

Informe de Proyecto Final

Recuperación de información y recomendaciones en la
web

Grupo 14 - 2015
Martin Santagata
Ari Chamlian
Guillermo Kuster
Gabriel Barbatto
Federico Martinez

Índice

[Introducción](#)

[Obtención y procesamiento de datos](#)

[Diseño y Arquitectura](#)

[Desarrollo](#)

[Características de la aplicación desarrollada](#)

[La aplicación cuenta con las siguientes funcionalidades:](#)

[Conclusiones y trabajo a futuro](#)

Introducción

En el presente trabajo se describe el proceso de resolución del proyecto final de la asignatura Recuperación de información y recomendaciones de la web.

Durante los últimos años se dio un auge en el surgimiento de los llamados datos abiertos. El concepto datos abiertos (open data, en inglés) es una filosofía y práctica que persigue que determinados tipos de datos estén disponibles de forma libre para todo el mundo, sin restricciones de derechos de autor, de patentes o de otros mecanismos de control. Estos datos son ofrecidos por diversas instituciones, organismos, empresas y particulares para que otras personas puedan darles alguna utilidad.

En la actualidad los datos abiertos son utilizados para realizar aplicaciones de uso cotidiano como por ejemplo las ya conocidas aplicaciones de servicio de autobús, los cuales mediante geolocalización nos indican en qué parada estamos y qué ómnibus pasan por dicha parada, así como también dentro de qué tiempo pasarán.

Si bien los datos abiertos nos permiten infinitas posibilidades de manipulación, existe un problema relacionado con la disposición de la información. El creador de los datos abiertos no tiene restricción en la forma de disposición de los datos, por lo que los mismos pueden estar en cualquier formato (xml, rdf, json, csv, etc). Además de que puede ser accesibles por diversos medios como archivos, web services, etc. Esto claramente influye en la aplicación que estamos realizando, ya que la misma deberá ser capaz de manejar diferentes fuentes de datos.

El proyecto que planteamos intenta ser una aproximación a los problemas, así como los beneficios que plantea el trato con los datos abiertos para realizar aplicaciones de interés general.

Obtención y procesamiento de datos

Para la obtención de las fuentes de datos, investigamos en los sitios oficiales del estado en busca de datos abiertos de interés turístico, así también como en otros sitios de interés general que se dedican a este rubro.

Es así como hicimos un relevamiento de información en dichos sitios y los resultados no fueron muy auspiciantes, era muy poca la información abierta que nos era de utilidad para nuestro planteo.

Es así, como finalmente pudimos obtener tres bases distintas de utilidad para nuestros objetivos. En todos estos casos realizamos el filtrado correspondiente para quedarnos con aquellos datos que necesitábamos.

Un problema que se nos planteó, era la necesidad de contar con la ubicación por coordenadas (latitud y longitud) de dichos sitios de interés para poder ubicarlos en el mapa google. De las tres bases obtenidas, tan solo una de ellas contaba con dicha información y hasta en algunas entradas esta era incorrecta. El dato que proporcionan la totalidad de estas bases, es la dirección por calles de tales sitios, con lo que debimos idear la solución al problema de las coordenadas.

A tales motivos, buscamos alguna herramienta que tomando como insumos la dirección del sitio nos brindara las coordenadas, y es así como consultando dimos con una aplicación web desarrollada por El Correo Uruguayo que funciona muy bien.

URL: <http://www.correo.com.uy/otrosdocumentos/pdf/serviciosweb/BusquedaDireccion.html>

De esta forma, es que utilizando el web service del correo obtuvimos y corregimos las ubicaciones de los sitios listados.

Las bases obtenidas fueron:

- Hoteles habilitados:

La fuente es el sitio de información oficial de AGESIC “datos.gub.uy”, y allí encontramos el listado de hoteles habilitados por el Ministerio de Turismo, actualizada a junio del 2013 y en formato CSV. Al listado se le realizó el filtro por departamento, específicamente los ubicados en Montevideo. El archivo contaba con las coordenadas de los sitios, aunque en muchos casos con errores, por lo que se requirió realizar una depuración de los datos utilizando la aplicación del Correo anteriormente mencionada.

