

Movie Discuss Chatbot

System Description

The goal of this project was to create a chatbot that is able to discuss a topic and remember things about the user. My focus was on movies when building this chatbot and the goal was to recommend movies to users based on their likes, dislikes, and search history. The chatbot was built using natural language processing techniques such as text tokenization, part-of-speech tagging, sentiment analysis, and intent classification. The chatbot utilized the TMDb API to fetch data about popular movies and use it as a knowledge base. The chatbot has a simple question-answering architecture that followed a rule-based system. It responds to user inputs by searching its knowledge base and returning relevant information to the user's input. The chatbot can handle a variety of user inputs including requests for movie recommendations, questions about genres, movie titles, and user preferences.

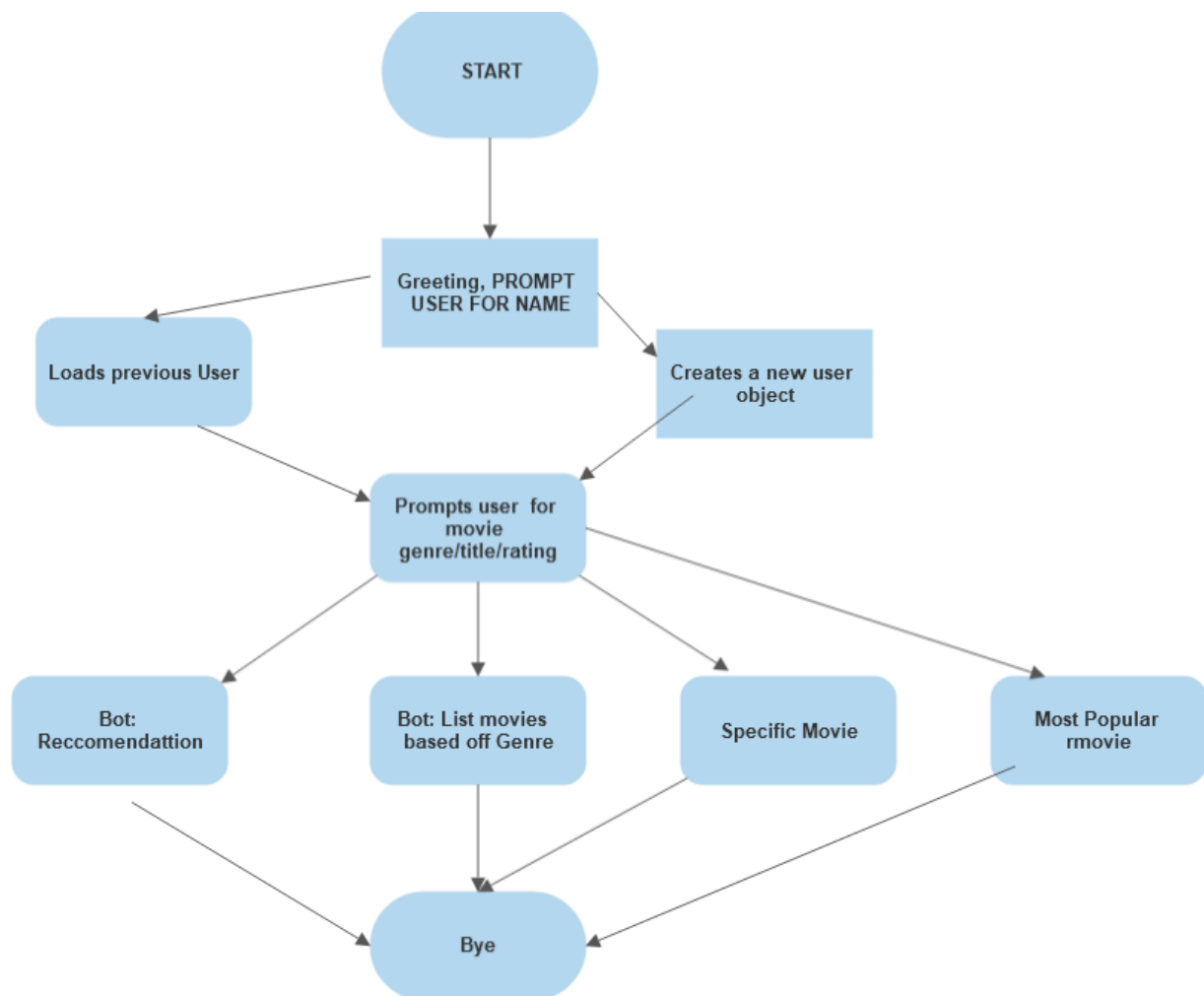
When a user inputs a message, the chatbot will preprocess the message by tokenizing the text and removing stopwords. The chatbot then uses part of speech tagging to identify any positive or negative words in the user's response. If any positive or negative words are identified, the chatbot labels the response as "like" or "dislike" which are then stored within the user's information. The chatbot then uses "sentiment analysis" to determine the overall sentiment score of the user's response. If the sentiment score is positive, then the chatbot checks if any preprocessed tokens match the knowledge base. If a match is found, the chatbot adds the item to the list of likes or dislikes depending on if the score is positive or negative.

