Kinsey Mellon

Chatbot Report

System Description:

The chatbot created for this project is made in Python from scratch. It starts by reading in the file containing the pure text from the La La Land Wikipedia page which it will make the knowledge base from. After reading all the text from the file the knowledge base is made by making a dictionary with the top 10 terms as keys and putting all the sentences that contain each top term into the specific words value list. After that there are multiple tuples made that contain key phrases to either look out for in the user response or to make responding easier for the bot. For example, there is a "Greeting_in" tuple that holds some common ways to say hello, a "Greeting_out" tuple that holds some ways for the bot to say hello, an "Opinion_in" list that just holds the word "you" in to check if the user ever asks about what the chatbot thinks, and a "Likes_Dislikes" tuple that holds a lot of common opinion oriented phrases or personal remark phrases such as "I like", "I don't like", "I dislike", "I think", "I want", and "I'm".

Before the chatbot begins a conversation it first asks for the user's name. The program checks if there is a pickle file that holds the user model and if it is empty. If it doesn't exist or it's empty a user model dictionary is made and then says hi to the user and adds them into the user model. If the pickle file exists and has user information in it, then the pickle file is opened and loaded into a user model dictionary. The user model is then checked for the inputted user name and if it is in the model than they are a returning user and don't need to be added into the model but if their name isn't in the model than their name is added into the model as the key with an empty list for the value waiting to be filled with any user information received.

Now going over how the chatbot works, the basic functionality it uses is a while loop. A flag is created and set to true, and the while loop continues until the criteria is met to make the flag false. Inside the while loop a Boolean variable "op" short for opinion is made and the user input is gathered and lower cased. The input is looped through to see if the user is asking the bot for its opinion, if it is then the flag is changed to false. Then the input is checked to see if the user made any personal comments/opinions and if they did to add what they said to the user model under the key with their name. Then the bot response is determined through a series of if statements. First if the user said "bye" then the flag is set to false, and the bot says bye and the program ends. If the user did not say bye, then if the user said any of the greetings in the greetings_in tuple the bot responds with a random choice from the greetings_out tuple. If the flag "op" was set to true, meaning the user asked the bot a personal question, the bot responds saying they don't know and asks what the user thinks to keep the conversation going. If none of the previous if statements have been met that means the bot can search the knowledge base for an answer. To do this the user input is added to the knowledge base of sentences and is passed to a function to find the correct response. The way the bot finds the most appropriate response is by using the TfidfVectorizer and cosine similarity, which are the NLP techniques used in this project. The knowledge base of sentences is fit and transformed with the tfidfvectorizer, where

the tokenizer is set to a defined function that goes through all the sentences to lemmatize them and get rid of any punctuation. The cosine similarity is then found between the user input and all the sentences in the knowledge base and made into a list. This list is flattened and sorted. The required similarity was found and if it's not zero then the sentence at the index of the highest cosine similarity is returned as the response. If the required similarity is zero, then a blank string is returned. If a blank string is returned, then if there are things in the user model for this user then the bot responds with one of the things from the user model about this user to attempt to keep the conversation going but if there isn't anything for the user in the user model then the bot responds with a prompt to say something else. If a sentence is returned, then the bot responds with that sentence. Once the user enters "bye" to end the conversation the user model is dumped to the pickle file.

Knowledge Base:


```
C:\Users\kinse\HLT\Scripts\python.exe C:/Users/kinse/Downloads/Fall22/HLT/ChatBot.py
Stone :  ['It stars Ryan Gosling and Emma Stone as a struggling jazz pianist and an aspiring actress, respectively, who meet and fall in love whil
movie :  ['About An American in Paris, Chazelle commented: "That\'s a movie that we just pillaged.', "Consequently, the movie was shot on celluloi
musical :  ['La La Land 2016 American romantic musical comedy-drama film by Damien Chazelle La La Land is a 2016 American romantic musical comedy-
Chazelle :  ['La La Land 2016 American romantic musical comedy-drama film by Damien Chazelle La La Land is a 2016 American romantic musical comedy
Gosling :  ['It stars Ryan Gosling and Emma Stone as a struggling jazz pianist and an aspiring actress, respectively, who meet and fall in love wh
Land :  ['La La Land 2016 American romantic musical comedy-drama film by Damien Chazelle La La Land is a 2016 American romantic musical comedy-dra
Hollywood :  ['That night, her roommates take her to a lavish party in the Hollywood Hills, promising her that someone in the crowd could jump-sta
love :  ['It stars Ryan Gosling and Emma Stone as a struggling jazz pianist and an aspiring actress, respectively, who meet and fall in love while
Mia :  ['While stuck in Los Angeles traffic, aspiring actress Amelia "Mia" Dolan has a moment of road rage directed at Sebastian "Seb" Wilder, a j
Bot: Hello! If you want to talk about La La Land the movie you came to the right place. When you want to exit say Bye
```