URL:

<https://catalogodatos.gub.uy/dataset/operadores-turisticos-habilitados/resource/c116a242-8d06-4dda-a1f1-2ef1a16835d1>

- Guías Turísticas:

La fuente al igual que los hoteles habilitados es el sitio oficial de AGESIC, allí encontramos el listado de guías turísticos habilitados por el Ministerio de Turismo, actualizada a junio del 2013 y en formato CSV.

Al listado se le realizó el filtro por departamento, específicamente los ubicados en Montevideo. Este archivo no contaba con las coordenadas de ubicación, por lo que fue necesario la utilización de la aplicación del Correo para obtener dicha información.

URL:

<https://catalogodatos.gub.uy/dataset/operadores-turisticos-habilitados/resource/5d465427-5ed2-4958-9b54-4f968178f94e>

- Museos:

La fuente de información de estos sitios es el portal web “museos.gub.uy” del Ministerio de Educación y Cultura, de donde se obtuvo el Registro Nacional de Museos, actualizada a octubre del 2014 y en formato CSV.

Al listado se le realizó el filtro por departamento, específicamente los ubicados en Montevideo. Este archivo no contaba con las coordenadas de ubicación, por lo que fue necesario la utilización de la aplicación del Correo para obtener dicha información.

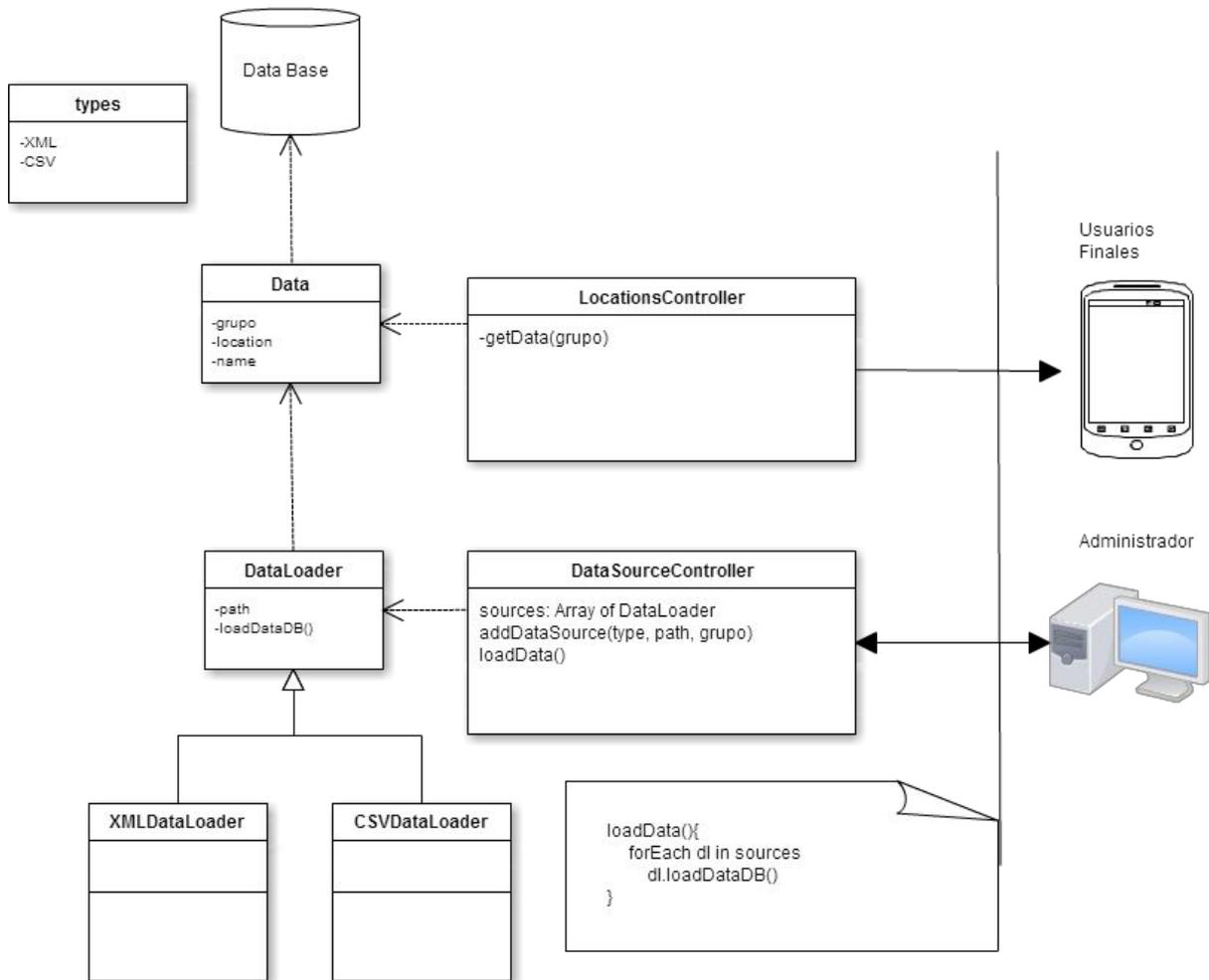
URL:

