

CSC498 – Introduction to Natural Language Processing

Homework 3: Conversational Agent

Due date: **Saturday November 28** - Midnight

Submission: Blackboard upload.

1. Build a Conversational Agent using Python.

In the present assignment, you are required to build a conversational agent, i.e. a chatbot, that we'll call **Bobbybot**. The assignment consists of the following modules:

Module #1: Load and Preprocess text data.

In the first module, go to https://en.wikipedia.org/wiki/Coronavirus_disease_2019 or to <https://en.wikipedia.org/wiki/Tennis> or pick any other Wikipedia text about a topic of your choice:

1. Copy some text from the page,
2. Paste it in a word document,
3. Save it as .txt file.
 - a. Save the file in the same directory where you saved your jupyter notebook for homework 3.
4. Preprocess and normalize your text.

Module #2: Greeting the bot.

In the second module you will create two lists of strings, one is called: `my_greetings` and the other is called `bot_response`:

1. List: `my_greetings` consists of a selection of greeting words you may say to the bot.
2. List: `bot_response` consists of a selection of replies returned by the bot, *randomly*.
 - a. Upon your input, the bot checks if your input exists in `my_greetings`, if it does, then it picks a random reply from `bot_response`.
3. At any time, if you input an erroneous text, the bot should reply: "I am sorry, I do not understand what you are asking me."

Module #3: Text Vectorization and Similarity Measure.

In this third module, you will ask the bot some questions about the topic you chose in Module 1. For example, if you chose the tennis page on Wikipedia, then suitable questions can be:

- Tell me about Roger Federer
- What is a grand slam tournament?

In this case, the bot will search the text for an appropriate answer to your question. To that end, you should:

1. Vectorize your question and text using TF-IDF technique.
2. Using Cosine similarity measure, return from the text the part of text that has the highest similarity measure with your question.
3. At any time, if you input an erroneous text, the bot should reply: *"I am sorry, I do not understand what you are asking me."*

Module #4: Speech Recognition.

In module 4, you will incorporate a speech recognition task.

Part I: To that end, you need to install on your system the **SpeechRecognition** package. Then:

1. Record your voice uttering a line of task such as a greeting: "Hello Bobby bot" or "What is Covid", or "tell me about the tennis grand slam" (depending on your text).
2. Load the mp3 voice into your jupyter notebook
3. Input it to the speech recognizer
4. Print the text transcription of the line you recorded.
5. Feed this line to your bot, i.e., call the method with the transcribed text, instead of user input (that you did in modules 2 and 3).

Part II: Then, use your microphone to talk directly and have the speech recognizer recognize your utterance. To that end you need to install:

1. portaudio
2. pyaudio
3. Installation instructions for Windows and Mac are found here:
<https://people.csail.mit.edu/hubert/pyaudio/>
4. The text you speak through the microphone, should be captured and sent to the bot, and the bot should reply (Module 3).

Module #5:

In this fifth module, you need to implement a text-to-speech task, where you output the response of the bot through your system's speakers. To this end you need to install:

1. gTTS package of pypi
2. Installation instructions for Windows and Mac are found here:
<https://pypi.org/project/gTTS/>

Module #6:

Finally, at any moment you type and/or speak “Bye” or “Bye Bobby”, the bot should reply with “Bye human” or “Take care” and close the input (you can no longer input text).

- ➔ In fine, when I run your code, I should be able to greet your Bobby bot or ask him a question, by inputting a text and/or by speaking through the microphone. Bobby bot should reply to me by text and/or through the speakers.
- ➔ All the best! Enjoy your first Chatbot!