



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería en Informática



TFG del Grado en Ingeniería Informática

**Development of a Chatbot for
Tourist Recommendations**



Presentado por Jasmin Wellnitz
en Universidad de Burgos — June 9, 2017
Tutor: Bruno Baruque Zanón



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería en Informática



D. Bruno Baruque Zanón, profesor del departamento de Ingeniería Civil, Lenguajes y .

Expone:

Que la alumna D. Jasmin Wellnitz, con DNI L28PGM5NZ, ha realizado el Trabajo final de Grado en Ingeniería Informática titulado Development of a Chatbot for Tourist Recommendations.

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

En Burgos, June 9, 2017

Vº. Bº. del Tutor:

D. nombre tutor

Resumen

Este proyecto consiste en desarrollar una aplicación de bote conversacional que es capaz de recomendar puntos de interés turísticos, personalizados a los gustos del usuario. Se accede al bote conversacional usando la aplicación de mensajería instantánea Telegram.

El proyecto se focaliza en desarrollar el lado servidor del interfaz de Telegram. Para ello se construye un servidor web Java que usa la plataforma de procesamiento de lenguaje natural api.ai para analizar la entrada del usuario. Luego la aplicación está desplegada en la plataforma-como-servicio Heroku.

Las recomendaciones dadas al usuario están basadas en las informaciones el usuario comparte con el bote conversacional y datos ya existentes de otros usuarios parecidos. El algoritmo de recomendación aplicado combina los dos enfoques de la teoría de recomendación más importantes: el filtrado colaborativo y basado en el contenido.

Además, se construye un base de datos geográfico usando PostGis, proporcionando las informaciones turísticas en las cuales las recomendaciones están basadas. En la versión de la aplicación presentada, se usa la información geográfica de Barcelona, España y por tal motivo, las recomendaciones proporcionadas están limitadas a esa ciudad.

Descriptores

bot conversacional, interfaz conversacional, Telegram, procesamiento de lenguaje natural, sistema de recomendación, base de datos geográfico, punto de interés...

Abstract

This project deals with the development of a chatbot application that is able to recommend tourist points of interest customized to the user's preferences. The chatbot can be accessed using the instant messaging app Telegram.

The project's main focus lies on the development of a server-side architecture to the Telegram interface. In order to do so, a Java web server is set up that uses the natural language processing platform api.ai to parse the user input. The application is deployed to the platform-as-a-service Heroku.

The recommendations given to the user are based on the information the user shares with the chatbot and already existing data of similar users. The applied recommendation algorithm combines the two most important approaches of recommendation theory, content-based and collaborative filtering.

Furthermore, a geographic database is set up using PostGis, providing the needed tourist information the recommendations are based on. In the presented version of the application, geographic information of Barcelona, Spain, is used and therefore limiting the provided recommendations to this city.

Keywords

Chatbot, conversational interface, Telegram, natural language processing, recommender system, geographic database, points of interest

Contents

List of Figures

List of Tables

Introduction

A chatbot is a computer program that interacts with its users through a conversational interface. They have been a topic of interest in the field of computer science for decades, yet, due to the rise of smart devices in the last couple of years, chatbots have received a great deal of new attention. The possibilities provided by the conversational interface were rediscovered and combined with the state of the art instant messaging methods.

This project concentrates on this upcoming trend by developing a chatbot that can be accessed through an instant messaging platform, in this case the messenger Telegram. Using the messenger's interface, the user interacts with the chatbot and is provided with tourist recommendations. The developed application consists of three principal components:

1. The chatbot which forwards the user input to the natural language processing platform `api.ai` and interprets the response.
2. The recommendation system which computes the personalized recommendation based on the user's preferences.
3. The geographic information database which makes the needed information for recommendation available.

In the following paper, the developed application is presented by examining the main objectives, underlying theories and way of proceeding as well as describing encountered challenges in the course of the project.