<http://www.museos.gub.uy/images/stories/doc/rnmoctubre2014.xls>

Finalmente, al obtener estos datos abiertos y lograr recopilar la información necesaria para utilizarlos según nuestros requerimientos, es que se normalizaron y se filtraron según la información que publicaríamos en la aplicación por medio de la construcción de las tablas de datos.

Diseño y Arquitectura

Presentamos a continuación el diseño junto con la arquitectura propuesta. Se tendrán dos interfaces, la interfaz móvil para uso de usuarios, y la interfaz web para uso del administrador. Los datos se persisten en una base de datos del servidor.



Desarrollo

La idea del proyecto entonces será obtener datos abiertos de diversas fuentes. Luego de obtenidos, se realizará una normalización de los mismos para poder tratarlos de manera homogénea en lo posible. Se procederá a su carga en una base de datos local y luego se realizara algún tipo de procesamiento con los mismos.

Para instanciar el problema, nos centraremos en obtener ubicaciones de lugares de diferente clase, ya sea puntos de interés cultural, hoteles, alojamientos, etc, para luego mostrarlos en un mapa de google.

Se buscará trabajar con los siguientes formatos de archivos: XML, CSV. Además se puede consumir alguna API web para obtener más datos. Una de las características que proponemos para la solución es que sea fácilmente extensible, de manera de poder agregar cualquier tipo de datos en un futuro. Además se desea poder agregar nuevas fuentes de datos en forma dinámica.

A continuación se listan las herramientas a utilizar para la implementación de la aplicación:

Para el backend, se utilizará Ruby on Rails. Ruby on Rails es un framework web open source escrito en el lenguaje de programación Ruby, orientado a objetos, ideado para programar de forma rápida y ágil siguiendo el paradigma de la arquitectura Modelo Vista Controlador (MVC). Rails se distribuye a través de [RubyGems](#), que es el formato oficial de paquete y canal de distribución de bibliotecas y aplicaciones Ruby.

Para el frontend, se utilizará el framework ionic para la realización de aplicaciones móviles. Está construido sobre apache cordova, una herramienta que permite el desarrollo de aplicaciones móviles mediante el uso de webviews. El programador desarrolla la aplicación utilizando sus conocimientos de desarrollo web, y la herramienta genera el código nativo necesario para correr la aplicación tanto en android como en iOS.

El frontend utiliza el framework javascript Angular el cual utiliza un patrón MVC. Esto permite modelar la aplicación front end de manera más ordenada y principalmente es muy fácil vincular la vista con los modelos con pocas líneas de código. Además se utilizó el la api de google maps de javascript para incluir el mapa donde se mostrarán los distintos puntos de interés.

Características de la aplicación desarrollada

Dado el creciente aumento del turismo, tanto a nivel local como internacional, el grupo decidió desarrollar una aplicación sobre este rubro.

También cada vez va en aumento el uso de smartphones y en particular de aplicaciones para los mismos, por lo que se decidió la realización de una aplicación móvil para Android.

