

Solución de alto rendimiento para el geoportal de turismo de la Comunidad Valenciana.

J. Carrasco Marimón⁽¹⁾, S. Belentani Cid⁽²⁾ y J. Sanz Salinas⁽¹⁾

⁽¹⁾ Prodevelop SL, Plaza Don Juan de Villarrasa 14-5, 46001, Valencia, {{carrasco|jsanz}}@prodevelop.es

⁽²⁾ Agència Valenciana de Turisme, Avenida de Aragón 30, 46021, Valencia, belentani_ser@gva.es

RESUMEN

La Agència Valenciana de Turisme ha emprendido el proyecto de actualización de la web de promoción de la Comunidad Valenciana, el portal <http://www.comunitatvalenciana.com>. Este portal es una referencia internacional para todos los visitantes de la Comunidad Valenciana y por tanto un proyecto clave en su promoción turística. En esta ponencia se presentará la elaboración de un geoportal turístico atractivo y preparado para soportar un número elevado de visitas.

En el contexto del proyecto se han abordado problemas como la visualización de información vectorial (puntos de interés) en un elevado número mediante la técnica de agregación o clustering. Por otro lado dicha información vectorial se procesa de tal forma que el visitante de la web obtiene un rendimiento en los tiempos de respuesta bastante elevado gracias al uso de técnicas de multirresolución en el visor web. La información de partida se migra a una base de datos espacial libre y se procesa para generar archivos en formato JSON.

Por otro lado, el geoportal ofrece un flexible motor de búsquedas, preparado igualmente para soportar una carga elevada de peticiones mediante el uso de la indexación con el soporte para consultas espaciales. Este motor de búsquedas igualmente se ha preparado para utilizarse como servidor para ofrecer toda la información del portal a través de Layar, un servicio de realidad aumentada para móviles. Este servicio está completamente basado en componentes libres como el framework Spring o el soporte de búsquedas Lucene.

Se presentará por tanto cómo se ha abordado la realización de una solución completa de presentación de información de un portal turístico de exigencias de rendimiento elevadas, centrando la atención en los componentes del servidor, todos ellos basados en software libre.

Palabras clave: *turismo, geoportal, rendimiento, indexación, JSON, multirresolución, nivel de detalle, clustering*

EL PORTAL DE TURISMO DE LA COMUNIDAD VALENCIANA, PROBLEMÁTICA Y SITUACIÓN PREVIA

Introducción

Una de las principales actividades económicas de la Comunidad Valenciana es el turismo. Como todo destino de prestigio internacional, la Comunidad Valenciana dispone de un portal dedicado a la promoción de las muchas alternativas que los visitantes tienen a su disposición. El portal <http://www.comunitatvalenciana.com> tiene por tanto como objetivo principal difundir la amplia oferta de productos, recursos y destinos que la Comunidad Valenciana ofrece a los turistas.

Este portal está compuesto por múltiples secciones dedicadas a los diferentes tipos de ofertas y destinos como alojamientos, agenda de eventos, equipamientos deportivos, etc. Por supuesto, el portal está traducido a múltiples idiomas: español, inglés, valenciano, inglés, francés, alemán, chino y japonés. El portal además aloja también una red de blogs, una biblioteca de contenidos multimedia (fotografías y vídeos) aportados por los propios usuarios del portal y dispone de una versión móvil y algunas otras aplicaciones para otras temáticas concretas.

The screenshot shows the main page of the Valencia Tourism Portal. At the top, there is a navigation bar with links for 'Inicio', 'RSS', 'Conselleria de Turisme', 'Newsletter', 'Mòbil', 'Mapa web', and a search bar. Below this is a large banner with the Valencia logo and the slogan 'Te doy todo'. The main content area is divided into several sections: 'Actualidad' (News) featuring a photo of people on a boat and a headline about the '7ª Feria Internacional de Turismo de la Comunitat Valenciana'; 'Mapa turístico' (Tourist Map) showing a map of Europe with various countries labeled; 'El tiempo' (Weather) for Vinaròs, showing a minimum of 19°C and a maximum of 25°C; and social media links for Facebook, Twitter, YouTube, and RSS. There is also a section for 'Descubre la Comunitat Valenciana con Óscar' and 'Castellón de la Plana'.

Figura 1: *Página principal del portal turístico de la Comunidad Valenciana*

La práctica totalidad de los contenidos alojados en el portal refieren a realidades georreferenciables en mayor o menor medida. Es por esta razón que el portal dispone de funcionalidades de presentación de dichos contenidos sobre cartografía, utilizando Google Maps como proveedor de la misma en forma de una página principal para navegar por las diferentes temáticas así como un elemento de página (*widget*) que localiza el elemento descrito en el territorio.

