CHATBOT ON JR.NTR

# Web crawling:

Web crawling is the process hyperlinks present in the url . initially I took 3 links related to jr.NTR.

start\_urls=['https://en.wikipedia.org/wiki/N.\_T.\_Rama\_Rao\_Jr.','https://en.wikipedia.org/wiki/N.\_T.\_Rama\_Rao\_Jr.\_filmography','https://en.wikipedia.org/wiki/List\_of\_awards\_and\_nominations\_received\_by\_N.\_T.\_Rama\_Rao\_Jr.']

max\_urls=15

then, I extracted all the links which are related to the topic present in those urls using crawl function using beautiful soup.

# Web scraping:

The purpose of web scraping is to extract all the data in the crawled urls and store the data in different files. The process is done using beautiful soup. After performing beautiful soup operation on the ‘response.text’ where response is the output of sending get request to the url, I extracted all the paragraphs using ‘soup.find\_all(‘p’)’. Then I extracted all the text from each paragraph. For every link , I created separate file and stored the data in ‘scraped\_texts\_NTR’ directory.

response = requests.get(url)  # Send a GET request to the URL

            soup = BeautifulSoup(response.text, 'html.parser')  # Parse the HTML content

            text=' '

            paragraphs=soup.find\_all('p')

            for para in paragraphs:

                text += para.get\_text().lower()

            # Write the extracted text to a file

            filename = os.path.join(output\_dir, f"file\_{i}.txt")

            with open(filename, 'w', encoding='utf-8') as f:

                f.write(text)

# Clean text:

The purpose of cleaning text is to tokenize the each file and to find most important terms. For this process, I created a separate file cleantext.py where I performed preprocessing of data in each file. first I performed tokenization, lemmatization on the tokens, and filtered the tokens which are not stop words and the tokens which are not string punctuations and stored in separate directory(cleaned\_texts\_NTR).

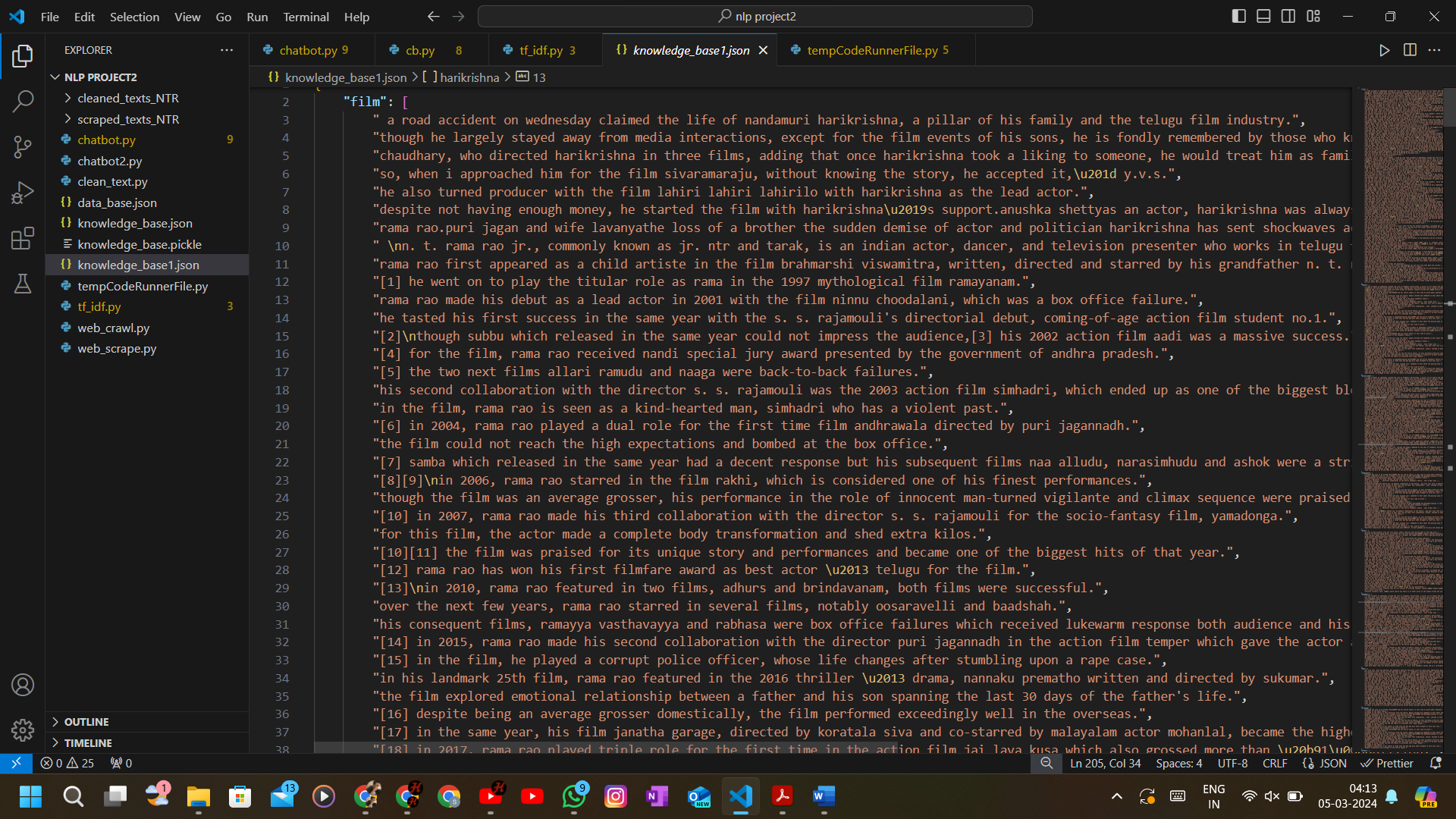
# Finding important terms:

First, I calculated the tf-idf scores of the preprocessed texts using tf-idf vectorizer.After calculating the scores,I sorted the terms based on the scores in descending order. And then selected the most important 15 terms based on the scores. In that important terms, I removed some unwanted words and added some keywords.

