COSC 310: ChatBot Assignment

Instructor: Dr. Bowen Hui

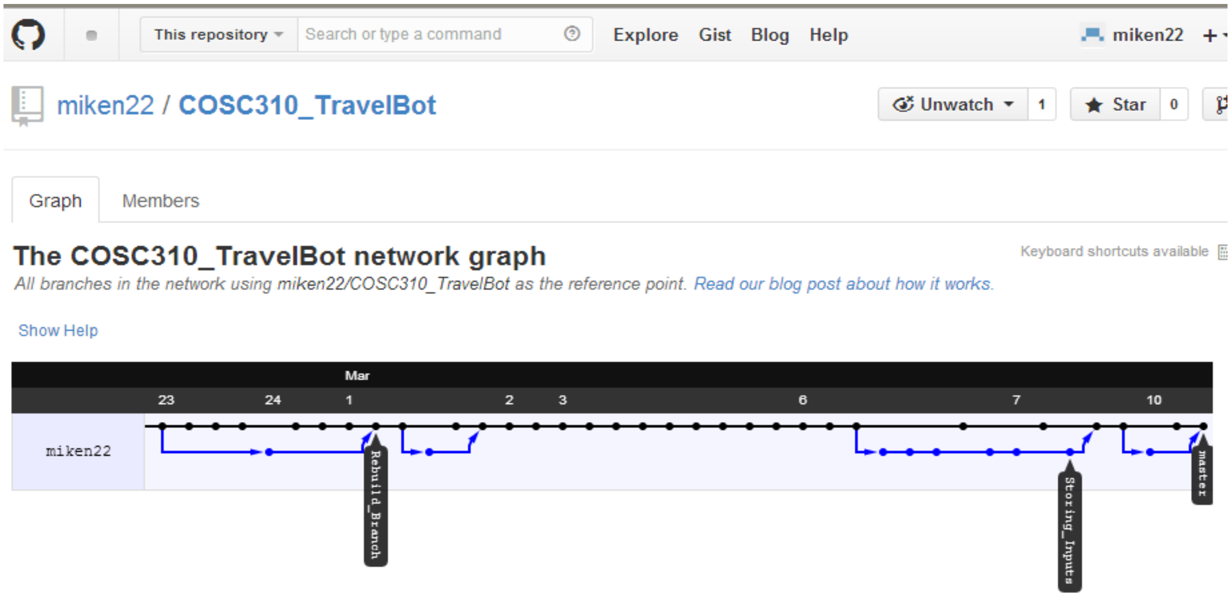
Date: March 14, 2014

By: Michael Nowicki

GitHub Repository:

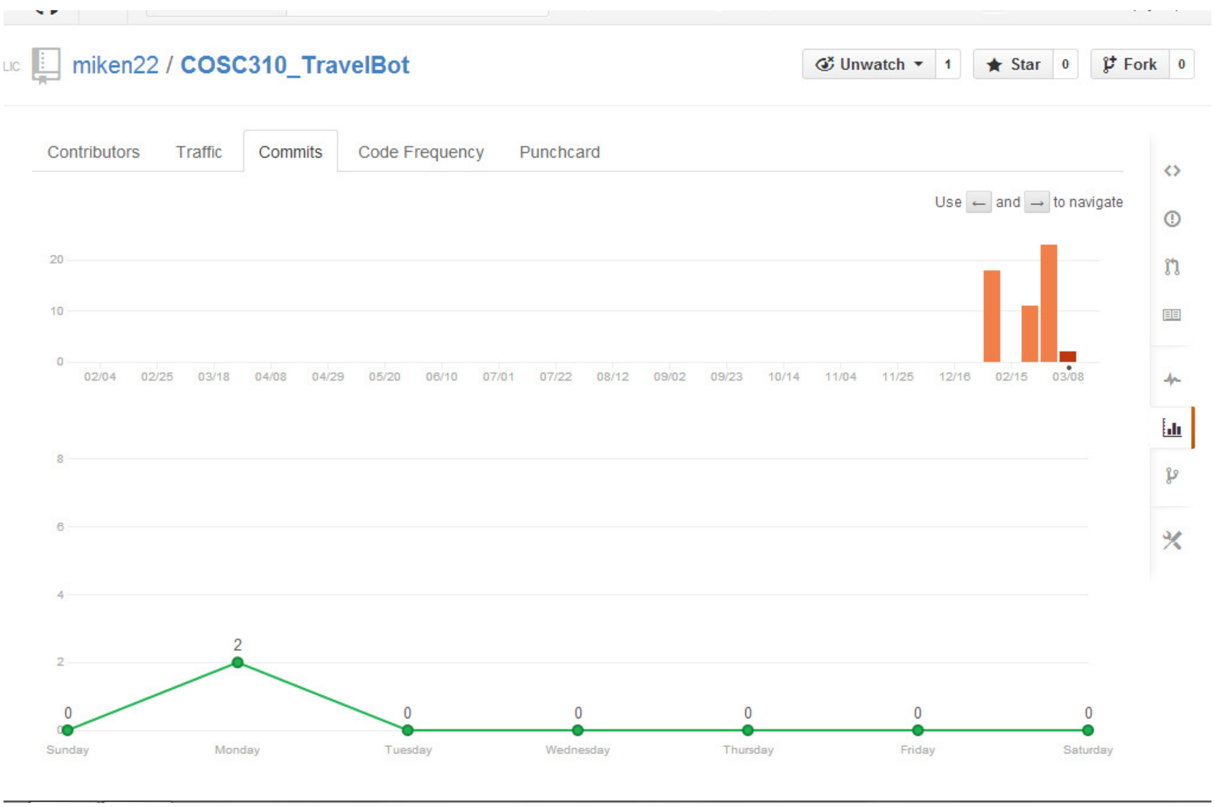
https://github.com/miken22/COSC310_TravelBot

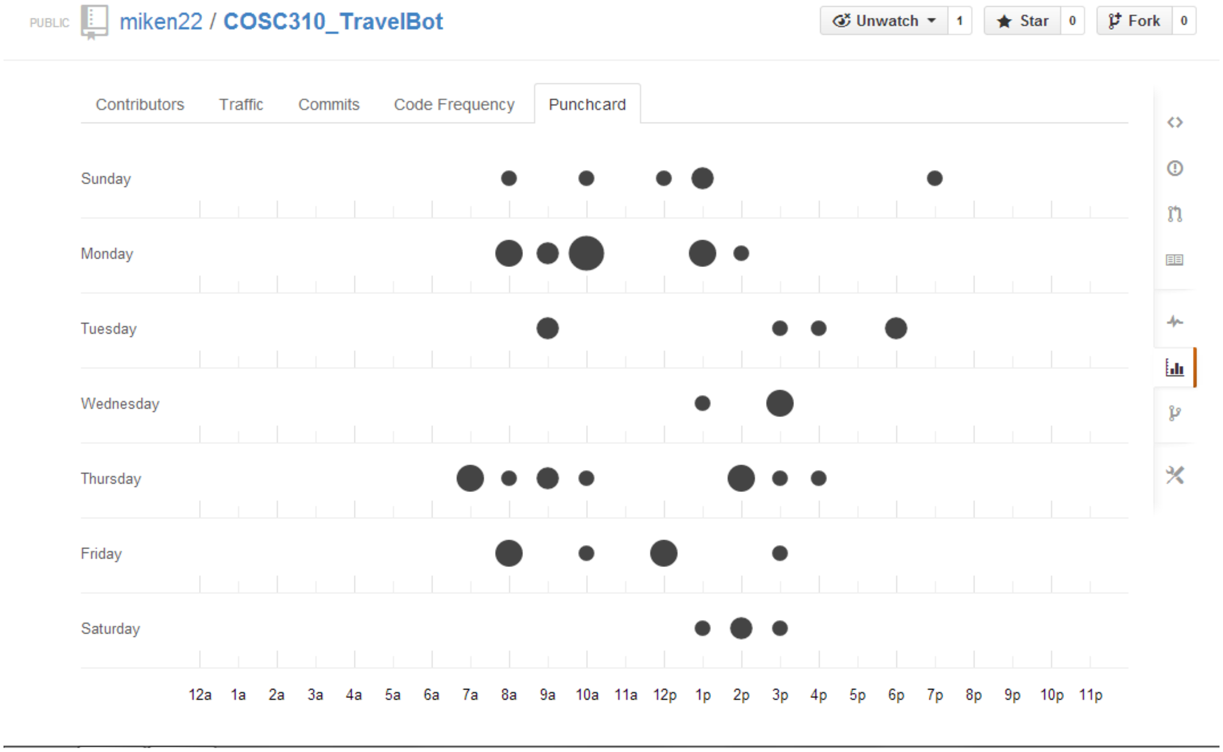
Github Graphs:



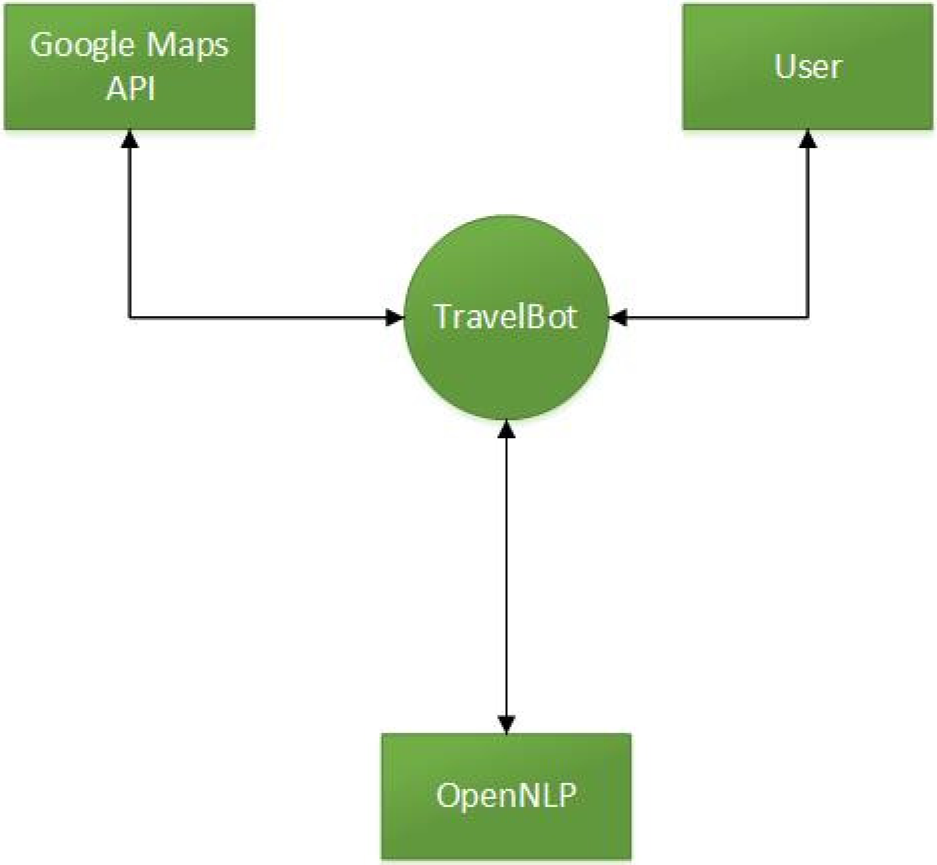
The COSC310_TravelBot network graph

All branches in the network using miken22/COSC310_TravelBot as the reference point. Read our blog post about how it works.




PUBLIC
miken22 / COSC310_TravelBot
Unwatch 1
Star 0
Fork 0
Contributors
Traffic
Commits
Code Frequency
Punchcard
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
12a
1a
2a
3a
4a
5a
6a
7a
8a
9a
10a
11a
12p
1p
2p
3p
4p
5p
6p
7p
8p
9p
10p
11p

