

CASEBOT : PROJECT1

REPORT PART -2

Name: Sahrudayi Caroline P.
NETID: SXP230055

- **System Description:**

- The chatbot, named CaSe, is a rule and retrieval based conversational agent designed to provide information about **Christopher Nolan and his movies, critic reviews and awards**. Some of his most notable movies are Oppenheimer, Interstellar, Inception, Dark Knight Trilogy, Batman begins, Dunkirk, Tenet, Memento which are handled by the chatbot.
- It utilizes Natural Language Processing (NLP) techniques to understand user input and generate appropriate responses. The NLP techniques used include named entity recognition (NER), and Information Retrieval.

- CaSebot uses the nltk tokenization and named entity recognition in spacy to determine what and where a proper noun might be, for the purpose of parsing a name from a sentence

- **Data Retrieval and Preprocessing:**

- The script connects to a SQLite database (database_final_latest.db) containing sentences associated with specific keywords.
- User input and sentences from the database are preprocessed to remove stopwords, punctuation, and to lemmatize the words using NLTK's WordNetLemmatizer and stopwords corpus.

- **Cosine Similarity Calculation:**

- The calculate_cosine_similarity function computes the cosine similarity between the preprocessed user input and preprocessed sentences retrieved from the database that have matching key words with terms in user input
- It first converts the text data into a matrix of TF-IDF (Term Frequency-Inverse Document Frequency) features using TfidfVectorizer from sklearn.feature_extraction.text.
- Then, it calculates cosine similarity between the user input and each sentence using cosine_similarity from sklearn.metrics.pairwise.

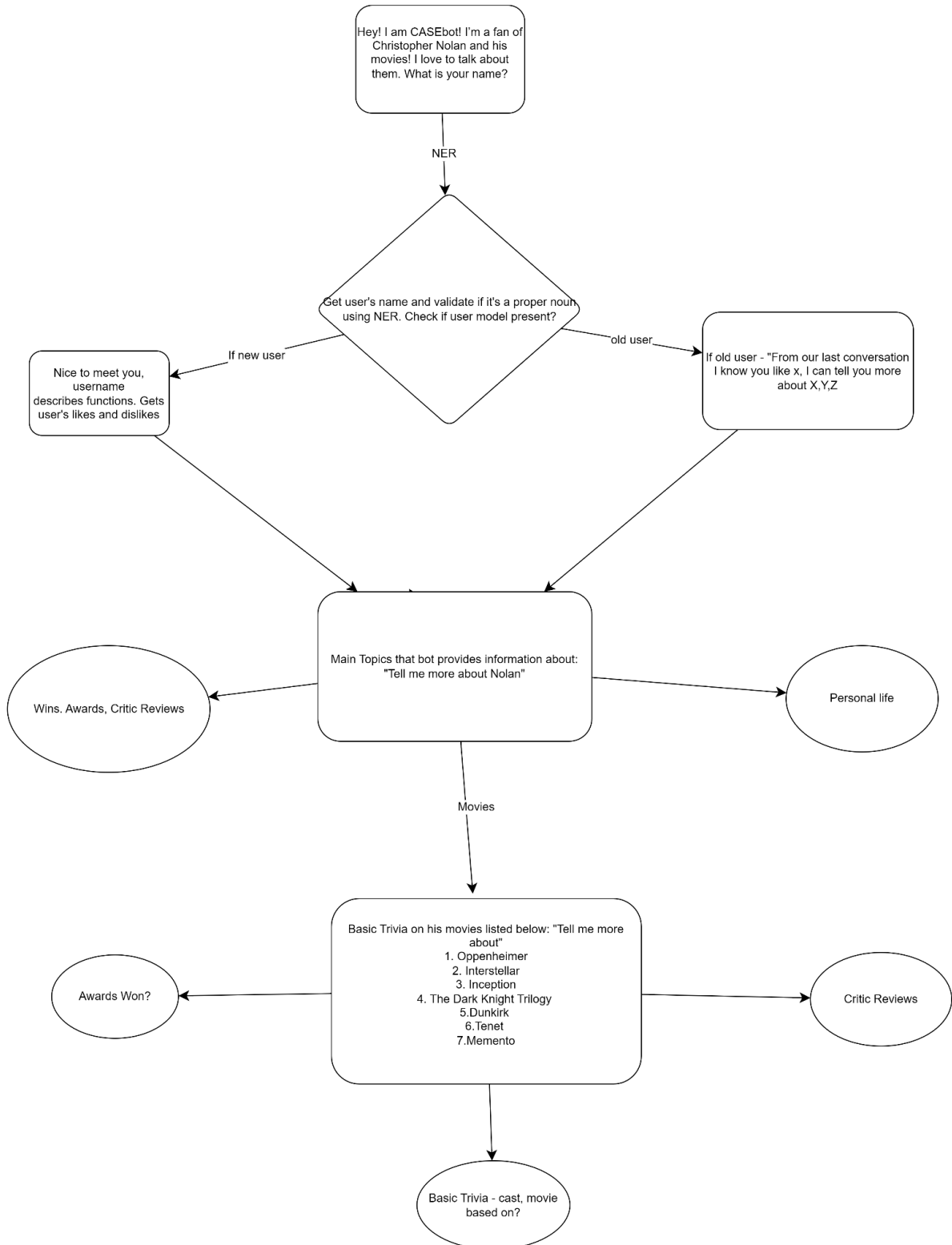
```
44
45 def preprocess(text):
46     #returns all tokens joined into a string
47     lemmatizer = WordNetLemmatizer()
48     stop_words = set(stopwords.words('english'))
49     tokens = word_tokenize(text.lower())
50     filtered_tokens = [lemmatizer.lemmatize(token) for token in tokens if token.isalpha() and token not in stop_words]
51     return ' '.join(filtered_tokens)
52
53 def calculate_cosine_similarity(user_input, sentences):
54     vectorizer = TfidfVectorizer()
55     tfidf_matrix = vectorizer.fit_transform([user_input] + sentences)
56     cosine_similarities = cosine_similarity(tfidf_matrix[0:1], tfidf_matrix[1:]).flatten()
57     return cosine_similarities
58
59 def remove_numbers_in_square_brackets(sentence):
60     return re.sub(r'\[[0-9]+\]', '', sentence)
61
62 def respond_to_input(user_input):
63     preprocessed_input = preprocess(user_input)
64     # print("preprocessed user input", preprocessed_input)
65     keywords = get_keywords_from_database()
66     matching_sentences = []
67     try:
68         # Retrieve all sentences that match each term in the user input with its corresponding keyword
69         for term in preprocessed_input.split():
70             if term in keywords:
71                 matching_sentences.extend(get_sentences_from_database(term))
72
73     # Calculate cosine similarity between user input and matching sentences
```