Como es de prever, el portal es un sitio web de una complejidad considerable, y su impulsor, la Agència Valenciana de Turisme ha dedicado un gran esfuerzo a facilitar la navegabilidad y mantener un diseño y usabilidad adecuados.

Fruto de este continuo esfuerzo por su mejora, este artículo presenta el proyecto de desarrollo de una nueva versión del visor de los contenidos georreferenciados tanto propios del portal como externos así como toda una serie de productos derivados que hacen los contenidos del portal más accesibles a los usuarios a través de otros medios además del propio portal, especialmente en dispositivos móviles.

Versión previa del geoportal

La versión existente del geoportal está basada en Google Maps pero su nivel de funcionalidad era reducido y adolecía especialmente de problemas de visualización a escalas pequeñas. Esto se debía a que no se estaba aplicando ningún tipo de filtro sobre los contenidos en función de la escala y era fácil para el usuario de la web encontrarse con un número excesivo de elementos que restaban mucha legibilidad al mapa además de ofrecer unos tiempos de respuesta exagerados.

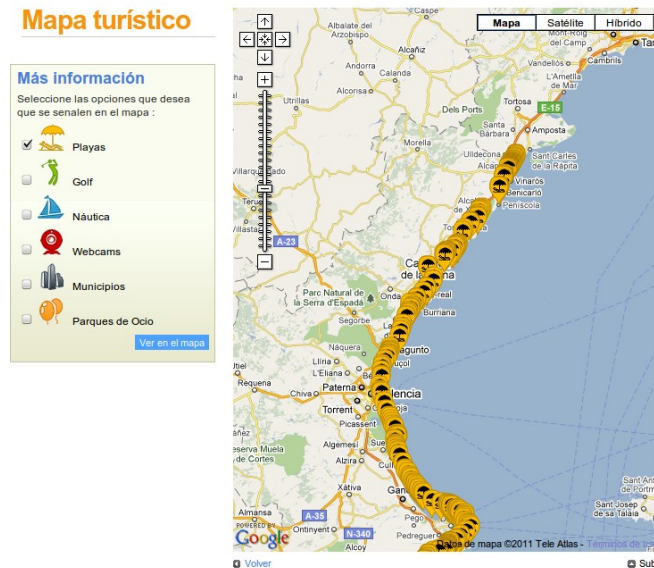


Figura 2: Saturación de elementos de tipo Playa en el visor

REQUISITOS Y CARACTERÍSTICAS DE LA SOLUCIÓN

El proyecto debía reunir los siguientes requisitos, a grandes rasgos:

- Mejora sustancial de la experiencia de usuario
- Claro enfoque hacia el mayor rendimiento posible para soportar una carga de usuarios elevada
- Debía seguir basándose en Google Maps como tecnología base de visualización de la información geográfica
- Permitir la realización de búsquedas por texto libre en la interfaz
- Desarrollo de interfaces de usuario de la información del geoportal para dispositivos móviles, incluyendo una capa en el servicio de realidad aumentada Layar
- El desarrollo no debe penalizar el rendimiento del sitio actual
- El desarrollo debe enfocarse al mayor desacoplamiento del portal actual ya que está ya prevista la migración del mismo

FUNCIONALIDADES

A continuación se detallan, desde un punto de vista funcional, los diferentes productos desarrollados que se pondrán a disposición de los usuarios del portal. De este modo, antes de entrar a ver en detalle la solución técnica que se ha abordado.

Geoportal

El principal producto del desarrollo es desde luego el geoportal. Este se muestra a los usuarios como una página del portal accesible desde la portada del sitio web. Basado en la tecnología de visualización Google Maps aparece como un visor en el

que el usuario tiene a su disposición la posibilidad de activar y desactivar diferentes «capas» temáticas tanto de contenidos del portal como también de contenidos externos publicados en diferentes redes sociales. Actualmente las redes sociales soportadas son Twitter¹, Minube², Foursquare³ y Flickr⁴. Es posible además realizar búsquedas por cualquier texto sin salir de la página, realizándose dichas búsquedas tanto en los contenidos del portal como también en las redes sociales soportadas.

