

Buscadores Geo-Espaciales. Una arquitectura posible.

José Eduardo Córcoles
Universidad de Castilla-La Mancha - España
corcoles@dsi.uclm.es

1. Introducción

Con el crecimiento experimentado en los últimos años de la Web se ha detectado una insuficiencia sustancial en los métodos disponibles para encontrar y manejar información. Los métodos tradicionales están limitados a la búsqueda o comparación de palabras clave. Esto no ofrece un soporte adecuado para realizar búsquedas sobre datos estructurados que permitan al usuario una mejor determinación de la información deseada. Por tanto, los usuarios obtienen mezclas no deseadas de información en su búsqueda en la Web [Córcoles, J.E. y González P.].

Para eliminar estas limitaciones se necesitan operaciones de más alto nivel que permitan comparar, consultar, analizar, combinar y/o integrar los datos de la Web con el fin de aplicar sobre ellos técnicas de discriminación y razonamiento más parecidas a la humana. En esta línea, la comunidad científica ha establecido los cimientos de lo que será el siguiente nivel de la Web, la Web Semántica [Berners-Lee, J. Hendler, J. and Lassila O].

Dentro de la Web Semántica, un dominio que requiere una especial atención es la semántica de la información geográfica. Actualmente, son pocos los motores de búsqueda (Google, MSN Search, Yahoo!, etc.) que interpreten que una página que hable sobre Albacete, también se está refiriendo, de una manera más general, a Castilla la Mancha o España. Los motores actuales únicamente examinan contenidos relevantes según unas palabras clave. Así, por ejemplo, un usuario que esté interesado, por ejemplo, en las reservas acuíferas que posee España, podría buscar en cualquiera de los buscadores "*reserva acuífera España*". Esta consulta, devolvería documentos que tuvieran las tres palabras, pero podrían existir documentos que hablaran de reservas acuíferas de algún lugar de España, sin la necesidad de tener la palabra clave España, por lo que estos documentos no se mostrarían como resultado de la búsqueda. Por otro lado, continuando con el ejemplo de la búsqueda

da anterior, podrían existir documentos que hablasen sobre lagos, ríos o manantiales de España y que no utilizaran las palabras clave *reserva acuifera*, por lo que estos documentos tampoco aparecerían como resultados de los buscadores actuales.

Existen otro tipo de buscadores que permiten realizar búsquedas con cierto criterio geográfico (Google Maps, Live Search Maps). En estos buscadores la palabra España del ejemplo anterior sí que abarcaría a todo el país. Pero estos buscadores se limitan al ámbito geográfico y no pueden realizar búsquedas sobre documentos. Un ejemplo de estos buscadores puede verse en la figura 1.

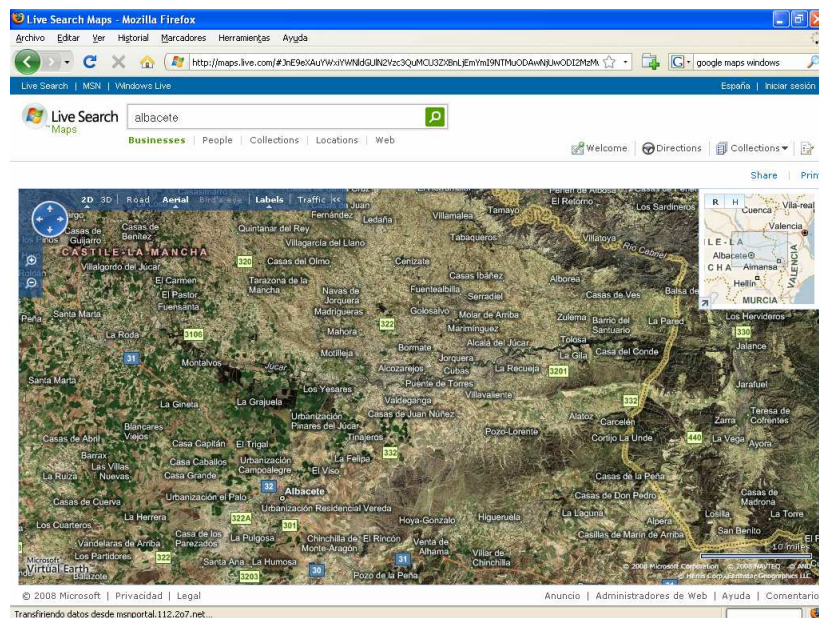


Figura 1. Aplicación Live Search Maps.