Project Objectives

This project aims to develop a chatbot that is able to present customized tourist recommendations to its users via an instant messaging application. In the presented version of the application, the messenger Telegram is used to interact with the user. The following steps and objectives must be realized to develop the chatbot:

- A Java web server is set up which is connected to Telegram through its webhook API to enable user interaction. The application is deployed to a platform-as-a-service cloud server to enable the webhook integration.
- Design of a conversational interface: the conversation flow between user and chatbot is mapped to the natural language processing platform `api.ai` which then parses the user input into formalized data. `api.ai` is accessed by the application through a REST-like HTTP endpoint. The parsed input is interpreted by the chatbot and triggers the desired behaviour, such as recommendation or storage of important user information.
- A recommender system must be implemented to provide personalized tourist recommendations. The recommender is based on data the user has shared with the chatbot and additional data of similar users. To overcome the problem of initially sparse user data, different recommendation methods are combined as well as retrieving existing user data from other sources and/or generating data.
- Retrieval of tourist data from the geographic information database OpenStreetMaps: the data is filtered so that only data of touristic importance is evaluated by the recommender and presented to the user. It is prepared in a way that similarity measures can be made between user interests and the point of interest inherent properties.

Theoretical Concepts

En aquellos proyectos que necesiten para su comprensión y desarrollo de unos conceptos teóricos de una determinada materia o de un determinado dominio de conocimiento, debe existir un apartado que sintetice dichos conceptos.

Algunos conceptos teóricos de L^AT_EX¹.

3.1 Secciones

Las secciones se incluyen con el comando `section`.

Subsecciones

Además de secciones tenemos subsecciones.

Subsubsecciones

Y subsecciones.

3.2 Referencias

Las referencias se incluyen en el texto usando `cite [?]`. Para citar webs, artículos o libros `[?]`.

3.3 Imágenes

Se pueden incluir imágenes con los comandos standard de L^AT_EX, pero esta plantilla dispone de comandos propios como por ejemplo el siguiente:

¹Créditos a los proyectos de Álvaro López Cantero: Configurador de Presupuestos y Roberto Izquierdo Amo: PLQuiz



Figure 3.1: Autómata para una expresión vacía

3.4 Listas de ítems

Existen tres posibilidades:

- primer ítem.
 - segundo ítem.
1. primer ítem.
 2. segundo ítem.

Herramientas	App	AngularJS	API REST	BD	Memoria
HTML5		X			
CSS3		X			
BOOTSTRAP		X			
JavaScript		X			
AngularJS		X			
Bower		X			
PHP			X		
Karma + Jasmine		X			
Slim framework			X		
Idiorm			X		
Composer			X		
JSON		X	X		
PhpStorm		X	X		
MySQL				X	
PhpMyAdmin				X	
Git + BitBucket		X	X	X	X
MikTeX					X
TeXMaker					X
Astah					X
Balsamiq Mockups		X			
VersionOne		X	X	X	X

Table 3.1: Herramientas y tecnologías utilizadas en cada parte del proyecto

Primer ítem más información sobre el primer ítem.

Segundo ítem más información sobre el segundo ítem.

•

3.5 Tablas

Igualmente se pueden usar los comandos específicos de \LaTeX o bien usar alguno de los comandos de la plantilla.

Tools and Technologies

Esta parte de la memoria tiene como objetivo presentar las técnicas metodológicas y las herramientas de desarrollo que se han utilizado para llevar a cabo el proyecto. Si se han estudiado diferentes alternativas de metodologías, herramientas, bibliotecas se puede hacer un resumen de los aspectos más destacados de cada alternativa, incluyendo comparativas entre las distintas opciones y una justificación de las elecciones realizadas. No se pretende que este apartado se convierta en un capítulo de un libro dedicado a cada una de las alternativas, sino comentar los aspectos más destacados de cada opción, con un repaso somero a los fundamentos esenciales y referencias bibliográficas para que el lector pueda ampliar su conocimiento sobre el tema.

Relevant Aspects of Project Development

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 (ED-VHV 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.

Related Work

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.

Conclusion and Future Work Lines

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.