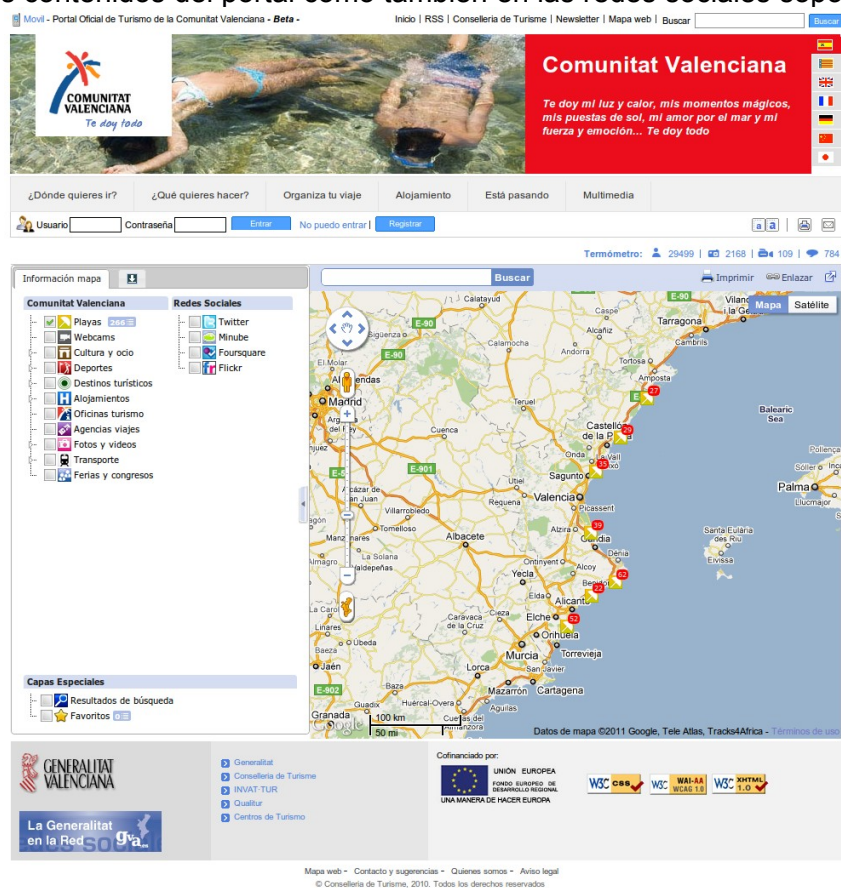


Figura 3: Interfaz del geoportal con la capa de playas activada, nótese la diferencia con la figura anterior

El portal permite calcular rutas entre puntos de interés o cualquier punto del mapa, utilizando para ello el servicio que Google Maps ofrece. Es posible igualmente imprimir en cualquier momento la cartografía visualizada así como las búsquedas o las rutas.

Como se puede ver en la figura anterior, en lugar de mostrar todos los puntos en el mapa, el usuario ve cómo los elementos se «agrupan» de forma que para ver en detalle dicho grupo de acercarse, manteniendo por tanto un número de elementos visualizados manejable. Igualmente, al cargar la capa se muestra junto a su nombre el número de elementos de dicha capa que se encuentran en pantalla. Este texto es clicable, mostrando un listado paginado con información de detalle. Cada elemento de estos listados puede clicarse, haciendo automáticamente zoom a la ubicación del mismo. Este elemento y el listado asociado además son sensibles a la navegación, de forma que se actualizan al desplazarse por el mapa.