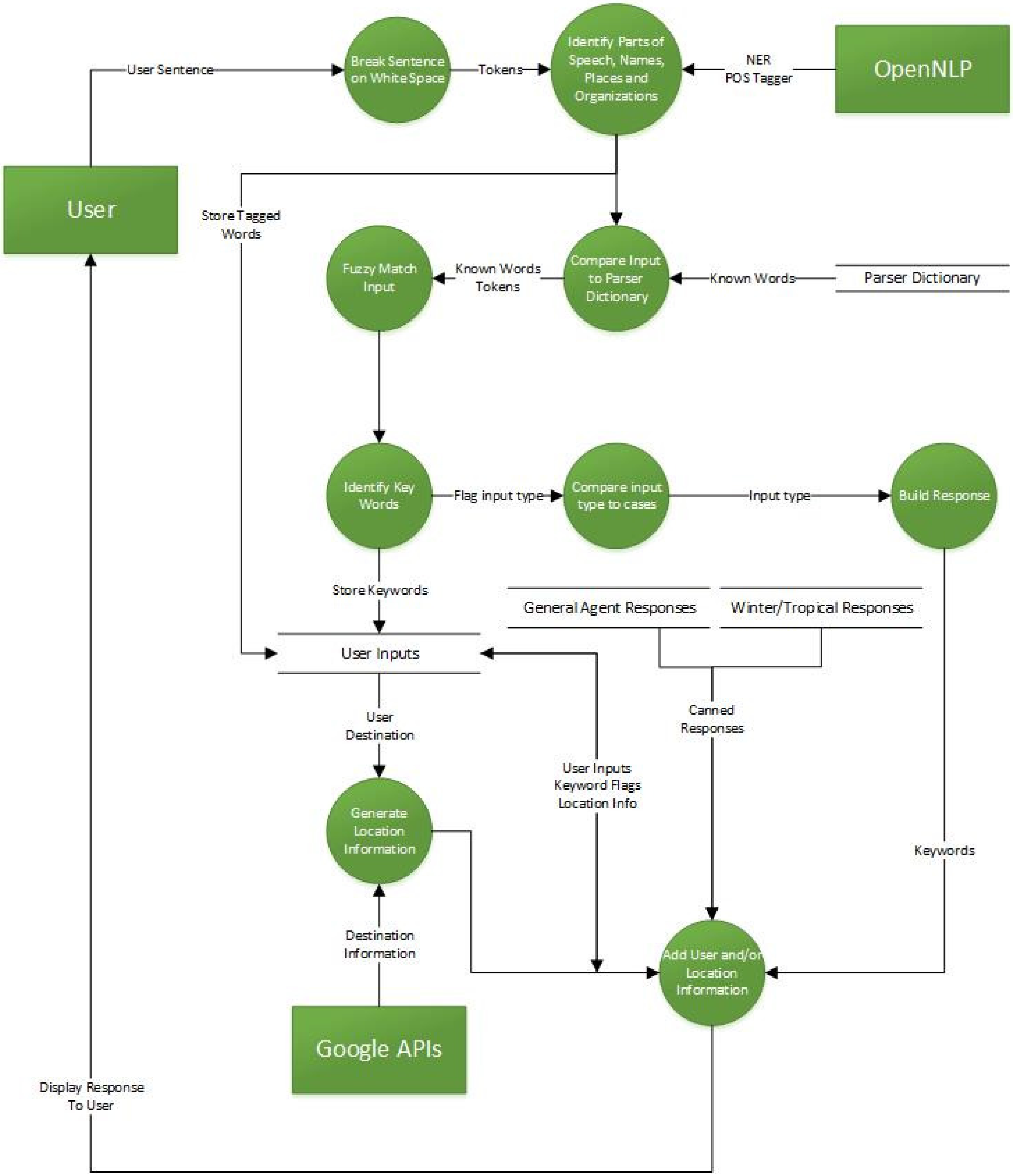
ChatBot
Level 0 DFD



Description:

This Level 0 DFD provides a high level overview of the TravelBot system. The central process is the TravelBot program which handles user input, parsing and spell checking, and response generation. There are three major external entities that the system must interact with to decide how to respond. The primary external entity is the User who provides the agent with input. The agent processes the input and returns a response to the user. The second external entity is the OpenNLP parsing toolkit. It provides the system with additional parsing tools such as the Parts of Speech Tagger and Named Entity Recognizers. These are used by the system to try and develop more intuitive responses by identifying keywords that have not been explicitly programmed into the agents' dictionary. The final external entity that interacts with the system are the Google APIs. The TravelBot, along with the JSON libraries included in the project, use the users information to perform Google queries in real-time during the conversation to generate dynamic responses.

