Drexel Chatbot

Tom Amon, Aaron Campbell, Daniel Fitzick, Shishir Kharel, Hoa Vu, Nanxi Zhang | Marcello Balduccini | Filippos Vokolos | Drexel University CCI | 2017

Abstract

A web-based API for a chatbot that is capable of answering natural English questions about Drexel University, using a neural network.

Introduction

Information about Drexel faculty and facilities are scattered across many websites. One typically has to use a search engine to find a webpage, and then search that webpage itself for the information they seek. Repeatedly searching for information like this is very tedious.

Drexel Chatbot solves this problem by simply answering the user's questions using natural language. For example:

- "What is Mongan's email?"
- "Where is Vokolos' office?"
- "Where is Rush?"

Information needed to answer these questions is first gathered into a database by means of webcrawling and datamining algorithms. When a user enters an English question, Drexel Chatbot interprets it, finds the information in the database, and forms a suitable answer in English.

Additionally, Drexel Chatbot provides an API, accessible for anyone to use to integrate the service into their own applications.

Features

Natural Language Processing

- Questions are asked in English
- Answers are given in English

API

- RESTful API
- Input URL parameters
- Output JSON
- Concurrent users

Interfaces

- Web
- Android
- SMS

Supported Questions

Faculty

- Email
- Website
- Title
- Office
- Phone Number

Faculty

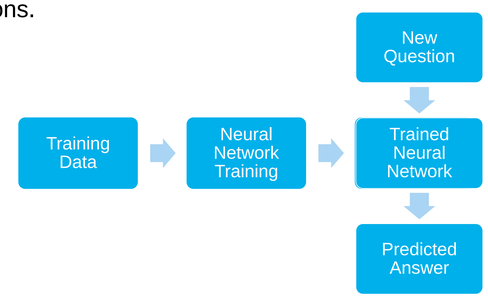
- Picture
- Department
- Education
- Publications
- Research Interests

Building

- Address
- Schedule
- Picture
- Website

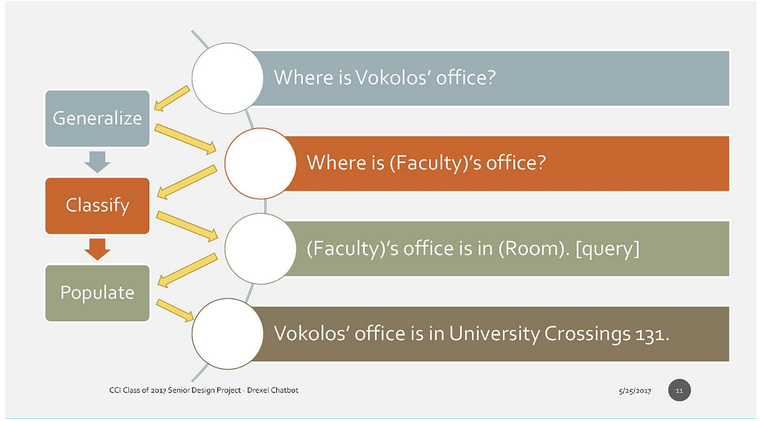
Training

To provide answers to questions in English, Drexel Chatbot uses a machine learning technique called neural network. A large set of questions with known answers was created as training data. This data is used to "teach" the neural network how to answer new questions.



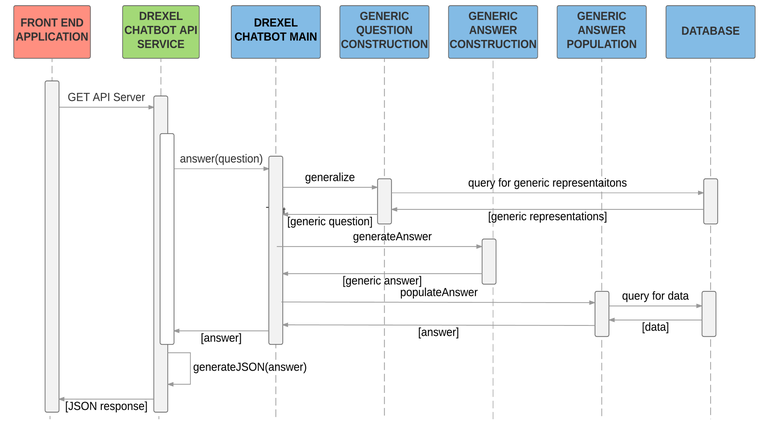
Question Answering

Instead of training the neural network on every possible question, Drexel Chatbot uses generalized questions and answers. The generalization process is carried out automatically by means of an ontology of concepts. Generalized questions do not contain specific subjects (e.g. a professor's name), but instead generic placeholders (e.g. "(faculty)"). This allows Drexel Chatbot to answer questions about all subjects in our database without training on each individual. It also simplifies the task of additional knowledge to the system, e.g. adding information about new faculty, since no further training is needed.



Generalize	Replaces keywords with generic representations, creating a generic question.
Classify	Uses trained neural network to predict a generic answer to a generic question.
Populate	Queries the database in order to make the generic answer specific to the original question.

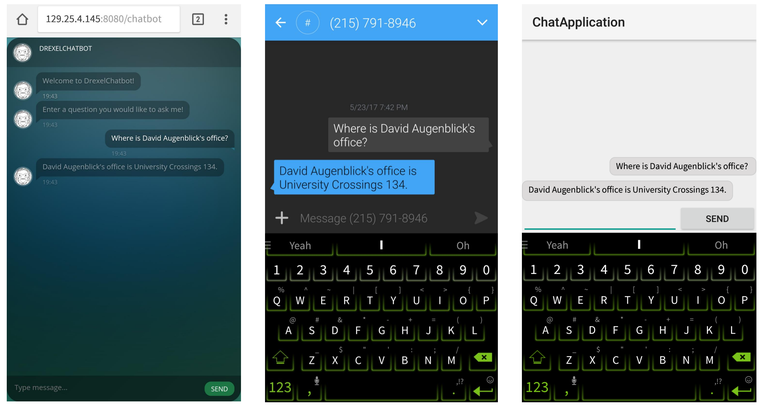
System Architecture



Information Extraction

Drexel Chatbot requires information from Drexel's websites to answer questions. Most of them organize information differently, which makes it impossible to use a single algorithm to collect all required data. As a result, Drexel Chatbot extracts the information from each website using a set of dedicated algorithms. This information is then stored in a database to be used by Drexel Chatbot.

Interface Screenshots



Limitations

Drexel Chatbot does not accept compound questions. In addition, each question is answered independently and previous answers do not provide context. Finally, Drexel Chatbot requires correct keywords to generalize sentences. If it fails to identify keywords, then it cannot answer the question.