1<http://twitter.com>

2<http://www.minube.com/>

3<http://foursquare.com/>

4<http://flickr.com>

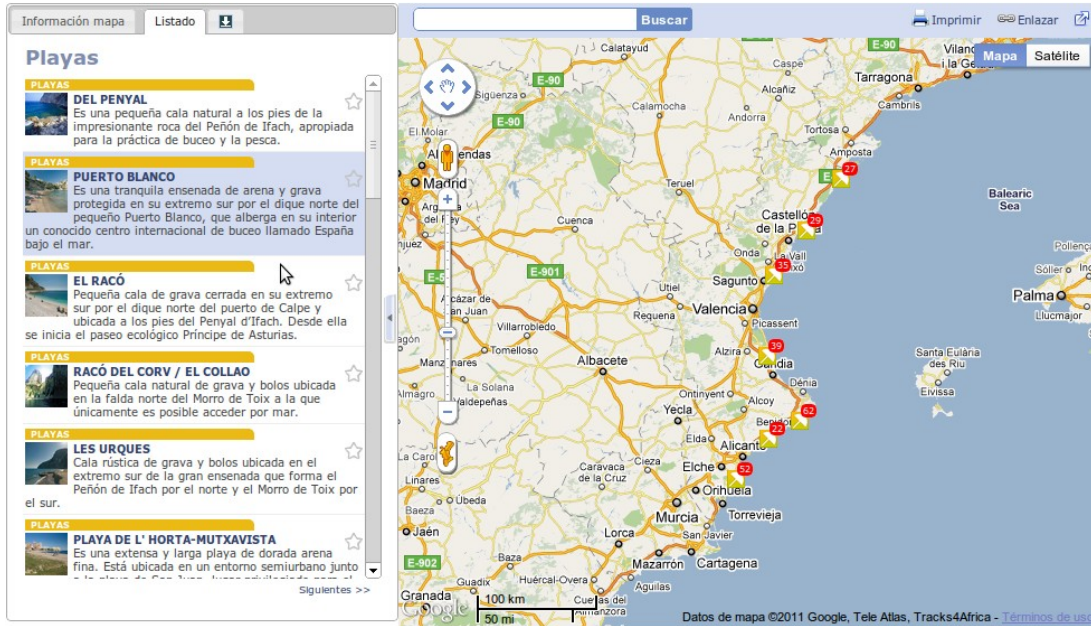


Figura 4: Listado de elementos visualizados

Al hacer clic sobre alguno de los elementos, tanto si es un contenido del portal como si es un contenido externo se muestra una ventana emergente (*pop-up*) que permite realizar diferentes acciones relativas al elemento como navegar a su ficha de detalle o a su web, usarlo en el cálculo de rutas, etc. Además se muestra un código QR⁵ que permite navegar a este elemento mediante un teléfono móvil como más adelante se detallará.



Figura 5: Búsqueda con contenidos del portal y externos (fotografías en flickr) así como la ventana emergente con información ampliada de uno de los elementos

Finalmente, el geoportal ofrece al usuario la posibilidad de descargar en su equipo contenidos en diferentes formatos para poder consumirlos fuera del propio portal. Se ofrece por un lado un archivo KML que permite navegar por los puntos de interés utilizando la misma técnica de visualización adaptada a la escala de visualización y por otro lado un archivo para utilizar en



Figura 6: Recursos de descarga para cada capa temática

⁵http://es.wikipedia.org/wiki/C%C3%B3digo_QR

navegadores GPS TomTom de modo que los usuarios pueden añadir a sus navegadores la información temática actualizada.

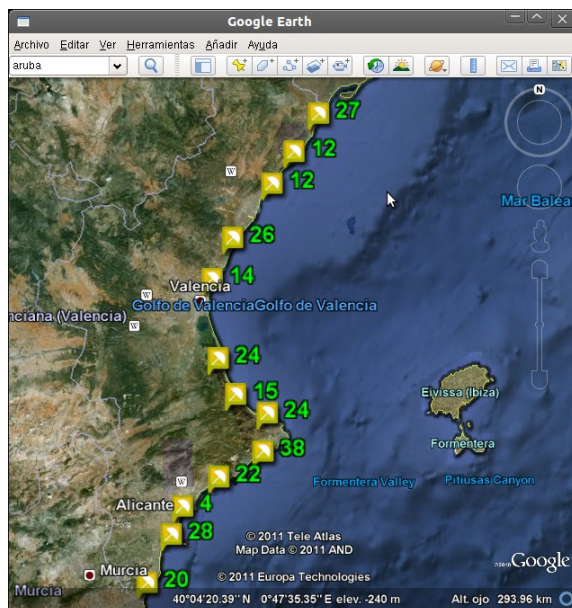


Figura 7: Visualización de la capa de playas en Google Earth aplicando el agrupamiento y visualización en función de la escala

Finalmente el visor del geoportal ofrece otras funcionalidades como son:

- Visualización a pantalla completa, con y sin el panel lateral de capas
- Generar un enlace permanente al contenido visualizado (permalink)
- Gestión de favoritos basado en cookies del navegador
- Búsqueda de todos los elementos que se encuentran alrededor de alguno de los puntos seleccionados indicando un radio
- Posibilidad de compartir un elemento de contenido en redes sociales
- Posibilidad de integrar el buscador del geoportal en los motores de búsqueda del navegador del usuario

Dispositivos móviles

Con la expansión en la adopción por parte de los usuarios de telefonía móvil de teléfonos con mayores funcionalidades y conexión a Internet (los llamados smartphones) la posibilidad de ofrecer contenidos de calidad a estos usuarios mediante interfaces de usuario específicas se ha hecho una realidad.

Así, se han desarrollado las siguientes funcionalidades:

- El código QR del visor web dirige al usuario a una página que permite navegar con la aplicación Google Maps del móvil a la ubicación del punto, mostrando además información de detalle en el propio móvil mediante una interfaz web adaptada a este dispositivo. Esta funcionalidad está disponible para teléfonos Android y iPhone
- Aplicaciones nativas para dispositivos Android y iPhone (disponibles en sus respectivas tiendas de aplicaciones) para navegar por las capas de contenidos del portal. Estas aplicaciones acceden a los contenidos en el portal por lo que siempre muestran contenidos actualizados.
- Se ha desarrollado una capa para el servicio de realidad aumentada en dispositivos móviles Layar.



