An Assistant for the CSE Departmental Website @ IIT Kanpur

Avideep Mukherjee (18111264)
Soumya Banerjee (18111271)

A Term Project Report for CS727
Group 14
Instructor: Prof. T. V. Prabhakar
November 24, 2018

Abstract

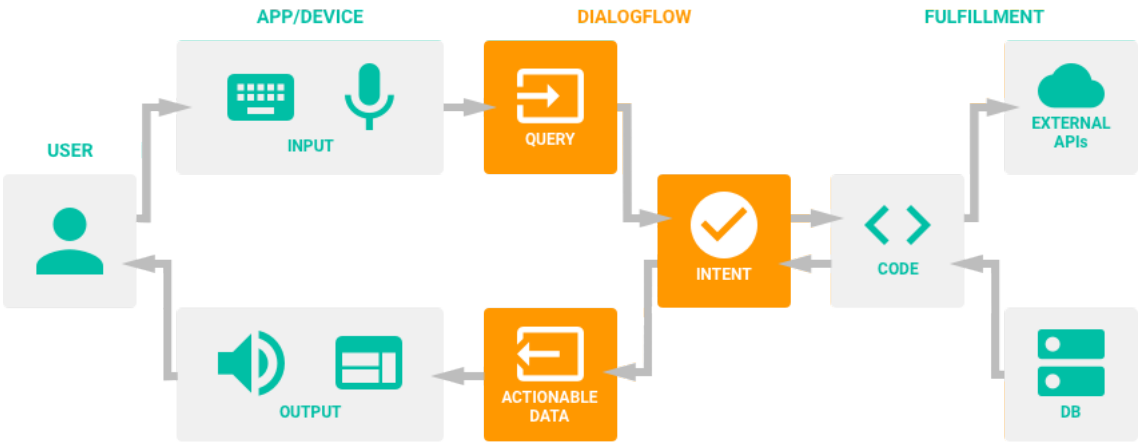
In this report we present **CSE-IITK-Assistant**, a bot that helps users who visit the CSE Departmental Website of IIT Kanpur (`https://cse.iitk.ac.in/`) derive information related to the department and are present in the website through a conversational question-answering mechanism. This assistant can answer most of the questions asked about the department and is developed using DialogFlow. The assistant is deployed here.

1 Introduction

Dialogflow[1] is a Google-owned developer of human–computer interaction technologies based on natural language conversations. It provides an online console to the developers where one can set the required intents and entities. The developer can also monitor what are the questions or messages that the users are asking across anywhere the assistant bot has been deployed on. A nice workflow of how the message passes from the end-users to the backend and gets answered back is demonstrated in Figure 1. As is clear from the picture the system has three major sections, namely, an end device - it may be a website or an app, the DialogFlow Agent and the webhook. These components are described in details later. In the next section different components of the dialogflow agent has been described.

[1] https://dialogflow.com/

Figure 1. Overview of the workflow of DialogFlow[a]



[a]Source: DialogFlow Documentation

2 The DialogFlow Agent

The bot that users' chat with in the end devices are termed as **Agents** in DialogFlow. The developer basically builds an Agent. Every DialogFlow Agent has a number of components. Following is a list of the components that are used for the development of this particular term project.

- ✓ Intents - Each time a user asks a question, the dialogflow agent captures that question and passes it to the concerned *intent*. An intent is created against every possible set of questions that a user can ask and is trained with a possible set of questions. The DialogFlow Agent will then train with these questions so that it can recognize any question that the users ask even if it is not exactly the one that is has been trained with. A list of intents that we have used for this project are shown in .

- ✓ Entities - Entities capture crucial part of the question into them that is integral for answering the question. For example, if a user wants to know the name of the course against a course code, it is unreasonable to write intents for every possible course codes. Instead, an entity is created that will capture the course code from a sentence, and the result can be returned to the user against that particular course code. A list of intents that we have used for this project are shown in .

- ✓ Knowledge[beta] - This feature is still in development version. Knowledge feature takes an html web page that has an FAQ in it and identify the questions and their answers and stores in the agent so that it can answer to the user accordingly if any question resembled to one of that from the FAQ. A list of intents that we have used for this project are shown in .

- ✓ Fulfillment - This is one of the most important components in DialogFlow. When the developer wants to reply to user answers that are specific to each question, Fulfillment is the choice to go with. Through Fulfillment, DialogFlow will send a POST request in the form of json to a webhook written by the developer. The webhook will capture the json, parse it and obtain key information the user wants and can provide answers in json form to the Agent based on those keys. The Agent will then forward this answer to the user.

3 Capabilities and Limitations

Here's an exhaustive list of topics that CSE-IITK-Assistant *can* answer as of now.

- ✓ About the Faculties, just their designations. And also list them according to designations (professor, assistant professor or associate professor) if asked.
- ✓ Names of the Ph.D. Students (year-wise) and Post-Doctoral Fellows.
- ✓ Links to other students' list pages.
- ✓ Research Areas and their associated faculties.
- ✓ Course names against codes and vice-versa.
- ✓ Links to some other important pages when asked accordingly (eg. Publications, webmail etc.).
- ✓ Small Talk.

