

Meghana Kambhampati, Alekhya Pinnamaneni

Professor Mazidi

CS 4395 - Human Language Technologies

15 April 2023

Chatbot Evaluation and Report

System Description:

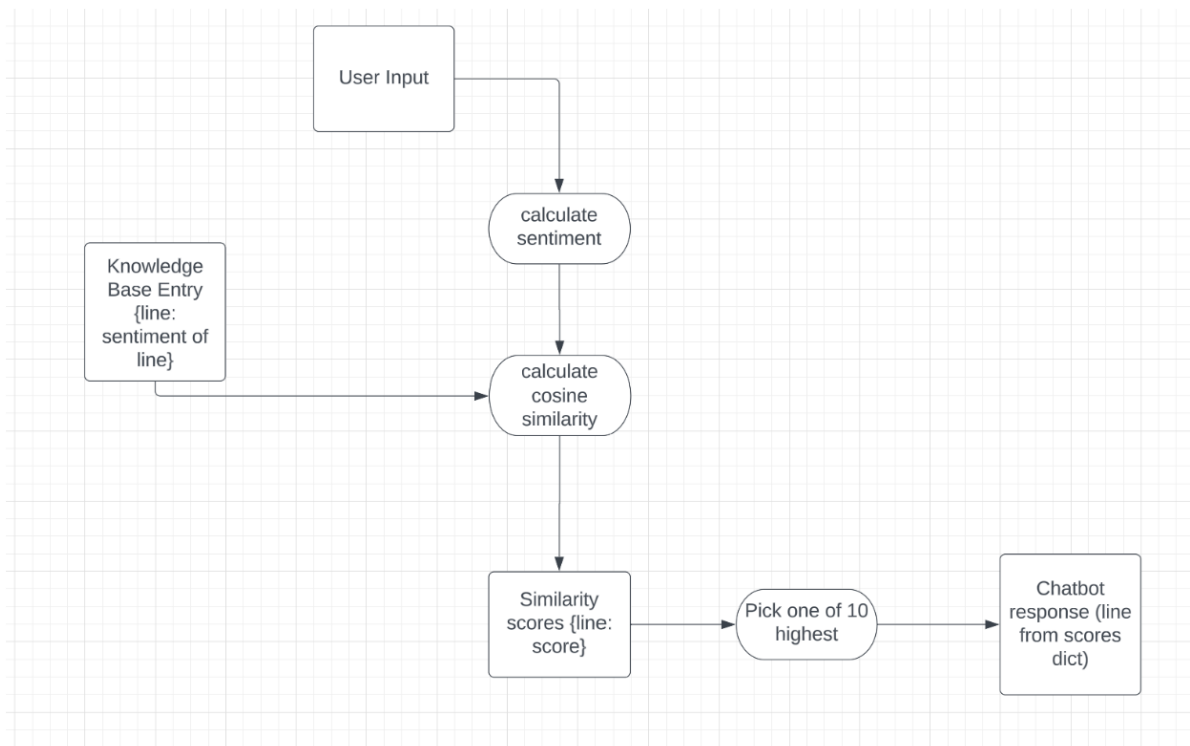
For this project, we wrote a chatbot from scratch that uses the stage play script for *Hamlet* from Project Gutenberg's website as a knowledge base to converse with the user. The first step in the project is to access the Project Gutenberg html page containing the eBook of *Hamlet* by sending a get request to the url. Once the page is retrieved, the BeautifulSoup library is used to create a soup object from the web page. Then, all elements of the html file that are not strings are removed, leaving only text. Next, all strings that are not visible to the user are removed. The next step is to clean up and tokenize the text. To clean up the text, the start and ending lines are specified to remove any irrelevant text before and after the eBook that is not a part of the play. Additionally, any lines containing headers, stage directions, and character names are removed, so that only the actual dialogue from the play is left in the text. This text is then tokenized into sentence tokens. The tokenized text is then used to create a knowledge base. The knowledge base is developed by conducting sentiment analysis on each sentence and storing the sentences with their corresponding sentiment score in a dictionary object. After the knowledge base is created, the chatbot then starts up and

asks the user for their name. Once the user inputs their name, the chatbot checks if a user model has already been created for this user and opens the user's file if it exists. If this user has not previously conversed with the chatbot before, a new user file is created. The chatbot then enters into a loop of outputting continual relevant responses to the user, until the user indicates that the session is concluding ('bye', 'farewell', 'goodbye', etc.). This loop contains code to formulate a response to the user's input. The first step in formulating the chatbot's response is to do sentiment analysis on the user's input using VADER. Then the program uses tf-idf to calculate cosine similarity, to find the best match for the user's input from the entries in the knowledge base. The entry is then output to the user, the user responds, and this loop continues until the user bids farewell to the chatbot. After the chatting session ends, all of the information gathered about the user during the session is written to the user's file and the program ends.

NLP Techniques Used:

Multiple NLP techniques were used in this project. First, the python library BeautifulSoup was used to extract text from the web page containing the *Hamlet* eBook. Nltk's sent_tokenize was used to tokenize the text into lines of dialogue to be used for developing the chatbot's knowledge base. Sentiment analysis with VADER was used to create the knowledge base and to measure the sentiment of the user's input. Cosine similarity and tf-idf were used to calculate similarity scores to select the most relevant response to the user's input.