Figura 8: Acceso en el móvil a la información de un punto y su localización en el mapa desde un código QR

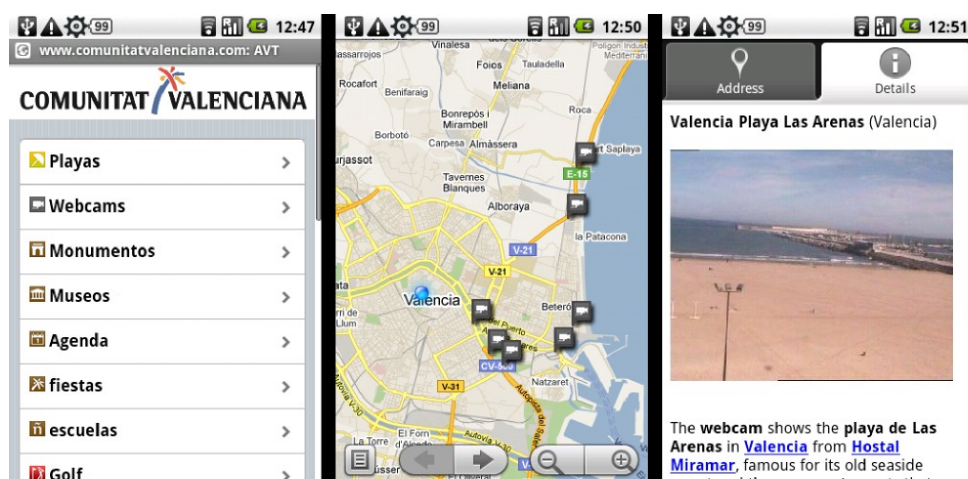


Figura 9: Aplicación nativa para dispositivos Android que muestra una capa temática (webcams) así como la información asociada a uno de los elementos



Figura 10: Visualización de puntos de interés cercanos en un entorno de realidad aumentada

ARQUITECTURA GENERAL DE LA SOLUCIÓN

Arquitectura previa

La arquitectura previa al inicio del proyecto se muestra en la figura siguiente. Básicamente se dispone de un balanceador de carga y caché de páginas web que redirige las peticiones cuando es necesario (invalidación de la caché, contenidos nuevos, etc.) a alguno de los nodos del cluster de servidores web que ejecutan el CMS Drupal con su propia caché. Todos los nodos del cluster comparten el mismo sistema de ficheros alojado en una máquina separada y atacan a la misma base de datos. Sobre esta base de datos hay configurada una base de datos replicada en tiempo real que hace las veces de copia de seguridad. Además del propio portal, como se ha comentado anteriormente existen otras aplicaciones secundarias que también se encuentran detrás del servidor de caché.

Esta arquitectura permite por un lado servir páginas web a los visitantes anónimos de forma ágil ya que en su gran mayoría se encuentran cacheadas y a la vez se acelera y soporta un elevado número de usuarios utilizando una granja de nodos del CMS, pese a que todos comparten una única base de datos. La arquitectura resultante intenta por tanto reducir el número de accesos a la base de datos al mínimo imprescindible para que el sitio responda con tiempos razonables.

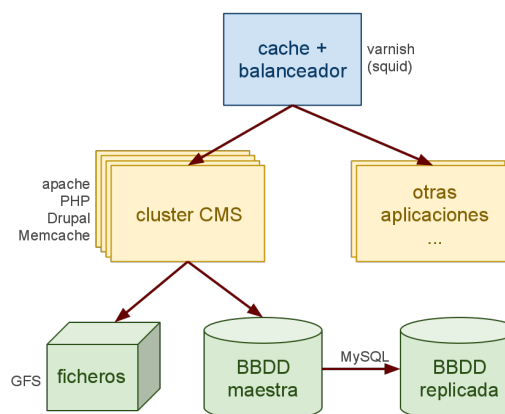


Figura 11: Arquitectura existente

Arquitectura del geoportal

Para cumplir con los requisitos del proyecto se planteó una configuración en la que se realiza un proceso de extracción y transformación periódica de la información almacenada en la base de datos replicada, migrando la información a una base de datos PostgreSQL/PostGIS para una vez actualizada la base de datos del geoportal realizar los procesos necesarios para regenerar tanto los archivos estáticos que van a dar soporte al mismo como los índices de la aplicación web JEE (Java Enterprise Edition) que responderá a diferentes tipos de peticiones de información, búsquedas, etc.

