Part 2 – Chatbot

**Summary:**

The chatbot part of the project involves accessing the knowledge base and responding to the user queries with the help of both hardcoded tailored responses as well as by parsing the knowledge base and returning the best matches.

**Program Flow:**

* Data Retrieval: It retrieves pre-processed data related to Keanu Reeves' movies and TV shows from pickles, the text data from all corpus files and the tf-idf dictionary.
* User Interaction: The program starts by interacting with the user, asking for their name, favorite color, and birth year and stores it in text and pickle files.
* Query Processing: The program processes user queries, applying NLP techniques to understand and respond to them.
* Response Generation: Based on the processed queries, the program generates and outputs appropriate responses.

**Function Descriptions:**

* retrieve\_data():

Loads pre-processed data from pickle files, including information about Keanu Reeves' movies, TV shows, and other text files.

* preprocess\_text(text):

Preprocesses a given text by tokenizing, converting to lowercase, removing stopwords, and keeping only alphanumeric tokens.

* cosim(user\_response, sentence):

Calculates the cosine similarity between a user's response and a sentence from the corpus to determine relevance.

* get\_sentences(query):

Retrieves relevant sentences from the corpus based on the user's query, using cosine similarity and keyword matching.

* movies\_print(user\_year, flag) and tvshows\_print(user\_year, flag):

Prints information about Keanu Reeves' movies or TV shows from a specific year or a random year if the user's input year is not available.

* response\_parse(user\_name, user\_year):

Parses the user's query and generates an appropriate response based on the pre-processed data and the user's input.

* create\_user\_def():

The main function that initiates user interaction, collects user information, and handles the conversation flow.

**NLP Tools Used:**

* **NLTK (Natural Language Toolkit):** Used for tokenization, stopwords removal, and other text processing tasks.
* **Autocorrect:** Used to correct spelling errors in user inputs. Additionally, the speller from this library is more accurate than Textblob’s autocorrect.
* **TF-IDF Vectorizer:** Used to convert text data into numerical form (vector) for calculating cosine similarity.
* **Cosine\_similarity:** Library to calculate the similarity between two vectors to return relevant sentences.
* **Sentiment Analysis (VADER):** Used to analyze the sentiment of user queries to generate positive or negative responses accordingly.
* **Pickle:** Used to store and retrieve pre-processed data for efficient access during runtime.

**Extra Adjustments:**

* Autocorrect changed Keanu’s name to ‘means’ so substituted it back with regex.
* Substituted the terms ‘mom’, ‘dad’, ‘birthday’ to ‘mother’, ‘father’, ‘born’ as the former terms are not available in the corpuses.
* Used tailored responses for questions about chatbot and basic questions about Keanu Reeves.

**Dialogue Tree:**

**A screenshot of a computer

Description automatically generated**

**Screenshots:**

* Usermodel text file:

**A screenshot of a computer

Description automatically generated** **A screenshot of a computer

Description automatically generated**

A screenshot of a computer

Description automatically generated

* Usermodel function interaction:

**A screenshot of a computer screen

Description automatically generated**

* Remembering user model:

A screenshot of a computer screen

Description automatically generated

* Tailored responses interactions and sentiment analysis score-based responses:

**A black screen with white text

Description automatically generated**

**A screenshot of a computer program

Description automatically generated**

* Queries that provide response with information extracted from the knowledge base.