Nos basamos solamente en la ciudad de Montevideo, en una primera etapa, la cual podría ser extensible a todo el país en etapas posteriores.

La aplicación va a permitir la visualización mediante mapas de Google, como ya fue mencionado, de puntos de interés turístico dentro de la ciudad. Los utilizados fueron museos, hoteles y guías turísticos.

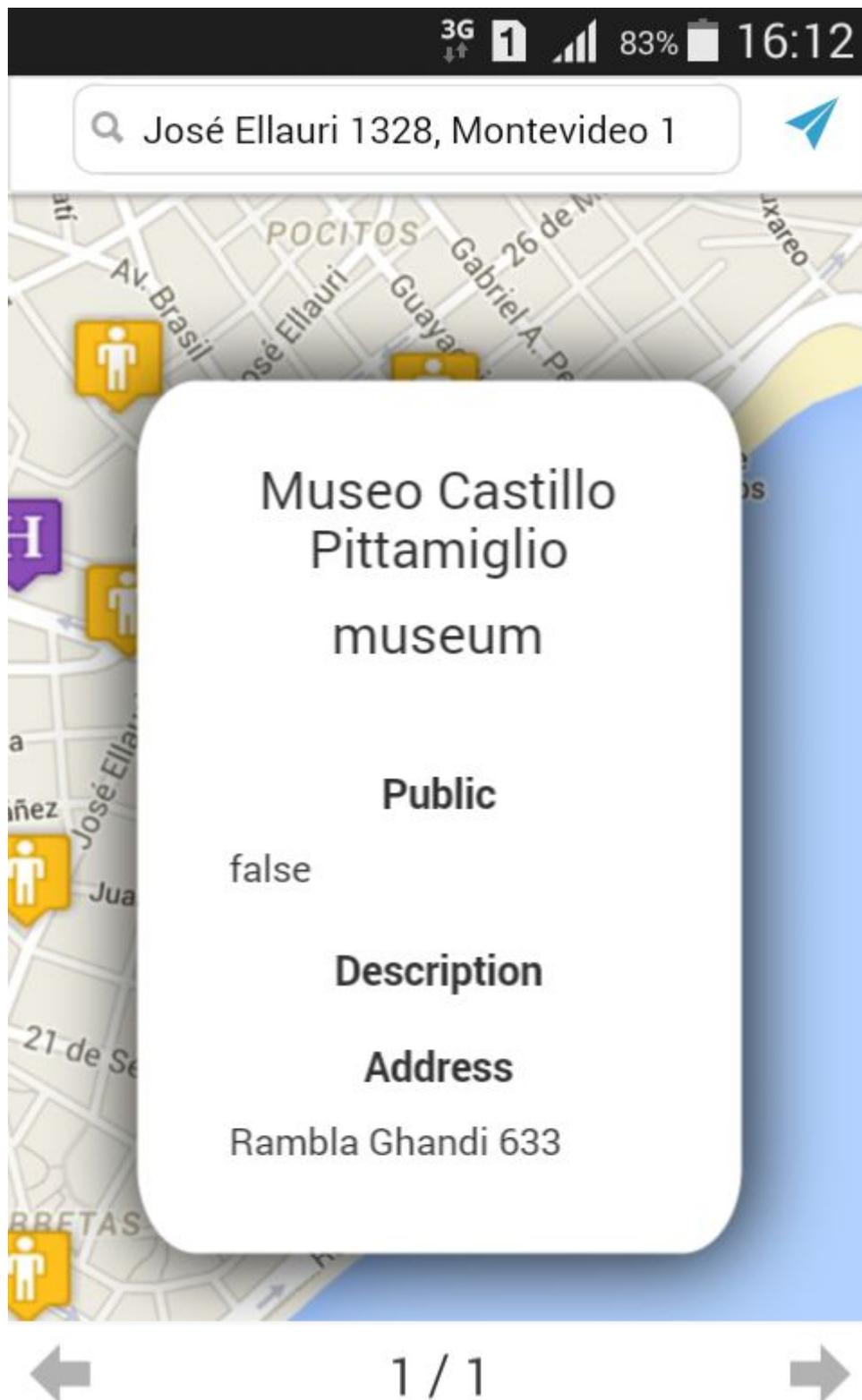
Al seleccionar el punto de interés deseado, se despliega un cuadro con información sobre dicho lugar, por ejemplo, teléfono, dirección, correo electrónico, entre otros tipos de datos. Es importante notar que no todas las fuentes de datos cuentan con la misma cantidad de datos.