Here are the possible questions that the Assistant *cannot* answer as of now.

- ✓ Count the number of students from each department.
- ✓ Awards received by faculties and students.
- ✓ Deeper details about students and faculties.
- ✓ Questions whose answers are found in non-crawlable webpages.

Figure 2. List of intents used

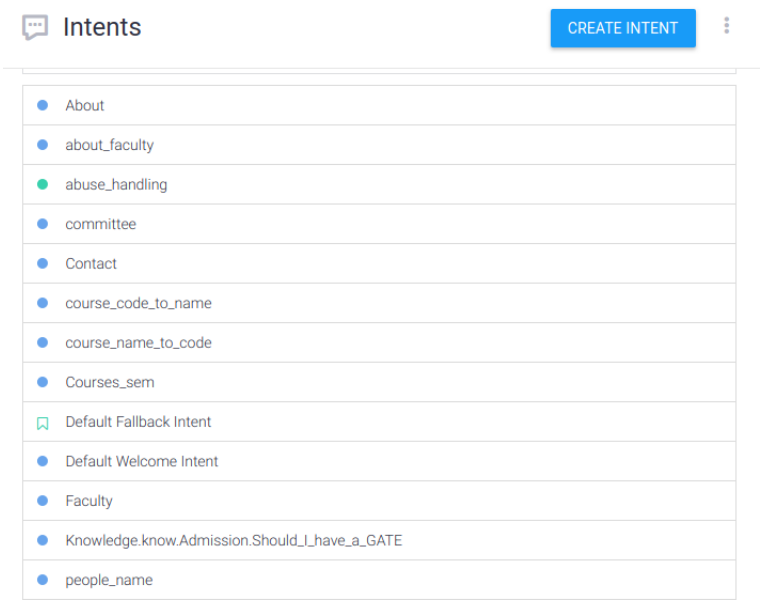


Figure 3. List of entities used

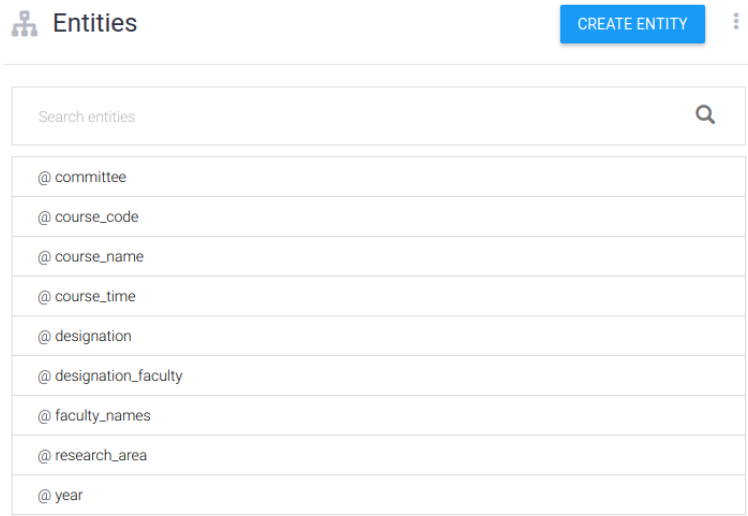
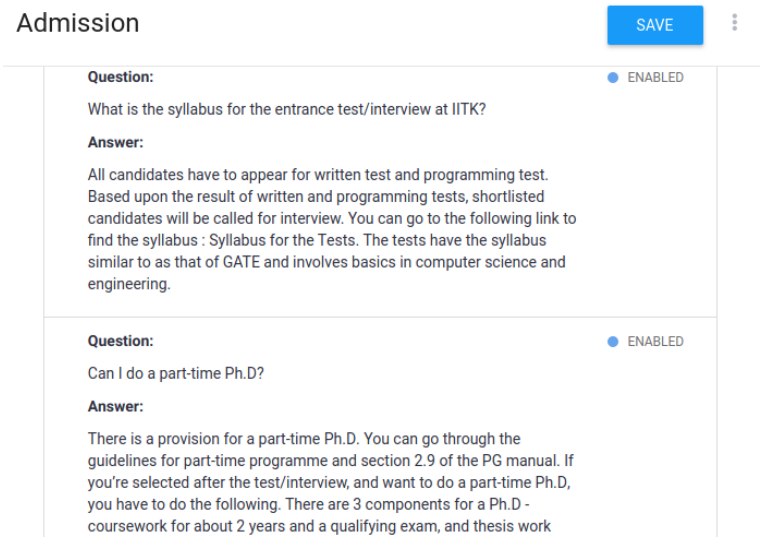


Figure 4. Snapshot of the knowledge that the agent generates from an FAQ



4 Conclusion

Sometimes the invoked intents are not what was intended. Failures are not automatically sent for training. Data from which the webhook is searching for answers is static. It retrieves information from already scraped `json` files. The Assistant cannot provide information about everything on the website still. Knowledge service from DialogFlow may provide better facilities once it has been deployed as a stable version. Extensive testing with human users is very much required and the agent is to be updated with the questions that it cannot answer it yet. The codes for this project can be found in this github link.