

EST 2016, 19° Concurso de Trabajos Estudiantiles

Geolocalización de instituciones educativas basada en datos abiertos

Mariano Emanuel Alejandro López¹, Pablo Ariel Silva¹, Alberto Agustín Meza¹, Nicolás Maximiliano Ledesma¹, Fernando Pedro Hernán Nuñez¹, Yanina Medida²

¹m_villa@hotmail.com, pabloariel01@hotmail.com, agus2k5@hotmail.com, ledesma.nicolas.maximiliano@hotmail.com, fer3061@gmail.com,

²yaninamedinaa@gmail.com, Profesor Adjunto, Departamento de Informática, Cátedra Taller de Programación I, Licenciatura en Sistemas de Información, Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, Argentina

Resumen. En el marco de una hackathon open data se desarrolló una aplicación de geolocalización de instituciones educativas basada en datos abiertos, denominada Mapa Educativo que permite ubicar y mostrar las escuelas comprendidas en un área seleccionada dentro de la provincia de Corrientes y brindar información sobre la misma. Para lograr el objetivo se utilizaron métodos y herramientas dados principalmente en la asignatura Taller de Programación I, orientada a la programación web, de la carrera Licenciatura en Sistemas de la Facultad de Ciencias Exactas y Naturales y de Agrimensura para la Universidad Nacional del Nordeste.

Palabras clave: Geolocalización, Datos Abiertos, Programación web MVC

1 Introducción

— Los datos abiertos son datos que pueden ser utilizados, re-utilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera de la que aparece [1].

La iniciativa de gobierno abierto fue impulsada durante la presidencia de Barack Obama, a inicios del 2009, con la finalidad de permitir a los ciudadanos acceder a datos generados por el gobierno. Esta modalidad se convirtió en una tendencia a nivel mundial. Argentina se unió en 2012 a la alianza mundial de gobierno abierto, con actividad escasa, pero se reforzó a inicios del corriente año con nuevas políticas gubernamentales, lo que estimuló la realización de múltiples campañas de gobierno abierto a nivel nacional [2].

Un ejemplo de datos abiertos en nuestro país se puede encontrar en la ciudad de Buenos Aires, que brinda la posibilidad de acceder a los mismos desde su página web oficial y en distintos formatos. Están disponibles además en el mismo sitio una serie de aplicaciones móviles que promueven la participación ciudadana [3]. Otra opción la ofrece la plataforma web y móvil de Barrios Activos que per-

mite participar e informarse acerca de la ubicación de distintos tipos de reportes hechos por los ciudadanos [4].

En la ciudad de Corrientes se organizó una Hackathon de Datos Abiertos [5] para obtener ideas y propuestas de los ciudadanos, principalmente de los sectores vinculados con las tecnologías de información (TI), incluyendo a estudiantes de diferentes niveles educativos, con el fin de desarrollar aplicaciones que aprovechen la disponibilidad de los datos y poner a disposición de los ciudadanos información útil y fácilmente accesible.

En el presente trabajo se describe la propuesta de explotación de los datos proporcionados por los Ministerios de Educación y de Planificación y Coordinación de la provincia de Corrientes, mediante una aplicación web que permite la geolocalización de las distintas instituciones educativas de la provincia, brindando información de las mismas.

2 Metodología

La aplicación se desarrolló cumpliendo las consignas del Hackathon: todos los proyectos debían ser dados de alta en la plataforma HackDash y enlazar al GitHub [6] correspondiente de cada equipo, disponible en este link: <https://github.com/agus2k5/opendata>. Se indicaron como criterios de evaluación de los trabajos los siguientes aspectos: originalidad de la idea, innovación del proyecto, utilización de framework, tener en cuenta aspectos de seguridad informática e interoperabilidad de la aplicación con otras APIs.

En función de estas directivas y de los datos disponibles se diseñó un sitio web denominado Mapa Educativo [7] en el cual se visualizan los establecimientos educativos comprendidos en un radio seleccionado, con la posibilidad de filtrar por tipo de dependencia de la institución (pública, privada subvencionada o privada no subvencionada). Una vez seleccionada la institución se muestra la información de la misma y es posible dejar un comentario accediendo por medio de una cuenta de Facebook.

Tratamiento de los datos:

Los datos abiertos puestos a disposición consistían en tres archivos con formato CSV, con información referida a los establecimientos educativos de la provincia de Corrientes, cursos, alumnos, tipo dependencia institucional, y su ubicación geográfica.