Por lo tanto, dentro de las búsquedas web existen retos que abordar y el ejemplo anterior es una muestra de la variedad de cuestiones que es necesario resolver para encontrar el conjunto de datos deseado dentro de un ámbito geográfico. También el ejemplo muestra como la carga para llevar a cabo las consultas recae enteramente sobre el usuario, lo cual puede producir, entre otras cosas, que haya información importante que se pierda en la consulta.

2. Diseño de la arquitectura global del sistema

En esta sección se describe la arquitectura propuesta para la implementación de un sistema que permita la búsqueda de información espacial basada en GSA de Google (Figura 2). Esta arquitectura utiliza los topónimos como elementos básicos de enlace de la información alfanumérica con la información espacial.

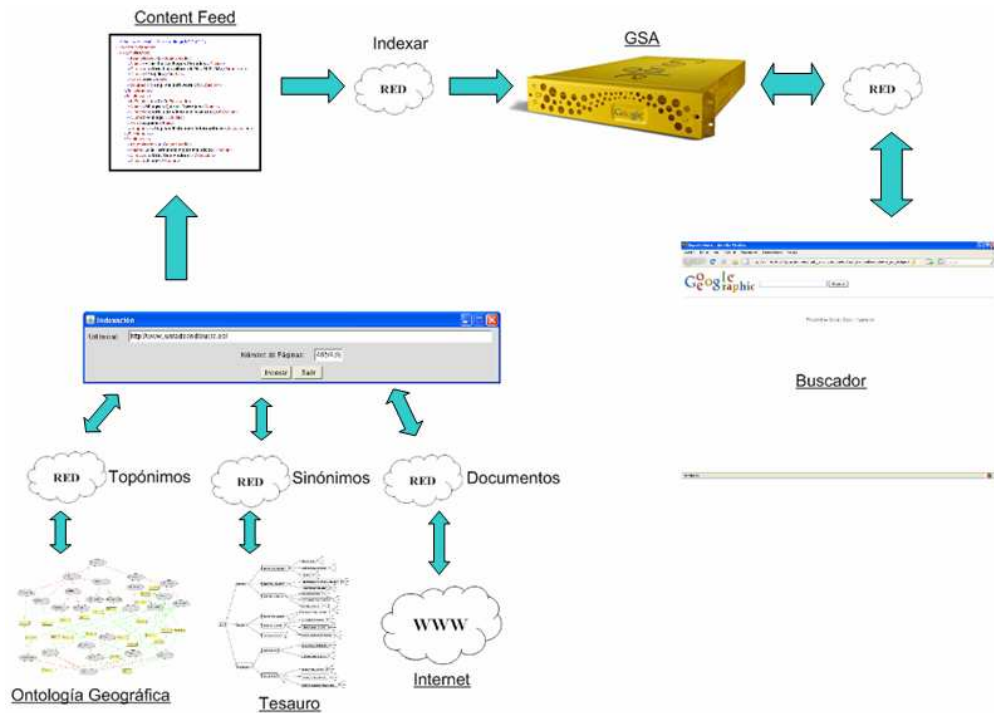


Figura 2. Arquitectura global de la aplicación.

El motor de búsqueda utilizado para indexar la información, tanto topográfica como semántica, ha sido Google Search Appliance (GSA). Se emplea el formato de resultados XML para poder personalizar la salida de los resultados, para ello, se debe aplicar una plantilla XSL, utilizando la transformación XSLT [W3C], al fichero que devuelve la GSA, con el fin de presentar los resultados al usuario en formato HTML. El funcionamiento de la aplicación se puede resumir así:

- Primero se rastrearán las páginas que se indexarán, para obtener la información que se desea ampliar semántica y topográficamente. Para este proceso se utiliza el rastreador WebSphinx.
- Una vez se tienen todas las páginas, se ampliará su información añadiendo sinónimos a aquellas palabras para las cuales el servicio web propor-

cionado por el tesoro GEMET sea capaz de encontrar términos semejantes.

- A continuación se obtiene la información topográfica de la página. Para esto se hace uso de la ontología geográfica GeoNames. Además, la información topográfica será ampliada gracias a los geo-microformatos en caso de que la página los utilice.
- Una vez se tiene toda la información necesaria, se crea el fichero *content feed* para la GSA [Google GSA].
- Una vez que ha sido creado el fichero *content feed* se debe enviar a la GSA, esto se hará a través de un formulario web, para permitir que el motor de búsqueda indexe dicho fichero y poder lanzar búsquedas sobre los nuevos datos.
- A continuación, una vez que se tienen los datos geográficos indexados en el motor de búsqueda ya se podrán lanzar consultas sobre dichos datos.
- Finalmente, mediante las consultas se obtendrá un fichero XML con los resultados. A este fichero habrá que aplicarle una plantilla XSL.

