: son siete tablas, una por cada tipo de unidad, cuyos campos son el código del documento y el código de la unidad espacial asociada.

Herramientas para la geocodificación

El proceso de consulta requiere que los documentos tengan un código espacial asociado. Por ello en primer lugar se deben geocodificar los documentos ya existentes en el fondo documental. Esto se lleva a cabo mediante un proceso automático de búsqueda de nombres de unidades en los títulos de los documentos mediante consultas sucesivas en lenguaje SQL. Este proceso aunque rápido, puede generar errores, que deben ser revisados y corregidos manualmente. Para esto se realiza una aplicación que permita visualizar automáticamente título de documento y sus unidades espaciales asociadas, para así corregir los códigos relacionados si fuese necesario. Mediante esta aplicación, desarrollada para el SID del IGME con las herramientas de POO de Visual Basic 6, se automatiza el proceso de revisión, posibilitando grandes velocidades de corrección.

También todo nuevo documento que llegue a un fondo documental debe ser geocodificado, algo que se consigue mediante una aplicación desarrollada también con Visual Basic 6 que hace que la codificación espacial sea una parte ineludible del proceso de dar de alta un nuevo documento.

Para evitar redundancias la geocodificación debe seguir una serie de criterios, así un documento sólo puede ser asociado a un tipo de unidad, la menor

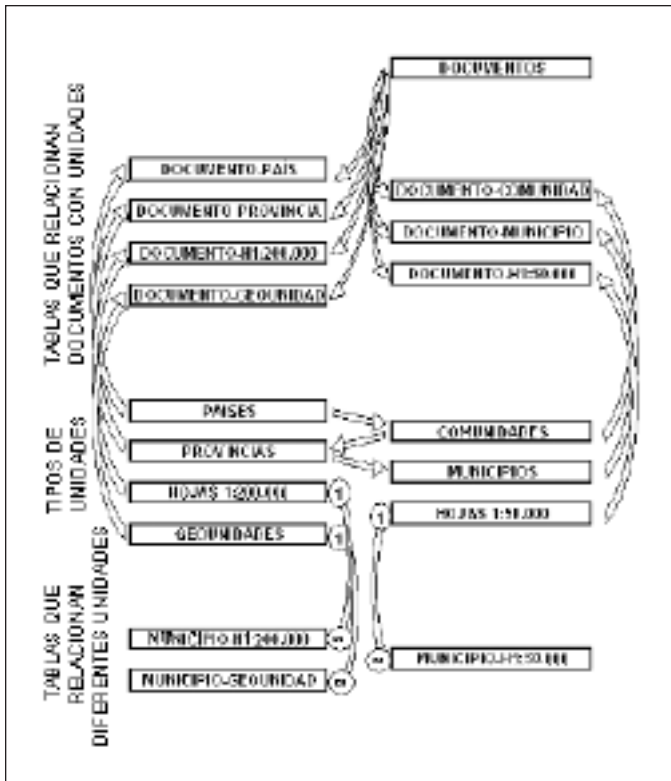


Fig. 5. Tipos de tablas y relaciones en una base de datos para sostener la geocodificación en el IGME
 Fig. 5. Kind of tables and relations in a database to maintain the geocodification in IGME

posible, para ganar resolución espacial. Aunque se puede asociar a muchas unidades dentro de un mismo tipo.

Aplicación de la geocodificación: proceso de consulta

Como fase final del diseño del sistema comienza el análisis y desarrollo del proceso de consulta, es decir, añadir a la consulta temática existente, la búsqueda mediante criterios espaciales. El resultado debe ser una consulta única, que integre la consulta temática y la consulta espacial. Las búsquedas espaciales deben ser flexibles, permitiendo al usuario optimizarlas según sus necesidades.

Las relaciones implícitas en la base de datos harán que una consulta espacial, realizada a través del nombre de alguna unidad concreta (de unos códigos), muestre los documentos asociados a la unidad espacial seleccionada, y también los documentos asociados a otras unidades relacionadas con la seleccionada.

Conclusiones

Las búsquedas temático-espaciales, útiles para sistemas documentales de muchas ciencias con significación espacial, pueden ser integradas en una única consulta.

Mediante un sistema de geocodificación de documentos, rediseñando una base de datos, pero sin alterar su estructura inicial, y desarrollando una serie de aplicaciones y objetos, cualquier fondo bibliográfico (como sucede con el SID del IGME), puede implementar búsquedas de tipo espacial con unas velocidades de respuesta óptimas para un medio como Internet.

Referencias

Bosque Sendra, J. *Sistemas de Información Geográfica*. Rialp, 1992, Madrid.
 Date, C.J. *Bases de datos. Una guía práctica*. Sistemas Técnicos de Edición, 1987, México.
 Moldes, F.J. *Tecnología de los sistemas de información geográfica*. Ra-ma, 1995, Madrid.
 Rigaux, P., Scholl, M. y Voisard, A. *Spatial Databases with application to GIS*. Morgan Kaufmann Publishers, 2002, San Francisco (USA).
 Rosenstein, M. *Estructuras de datos. Un enfoque práctico*. Anaya, 1990, Madrid.
 Toro, M., Consuegra, J.G^a. y Piattini, M. (coord). *Sistemas de información Geográfica. Una aproximación desde la Ingeniería del Software y las Bases de Datos*. Dintel, 2001, Madrid.

Recibido: diciembre 2005
 Aceptado: julio 2006