En primer lugar se realizó un análisis de los archivos para idear la aplicación, decidir qué datos utilizar, y de qué forma mostrarlos. Se optó por hacer uso de la georeferencia para ubicar a las instituciones en el mapa.

Para obtener la base de datos, se realizó un proceso de transformación de los archivos CSV al gestor MySQL mediante la herramienta Pentaho Kettle. Durante el tratamiento de los datos se realizaron diversas tareas, tales como, redefinición

de claves foráneas, normalización de campos de tipo Date, corrección de los campos latitud y longitud (los cuales fueron entregados intencionalmente de forma inversa para que los equipos se enfrentaran con cierta complejidad), entre otros.

Desarrollo de la aplicación

Una vez generada la base de datos, y superadas las dificultades de su limpieza y estructura, se procedió a la codificación de la aplicación, optando por el modelo vista controlador (MVC), y siguiendo con el mapeado a objetos de la base de datos, para lo que se utilizó el framework Hibernate ORM.

Luego se vinculó la aplicación con la API de Google para visualizar los datos en un mapa. Se hizo uso de la fórmula de Haversine para permitir al usuario seleccionar un radio en el mapa en el cual mostrar los establecimientos cercanos. A continuación se creó una API REST propia utilizando el patrón de diseño Criteria para alimentar el mapa únicamente con los datos comprendidos en el radio seleccionado.

Una vez funcionando la aplicación, se completó la información de los puntos del mapa que indican los establecimientos y se creó una vinculación con Facebook para permitir a los usuarios iniciar sesión y publicar un comentario sobre la institución educativa.

Paralelamente se trabajó con otras APIs: Google Places para mostrar sitios de interés (estaciones de policías, hospitales, bomberos, etc.), Google Charts para la generación del gráfico de torta de la distribución de escuelas según dependencia y Google Visualize para la creación de mapas de calor.

Para la aplicación se optó por no incorporar un sistema de login, ya que el objetivo era mostrar datos abiertos dispuestos por el gobierno. Pero si es posible realizar comentarios por medio de una cuenta de Facebook, permitiendo tanto identificar a los usuarios, personificar dichos comentarios y dar la posibilidad a las autoridades u otros de contactarse con el usuario en cuestión. Como funcionalidades adicionales, se pueden filtrar los comentarios, creando listas negras de palabras que no se desea que se muestren, evitando así entradas indeseadas u ofensivas.

Por otra parte, al utilizar Hibernate, se hace uso de HQL lo que dificulta el intento de realizar inyecciones SQL (método de infiltración de código intruso a través de consultas), a la vez que las mismas no van directamente a la base de datos sino a los objetos, por lo que la mayoría del código malicioso es bloqueado.

Spring también trata funcionalidades de seguridad mediante el módulo Spring Security, el cual incorpora características básicas implementadas por defecto, aportando una capa de seguridad extra a la aplicación.

Tecnologías Utilizadas

Para el desarrollo de la aplicación se utilizó una diversidad de tecnologías, entre las que se destacan:

Lenguajes de programación:

Java. La base de la aplicación fue programada en Java, utilizando el framework Spring Web MVC.

Javascript. Permitiendo que el sitio sea dinámico, el cual también permitió el uso de distintas APIs y mostrar información de los establecimientos sin necesidad de cargar otra página por medio de AJAX.

HQL. Se creó un modelo de objetos de la base de datos, donde se utilizó el framework Hibernate ORM, que posee su propio lenguaje de consulta: Hibernate Query Language.

JSTL. También se utilizó un lenguaje de marcado para sitios JSP, que tiene incorporadas funcionalidades básicas propias, lo que brinda la posibilidad de utilizar código java dentro de los documentos XML.

API's utilizadas:

Facebook. Se utilizó la API de Facebook para permitir a las personas conectarse con sus respectivas cuentas y realizar comentarios sobre los establecimientos, y que estos queden vinculados a los mismos.

Google. Se hizo uso de distintas librerías de la API de Google Maps. Se utilizó esta API para visualizar un mapa y permitir al usuario localizarse geográficamente para determinar las escuelas cercanas, como así también la posibilidad de solicitar a dicha API que ubique en el mapa una dirección solicitada por el usuario.

La librería Places. Para permitir al usuario localizar en el mapa sitios de interés, mencionados anteriormente.