ChatBot
Level 1 DFD



Description:

This Level-1 DFD demonstrates how the agent processes each turn of user dialogue. The process begins by having the user supply the agent with a given sentence. Once the agent receives the input its first process is to begin breaking the sentence into tokens. This is done first by splitting each word on the whitespace (spaces) in the text and keeping each individual word in a collection of tokens. Once the sentence has been tokenized the Parser then uses the POS Tagger to flag words which may represent proper nouns, which the Named Entity finders use to try key words not explicitly programmed in the agents' dictionary.

Once the sentence has been parsed by the OpenNLP parsers the input is then compared to against the agents' known words to try and identify key words. This is done by taking the token collection and comparing each word to a word in the given dictionary list which is supplied from the stored data lists. While this is happening a process called Fuzzy Matching tries to determine if what the user has input appears similar to a word in the list within a given tolerance. This process is the agents attempt to identify misspelled words in the users input.

The parser then tries to identify key words in the sentence. If the users' sentence contains a key word, or the OpenNLP parsers have found a key word, the input is stored with the agent. Depending on which key words are identified in the sentence the parser flags the input as being of a certain type. These types tell the agent which response will be most appropriate for the given input. The input type is passed to the response building process which uses the type as a flag for a larger case statement. Once the agent has determined which case matches the input type it starts to build a response. Sometimes some of the stored user information may be passed along to start building the responses.

When the agent builds its response it starts with the general framework of the answer as they are. Some responses have keywords within them to flag for the responder that the agent needs to replace the word with user input or Google query results. This process relies on getting information from the stored responses, stored user inputs, and location information as provided by the Google APIs. As the DFD also shows, some of the input is also stored back into the User Inputs based on the keywords that the parser identifies which the agent may use later in the conversation. Once all of this information has been brought together and all the sentence flags have been replaced the agent finally has a response which is displays for the user.

Sample Dialogue:

Travel Bot: Hi! Welcome to our travel center. We help arrange trips to Mexico and ski resorts in BC and Alberta. How can I help you today?

User: hey i'm michael
TravelBot: Greetings, welcome to my travel agency Michael.
User: I want to go to Calgary
Travel Bot: Calgary is very nice!
User: how far is calgary
Travel Bot: The distance between Calgary and Kelowna is 605.15km
User: can I drive?
Travel Bot: You can if you want.
Driving to Calgary, from Kelowna would cost approximately $299.58.
User: can i fly
Travel Bot: Searching...
I can certainly help you with that. All our flights are direct to Calgary,AB.
Flying to Calgary,AB from Kelowna would cost approximately $207.24
User: what activities are there?
Travel Bot: While you are there you could rent a snowmobile, or you could go skiing
User: is there a library there?
TravelBot: You could try 'Millarville Community Library', it looks perfect for you.
User: how can I get around?
Travel Bot: We could help set you up with a rental car if you'd like. Be sure to drive carefully on the mountain roads!
User: Can I get around another way?
Travel Bot: If you want to get between cities Greyhound is the way to go.
User: actually, I want to go to Revelstoke
Travel Bot: Everybody loves Revelstoke!
User: how far is revelstoke?
Travel Bot: The distance between Revelstoke and Kelowna is 197.43km
User: are there good restaurants there?
Travel Bot: Determining what's best for you.
A very popular place is The Village Idiot.
User: are there any activities there?
Travel Bot: While you are there you could rent a snowmobile, or you could go heliskiing
User: can you book me a hotel?
Travel Bot: We offer a wide variety of accomodations. Do you have a budget?