La aplicación cuenta con las siguientes funcionalidades:

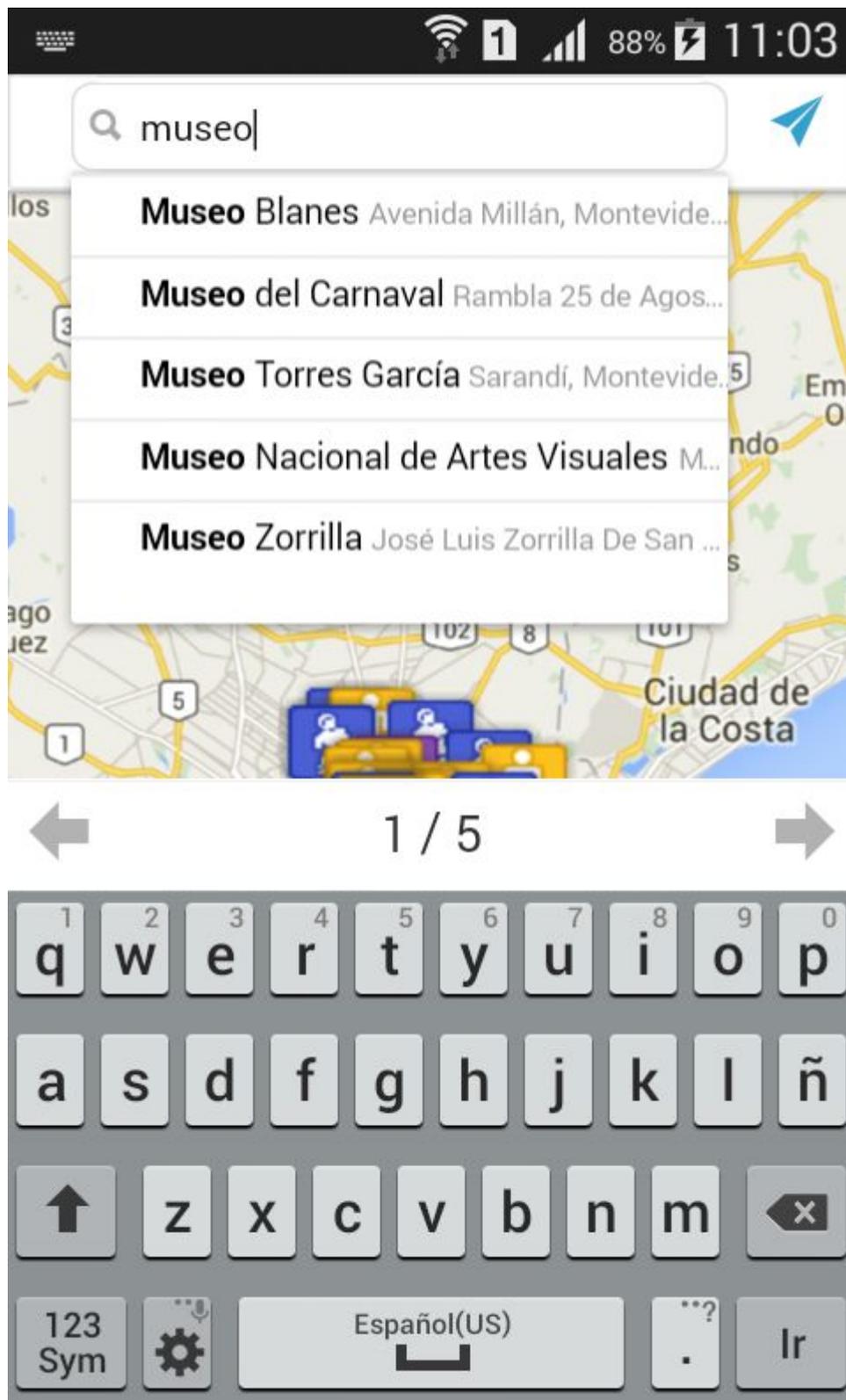
- Se muestran los lugares en el radio en el que se visualiza el mapa. Se realiza paginado de acuerdo a la cantidad de resultados obtenidos.