To build the knowledge base for this chatbot I scraped all of the text from the La La Land Wikipedia page and stored it in a file. The above picture is what the knowledge base I used for the chatbot looks like. It's a dictionary with 10 keys that represent the most frequent words from the La La Land Wikipedia page. The values of each key are all the sentences from the Wikipedia page that contain that key word.

Kinsey Mellon

Sample Dialog Examples:

```
C:\Users\kinse\HLT\Scripts\python.exe C:/Users/kinse/Downloads/Fall22/HLT/ChatBot.py
Bot: Hello! If you want to talk about La La Land the movie you came to the right place. When you want to exit say Bye.
Bot: First, what's your name?
kinsey
Welcome back kinsey ! What do you want to talk about this time?
what is la la land
Bot:  La La Land in particular is inspired by his experience of moving from the East Coast with preconceived notions of what L.A. would be like, "
who is in the cast
Bot:  [8]
Chazelle cast Gosling and Stone immediately after Summit bought the film.
who does emma stone play
Bot:  [8]
Emma Stone plays Mia, an aspiring actress in Los Angeles.
who does ryan gosling play
Bot:  [22]
Ryan Gosling plays Sebastian, a jazz pianist.
bye
Bye!
```

The above dialog shows a returning user asking the chatbot some basic questions about La La Land.

```
Bot: Hello! If you want to talk about La La Land the movie you came to the right place. When you want to exit say Bye.
Bot: First, what's your name?
lucy
Welcome back lucy ! What do you want to talk about this time?
vats
Bot: You said this ealier ' i like the weather today ' let's expand on that.
```

The above dialog shows what happens when a returning user starts to talk about something that is outside the knowledge of the chatbot. The chatbot goes into its user model to find something about that user that it can bring up to keep the conversation going.