Es decir, a partir de la información de la base de datos replicada se realiza un proceso ETL que actualiza la base de datos geoespacial. A continuación por un lado una aplicación java J2SE regenera los recursos estáticos que servirán la cartografía del geoportal, así como los recursos de descarga y por otro lado se reiniciarán los índices de la aplicación JEE que responderá a diferentes tipos de peticiones.

La integración con el portal actual se realiza añadiendo un pequeño módulo en el portal que dispone del código mínimo para inicializar el geoportal, estando por tanto la mayor parte de los contenidos tanto de lógica JavaScript como las imágenes, iconos y demás recursos en la máquina que alberga el geoportal.

Por supuesto todos estos contenidos se encuentran detrás del servidor con el acelerador web, por lo que se puede ajustar al máximo la política de qué recursos se deben cachear en el acelerador (archivos JavaScript, imágenes, etc.) y cuáles deben solicitarse al servidor (servidor JEE).

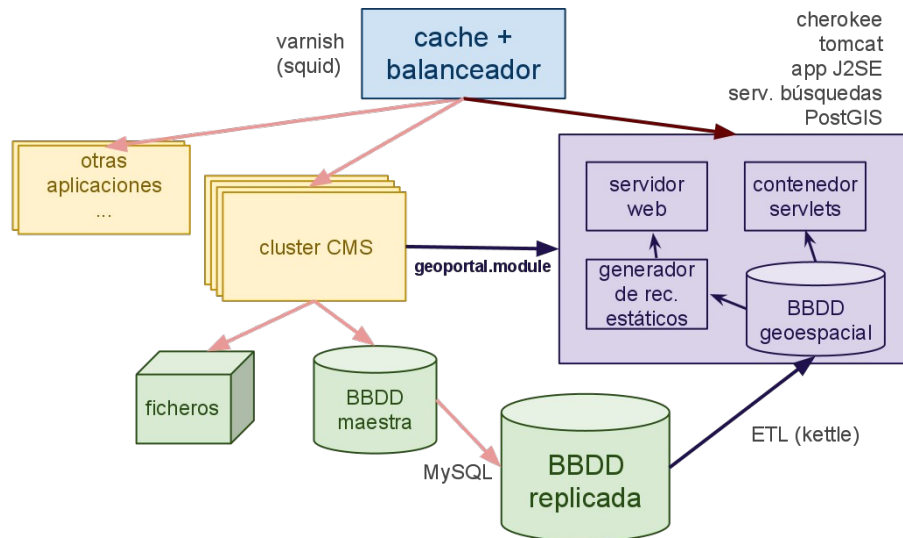


Figura 12: El geoportal integrado en la web

Esta arquitectura asegura por un lado la máxima independencia del portal actual, evitando además penalizar en ningún modo el rendimiento del mismo y por otro lado unos tiempos de respuesta elevados gracias a las técnicas de pregeneración de recursos y de indexación de contenidos. El geoportal no hace llamadas a la base de datos salvo al regenerarse los contenidos por lo que se elimina este punto como posible cuello de botella en entornos de un elevado número de usuarios.

INTERFAZ WEB DEL GEOPORTAL

Las principales tecnologías implicadas en la interfaz web del geoportal son:

- **Google Maps**⁶ como herramienta de visualización de la información geográfica. Se ha utilizado la reciente versión 3 de su API, permitiendo la visualización del geoportal en dispositivos móviles (independientemente de las aplicaciones específicas desarrolladas) y por su mejor rendimiento general frente a la versión anterior de la API.
- **JQuery**⁷ como biblioteca de manipulación de eventos, peticiones AJAX, etc
- **JQuery-UI**⁸ como biblioteca para la construcción de los elementos de la interfaz, pestañas, árbol de elementos, etc
- **Yahoo Query Language**⁹ para realizar peticiones a aquellos servicios de redes sociales que no permitían utilizar el protocolo JSONP, evitando así la necesidad de un proxy en nuestro servidor.

El portal conecta mediante peticiones AJAX no sólo con el conjunto de información generada a partir del contenido de la web sino también con varias redes sociales que se han considerado que aportan contenido de interés bien en forma de comentarios geolocalizados (twitter), en forma de fotografía (flickr) y finalmente de descripción de lugares de interés de todo tipo (minube, foursquare).

La interfaz se ha desarrollado en su mayor parte mediante código JavaScript, así como algunas hojas de estilo (CSS) y por supuesto toda una amplia colección de imágenes, incluyendo una iconografía diseñada para el geoportal. Se ha desarrollado un pequeño módulo de Drupal que genera una nueva dirección en el portal que carga el HTML necesario para realizar toda la carga del código de la interfaz del geoportal.