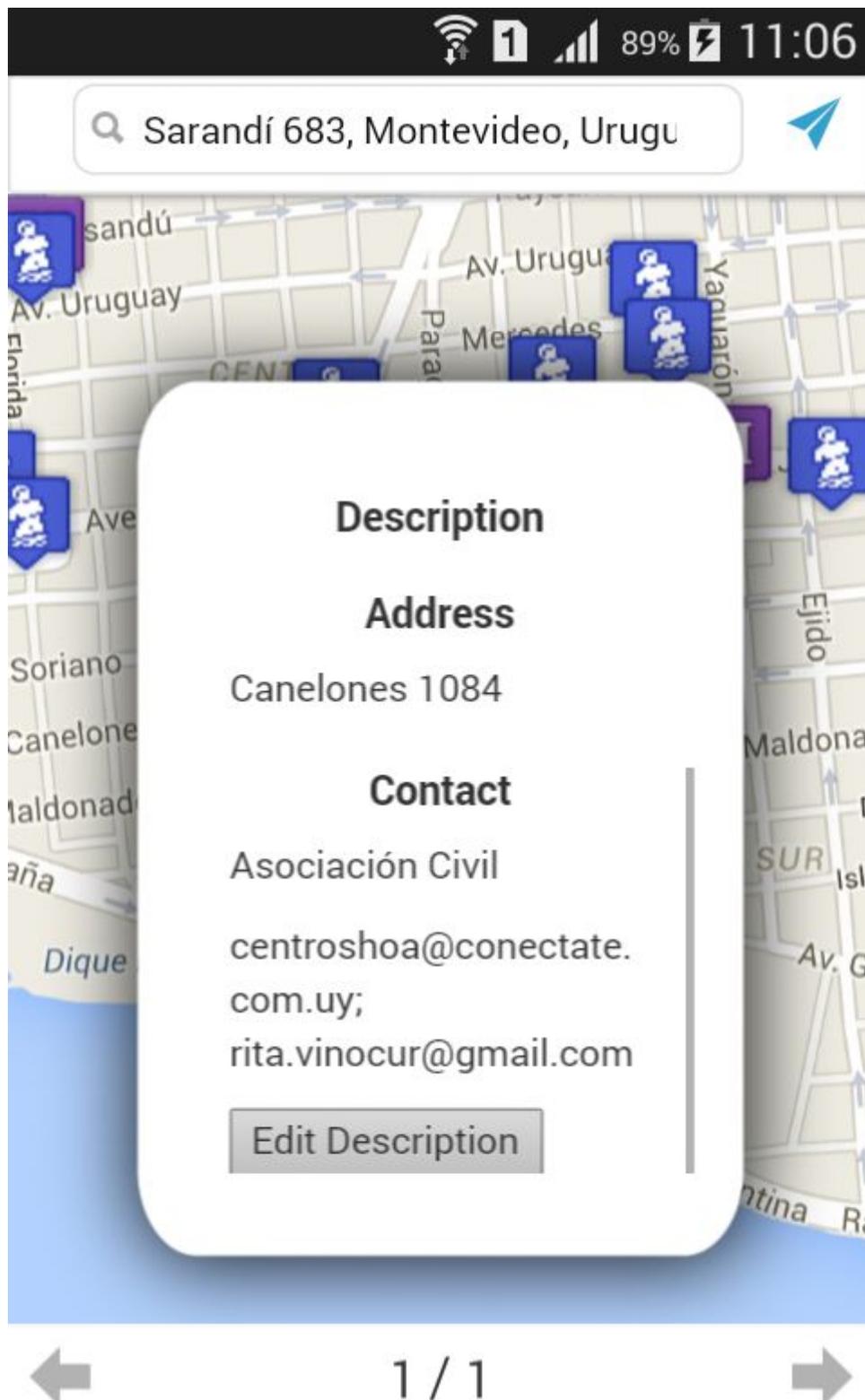
- Al seleccionar el lugar de preferencia, se despliega información del mismo



