SRS OF THE COLLEGE

CHATBOT

# **Introduction**

## Purpose of Document

This document will provide all of the requirements for the project Drexel Chatbot. It will serve as a reference for developers and customers during the development of the final version of the system.

## 1.2 Project Scope

Drexel Chatbot (Drexel natural language query service) is an AI chatbot that receives questions from users, tries to understand the question, and provides appropriate answers. It does this by converting an English sentence into a machine-friendly query, then going through relevant data to find the necessary information, and finally returning the answer in a natural language sentence. In other words, it answers your questions like a human does, instead of giving you the list of websites that may contain the answer. For example, when it receives the question "What time does the gym close today?", it will give a response “The gym closes at 10pm today.”

The main objective is creating a Web API, and sample web, mobile, and text messaging interfaces that demonstrate the use of the API.

The goal is to provide Drexel students and faculty a quick and easy way to have their questions answered, as well as to offer other developers the means to incorporate Drexel Chatbot into their projects.

## 1 .3 Overview of Document

**1.Revision History:**​ Provide the date of, reason for, and people who were involved with the modification of this document.

**2.Introduction:**​ Provide an overview of the application, explain the objectives and goal of the project and describe the document structure.

**3.Overall Description:** Provide the specification of the system model, the classes model​ and the main constraints.

**4.Functional Requirements:** Provide the analysis of the requirements by feature.​ **5.Non functional requirements:** Provide some other constraints that affect​ performance, safety and security.

**6.Use Cases:** ​Provide possible scenarios where the user interacts with the Web API and sample applications.

**7.Glossary:**​ Definitions of terms used.

# **Description**

## 2.1 Product Perspective

Most of the search engines today, like Google, use a system (The Pagerank Algorithm) to rank different web pages. When a user enters a query, the query is interpreted as keywords and the system returns a list of highest ranked web pages which may have the answer to the query. Then the user must go through the list of webpages to find the answer they are looking for. Drexel Chatbot, however, will try to understand the query and provide a definitive answer.

There will be four main units to the system working together to understand the question and return an appropriate answer:

* Generic question construction - capable of taking a natural language question and making it more generic.
* Generic answer construction - capable of taking a generic question template and providing a generic answer template.
* Generic answer population - capable of taking a generic answer template and populating it with information from the database to form an answer.
* Information extraction - capable of finding information through structured or unstructured websites, and storing that information in a database.

## 3.2 Product Features

The major features for Chatbot will be the following:

* **Web API:** An API call will include a question in the form of a query string​ URL parameter and the service will reply in GUI.
* **Natural Language Processing:** The system will take in questions written​ in standard English.
* **Natural Language Responses:** The answer to the question will be written in standard and understandable English.
* **Information Extraction:** There will be a database containing all the information needed, populated using information extraction techniques.

## 3.3 User Classes and Characteristics

The two classes of users for this system are described below:

* **API users**

API users consist of application developers who want to incorporate Drexel Chatbot API into other software applications.

* **Mobile app/Web app/SMS users**

These users consist of non-technical users who want to get answers for their questions. These users ask questions and get answers with mobile, web, or text messaging interfaces. This class of users include Drexel’s current and prospective students, teaching faculty, and staff.

## **4.Functional Requirements**

## A. User Interfaces

### 1. Website and mobile application GUI

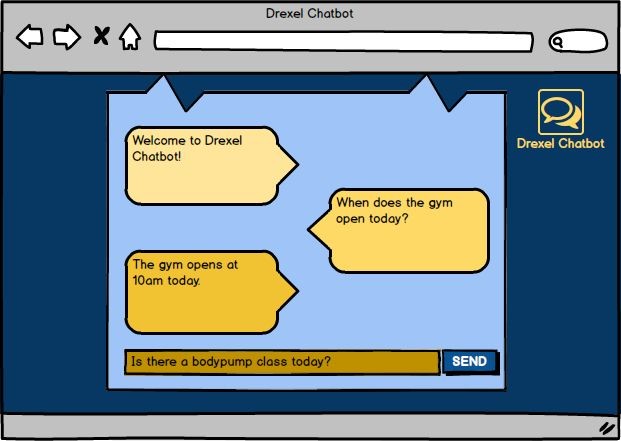


Figure 1: Mockup of the Web user interface

### B. SMS

R4.7.2.3 If a question is not understood by our API, the system will send a​ text containing an example question after the text with the API response.

