



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería en Informática



TFM del Máster en Ingeniería
Informática

Sistema para el registro y análisis
de trayectorias semánticas



Presentado por David Moreno del Hoyo
en Universidad de Burgos — 10 de febrero de 2017

Tutores: Bruno Baruque Zanón y
Santiago Porras Alfonso



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería en Informática



D. Bruno Baruque Zanón, profesor del departamento de , .

Expone:

Que el alumno D. David Moreno del Hoyo, con DNI 71305821-W, ha realizado el Trabajo final de Máster en Ingeniería Informática titulado Sistema para el registro y análisis de trayectorias semánticas.

Y que dicho trabajo ha sido realizado por el alumno bajo la dirección del que suscribe, en virtud de lo cual se autoriza su presentación y defensa.

En Burgos, 10 de febrero de 2017

Vº. Bº. del Tutor:

Vº. Bº. del co-tutor:

Bruno Baruque Zanón

Santiago Porras Alfonso

Resumen

En este primer apartado se hace una **breve** presentación del tema que se aborda en el proyecto.

Descriptores

Palabras separadas por comas que identifiquen el contenido del proyecto Ej: servidor web, buscador de vuelos, android . . .

Abstract

A **brief** presentation of the topic addressed in the project.

Keywords

keywords separated by commas.

Índice general

Índice general	III
Índice de figuras	IV
Introducción	1
1.1. Estructura del documento	1
Objetivos del proyecto	3
2.1. Objetivos generales	3
2.2. Objetivos técnicos	5
2.3. Objetivos personales	6
Conceptos teóricos	7
3.1. Limpieza de datos	7
3.2. Enriquecimiento semántico	8
3.3. Anotación de trayectorias	8
3.4. Consultas por anotaciones	8
3.5. Descubrimiento de conocimiento	8
Técnicas y herramientas	9
Aspectos relevantes del desarrollo del proyecto	12
Estado del arte	13
Conclusiones y Líneas de trabajo futuras	14
7.1. Conclusiones	14
7.2. Líneas de trabajo futuras	14
Bibliografía	15

Índice de figuras

2.1. Puntos de Interés señalados en el mapa	4
2.2. Ruta semántica sobre París	4

Introducción

Realizar una visita turística a una localidad desconocida puede resultar todo un reto. Ya sea por desconocer cómo es el lugar, no saber el tiempo que implica recorrer la distancia entre los puntos de interés, como por no desenvolverse con soltura en la lengua hablada en dicho lugar o simplemente por la falta de tiempo a la hora de preparar la visita.

Realizando una sencilla búsqueda por Internet se podrían localizar los lugares más importantes de la localidad elegida para ser visitada. Si se invirtiesen unos minutos más, se podrían localizar distintos museos, monumentos o plazas de interés. Añadiendo un tiempo adicional, se podrán localizar restaurantes, bares, etc. que pudiesen estar en lugares cercanos a los sitios turísticos seleccionados. Y si se dedicasen unos instantes más, se podría calcular una ruta adecuada para poder visitar todos esos lugares en el tiempo que la persona, cabe llamarla turista, puede dedicar a la localidad elegida anteriormente.

Aunque parece un espacio de tiempo relativamente corto, la planificación de una visita a un lugar desconocido puede llevar una cantidad de tiempo elevada. En la actualidad, invertir un tiempo elevado en esta tarea puede resultar complicado y, por eso, el presente Trabajo de Fin de Máster plantea una solución adecuada a este problema tan común.

La aplicación planteada en este trabajo trata de evitar la necesaria búsqueda de lugares de interés plantando una ruta adecuada al tiempo que puede dedicar al destino turístico seleccionado.

1.1. Estructura del documento

La estructura del presente documento será la siguiente:

Memoria

En la parte correspondiente a la memoria se encontrarán las siguientes secciones:

1. **Introducción:** en el primer apartado de la memoria se describe el contexto en el que el proyecto ha sido llevado a cabo.
2. **Objetivos del proyecto:** el siguiente apartado detalla los objetivos marcados a la hora de realizar este proyecto. Se distingue entre objetivos generales, técnicos y personales.
3. **Conceptos teóricos:** a continuación, se realiza una descripción de los conceptos necesarios para la realización del proyecto.
4. **Técnicas y herramientas:** en este apartado se detallan las herramientas usadas para la realización de este proyecto.
5. **Aspectos relevantes del desarrollo del proyecto:** en este punto se detallan los aspectos más importantes o interesantes del desarrollo del proyecto resumiendo la experiencia práctica del mismo.
6. **Estado del arte:** antes de finalizar, se realizará una comparación con otros trabajos o proyectos relacionados con este.
7. **Conclusiones y líneas de trabajo futuras:** por último, en el apartado de conclusiones, se incluirá un conjunto de ideas resultantes de la experiencia recabada durante el desarrollo de este proyecto.