También se incorporó la librería Visualize de Google, que permitió crear mapas de calor, resaltando los sitios donde la concentración de escuelas es mayor.

Además se utilizó Google Charts para graficar la cantidad y el porcentaje de establecimientos dentro del área seleccionada, teniendo la posibilidad de discriminarlos por dependencia.

JQuery. Se combinaron funciones de JQuery para acceder a objetos del DOM.

Materialize. También se introdujo el framework de CSS creado por Google, permitiendo que el sitio sea responsive, a la vez que se aplicaron estilos a la misma.

Servidores utilizados:

Apache Tomcat. Se desplegó la aplicación sobre Apache Tomcat como servidor JSP para poner en funcionamiento el sitio.

Base de datos:

MySQL. Como gestor de base de datos se implementó MySQL.

Otras Herramientas:

Pentaho Kettle. Para realizar el proceso ETL y generación de entidades se utilizó la herramienta Pentaho, permitiendo migrar los archivos CSV a la base de datos.

GitHub. Es un servicio comercial de alojamiento de repositorios Git "Sistema de Control de Versiones (SCV) distribuido" remotos.

3 Descripción de la aplicación

La aplicación fue desarrollada durante la realización de la Hackathon los días 5 y 6 de mayo de 2016, y se encuentra disponible en: <http://mapaeducacional-mapeunne.rhcloud.com/OpenData/>. La misma es de acceso público, en primera instancia la página solicita compartir nuestra ubicación con un mensaje como se observa en la Fig.1, con el fin de realizar el posicionamiento en el mapa.

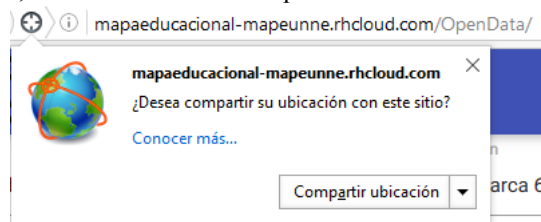


Fig. 1. - Compartir ubicación

La página también permite ingresar la ubicación de forma manual, para lo cual requiere; calle, altura, provincia y ciudad.

Una vez ubicada la posición en el mapa, se visualizan las instituciones educativas en un radio de hasta 4 km tomando como centro la ubicación ingresada.

En las Fig. 2 y 3 se observan los establecimientos educativos de Corrientes Capital y de la ciudad de Goya, con un radio de 2.4 km desde 9 de Julio 1449 y 1.8km desde Catamarca 653 respectivamente.

