

CHATBOT PROJECT 1

NLP 6320

AXG230010

About the Project:

The purpose of this project is to create a David Fincher fan chatbot. David Fincher is a highly regarded filmmaker in Hollywood and hopefully my chatbot can help you learn a few things about him in case you didn't know earlier.

Part 1: Web Crawler:

- For part 1, to start off with, in the '**crawl.py**' file, I have taken 2 starter URLs related to David Fincher. Then, functions are defined to ensure that the URLs crawled to are outside of the starting URL and to ensure that links ending with ".php", ".pdf", ".jpg", etc., are not considered. The URLs are stored in the variable 'result', which is a list of all the URLs. Once the URLs are collected, their data is scraped and stored in separate files. For this project, I have considered a maximum of 25 URLs, but a couple of them denied access to be scraped from. So, I'd arrived at 23 URLs' data being scraped into 23 files.
- Now, these 23 files are then preprocessed in the '**clean.py**' file, where URLs, and non-alphanumeric characters are removed from the raw text. Upon cleaning the text files, they are combined into a single text file. This is done to extract the most important terms across all the files. This can be seen in the '**combine_files.py**' file. Then, in the '**important_terms.py**' file, the 40 most important terms are found. Here, TF-IDF is used to find them.
- Upon finding the important terms, I have manually picked 12 of them which I felt were the most relevant to the topic. Then, for the knowledge base creation, the file '**kb_create**' is run, which initializes an empty dictionary with these terms as keys. For each important term, it loops through the sentences in the combined file and adds those with the term present in it to the list under that key. The knowledge base is pickled into a file, which is then loaded into the Chatbot for part 2. A readable version of the knowledge base can be viewed in the file '**kbdict.txt**'. It contains the 12 important terms as keys and sentences under them are the keys' values.

The important terms of the project are:

1. Film
2. Music
3. Writer
4. Adaptation
5. Thriller
6. Noir
7. Reznor

8. Alien
9. Propaganda
10. Award
11. Zuckerberg
12. Sorkin

Below I'm attaching a screenshot of the important terms calculated using TF-IDF.