Anexos

Los anexos quedan divididos en las siguientes secciones:

- **Plan de proyecto:**
- **Especificación de requisitos:**
- **Especificación de diseño:**
- **Documentación técnica de programación:**
- **Documentación de usuario:**

Objetivos del proyecto

A continuación se detallan los objetivos que se han llevado a cabo en este proyecto. Se han dividido en tres secciones, la primera trata de los objetivos generales (qué se ha querido implementar), en la segunda se tratan los objetivos técnicos (qué herramientas se han usado) y por último se abordan los objetivos personales (por qué se escogió este proyecto).

2.1. Objetivos generales

En este Trabajo de Fin de Máster se persigue dotar a las rutas que una persona puede seguir en sus días de vacaciones de un significado semántico. Es decir, se pretende asignar una palabra clave a posiciones significativas dentro de la ruta que dicha persona ha seguido o está siguiendo en un momento determinado. Enfocándose en una visita turística, estas palabras clave estarán relacionadas con lugares turísticos de la localidad visitada, por ejemplo, su plaza mayor, su catedral o sus iglesias, sus museos o edificios de interés, etc. No obstante, se tienen en cuenta otros lugares como pueden ser centros comerciales, bancos, o restaurantes. Estos puntos o zonas reciben el nombre de Puntos de de Interés (PDI o POI, por sus siglas en inglés).

PDI (Point de Interest)

Como se ha descrito anteriormente, un Punto de Interés (PDI) es un lugar especificado mediante coordenadas geográficas que es interesante o útil para un gran número de personas. Este término es usado de forma general en aplicaciones software de cartografía o navegadores GPS.

Un PDI constará, como mínimo, de latitud y longitud pero también se suele incluir cierta información como un nombre, su descripción, la altitud o un número de teléfono. Cada PDI puede llevar asociado un cierto icono representativo usado por las aplicaciones para dotar de una mayor facilidad

de uso al usuario. La Figura 2.1 muestra un ejemplo de estos PDI sobre un mapa de Open Street Maps. En este caso, estos puntos marcan un hospital, un parking, un taller de automóviles, diversas paradas de autobús, bancos e, incluso, semáforos.

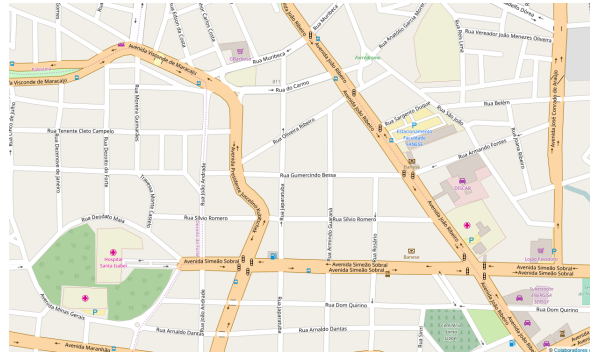


Figura 2.1: Puntos de Interés señalados en el mapa

Una vez se conocen los PDI por los que una persona ha pasado, se pueden reconocer los más significativos y asignar a su ruta. En este momento se puede decir que la trayectoria seguida es una trayectoria semántica. De esta forma se pasará de una serie de coordenadas (latitud, longitud) a un recorrido cíclico como el de la Figura 2.2. En este recorrido se observa la salida desde el Hotel Zola, pasando por lugares de interés como la Torre Eiffel, el Palacio Borbón o el Museo del Louvre, terminando en el Restaurante *Babylone* antes de regresar al punto de origen.

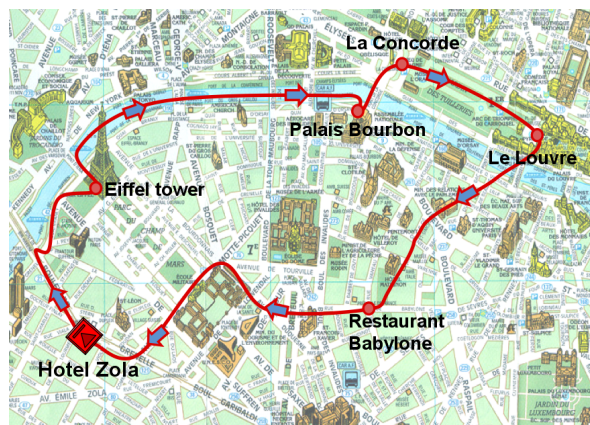


Figura 2.2: Ruta semántica sobre París

Siguiendo la filosofía de Open Street Maps, se pueden distinguir los PDI situados sobre una zona concreta dependiendo del tipo o del lugar al que

hacen referencia. Dentro de estos mapas, estas categorías son denominadas “Amenities”.