La figura 3 muestra la pantalla principal de la aplicación.

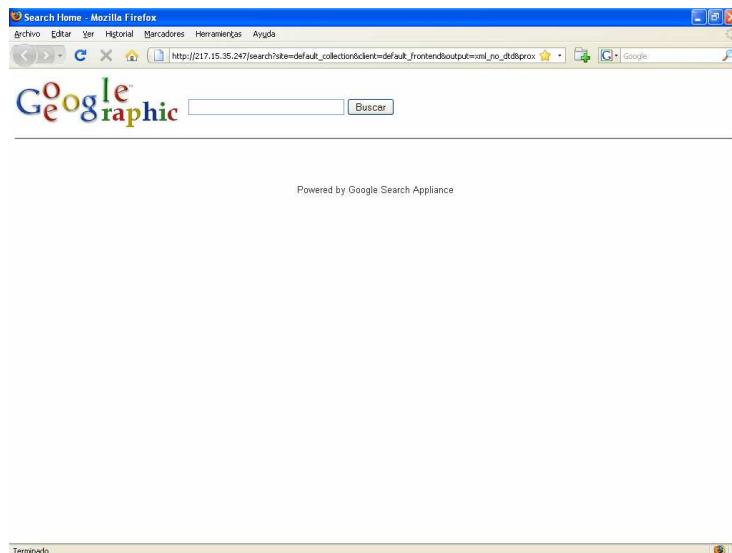


Figura 3. Página principal del buscador.

Una vez que el usuario se encuentra conectado a la página del buscador, ya podrá teclear la cadena de búsqueda deseada en el cuadro de texto que aparece en la página web. A continuación, se debe pulsar sobre el botón

Buscar o pulsar la tecla intro para dar comienzo la búsqueda. Para un ejemplo en el que se realiza una consulta con las palabras clave "denuncia" "ecológica" "Castilla la Mancha", el resultado se muestra en la siguiente figura 4.

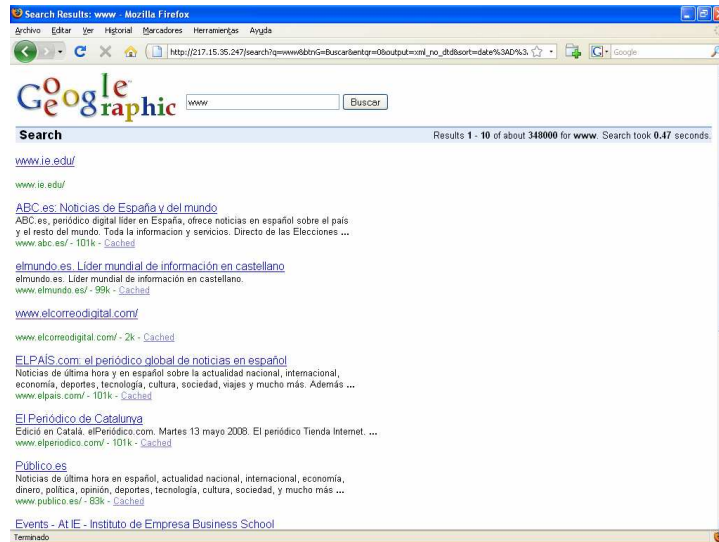


Figura 4. Página de resultados.

3. Bibliografía

Berners-Lee, J. Hendler, J. and Lassila O. "The Semantic Web". Scientific American, vol. 184, no. 5. 2001.

Córcoles, J.E. y González P. "Integrating gml resources and other web resources". 1st International Workshop on Geographic Information Management (GIM'04) in Conjunction with DEXA'04. Zaragoza. Spain. IEEE Computer Society Press (2004)

Google GSA. "Web con información corporativa de Google". WGoogleCorp. <http://www.google.es/intl/es/corporate/index.html>

W3C. XSL Transformation (XSL T). 1999. <http://www.w3c.es/>

SOCIEDAD DE LA INFORMACION

www.sociedadelainformacion.com

Edita:



Director: José Ángel Ruiz Felipe

Jefe de publicaciones: Antero Soria Luján

D.L.: AB 293-2001

ISSN: 1578-326x