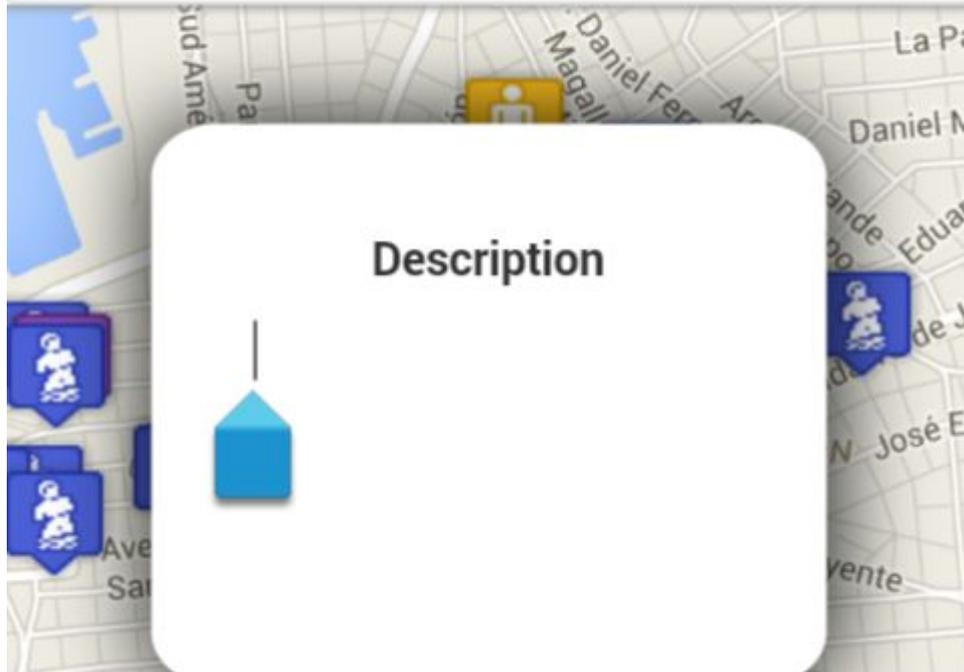
- Se pueden realizar búsquedas por calles o por nombre del lugar de interés.



- Se permite la edición en los datos de cualquiera de los lugares indicados en el mapa.



Search bar with magnifying glass icon and a blue paper airplane icon to the right.



1 / 2



Conclusiones y trabajo a futuro

Se realizó una aplicación dedicada al turismo, la cual para el alcance del proyecto es limitada en cuanto a las funcionalidades que podría brindar (por ejemplo agregar una imagen).

Por el diseño de la misma, sin embargo, sería fácilmente extensible en cuanto a tipo de datos con los que contar y desplegar, así como implementación de nuevas funcionalidades. Además se podría desarrollar para otros sistemas operativos, por ejemplo, iOS o Windows Phone.

El único impedimento para la adición de nuevos datos es la disponibilidad de datos abiertos sobre los mismos y en caso de que se disponga de los mismos, puede ocurrir que no estén actualizados al año corriente y no dar información del todo correcta.

Además, en el caso de contar con los datos, la mayor complicación se da por el lado de la normalización de los mismos, y del tratamiento de ciertas palabras que contienen caracteres que pueden causar problemas a la hora de su procesamiento (palabras con tildes o con letra ñ).

También es importante notar que se puede causar problemas con los formatos en los que estén los archivos (hablando de csv, xls yxlsx), ya que por ejemplo, se usan versiones un poco más antiguas y más nuevas para abrir ciertos archivos puede generar la incorrecta lectura de los mismos. Para clarificar lo descrito, suponga que se quiere abrir un archivo .docx en Microsoft Office 2003 en vez de abrirlo en Office 2013.

Finalmente, se puede afirmar que la herramienta, cuenta con un gran potencial a futuro, en caso de que el proyecto se quisiera extender o mejorar.