Chatbot code implementation of data retrieval from the knowledge base

- **Response Generation:**

- The function `respond_to_input` processes the user input, retrieves relevant sentences from the database based on keywords, calculates cosine similarity, and selects the most similar sentence as the response.
- If there's an error (e.g., no matching sentences or `ValueError`), it generates a random response using the `generate_random_response` function.

Dialog Tree:



Sample Dialog Interactions:

You: Can you tell me your name?

I am Priyanka

I'm unable to get your name from your input. Can you please tell me your name and just enter your name by itself?

You: Priyanka

Nice to meet you, Priyanka! As a fan bot of Nolan, I can tell you information about Christopher Nolan and his movies, famous characters, critic reviews and awards. Some of Nolan's most notable movies are Oppenheimer, Interstellar, Inception, Dark Knight Trilogy, Dunkirk, Tenet, Memento. I'd like to understand your interests first!

What are the topics you like from the information I can provide? (Enter as comma-separated list): Oppenheimer, Dunkirk

What are the topics you dislike from the information I can provide? (Enter as comma-separated list) memento

I see that you like Oppenheimer, Dunkirk. I can tell you more about Nolan, his movies, awards and critic reviews. Ask away!

You: tell me about Nolan

CaSeBot: Christopher Edward Nolan CBE (born 30 July 1970) is a British and American filmmaker. He gained widespread recognition for directing blockbusters like "Oppenheimer", "Interstellar", "Inception," and "The Dark Knight Trilogy". Nolan's work often explores themes of time, identity, and morality, earning him critical acclaim and a dedicated fanbase.

You: how old is Nolan?

CaSeBot: Christopher Edward Nolan is 53 years

You: how many movies has Nolan directed?

CaSeBot: Christopher Nolan has directed twelve (12) movies

You: who is he married to?

CaSeBot: Could you please provide more information?

You: who is Nolan married to?

CaSeBot: Nolan is married to Emma Thomas, whom he met at University College London when he was 19.

You: who is Nolan's father?

CaSeBot: His father, Brendan James Nolan, was a British advertising executive who worked as a creative director.

You: what was Nolan's first big film?

CaSeBot: Nolan on the jump from his first film to his second (2012) Following's success afforded Nolan the opportunity to make Memento (2000), which became his breakthrough film.

You: what was Nolan's last movie?

CaSeBot: Christopher Nolan has directed twelve (12) movies

You: what was Nolan's last film?

CaSeBot: Oppenheimer is the last recent film directed by Nolan. It was his 12th film, a biopic based on J. Robert Oppenheimer (played by Cillian Murphy) and his role in the development of the atom bomb.

Hello! I'm CaSe, a Christopher Nolan fan bot!

You: Can you tell me your name?

I am Priyanka

I'm unable to get your name from your input. Can you please tell me your name and just enter your name by itself?

You: Priyanka

CaSeBot: Welcome back, Priyanka!

I see that you like Oppenheimer, Dunkirk. I can tell you more about Nolan, his movies, awards and critic reviews. Ask away!

You: when was Oppenheimer released?

CaSeBot: Oppenheimer was released theatrically on July 21, 2023, by Universal Pictures.

You: what book is Oppenheimer based on?

CaSeBot: Oppenheimer is based on the 2005 book American Prometheus by Kai Bird and Martin J. Sherwin.

You: Tell me about Inception?