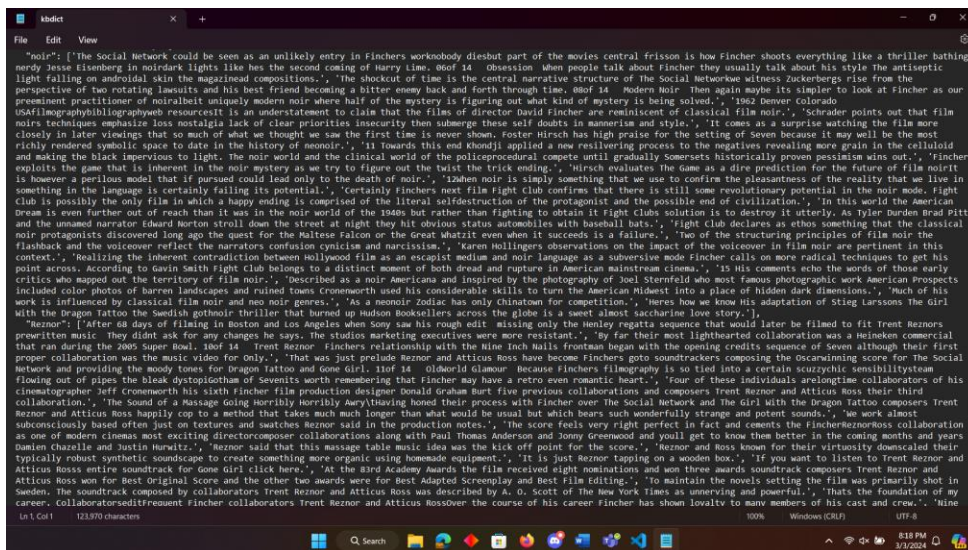
```
C:\Windows\System32\cmd.exe X + v
C:\Users\abhir\OneDrive\Desktop\Crawler>python important_terms.py

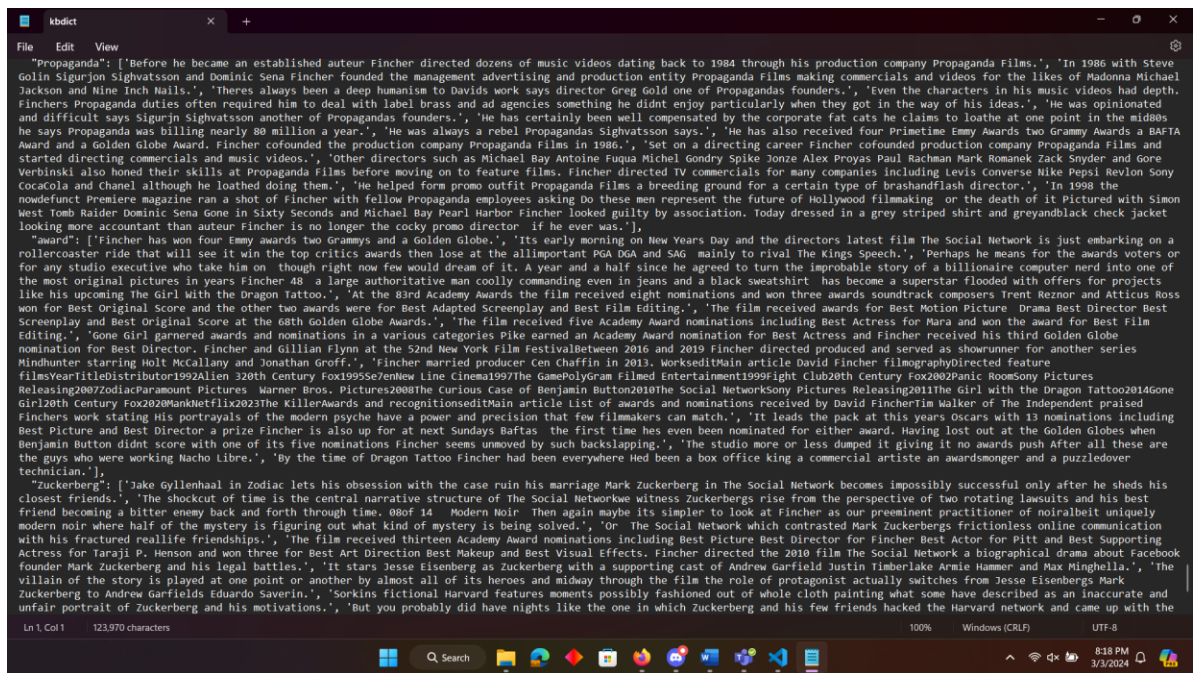
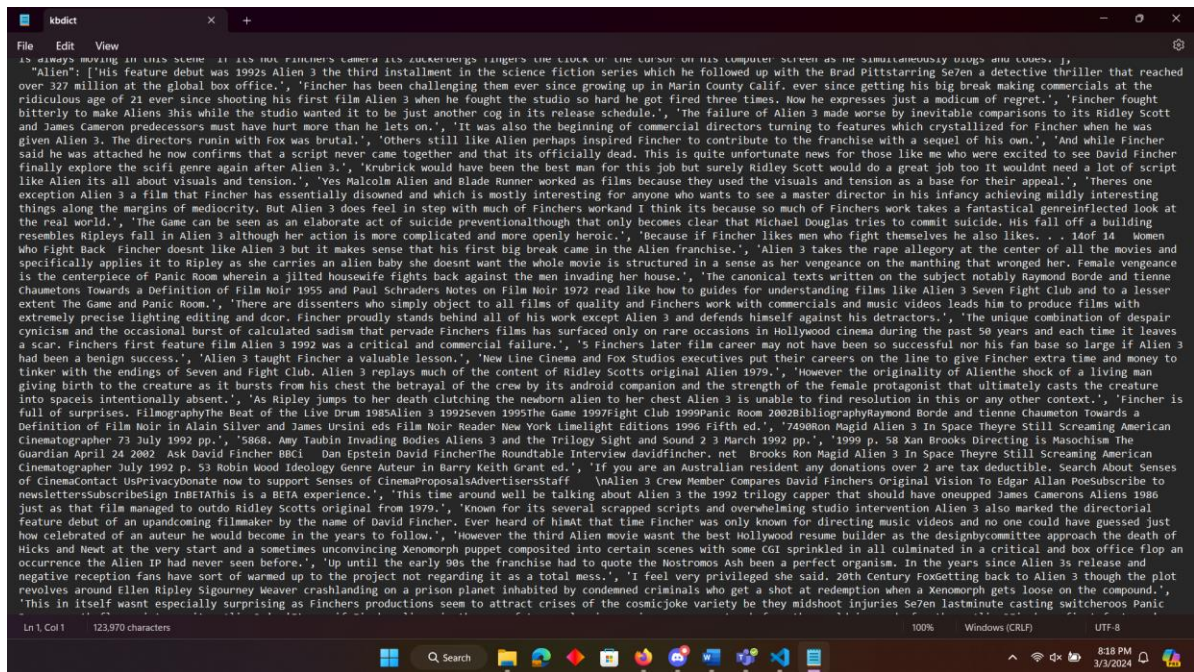
Top important terms using TF-IDF for the cleaned text file:
fincher: 0.4036882818585491
film: 0.27439687537642216
david: 0.1902485808260986
original: 0.1554915627133059
like: 0.15274759395954166
finchers: 0.10268395257395
one: 0.13994240644197528
retrieved: 0.13171050018068262
archived: 0.1289665314269184
films: 0.12530790642189943
movie: 0.12164928141688047
first: 0.11799065641186152
january: 0.10975875015856885
best: 0.10792943764805937
new: 0.09695356263308249
would: 0.08788780812045569
make: 0.0859776861794056
girl: 0.08414837511543612
also: 0.08323371886418138
director: 0.08048975011041716
movies: 0.07957509385916242
time: 0.07774578135655294
music: 0.07580181268288871
people: 0.07580181268288871
get: 0.0722578438491245
us: 0.07134318759786976
hollywood: 0.078428513461582
story: 0.06951387589536826
news: 0.06859921884410554
social: 0.06768456259285079
way: 0.06768456259285079
work: 0.06585525089034131
video: 0.06490859383906657
scene: 0.0640259375873183
dont: 0.06219662588532235
fight: 0.061281968834067686
going: 0.061281968834067686
says: 0.06036731258281287
```