Lastly, the chatbot then uses intent classification to determine the user intent. The chatbot is rule-based so it's trained on a set of predefined intents. This includes the recommendations, likes, dislikes, and user search history. The chatbot will search the knowledge base to return a list of information regarding the genre the user selected while responding to the user's intent.

The NLP techniques used when building this chatbot were:

- Text tokenization – The chatbot tokenized the user inputs words by splitting it.
- Parse User Response
- Part-of-Speech tagging: The chatbot used the 'pos_tag' function from the NLTK library to tag parts of speech from the input. This allowed the chatbot to identify information from positive and negative based on the user's response.
- Sentiment Analysis: Utilized the TextBlob library to perform sentiment analysis on the user input. The chatbot calculates the sentiment score for each token in the input and then calculates the overall sentiment.
- Lemmatization: Grouped together words with the same base meaning
- Classification: the chatbot uses a rule-based pattern that classifies the user's intents. It's trained on a set of rules and searches the patterns in the input to match the intent

Dialog Tree Diagram



Alan Perez
AXP200075
CS4395.001

Sample Dialog Interactions

Hi, I'm MovieBot, I love to talk about movies! What is your name?

You: Jacky

Sorry, I didn't catch your name. Can you please tell me your name again?

You: My name is Jacky

Hello, Jacky. What genre of movies do you enjoy?

You: I really love drama

Added drama to likes

sentiment score 0.7

I recommend you to watch the following Drama movies:

- Creed III

- The Park

- Prizefighter: The Life of Jem Belcher

- 13 Exorcisms

You: can you give more information about THE Park

sentiment score 0.0

Here's more information about the park:

Release Date: 2023-03-02

Genres: Action, Drama, Horror, Science Fiction, Thriller

Overview: A dystopian coming-of-age movie focused on three kids who find themselves in an abandoned amusement park, aiming to unite whoever remains. With dangers lurking around every corner, they will do whatever it takes to survive their hellish Neverland.

Vote Average: 5.75

Any other genres(action, thriller, horror, comedy,drama, family, fantasy, Science Fiction, adventure) you'd be interested in?

You: Thanks

sentiment score 0.2

Alan Perez
AXP200075
CS4395.001

```
Hi, I'm MovieBot, I love to talk about movies! What is your name?  
You: Jacky  
Sorry, I didn't catch your name. Can you please tell me your name again?  
You: My name is Jacky  
Hello, Jacky. What genre of movies do you enjoy?  
You: I really love drama  
Added drama to likes  
sentiment score 0.7
```

```
I recommend you to watch the following Drama movies:  
- Creed III  
- The Park  
- Prizefighter: The Life of Jem Belcher  
- 13 Exorcisms  
You: can you give more information about The Park  
sentiment score 0.0  
Here's more information about the park:  
Release Date: 2023-03-02  
Genres: Action, Drama, Horror, Science Fiction, Thriller  
Overview: A dystopian coming-of-age movie focused on three kids  
Vote Average: 5.75
```

```
Any other genres(action, thriller, horror, comedy,drama, family, fantasy, Science Fiction, adventure)  
You: Thanks  
sentiment score 0.2  
Anytime!  
Any other genres you'd be interested in?  
You: Yes, can you recommend a movie  
sentiment score 0.0  
Here's more information about the super mario bros. movie:  
Release Date: 2023-04-05  
Genres: Animation, Adventure, Family, Fantasy, Comedy  
Overview: While working underground to fix a water main, Brooklyn plumbers—and brothers—Mario and Luigi  
Vote Average: 7.504  
Any other genres(action, thriller, horror, comedy,drama, family, fantasy, Science Fiction, adventure)  
You: Thanks!  
sentiment score 0.2  
No problem.  
Any other genres you'd be interested in?  
You: Which movie had the highest rating
```

Alan Perez
AXP200075
CS4395.001

```
You: Which movie had the highest rating  
sentiment score 0.0  
Here's more information about the super mario  
Release Date: 2023-04-05  
Genres: Animation, Adventure, Family, Fantasy  
Overview: While working underground to fix a  
Vote Average: 7.504  
Any other genres(action, thriller, horror, co  
You: bye  
Goodbye!
```

Appendix Knowledge Base

For the knowledge base, I used TMDB API to store movie data as my knowledge base. It includes the following information for each movie. I then used intents to determine the output of what the bot would say based on the response from the user.