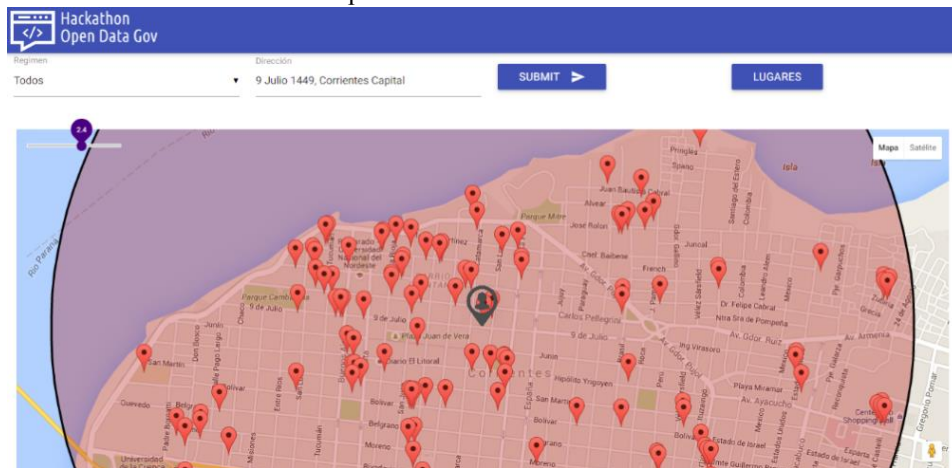


Fig. 2. - Instituciones educativas Corrientes Capital

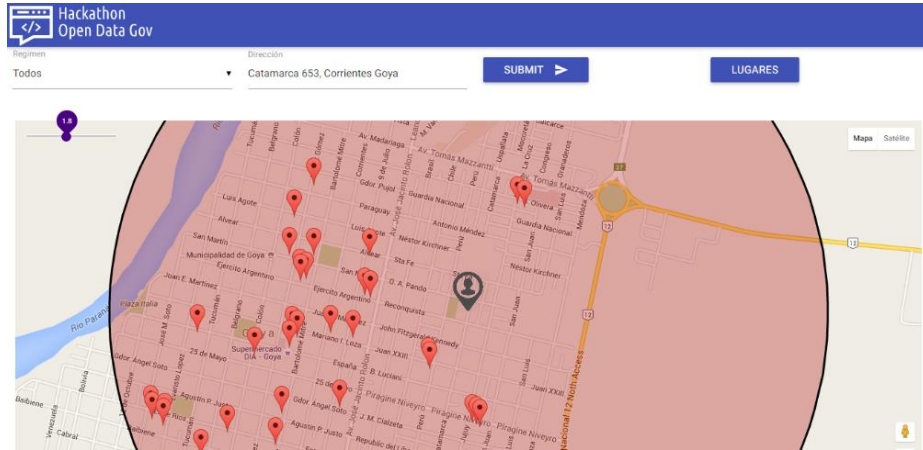


Fig. 3. - Instituciones educativas Corrientes Goya

La página también cuenta con la opción de seleccionar la dependencia de las instituciones para su filtrado (Todos, Público, Privado Subvencionado, Privado No Subvencionado) como se observa en la Fig.4.

Al pie de la página se encontrará un gráfico con la distribución de los establecimientos según la dependencia, como se observa en la Fig.5.

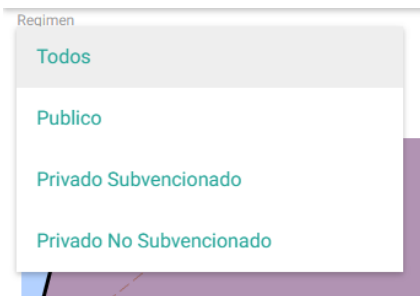


Fig. 4. - Dependencia de las instituciones

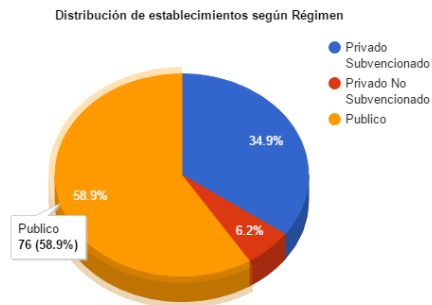


Fig. 5. - Distribución de establecimientos

Al hacer clic sobre alguna institución, se desplegará un menú con datos de la misma, tales como: Nombre, Cue-Anexo (código de identificación del establecimiento), Departamento, Localidad, Jurisdicción, Cursos (si los posee), y una opción para dejar comentarios a través de Facebook. Lo descripto se puede observar en la Fig.6.



Fig. 7. Menú de datos de la institución



Fig. 6. Comentar con Facebook

Al hacer clic en “Comentar”, se abrirá una nueva pestaña con los datos de la institución, y los comentarios que posee la misma, con la opción de dejar nuevos comentarios, el único requerimiento es poseer una cuenta de Facebook. En la Fig.7 se observa un ejemplo de publicar comentario para la Escuela Técnica “Juana Manso”.

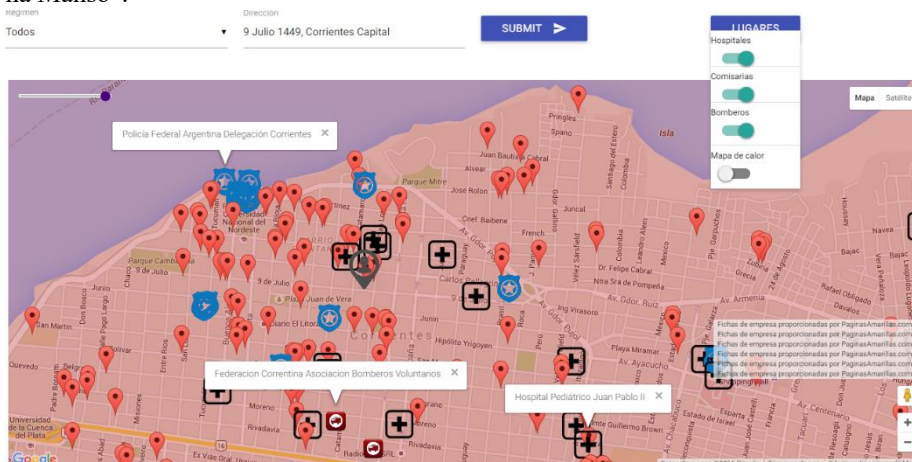


Fig. 8. - Hospitales, Comisarias, Bomberos

En la Fig. 8 se observa otra funcionalidad de la aplicación, que permitiría solapar las instituciones educativas junto a otras instituciones (hospitales, comisarias y bomberos).