Trayectoria semántica

La consecución de distintos PDI sobre una misma ruta, define una trayectoria semántica. Esta trayectoria semántica puede estar formada por los puntos mencionados anteriormente, sobre el mapa del centro de París.

Análisis de trayectorias semánticas

El interés de una trayectoria semántica radica en los tipos de Puntos de Interés visitados por una persona que puede ser considerada turista en una localidad. De esta forma se puede categorizar cada ruta por los sitios de interés visitados. A continuación se muestra un ejemplo sencillo.

Una persona que realiza una visita turística en una ciudad puede estar interesada en visitar su catedral, sus iglesias o monasterios más importantes, etc. Englobando todo tipo de edificios destinados al culto religioso. Sin embargo, otra persona puede realizar una visita a los museos o monumentos de arte moderno combinándola con restaurantes de “autor”. Cada persona puede que tenga un ritmo diferente y sus visitas se duren un periodo de tiempo diferente.

En estos dos casos, es evidente, que ambas personas buscan lugares de interés turístico completamente diferentes. Seguramente estos patrones se repitan en las distintas visitas turísticas que realicen a otras localidades del país o, incluso, de otros países.

Habiendo realizado un análisis de las rutas o trayectorias que cada persona sigue, es posible recomendar a dicha persona una ruta adecuada en una ciudad que no haya visitado. De esta forma se ahorrará un tiempo valioso durante la preparación del viaje, que podrá ser aprovechado en la visita de los lugares seleccionados.

2.2. Objetivos técnicos

A la hora de implementar los objetivos marcados en este Trabajo de Fin de Máster, se han debido resolver distintas cuestiones técnicas. En esta sección se tratarán estos objetivos.

PostGIS

PostGIS es una extensión espacial para las bases de datos relacionales PostgreSQL. Añade soporte para objetos geográficos permitiendo realizar consultas espaciales en SQL.

Este soporte se encuentra respaldado mediante tres características únicas de esta extensión: tipos de datos espaciales, índices espaciales y funciones que actúan sobre ellos. Esta información espacial será almacenada en una columna de tipo “GEOMETRY”.

Cabe destacar que PostGIS ha sido certificado por el Open Geospatial Consortium (OGC) garantizando la interoperabilidad con otros sistemas.

PostgreSQL

Viendo la dependencia que PostGIS tiene sobre PostgreSQL, ha sido necesario el uso de este Sistema Gestor de Base de Datos para dar soporte a la extensión comentada. El uso de este SGBD no entraña una complicación superior o diferente a la del uso de cualquier otro sistema.

QGIS

QGIS es un Sistema de Información Geográfica de código libre. Permite manejar distintos tipos de objetos mediante las bibliotecas GDAL y OGR además de comunicarse con Bases de Datos. La interfaz de QGIS permite visualizar mapas de Open Street Maps, situar rutas sobre dichos mapas, consultar rutas a PostGIS, descargar ficheros de extensión OSM. Estos ficheros contienen todo tipo de información sobre una sección determinada de un mapa. Por ejemplo, se puede extraer todo tipo de información de las calles de una ciudad concreta, sobre sus POI, etc.

Biblioteca “tSemanticas”

La biblioteca que acompaña a este trabajo ha sido desarrollada en Java.

2.3. Objetivos personales

El presente Trabajo de Fin de Máster ha supuesto un reto personal puesto que en el momento de comenzar el mismo no se contaba con los conocimientos suficientes como para llevarlo a cabo. Durante los meses en los que se ha trabajado sobre este documento, se ha realizado un esfuerzo por aprender las tecnologías necesarias así como todos los conceptos teóricos asociados a las mismas.

Conceptos teóricos

Para poder validar los datos de trayectorias en crudo obtenidos de distintos usuarios se han de seguir una serie de pasos que permitan limpiarlos y transformarlos en información útil para el sistema. Esta serie de pasos será descrita en esta sección de la memoria.