['film', 'family', 'harikrishna', 'condolence', 'actor', 'nandamuri', 'year', 'strength', 'action', 'soul', 'kalyan ram', 'NTR', 'lakshmi pranathi', 'nandamuri taraka rama rao']

# Knowledge base:

After finding most important terms, I created knowledge base based on the important words and the scraped files.



# CHATBOT:

## description:

I created a chatbot named Loki which is designed to provide the information about an actor Jr.NTR, who is known for his performances in films like ‘RRR’.

## NLP Techniques:

Word tokenization: splitting the text into list of words

Sentence tokenization: splitting text into list of sentences

I used these nlp techniques in preprocessing of the text

Cosine similarity:finding the similarity of two documents

In this project, cosine similarity is used to find the sentence which is relevant to user query.

CHATBOT FLOWCHART

Step1: check user in database if yes proceed

Else create new user

Step2: check if query is bye if true exit

Else

If query is in greeting call greeting function

Else if query matches with welcome messages give the value related to it

Else call respond function and print the output

Step3: give another query and go to step2

One or more sample dialog interactions

Hello, my name is Loki,the chatbot designed to provide the information about jr.ntr . What is your name? : harsha

welcome back harsha.

Loki: Hi harsha, what do you want to know about ntr?. If you want to exit, type Bye. :

You:hi how are you

Loki: hi there harsha

You:how are you

Loki: I am fine. Thankyou for asking

You:how are you doing

Loki: I am fine. Thankyou for asking

You:who is ntr

Loki:

when harikrishna opted for a second marriage with shalini (jr ntr’s mother), ntr didn’t accept initially.

You:awards of ntr

Loki:

one of the highest paid telugu film actors, rama rao has won several accolades, including two filmfare awards, two state nandi awards, and four cinemaa awards.

You:what is the highest grodssing film of tarak

Loki:

the film registered the highest opening for a tollywood film in 2016, and became the highest grossing telugu film of the year.

You:bye

Loki: This is Loki signing off. Bye, take care harsha

"kalyan ram": [

        "my condolences to balakrishna garu, tarak, kalyan ram & the whole nandamuri family.",

        "deepest condolences to tarak, kalyan ram and their entire family.",

        "deepest condolences to tarak sir kalyan ram sir and the entire family..at this extremely difficult time lets pray god to grant them the peace n strength that they need to get through this.",

        "strength to tarak, kalyan ram and the family to withstand this tragic moment.",

        "my condolences to balakrishna garu, tarak, kalyan ram & the whole nandamuri family.",

        "deepest condolences to tarak, kalyan ram and their entire family.",

        "deepest condolences to tarak sir kalyan ram sir and the entire family..at this extremely difficult time lets pray god to grant them the peace n strength that they need to get through this.",

        "strength to tarak, kalyan ram and the family to withstand this tragic moment.",

        "[11] he is the half-brother of actor and producer, nandamuri kalyan ram,[12] the nephew of actor and politician nandamuri balakrishna and former chief minister of andhra pradesh, n. chandrababu naidu, and the cousin of actor taraka ratna and politician nara lokesh.",

        "arts banner with his half-brother nandamuri kalyan ram as the producer.",

        "[46] both these films are being produced under n. t. r. arts, with nandamuri kalyan ram as a co-producer."

    ],

    "lakshmi pranathi": [

        "[54]\nrama rao married lakshmi pranathi, the daughter of realtor and businessman narne srinivasa rao."

    ],

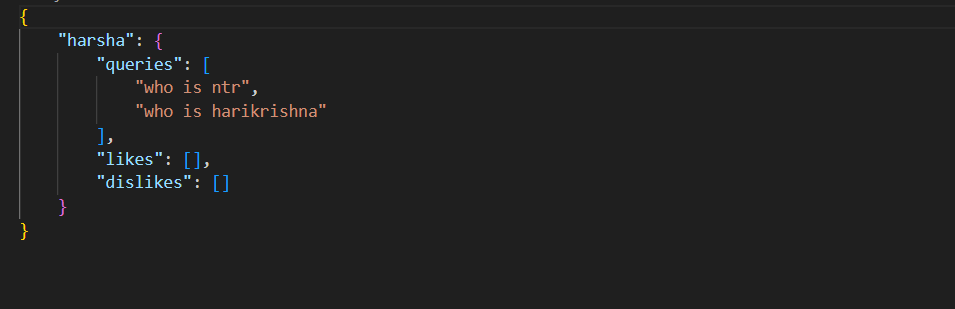
    "nandamuri taraka rama rao": [

        " \nnandamuri taraka rama rao jr. (born 20 may 1983), also known as jr. n.t.r."

This is the sample knowledge base created

An appendix for sample user models that were created

The following is the appendix for the sample user model which stores user queries,likes and dislikes



Strengths of chatbot:

* Prompt response time and good performance
* Strong natural language understanding capabilities

Weakness of chatbot:

* Irrelevant responses
* Lack of support for other languages

Evaluation :