6<http://maps.google.es/support/bin/answer.py?answer=7060&topic=10778>

7<http://jquery.com/>

8<http://jqueryui.com/>

9<http://developer.yahoo.com/yql/>

SERVIDOR

Base de datos y ETL

Diariamente se ejecuta un componente en el servidor que realiza la extracción, transformación y carga (Extract, Transform, Load). El componente es un proyecto de software libre llamado Kettle, que ofrece no solo un componente ejecutable como aplicación de consola (J2SE) sino también un interfaz para diseñar los flujos de trabajo en una aplicación de escritorio y un componente JEE para desplegar los flujos en un servidor de aplicaciones. Kettle¹⁰ es un componente de la suite de software libre para lo que se conoce como Business Intelligence llamada Pentaho¹¹ que incluye otros proyectos para la realización de informes, minería de datos, etc.

Fruto del proceso ETL se dispone de un modelo de datos completo en la base de datos geoespacial, con una tabla por cada capa temática adaptando la información en dicha tabla a cada una de las capas. Estas tablas se unen formando una vista homogénea de información que más tarde será consultada por el servidor de búsquedas.

Generación de recursos estáticos

Una vez se ha actualizado la base de datos se arranca el proceso de generación de recursos estáticos. Este componente se ha desarrollado por completo usando software libre, en concreto el componente depende de:

- **Java Topology Suite**¹² para la realización de las operaciones geométricas
- **JDBC** para el acceso a la base de datos, utilizando el driver de PostgreSQL/PostGIS
- **Apache Velocity** como motor de plantillas para la generación de archivos en diferentes formatos (JSON, KML, etc.)

El proceso más importante de esta fase es la generación de los archivos JSON (JavaScript Object Notation) para el geoportal. Estos archivos además de un archivo de metadatos permiten al cliente recibir clusters de información en lugar de recibir toda la información de una vez saturando la visualización al usuario. Al recibir un cluster el mapa únicamente pinta un símbolo e indica mediante un texto el número de elementos agrupados bajo ese punto. Para poder «explotar» ese punto el usuario debe descender, solicitándose un nivel más de detalle pudiendo recibir todos los puntos separados o bien si se supera una cierta densidad se volverá a recibir un cluster.

Se ha utilizado un algoritmo sencillo *bottom-top*, agrupando pares de nodos. En cada iteración se agrupan los dos más próximos. Antes de empezar, se ordenan aleatoriamente los nodos. Para generar los tiles, se parte del nodo raíz y se expande hasta la distancia que genere la separación deseada en un tile de 256 pixels.

Es decir, se implementa una especie de «tileado vectorial» que a menor escala agrega una mayor cantidad de puntos asegurando una densidad de puntos adecuada en el mapa. Cada cluster dispone de la información mínima para poder mostrar la ventana emergente con nombre y fotografía, así como un identificador del punto de interés.

Además de los recursos para visualizar en el geoportal, aplicando el mismo procedimiento se generan archivos KML para Google Earth. Estos archivos, uno por cada capa, enlazan con los sucesivos niveles de detalle automáticamente, de forma que el usuario al acercarse en el globo va descargando archivos KML. Este procedimiento, en la terminología de Google Earth se conoce como Super-Overlays¹³. Finalmente, también a partir de cada capa temática se generan archivos con toda la información contenida en un único fichero, tanto en formato KML como en formato

10<http://kettle.pentaho.com/>

11<http://www.pentaho.com/>

12<http://www.vividsolutions.com/jts/JTSHome.htm>

OV2. El primero será consumido por las aplicaciones móviles y el segundo se ofrecerá a los usuarios para su libre descarga y consumo en sus receptores GPS.

Servicio de búsquedas

Uno de los requisitos del proyecto es disponer de un servicio de búsquedas. La aplicación desarrollada se ejecuta en un contenedor de servlets, es por tanto una aplicación java para web (JEE). Las tecnologías empleadas para el desarrollo de la aplicación, de nuevo son todas libres:

- **Spring**¹⁴ como framework principal que proporciona funcionalidades de alto nivel para el acceso a datos de forma transaccional y mapeo a vistas mediante el paradigma MVC (Modelo-Vista-Controlador).
- **Apache Velocity**¹⁵ como motor de plantillas para la generación de salidas en diversos formatos, principalmente JSON (JavaScript Object Notation) pero también en KML y RSS.
- **Apache Lucene**¹⁶ como motor de búsquedas e indexación. Este componente permite construir índices de forma eficiente y ofrece también una completa interfaz de aplicación para realizar todo tipo de consultas, filtrados y ordenaciones de los datos almacenados en el índice, incluyendo búsquedas de texto libre.
- **Lucene spatial**¹⁷ como el componente que amplía Lucene para dar soporte al indexado de información espacial, permitiendo generar filtros basados en distancia. Estos filtros serán necesarios para atender las peticiones Layar así como peticiones de cercanía a un punto de interés en el geoportal.