Kinsey Mellon

```
Bot: Hello! If you want to talk about La La Land the movie you came to the right place. When you want to exit say Bye.
Bot: First, what's your name?
katie
Hey katie ! Let's get started!
is la la land a good movie
Bot:  La La Land
2016 American romantic musical comedy-drama film by Damien Chazelle
La La Land is a 2016 American romantic musical comedy-drama film written and directed by Damien Chazelle.
did it win awards
Bot:  [65] The next week, following its six Oscar wins, the film grossed $3 million.
who is damien chazelle
Bot:  La La Land
2016 American romantic musical comedy-drama film by Damien Chazelle
La La Land is a 2016 American romantic musical comedy-drama film written and directed by Damien Chazelle.
do you like it
Bot: I'm not sure, how about you?
i think it was ok
Bot:  [17]
Chazelle said that the romantic dinner that Sebastian prepared for Mia was "one of the scenes that I think I wrote and rewrote and rewrote more
bye
```
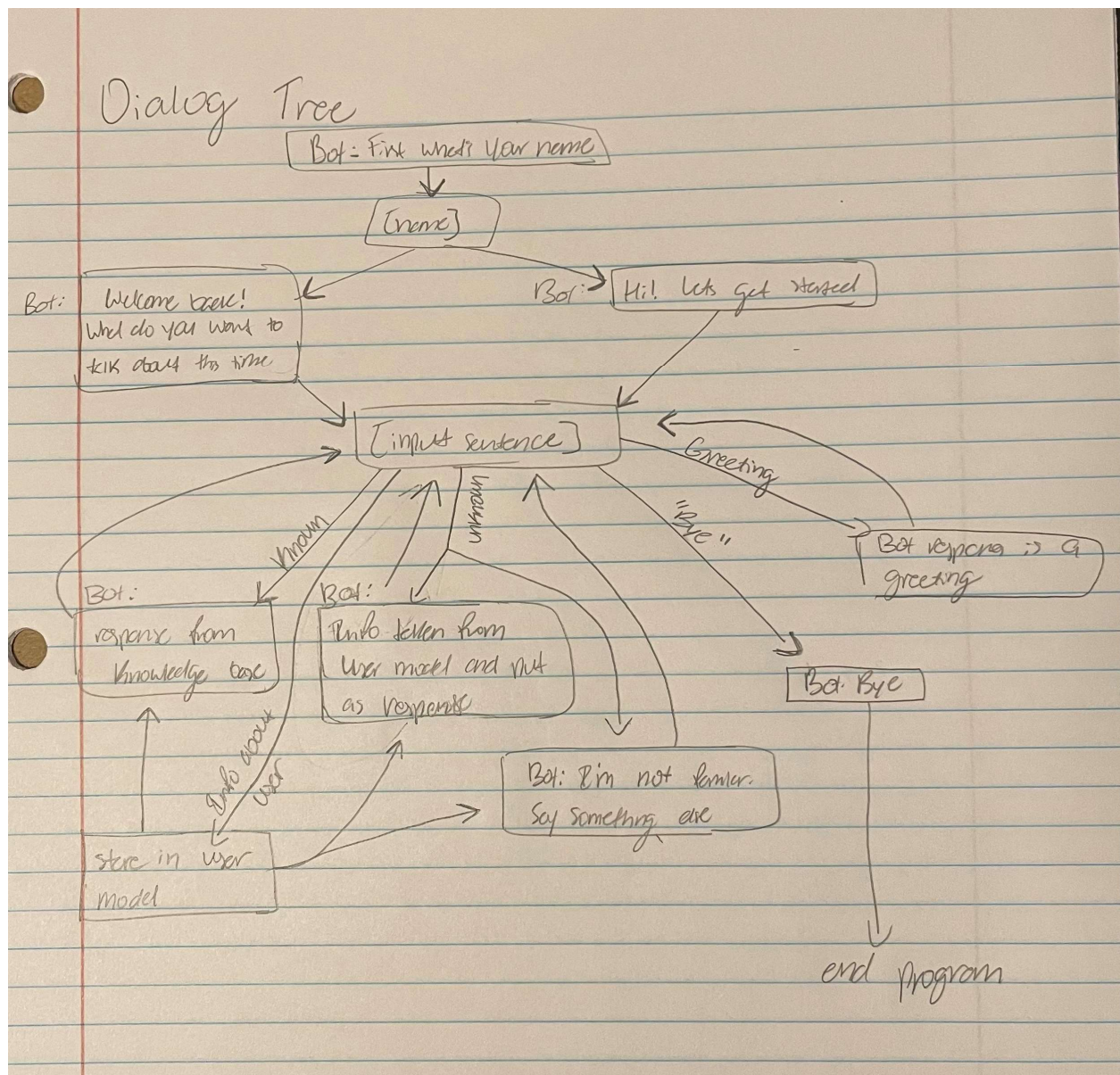
This dialog shows a new user asking the bot some questions about La La Land which it answers, but they also ask what the chatbot thinks and the response of the chatbot when asked about its own opinions is to turn the question back to the user since the chatbot has no real opinions.

User Model:

```
kinsey :  ["i don't like emma stone", 'i like musicals', 'i like ryan gosling']
steve :  ["i don't like musicals", "i'm 45 years old", 'i dislike music', 'i like sports', 'i think musicals are dumb']
lucy :  ['i like the weather today']
```

This is a tiny snippet of the user model. When the chatbot encounters a new user, their name gets added into a dictionary and anything they say about their likes or dislikes or about themselves gets added. At the end of the program the dictionary gets dumped to a pickle file so when the chatbot is run again it's with the updated user model so if there is a returning user any new information is added into the right key in the dictionary.

Kinsey Mellon

Dialog Tree/Diagram:

Chatbot Evaluation:

      This chatbot project is a good example of a simple conversational chatbot. It has a very basic implementation but is executed well and does the job of a chatbot although it's not the most intelligent or high end chatbot. One of the biggest strengths of this chatbot is how it chooses what sentence to respond to the user's input. Since the tfidf (term frequency-inverse document frequency) is collected for each sentence, the chatbot is able to accurately compare the

Kinsey Mellon

input sentence to the knowledge base sentences and choose the most appropriate one. Due to the accuracy of picking the most appropriate response the chatbot is able to have reasonably smooth conversations about the topic of La La Land. There are definitely some weaknesses in this project though. For example, how the user model information is added into the conversation is one. The only time that information is taken from the user model and used in the dialog is if the user says something that is not anywhere close to the information in the knowledge base, then the chatbot prints out something from the user model and asks the user to talk about that instead. Although this is a good way to keep the conversation going when the user says something that the chatbot doesn't know how to respond to, it's not guaranteed to be any better of a response than just a random fact from the knowledge base. I would say a weakness in general is the use of the user model, I think the way it's created and checked throughout the program is good but only using it in that one instance isn't a very effective way to use a user model.