4 Conclusiones y trabajos futuros

La elaboración de la aplicación en las condiciones impuestas por la competencia, constituyó una experiencia muy enriquecedora, al explotar los datos abiertos, y plantear e implementar la interoperabilidad de muchas tecnologías.

El desafío de manipular datos abiertos implicó el tratamiento de un gran volumen de datos reales, que presentaron diversas dificultades que fueron superados satisfactoriamente gracias al trabajo en equipo y utilización de distintas herramientas como ser Pentaho.

Durante el desarrollo del sitio web, surgió una problemática de tiempos de respuesta. Como se ha mencionado anteriormente, se utilizó Hibernate como ORM para la transformación de la base de datos relacional a objetos en tiempo de ejecución.

El alto volumen de datos, que no fue considerado a priori, afectó la performance, haciendo que una petición web de consulta de establecimientos cercanos a una coordenada geográfica demorara aproximadamente 3 minutos. Esta demora en gran parte era causada por el gran trabajo que debía realizar el motor de base de datos para la manipulación de esos volúmenes junto al trabajo del ORM. Ante esta problemática se optó por seguir utilizando el ORM dado que no se contaba con el tiempo necesario para reemplazarlo, pero esta decisión obligó a optimizar a las consultas a la base de datos y re-configurar el ORM, cambiando el tamaño del buffer y tamaño de las transacciones. Al finalizar las optimizaciones se logró reducir el tiempo de respuesta a 20 – 30 segundos. La experiencia dejó como lección aprendida que para el tratamiento de grandes volúmenes de datos se optará por utilizar soluciones orientadas a Big Data.

También cabe destacar que se profundizó la experiencia con la utilización de APIs, muchas de ellas utilizadas son del código del lado del cliente y son ejecutadas de forma asíncrona, por lo que se requiere tener cuidado con el flujo de interacción para que todas ellas funcionen sin problemas entre sí.

Como trabajo futuro se prevé la utilización de la API “Places” de Google, la cual permite solicitar información geográfica de otros establecimientos a Google. En base a esta información se podría localizar puntos que se encuentren exentos de establecimientos u otros sitios de interés dentro de un radio. También se desea seguir profundizando en la aplicación de las APIs sociales, para que los comentarios puedan ser publicados desde otras redes. El almacenamiento de esta información posibilitaría aplicar técnicas de minería de datos sobre los comentarios de los ciudadanos respecto de los establecimientos educativos, información que podría ser útil para la gestión gubernamental de las mismas.

Otra posibilidad es acceder a información censal sobre abandono escolar, y crear un mapa de calor con Google Visualice de la ubicación de hogares donde exista el caso, y superponerlo a las escuelas más cercanas, para evaluar si la lejanía podría ser un factor que propicia esta situación.

La herramienta tiene el potencial de ofrecer valiosa información para las autoridades gubernamentales vinculadas con la educación al geolocalizar las instituciones educativas e integrarlas con información del contexto para analizar distintas problemáticas sociales que afectan a los resultados educativos.

5 Referencias

1. «Open Data Handbook,» [En línea]. Available: <http://opendatahandbook.org/guide/es/what-is-open-data/>.
2. «Argentina.gob.ar,» [En línea]. Available: <http://www.argentina.gob.ar/mesasogp>.
3. «Buenos Aires Ciudad,» [En línea]. Available: <http://www.buenosaires.gob.ar/aplicacionesmoviles>.
4. «Barrios Activos,» [En línea]. Available: <http://barriosactivos.com/>.
5. «Ministerios de Coordinación y Planificación de Corrientes,» [En línea]. Available: <http://mcp.corrientes.gov.ar/noticia/apertura-de-la-maraton-gobierno-abierto-hackathon-2016?region=true>.
6. F. J. Lopez, R. B. Pellicer, J. N.-I. M A. Latre y F. J. Zarazaga-Soria, «GitHub como herramienta docente,» de *Jornadas de Enseñanza Universitaria de la Informática*, 2015.
7. «Mapa Educativo,» [En línea]. Available: <http://mapaeducacional-mapeunne.rhcloud.com/OpenData/>.