A black screen with white text

Description automatically generated

A screenshot of a computer screen

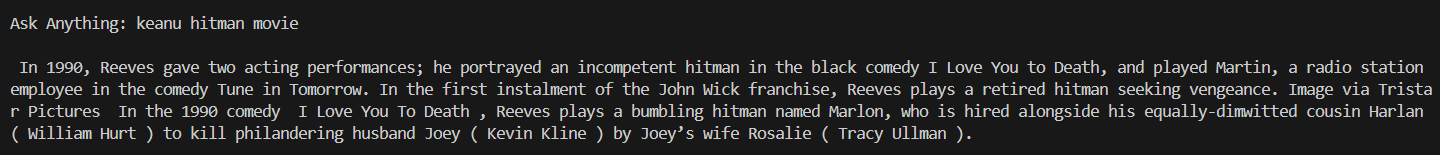
Description automatically generated

A black background with text

Description automatically generated

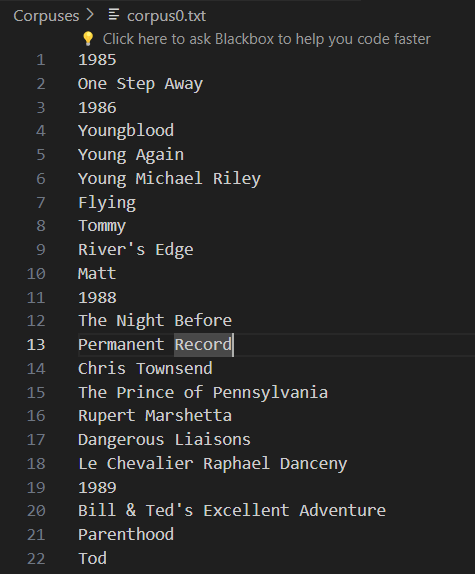
A black background with white text

Description automatically generated



**Knowledge base Appendix:**

* corpus1.txt:

* List of documents in a dictionary with keys- ‘doc1’, ‘doc2’ etc and sentence tokenized

A screen shot of a computer screen

Description automatically generated

* movies.pickle

A screen shot of a computer

Description automatically generated

* tvshows.pickle

A screen shot of a computer

Description automatically generated

* Important words list:

A black background with white text

Description automatically generated

**Evaluations:**

**Strengths and Weaknesses:**

Weaknesses:

1. Does not produce output if the query contains only stop words.
2. Very limited questions are answered.
3. Random sentences are generated along with relevant ones.

Strengths:

1. Produces answer to query with good enough similarity.
2. Answers basic questions perfectly.
3. Notices user sentiments and produces tailored empathetic answers.
4. Response time is pretty good with less than 10 seconds.
5. Output format is readable and does not generate non-alphanumeric characters.

**Likert Survey:**

**Person 1:**

1. How well did the chatbot understand your questions?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Very well | Well | Neutral | Unwell | Very unwell |

1. How relevant were the chatbot's responses to your queries?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Very relevant | Relevant | Neutral | Irrelevant | Very irrelevant |

1. How satisfied are you with the speed of the chatbot's responses?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Very Satisfied | Satisfied | Neutral | Not Satisfied | Very Unsatisfied |

1. How helpful was the chatbot in providing information about Keanu Reeves?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Very helpful | Helpful | Neutral | Unhelpful | Very unhelpful |

1. How natural did the conversation with the chatbot feel?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Very natural | Natural | Neutral | Unnatural | Very unnatural |

**Person 2:**

1. How well did the chatbot understand your questions?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Very well | Well | Neutral | Unwell | Very unwell |

1. How relevant were the chatbot's responses to your queries?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Very relevant | Relevant | Neutral | Irrelevant | Very irrelevant |

1. How satisfied are you with the speed of the chatbot's responses?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Very Satisfied | Satisfied | Neutral | Not Satisfied | Very Unsatisfied |

1. How helpful was the chatbot in providing information about Keanu Reeves?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Very helpful | Helpful | Neutral | Unhelpful | Very unhelpful |

1. How natural did the conversation with the chatbot feel?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Very natural | Natural | Neutral | Unnatural | Very unnatural |

**Person 3:**

1. How well did the chatbot understand your questions?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Very well | Well | Neutral | Unwell | Very unwell |

1. How relevant were the chatbot's responses to your queries?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Very relevant | Relevant | Neutral | Irrelevant | Very irrelevant |

1. How satisfied are you with the speed of the chatbot's responses?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Very Satisfied | Satisfied | Neutral | Not Satisfied | Very Unsatisfied |

1. How helpful was the chatbot in providing information about Keanu Reeves?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Very helpful | Helpful | Neutral | Unhelpful | Very unhelpful |

1. How natural did the conversation with the chatbot feel?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Very Natural | Natural | Neutral | Unnatural | Very unnatural |