3.1. Limpieza de datos

Debido a fallos hardware o software, pueden capturarse rutas que no correspondan con la realidad. Por ejemplo, el receptor de GPS del dispositivo usado para la recolección de datos puede dejar de recibir señal durante unos minutos y provocar una unión entre un punto “x” y otro “y” habiendo perdido datos relevantes en dicho intervalo de tiempo. Esto provocará una posible línea recta entre coordenadas que no deberían estar relacionadas entre sí haciendo que la ruta pierda información sobre los puntos recorridos. Para solventar estos posibles fallos, se hará una limpieza de los datos recibidos en el servidor y de los ya almacenados en el mismo.

Para ello se calculará la mediana entre los puntos de una misma ruta y, aplicando, un porcentaje de margen de error, se validará que la distancia entre los puntos no varíe de forma significativa. En caso de detectarse algún posible problema se procederá a validar la ruta sin la coordenada tomada como errónea. En caso contrario, la ruta quedará invalidada por completo y no será usada por el sistema.

Segmentación de la trayectoria

Lo primero a hacer es dividir la trayectoria seguida en “movimiento” y “paradas”. Un periodo de movimiento es el tiempo en el que la persona se encuentra moviéndose entre dos ubicaciones. Un periodo de parada es el tiempo en el que la persona se encuentra viendo un lugar de interés.

3.2. Enriquecimiento semántico

Los datos obtenidos por el usuario solo incluyen posiciones geográficas tomadas en un instante determinado, es decir, solo se tienen datos como la latitud y longitud y el espacio temporal en el que han sido tomados. En esta etapa se incluye información adicional a estos datos.

3.3. Anotación de trayectorias

Una vez conocidas las paradas que ha realizado la persona, se pueden asociar los Puntos de Interés a cada parada y así conocer los lugares en los que el usuario ha dedicado su tiempo.

3.4. Consultas por anotaciones

En este momento es posible realizar peticiones al sistema según lugares.

3.5. Descubrimiento de conocimiento

Extracción de patrones generalistas a partir de las consultas realizadas.

Técnicas y herramientas

Esta sección presenta un breve resumen de las técnicas y herramientas que han sido usadas para llevar a cabo este trabajo.

GPS y GLONASS

El Sistema de Posicionamiento Global (GPS) es un Sistema Global de Navegación por Satélite (GNSS) siendo conocido y usado ampliamente. Este sistema permite posicionar un dispositivo (terminal móvil, dispositivos GPS, etc) de una forma rápida y precisa (con un error no mayor a uno dos metros).

El sistema GPS cuenta con 24 satélites orbitando la Tierra a 20.000 Km de altura. Aunque inicialmente fue desarrollado por el Departamento de Defensa de los Estados Unidos (únicamente disponible para operaciones militares), ahora es un sistema de uso libre.

El sistema GLONASS (Global'naya Navigatsionnaya Sputnikovaya Sistema) es el sistema de posicionamiento desarrollado por la Unión Soviética para competir contra el GPS americano. Este sistema cuenta con 31 satélites (no todos en activo) orbitando a 19.100 Km de altura. En 2007 la Federación de Rusia elimina todas las restricciones para usos comerciales, siendo su precisión similar a la del sistema americano.

Esta apertura ha posibilitado la inclusión de ambos sistemas en todo tipo de terminales móviles y "GPS" posibilitando crear dispositivos que puedan apoyarse en uno u otro sistema para lograr una precisión todavía mayor.

En 2018 se espera que el Sistema Galileo (el sistema GNSS europeo) esté operativo contando con 27 satélites. Actualmente se cuenta con 3 satélites operativos y otros 6 en fase de pruebas.

Open Street Maps

Open Street Maps (OSM) es un proyecto colaborativo que permite crear y editar mapas. Estos mapas son de uso libre y gratuito para todos los usuarios, siendo solo los usuarios registrados los que pueden crear dichos mapas.

La creación de los mapas se lleva a cabo mediante la recogida de coordenadas geográficas usando dispositivos móviles. Estos dispositivos móviles obtendrán su ubicación gracias a sistemas GPS o Glonass o al uso de wifi o a la ubicación proporcionada por el operador de telefonía en caso de que el dispositivo cuente con una tarjeta SIM.

Los mapas son distribuidos bajo licencia ODbL (Licencia Abierta de Bases de Datos).