- Title
- Overview
- Genres
- Release Date
- Vote Count
- Vote Average

Alan Perez
AXP200075
CS4395.001

	id	title	release_date	genres	overview
1	502356	The Super Mario Bros. Movie	2023-04-05	<div><div>[-] genres [5]</div><div><div>1</div>Animation</div><div><div>2</div>Adventure</div><div><div>3</div>Family</div><div><div>4</div>Fantasy</div><div><div>5</div>Comedy</div></div>	While working underground
2	594767	Shazam! Fury of the Gods	2023-03-15	<div><div>[-] genres [3]</div><div><div>1</div>Action</div><div><div>2</div>Comedy</div><div><div>3</div>Fantasy</div></div>	Billy Batson and his foster
3	76600	Avatar: The Way of Water	2022-12-14	<div><div>[-] genres [3]</div><div><div>1</div>Science Fiction</div><div><div>2</div>Adventure</div><div><div>3</div>Action</div></div>	Set more than a decade a
4	677179	Creed III	2023-03-01	<div><div>[-] genres [2]</div><div><div>1</div>Drama</div><div><div>2</div>Action</div></div>	After dominating the boxin
5	700391	65	2023-03-02	<div><div>[-] genres [4]</div><div><div>1</div>Science Fiction</div><div><div>2</div>Adventure</div><div><div>3</div>Thriller</div><div><div>4</div>Action</div></div>	65 million years ago, the c
6	638974	Murder Mystery 2	2023-03-28	<div><div>[+] genres [3]</div></div>	After starting their own de

```
{
  "id": 502356,
  "title": "The Super Mario Bros. Movie",
  "release_date": "2023-04-05",
  "genres": [
    "Animation",
    "Adventure",
    "Family",
    "Fantasy",
    "Comedy"
  ],
  "overview": "While working underground to fix a water main, Brooklyn plumbers—and brothers—Mario and Luigi are transported down a mysterious pipe and wander into a magical new world. But when the brothers are separated, Mario embarks on an epic quest to find Luigi.",
  "vote_average": 7.504,
  "actors": []
},
{
  "id": 594767,
  "title": "Shazam! Fury of the Gods",
  "release_date": "2023-03-15",
  "genres": [
    "Action",
    "Comedy",
    "Fantasy"
  ],
  "overview": "Billy Batson and his foster siblings, who transform into superheroes by saying \"Shazam!\", are forced to get back into action and fight the Daughters of Atlas, who they must stop from using a weapon that could destroy the world.",
  "vote_average": 6.972,
  "actors": []
},
```

Appendix Sample User Models

- User 1:
 - Name: Alan
 - Likes: action, drama, horror, fantasy
 - Dislikes: None
 - Search_History: [{ 'question': False, 'response': 'I like action' }, { 'question': False, 'response': 'Tell me more about BBlack Panther' }, { 'question': False, 'response': 'I like adventure' }]

- User 2:
 - Name: Jacky
 - Likes: drama
 - Dislikes: None
 - Search_History: [{ 'question': False, 'response': 'I really love drama' }, { 'question': False, 'response': 'can you give more information about THe Park' }, { 'question': False, 'response': 'Thanks' }]
- User 3:
 - Name: Jacky
 - Likes: drama
 - Dislikes: None
 - Search history: [{ 'question': False, 'response': 'I like drama' }, { 'question': False, 'response': 'thanks' }, { 'question': False, 'response': 'I effn hate drama' }, { 'question': False, 'response': 'I hate horror fantasy' }]

```
User: Arsen  
Search history: [{ 'question': False, 'response': 'I like drama' }, { 'question': False, 'response': 'thanks' },  
Likes: []  
Dislikes: ['drama', 'fantasy']
```

Evaluations:

The chatbot is a mixture of a movie recommendation system where the bot and user can discuss movies to some extent. The goal was to add a lot more features, but there were a lot of blockers during the development of this. This bot used natural language processing techniques to handle user input and provide appropriate dialog and recommendations. Overall, the chatbot performed all right when testing and evaluating.

Some of the strengths that this chatbot has it has a user-friendly interface since it allows the user to communicate with the chatbot via console. The chatbot is able to identify the intentions with a decent level of accuracy when using techniques such as matching and sentiment analysis. The knowledge base is also another strong point since there's a variety of movies with different genres for users to select from

and talk about. The chatbot is able to recommend based on the user's preferences and past searches.

Some of the weaknesses that the chatbot has is the knowledge base is only limited to movies available to TMD. The chatbot is unable to handle complex inputs and is unable to have a comprehensive dialog with the user due to the quality of training data used to create the chatbot.

To further evaluate the chatbot I conducted a survey to test the chatbot. The results of the survey were positive in the sense that the chatbot was easy to navigate and use, the accuracy of recommendations, and the sentiment analysis when users input their queries. Although the chatbot has strict limitations at this moment it can be further improved to handle complex queries and comprehensive dialog.