CaSeBot: Inception is a 2010 science fiction action film written and directed by Christopher Nolan, who also produced the film with Emma Thomas, his wife.

You: The main cast in Inception?

CaSeBot: The main cast includes Leonardo DiCaprio, Joseph Gordon-Levitt, Ellen Page - Skilled thieves who enter the dreams of others to steal secrets.

You: how many awards did Inception win?

CaSeBot: Many critics and media outlets included Inception in their rankings of the best films of the 2010s.

You: Critic reviews about Inception?

CaSebot: Many critics and media outlets included Inception in their rankings of the best films of the 2010s.

You: what is a dream in inception?

CaSebot: Roger Ebert of the Chicago Sun-Times awarded the film a full four stars and said that Inception "is all about process, about fighting our way through enveloping sheets of reality and dream, reality within dreams, dreams without reality.

Appendix:

Knowledge Base: Includes information about Christopher Nolan, his movies, characters, awards, and critical reviews

	sentence	keyword ▼1
	Filter	Nolan✕
1	Christopher Edward Nolan CBE (born 30 July 1970) is a British and American filmmaker....	nolan
2	Known for his Hollywood blockbusters with complex storytelling, Nolan is considered a leading filmmaker of the 21st century....	nolan
3	Nolan developed an interest in filmmaking from a young age...	nolan
4	Nolan has made some of the most influential and popular films of his time....	nolan
5	Parting ways with longtime distributor Warner Bros. Pictures, Nolan signed with Universal Pictures for the biopic Oppenheimer (2023), which won him a Golden Globe Award and a BAFTA Award for Best Directio...	nolan
6	Christopher Edward Nolan was born on 30 July 1970, in Westminster, London....	nolan
7	His father, Brendan James Nolan, was a British advertising executive who worked as a creative director....	nolan
8	Growing up, Nolan was particularly influenced by the work of Ridley Scott and the science fiction films 2001: A Space Odyssey (1968) and Star Wars (1977)....	nolan
9	Nolan began making films at the age of seven, borrowing his father's Super 8 camera and shooting short films with his action figures....	nolan
10	Between 1981 and 1983, Nolan enrolled at Barrow Hills, a Catholic prep school in Weybridge, Surrey....	nolan
11	Nolan and Roko co-directed the surreal 8 mm Tarentella (1989), which was shown on Image Union, an independent film and video showcase on the Public Broadcasting Service....	nolan
12	Nolan was president of the Union's Film Society, and with Emma Thomas (his girlfriend and future wife) he screened feature films in 35mm during the school year and used the money earned to produce 16 mm...	nolan
13	After earning his bachelor's degree in English literature in 1993, Nolan worked as a script reader, camera operator and director of corporate films and industrial films....	nolan
14	Funded by Nolan and shot with the UCL Union Film society's equipment, it appeared at the Cambridge Film Festival in 1996 and is considered one of UCL's best shorts....	nolan
15	Nolan filmed a third short, Doodlebug (1997), about a man seemingly chasing an insect with his shoe, only to discover that it is a miniature of himself....	nolan
16	Nolan and Thomas first attempted to make a feature in the mid-1990s with Larry Mahoney, which they scrapped....	nolan
17	Shortly after abandoning Larry Mahoney, Nolan conceived the idea for his first feature, Following (1998), which he wrote, directed, photographed and edited....	nolan
18	Following won several awards during its festival run and was well received by critics who labelled Nolan a majorly talented debutant....	nolan
19	Nolan on the jump from his first film to his second (2012) Following's success afforded Nolan the opportunity to make Memento (2000), which became his breakthrough film....	nolan
20	Jonathan worked the idea into a short story, "Memento Mori" (2001), and Nolan developed it into a screenplay that told the story in reverse....	nolan
21	In the book The Philosophy of Neo-Noir, Basil Smith drew a comparison with John Locke's An Essay Concerning Human Understanding, which argues that conscious memories constitute our identities – a theme ...	nolan

Sample User Models: Profiles of users with different interests, dislikes and preferences.

Name: Priyanka

Likes: Oppenheimer, Dunkirk

Dislikes: memento

User Model - Priyanka.txt

Evaluations and Analysis:

Strengths:

- Ability to understand user interests and provide relevant information.across different interests- nolan and basic questions about 7 of his movies are covered in the corpus & knowledge base
- Dynamic dialog management for engaging conversation flow.
- Detailed knowledge base with accurate information about Nolan and his movies.

Weaknesses:

- Limited ability to handle ambiguous or complex queries.
- Lack of support for contextual understanding beyond current conversation.
- Reliance on pre-defined dialog tree limits flexibility in handling unexpected user inputs. Bot generates a random response but breaks the conversation flow and drifts from the topic

Survey:

Survey on Chatbot Experience

Thank you for taking the time to provide feedback on your experience with our chatbot, CaSe! Please rate the following statements based on your interaction with the chatbot:

	Priyanka D	Riya C	Yogeswari S
Overall Satisfaction	4	5	5
Ability to hold Conversation	5	4	5
Ease of use	5	5	5
Any additional Comments	I liked that I can ask the bot about various Nolan movies	Can handle context better	Made me interested to watch interstellar again