Los datos se almacenan siguiendo el datum WGS84 (World Geodetic System 1984, constituido por parejas latitud-longitud) usando la proyección de Mercator (definida en 1569 por Gerardus Mercator). La información primitiva consta de:

- Nodos: son posiciones geográficas concretas.
- Vías: una lista ordenada de nodos constituye una vía. Esta puede ser una polilínea o un polígono.
- Relaciones: son grupos de vías, nodos y relaciones que pueden ser agrupadas ya que contienen propiedades comunes (por ejemplo, .
- Etiquetas: son usadas para almacenar metadatos sobre los objetos del mapa, constan de una pareja clave-valor.

PostgreSQL 9.6

PostgreSQL es un Sistema Gestor de Bases de Datos (SGBD) relacional, orientado a objetos y libre, distribuido bajo licencia PostgreSQL License. Actualmente se encuentra en su versión 9.6 siendo la usada en este trabajo.

PostGIS

PostGIS es una Base de Datos espacial que expande las posibilidades de PostgreSQL añadiendo soporte a objetos geográficos permitiendo consultas de localización. PostGIS ha sido diseñado y desarrollado por la empresa Refraction Research siendo distribuido bajo licencia GNU (GPLv2).

Osmosis

Git

Git ha sido el sistema de control de versiones elegido y como plataforma web se hará uso de GitHub.

GitHub (aplicación de escritorio)

La aplicación de escritorio de GitHub permite mantener sincronizado el proyecto con la nube. Permite añadir, crear o clonar un repositorio existente. También facilita una visión del proyecto en la ventana "History". Dentro de la ventana "Changes" se pueden ver los ficheros que han sido modificados y necesitan de un commit para ser sincronizados con el repositorio local. Posteriormente, se puede sincronizar el proyecto con la plataforma web clicando sobre el botón "Sync".

ZenHub

El proyecto se basará en Scrum (marco de desarrollo ágil) y para permitir la creación de sprints, tareas, etc, se hará uso de ZenHub. ZenHub es un plugin que añade funcionalidad extra a la plataforma de GitHub.

L^AT_EX

Se ha usado L^AT_EX para la realización de la documentación del presente Trabajo de Fin de Máster. El contenido ha sido editado con TexMaker en su versión 4.5.

RouteConverter

RouteConverter es un software que permite mostrar una ruta (o traza) que ha seguido un usuario de Open Street Maps sobre un mapa. Para ello será necesario descargar dicha traza y cargar el fichero resultante (extensión gpx) en el programa. Si los datos contenidos en el fichero han sido formateados de forma correcta, será posible visualizar sobre el mapa la ruta o rutas seguidas.

Aspectos relevantes del desarrollo del proyecto

Este apartado pretende recoger los aspectos más interesantes del desarrollo del proyecto, comentados por los autores del mismo. Debe incluir desde la exposición del ciclo de vida utilizado, hasta los detalles de mayor relevancia de las fases de análisis, diseño e implementación. Se busca que no sea una mera operación de copiar y pegar diagramas y extractos del código fuente, sino que realmente se justifiquen los caminos de solución que se han tomado, especialmente aquellos que no sean triviales. Puede ser el lugar más adecuado para documentar los aspectos más interesantes del diseño y de la implementación, con un mayor hincapié en aspectos tales como el tipo de arquitectura elegido, los índices de las tablas de la base de datos, normalización y desnormalización, distribución en ficheros³, reglas de negocio dentro de las bases de datos (EDVHV GH GDWRV DFWLYDV), aspectos de desarrollo relacionados con el WWW... Este apartado, debe convertirse en el resumen de la experiencia práctica del proyecto, y por sí mismo justifica que la memoria se convierta en un documento útil, fuente de referencia para los autores, los tutores y futuros alumnos.

Estado del arte

Este apartado sería parecido a un estado del arte de una tesis o tesina. En un trabajo final grado no parece obligada su presencia, aunque se puede dejar a juicio del tutor el incluir un pequeño resumen comentado de los trabajos y proyectos ya realizados en el campo del proyecto en curso.

Conclusiones y Líneas de trabajo futuras

Todo proyecto debe incluir las conclusiones que se derivan de su desarrollo. Éstas pueden ser de diferente índole, dependiendo de la tipología del proyecto, pero normalmente van a estar presentes un conjunto de conclusiones relacionadas con los resultados del proyecto y un conjunto de conclusiones técnicas. Además, resulta muy útil realizar un informe crítico indicando cómo se puede mejorar el proyecto, o cómo se puede continuar trabajando en la línea del proyecto realizado.

7.1. Conclusiones

7.2. Lineas de trabajo futuras

Bibliografía

- [1] John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, 1992.
- [2] Wikipedia. Latex — wikipedia, la enciclopedia libre, 2015. [Internet; descargado 30-septiembre-2015].