Dialog Logic:



Sample Dialog 1:

BOT:

What is thy name?

USER:

Bob

BOT:

How fares my lord?

USER:

I'm great!

BOT:

In the most high and palmy state of Rome, A little ere the mightiest Julius fell, The graves stood tenantless and the sheeted dead Did squeak and gibber in the Roman streets; As stars with trains of fire and dews of blood, Disasters in the sun; and the moist star, Upon whose influence Neptune's empire stands, Was sick almost to doomsday with eclipse.

BOB:

Are you ok

BOT:

But, woe is me, you are so sick of late, So far from cheer and from your former state, That I distrust you.

BOB:

Wow what did I do :(

BOT:

If he be now return'd, As checking at his voyage, and that he means No more to undertake it, I will work him To exploit, now ripe in my device, Under the which he shall not choose but fall; And for his death no wind shall breathe, But even his mother shall uncharge the practice And call it accident.

BOB:

So this sounds a little murder-y

BOT:

Murder!

BOB:

Riiiiiiight ok then

BOT:

Up, sword, and know thou a more horrid hent: When he is drunk asleep; or in his rage,
Or in th'incestuous pleasure of his bed, At gaming, swearing; or about some act That
has no relish of salvation in't, Then trip him, that his heels may kick at heaven, And that
his soul may be as damn'd and black As hell, whereto it goes.

BOB:

Goodbye

Fare ye well, Bob

Sample User Model 1 (Bob):

[[["I'm great!", 0.6588], ["Are you ok", 0.296], ["Wow what did I do :(", 0.2263], ["So this
sounds a little murder-y", 0.0], ["Riiiiiiight ok then", 0.296]]

Sample Dialogue 2:

BOT:

What is thy name?

USER:

Meghana

BOT:

How fares my lord?

USER:

I'm ok

BOT:

And your water is a sore decayer of your whoreson dead body.

MEGHANA:

Excuse me? How rude

BOT:

But thou wouldst not think how ill all's here about my heart: but it is no matter.

MEGHANA:

I do not care

BOT:

Give me that man That is not passion's slave, and I will wear him In my heart's core, ay,
in my heart of heart, As I do thee.

MEGHANA:

Flattery will get you everywhere

BOT:

Who would these fardels bear, To grunt and sweat under a weary life, But that the dread
of something after death, The undiscover'd country, from whose bourn No traveller
returns, puzzles the will, And makes us rather bear those ills we have Than fly to others
that we know not of?

MEGHANA:

I don't follow

BOT:

Follow her close; give her good watch, I pray you.

MEGHANA:

WHO?

BOT:

Peace, who comes here?

MEGHANA:

Meghana

BOT:

But that I love thee best, O most best, believe it.

MEGHANA:

Are you flirting with me

BOT:

But, woe is me, you are so sick of late, So far from cheer and from your former state,
That I distrust you.

MEGHANA:

Ok that was unexpected

BOT:

You must not put another scandal on him, That he is open to incontinency; That's not
my meaning: but breathe his faults so quaintly That they may seem the taints of liberty;
The flash and outbreak of a fiery mind, A savageness in unreclaimed blood, Of general
assault.

MEGHANA:

Goodbye

Fare ye well, Meghana

Sample User Model 2 (Meghana):

```
[["I'm ok", 0.296], ["Excuse me? How rude", -0.4019], ["I do not care", -0.3875], ["Flattery will get you everywhere", 0.1027], ["I don't follow", 0.0], ["WHO?", 0.0], ["Meghana", 0.0], ["Are you flirting with me", 0.2023], ["Ok that was unexpected", 0.296]]
```

Knowledge Base:

The knowledge base is a dictionary, storing each dialogue (in order from the first to the last dialogue) as the key and its corresponding sentiment score as the value. (e.g. 'line': -0.87)

```
{  
    'A platform before the Castle.': 0.0,  
    'Elsinore.': 0.0,  
    ...,  
    'Heaven secure him.': 0.6908,  
    ...,  
    '[ A dead march. ]': -0.6486,  
    '[ Exeunt, bearing off the bodies, after which a peal of ordnance is shot off. ]': 0.0  
}
```

Evaluation:

The chatbot does a good job of staying consistent with *Hamlet's* tone of existential dread throughout the conversation, due to its heavy focus on sentiment analysis. The use of tf-idf and cosine similarity allow the chatbot to reciprocate the sentiment and mood of the user efficiently and accurately. The chatbot follows a simple algorithm to

decide which response from the knowledge base will be the most appropriate and relevant. However due to the simple nature of the chatbot's algorithm, some of the chatbot's responses can occasionally be puzzling to the user. Additionally since the knowledge base is entirely made up of dialogue from *Hamlet*, the chatbot struggles with carrying a conversation about a specific topic. Although the chatbot does not have deep knowledge on specific topics, it is a great tool for general chatting and amusing small talk in Shakespearean English.