**P**

# 5. Non-functional Requirements

5.1 API

### 5.1.1 Modularity

R5.1.1.1 The system will be designed in such a way that the algorithms​ for the four main units will be able to be easily swapped out. **Priority 1**​

### 5.1.2 Accuracy

R5.1.2.1 The overall accuracy of the Web API’s response will be​ measured using a developer-made testing set. **Priority 1**​

R5.1.2.2 The overall accuracy is calculated by dividing total number of​ correct answers by the number of questions asked. **Priority 1**​

R5.1.2.3 The accuracy of the ​ Generic Question Construction​ part will be​ close to 80%. ​**Priority 2**

R5.1.2.4 The accuracy of the ​ ​Generic Answer Construction unit will be​ close to 70%. ​**Priority 2**

R5.1.2.5 The accuracy of the ​ ​Generic Answer Population unit will be​ close to 70%. ​**Priority 2**

The accuracy for each supported topic will be as follows:

R5.1.2.6 ​ ​Drexel facilities’ locations and schedules will have accuracy greater than 70% **Priority 2**​

R5.1.2.7 ​ ​Drexel staff’s office locations, contact information, and positions will have accuracy greater than 70% **Priority 2**​

R5.1.2.8 ​ ​Drexel policies including academics, admissions, information technology etc. will have accuracy greater than 70%. **Priority 2**​

R5.1.2.9 ​ On-campus dining locations, hours, food types, etc. will have​ accuracy greater than 50%. **Priority 2**​

R5.1.2.10 ​ ​Food trucks’ general locations, hours, and food types on Drexel

Campus will have accuracy greater than 40%. **Priority 3**​

R5.1.2.11 ​ ​Official events listed on Drexel website, location and hours will have accuracy greater than 40%. ​**Priority 3**

### 5.1.3 Fast Response

R5.1.3.1 The average time for the server to respond, over the question​ testing set, will be less than or equal to 2 seconds. **Priority 2**​

### 5.1.4 Security

R5.1.4.1 The connection between the Web API and the programs will use​ HTTPS, for security. ​**Priority 3**

## 5.2 Web interface/Mobile application

### 5.2.1 Ease of Use

R5.2.1.1 A new user will make less than 3 mistakes in 5 minutes after 5​ minutes of use. ​**Priority 1**

# 6. Use Cases

## 6.1 Use Case Flow

### 6.1.1 API

**Precondition:** The server that is running the API is online.​

**Main Flow:** ​The user sends their question as a URL parameter to our API’s url.

**Postcondition:** The user receives the answer to their question in JSON.​

### 6.1.2 Web/mobile application

#### 6.1.2.2 Entering a question

**Preconditions:** The user starts the mobile application or web​ application.

**Main Flow:** The user enters the question in the text box​

**Postcondition:** The text box will show the question entered.​

#### 6.1.2.2 Sending a question

**Precondition**:​ Some texts exist in the Text Field box.

**Main Flow**​: User clicks the send button

**Postcondition**:​ The texts in the Text Field box appear in the chat window, and the box is cleared out. After sometime, some texts generated by the software appear in the chat window.

### 6.1.3 SMS application

#### 6.1.3.1 Sending a question

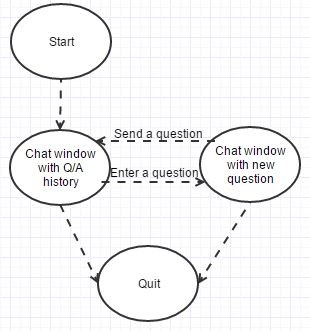
**Precondition**:​ User can send and receive SMS messages

**Main Flow**​: Create and send a text message to the designated

Drexel Chatbot number

**Postcondition**:​ The answer to the user’s question is sent as a SMS to the number that sent the question.

## 6.2 Activity Diagram



# 7. Glossary

* Chatbot: An interface, usually text based, specializing in the mimicry of natural language conversation. AKA “artificial conversational entity.”
* GUI: Graphic User Interface, a type of user interface that allows users to interact with the software through graphical icons (e.g. buttons, etc.).
* HTML: Hypertext Markup Language, a standardized system for tagging text files to achieve font, color, graphic, and hyperlink effects on webpages.
* JSON: JavaScript Object Notation, a data-interchange format that is commonly used in exchanging data over the Internet.
* Pagerank: ​PageRank is an algorithm used by Google to rank websites. It works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites.
* SMS: Short Message Service, the text messaging protocol of cellular telephones.
* Standard English: the language that can be understood by English-speaking high school graduates.
* URL: Uniform Resource Locator, an address to a resource on the Internet.
* URL parameter: parameters whose values are set in a webpage’s URL.
* Web API: an application programming interface (API) for either a web server.
* Web scraping: web scraping is a technique employed to extract large amounts of data from websites whereby the data is extracted and saved to a local file in your computer or to a database.