El servicio es usado ampliamente a lo largo de toda la aplicación, ya que ofrece diversos modos de consulta que son aprovechados por la interfaz para recabar de forma dinámica información contextual a la ubicación de visualización gracias a la búsqueda mediante punto y radio, información de un punto concreto a partir de un identificador, generación del KML a partir del código QR para dispositivos móviles, etc.

El mismo componente atiende también a las peticiones Layar, cumpliendo el mínimo de su API documentada para poder dar funcionalidad de navegación en realidad aumentada, así como la configuración del rango de visualización alrededor del usuario y la posibilidad de buscar por un cierto texto.

Dado que este es el único componente que implementa lógica en el servidor (recordemos que el resto son recursos estáticos) se han realizado pruebas de rendimiento del servicio utilizando JMeter¹⁸ como banco de pruebas para lanzar peticiones aleatorias geográficamente de forma intensiva. Se muestra una prueba realizada a una hora sin demasiado tráfico en el portal de forma que se acerca bastante al posible rendimiento de la aplicación. Se muestran diferentes iteraciones de peticiones de búsqueda por punto y radio, donde tanto las coordenadas geográficas como el radio de búsqueda son aleatorios.

13http://code.google.com/intl/es/apis/kml/documentation/kml_21tutorial.html#superoverlays

14<http://www.springsource.org/about>

15<http://velocity.apache.org/>

16<http://lucene.apache.org/java/docs/index.html>

17http://lucene.apache.org/java/3_0_3/api/contrib-spatial/index.html

18<http://jakarta.apache.org/jmeter/>

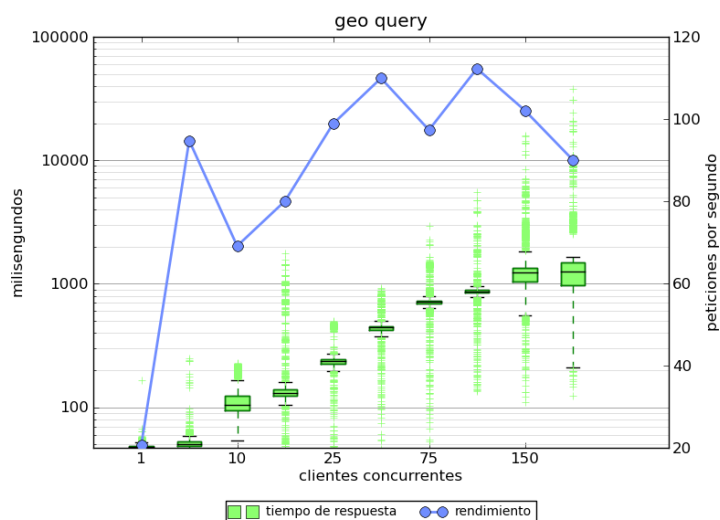


Figura 13: Ejemplo de ejecución de un plan de pruebas

Como se puede apreciar, el rendimiento general del servidor es aceptable con un número de usuarios concurrentes alrededor del centenar, siendo en general todas las peticiones atendidas en menos de un segundo.

CONCLUSIONES

Se ha presentado una completa solución a un proyecto ambicioso y de gran visibilidad. La Agència Valenciana de Turisme cuenta con un geoportal para distribuir los contenidos georreferenciados de su portal con una interfaz moderna y con un número elevado de funcionalidades, adaptado a las necesidades de un ambiente en el que son importantes tanto los contenidos propios como los producidos por la comunidad de usuarios a través de redes sociales.

El uso de tecnologías libres para el tratamiento previo de la información para ofrecer contenidos estáticos o bien indexados de forma que evitemos en lo posible exponer el acceso a la base de datos ha demostrado ser un acierto, ofreciendo el geoportal unos tiempos de respuesta al usuario francamente satisfactorios.

La arquitectura propuesta podría sin lugar a dudas optimizarse ampliando el número de nodos del geoportal (atacando todos a la misma base de datos, por ejemplo) y realizando un mayor afinamiento de todos los componentes del sistema. Hay todavía por tanto mucho margen para la mejora del rendimiento de los diferentes componentes tanto del servidor de búsqueda como del servidor web que da acceso a los recursos estáticos si el portal aumentara el tráfico en los próximos años por lo que se puede concluir que la AVT dispone de un geoportal que podrá crecer con su sitio web.