User: I can afford $60
Travel Bot: Searching for the best accommodations that match you budget.
In that price range there is the Sandman Hotel Revelstoke for only $55 a night.
User: what's the weather like right now?
Travel Bot: It is currently -4.0 degrees C in Revelstoke and the sky is clear.
User: how can I get around there?
Travel Bot: If you want to get between cities Greyhound is the way to go.
User: is there an aquarium there?
TravelBot: Sorry, there are no matches for 'aquarium'. Would you like me to search for something else?
User: No. Actually, I want to go to Cancun
Travel Bot: Everybody loves Cancun, Mexico!
User: what's the weather like there?
Travel Bot: It is currently 25.37 degrees C in Cancun, Mexico and the sky is clear.
User: can you book me a hotel?
Travel Bot: Were you interested in a family resort, or one of our luxury offerings?
User: I'll spend $100 a night.
Travel Bot: Searching for the best accommodations that match you budget.
I suggest BeachScape Kin Ha Villas. It just takes seconds to walk from your room to the beach!
User: what activities are there?
Travel Bot: While you are there you could swim with dolphins, or you could see some of the historical landmarks
User: how can i get around?
Travel Bot: If you want to get between cities coach bus is the way to go.
User: are there any popular restaurants?
Travel Bot: Well let me find out...
There may be some local places, but nothing outstanding.
User: can I go on a cruise?
Travel Bot: Searching...
Yes! The Maya Riviera is one of the most popular cruise destinations.
User: can i fly there?
Travel Bot: Searching...
I can certainly help you with that. All our flights are direct to Cancun.
Flying to cancun, mexico, from Kelowna would cost approximately $2157.65.
User: can I drive?
Travel Bot: You can if you want.
Driving to cancun, mexico, from Kelowna would cost approximately $3118.99.
User: ok thanks
Travel Bot: Of course, anytime.
User: bye
TravelBot: Goodbye Michael.

Limitations:

- One of the limitations of the program is how input is provided by the user. Sentences must be phrased in a manner that uses the keywords from the parser dictionary to identify what the user has said.
- The agent is only able to handle conversations in English.
- The agent is only able to maintain a conversation about the destinations in its dictionary.
- Users must ask the agent direct questions
- Sometimes there are issues interacting with Google's APIs. For example, occasionally their application accepts searches for "Revelstoke", however most of the time it is not specific enough so when the agent see's "Revelstoke" it converts it to "Revelstoke,BC". This normally works, however when the API decides that just "Revelstoke" is OK the agent crashes because the API does not know how to search "Revelstoke,BC". If that happens when running the agent switch to another city and the problem will not happen.
- The en-pos-maxent.bin file needs to be included and stored along with the executable JAR.
- Can only do searches on certain keywords. The input must be close in spelling to how they appear in the list (the spell checker compensates for differences such as '_')
    - Keywords: "airport", "amusement_park", "aquarium", "bar", "café", "campground", "casino", "clothing_store", "convenience_store", "department_store", "grocery_or_supermarket", "hospital", "library", "liquor_store", "lodging", "movie_theater", "museum", "night_club", "park", "parking", "restaurant", "shopping_mall", "spa", "stadium", "travel_agency", "university", "zoo"

Other Sample Dialogues:

These short conversations demonstrate some of the limitations of the TravelBot program. Below is an example of when the user asks the agent too general of a question. The agent indicates that it does not know how to respond to the given sentence.

> User: where can I go there
> Travel Bot: Sorry, I don't know what to say to that.

In this example the user has prompted the agent to respond to a question, however the agent has not been provided with any context about where the question is referring to.

> User: What's the weather there?
> Travel Bot: Sorry, you have to decide where you want to go before we can talk about weather.
> User: can afford
> Travel Bot: Sorry, you need to tell me more about what you mean by 'can afford '

The Google Maps API does not have a lot of information for nearby searches in Canadian and Mexican locations. The query feature does work, however because of the limited amount of information there may only be a few, if any, returned search results. There may be a lot of "could not find what you were looking for" type responses depending on the city you are talking about. (Calgary has the most working results)

<u>When it works:</u>
User: I want to go to Calgary
…
User: is there a library?
Travel Bot: Yes, I've found a place called 'Millarville Community Library' that might be just what you're looking for.
User: How about a liquor store?
Travel Bot: You could try 'Alberta Gaming & Liquor Commission', it looks perfect for you.

<u>When it doesn't work:</u>
User: is there an aquarium?
Travel Bot: Sorry, I could not find anything matching 'aquarium', would you like to try again?
User: is there a department store?
Travel Bot: Sorry, there are no matches for 'department store'. You can enter another search if you would like.

Extractable Features:

1. Fuzzy Matching for Spell Checking
2. Interfacing with Googles' Maps and Search API's – Dependent on the JSON libraries
3. POS Tagging user input, flagging non-hardcoded keywords
4. Saving conversation histories
5. Tokenizing sentences based on a Regular Expression class