The knowledge base screenshots can be seen below. These are some images taken from the readable version of the knowledge base.

```
kbdict
File Edit View

"film": ["His latest film The Killer is an action thriller starring Michael Fassbender as an assassin who finds himself at the center of an international manhunt after a job goes awry.",
'It's early morning on New Year's Day and the directors latest film The Social Network is just embarking on a rollercoaster ride that will see it win the top critics awards then lose at the
allimportant PGA DGA and SAG mainly to rival The Kings Speech.', 'Fincher has immersed himself in every detail of its geography just as he has Stieg Larssons bestseller on which the film
is based.', 'Fincher has been challenging them ever since growing up in Marin County Calif. ever since getting his big break making commercials at the ridiculous age of 21 ever since
shooting his first film Alien 3 when he fought the studio so hard he got fired three times. Now he expresses just a modicum of regret.', 'Critics panned it and the film made a disastrous
$5.5 million domestically quite a contrast to the success of Fincher videos such as Madonnas Express Yourselfand Aerosmiths Janies Got a Gun. He was very successful at a young age and
had people try to take it away from him and he knows what that feels like his Social producer Scott Rudin says.', 'The son of a Life magazine journalist and a mentalhealth nurse it was his
father who introduced him to film.', 'They're firing blanks at each other and riding on horseback and traveling to these locations, and theres Katharine Ross who would want to do anything
else six months later Finchers parents bought him a Super 8 camera and he started making his own films.', 'Asked whether he has disowned the film he says I never owned it and it has clearly
soured him on studios though he protests I like studios.', 'Hes an extraordinary filmmaker. Even Mechanic says Finchers reputation is exaggerated.', 'By that time Fincher had redeemed
himself with Seven1993 which in many ways defined his work dark visually audacious emotionally distant. I was the first person hired for the film Morgan Freeman recalls and I remember him
talking to me about the process he wanted to use to make the film dark.', 'At the same time he is very collaborative. Fincher followed Seven with the Michael Douglas thriller The Game1997
then after Fight Club he moved on to the Jodie Foster starrer Panic Room2002 and Zodiac2007 an adaptation of Robert Graysmiths books about the Bay Area killer who was active when Fincher
grew up. The film made only 33 million in North America but was one the bestreviewed pictures of the year.', 'With it Fincher solidified his reputation as a supremely talented filmmaker if
sometimes arrogant whose work was defined by its brilliance and also coldness. And so it came as a surprise to many Fincherfanatics yes theres a website of that very name when he took on
The Curious Case of Benjamin Button2008 Eric Roths adaptation of F. Scott Fitzgeralds slender short story.', 'Its more adult. More adult than the films Fincher loved in his youth like
Jawsand Star Wars I loved Luke Skywalker and I loved Darth Vader and I loved watching them work it out he reflects.', 'The film drew mixed reviews but reminded the industry that superstars
like Brad Pitt who also appeared in Seven and Fight Club often delivered their best performance for Fincher.', 'And yet its success was bittersweet not because the film won only three
Oscars but because it left him utterly inebriated. It took six years to get that through the starting gate he sighs.', 'Following the usual back and forth the studio wanted to make
the film for 25 million Fincher insisted on 42 million shooting began three months later.', 'After 68 days of filming in Boston and Los Angeles when Sony saw his rough edit missing only
the Henley regatta sequence that would later be filmed to fit Trent Reznors prewritten music. They didnt ask for any changes he says. The studios marketing executives were more
resistant.', 'Now hes back in the Sony saddle with Dragon Tattoo. Its two weeks after our first meeting and were in a tiny trailer at Paramount where Fincher is filming Dragonone six
soundstages even more than Button and is huddled with his longtime companion and producer Sean Chaffin.', 'With a shoot that includes locations in Stockholm Zurich and England he still
has 90 days of filming ahead.', 'And finally came the reality of filming in Sweden.', 'Recently Fincher sat down to answer questions for The Guardian and the filmmaker was asked about the
project. I was asked if I might be interested in the first SpiderMan. . . Im not interested in doing A Superhero.', 'Ive also read elsewhere rumors that Fincher wanted Edward Norton to
play Peter Parker not sure if it is true or not. As much as I love David Finchers incredible films and as much as Im curious to see what he would have done Im kinda glad he never had the
chance.', 'His collaborations with Brad Pitt have given us two of the most iconic movies of the 1990s namely Se7en and Fight Club whileThe Curious Case of Benjamin Button offered
jawdropping visual effects that garnered several welldeserved Academy Award wins as well as a Best Director nomination for Fincher himself. READ MORE David Fincher Discover The Hidden
Visual Effects in His Greatest Films Watchconsidering his storied film career its perhaps no surprise that some incredible films have had a great influence on Finchers career over the
years and have been imprinted in many of his films.', 'In an update from Freeman last year he confessed that its a very intellectual science fiction film a very difficult book to translate
cinematically.', 'If not Fincher then who elsefind more posts Movie News Opinions28 CommentsIm not sure this isnt a good thing for scifi films visually Im sure this could be quite
stunning just think Megahalo but apart from a bunch of people wandering around an empty tube not a great deal happens in Rama.', 'It works very well as a book but as a film well. . . some
things just dont translate. . . but then. . . Sunshine worked well. . . Dr. Duvel on Oct 13 200821 disagree.', 'This probably would be like Finchers excellent Zodiac a fantastic film
with a conclusion that feels somewhat empty.', 'That film deserved all the accolades it got but it didnt exactly light the box office on fire.', 'And it doesnt help that weak films like
Independence Day have ripped off so much from Rama and stolen a lot of its thunder.', 'Rama would be a fantastic event for film fans but not such a big deal for movie fans.', 'It would
probably be the Ang Lees Hulk of todays scifi films. kevjohn on Oct 13 20084How exactly did Independence Day rip off RamaID4 was just War of the Worlds with a computer virus instead of a
biological virus as the key plot point.', 'I think Rama could work as a film but probably taking books 1 2 together to make the first and then if it was a success make a sequel from 3
4.', 'Yes Malcolm Alien and Blade runner worked as films because they used the visuals and tension as a base for their appeal.', 'The discovery that Rama is this immense object that its
not a natural phenomenon and is transiting our solar system with a definite purpose one that lies far beyond us is another source that has to be used to draw the audience into the films
```



Part 2: Chatbot:

System Description:

The chatbot takes the data from the pickled knowledge base. The chatbot prompts for the user's name and asks them whether they like Fincher or know about him. Then, they are asked if they know any movies by Fincher. This starts the conversation before giving some general topics that they can learn about Fincher. More on the dialog flow is covered in the next section.

Based on the user input, the chatbot runs a similarity search that fetches the most relevant answer to the input. The users' data is stored in a pickled file, which stores their name, preference on Fincher, and their favorite Fincher movie name. This enables the chatbot to remember user information so when a user returns to the chatbot, it doesn't need to ask them about their thoughts on Fincher again and can go straight to the answering for questions. If the chatbot cannot find a relevant response to the query, it responds with a generic sentence saying it doesn't have any information on that topic.

An example is shown below.

```
Fin: Videos before entering filmmaking, and has a longtime collaboration with composer Trent Reznor from the Nine Inch Nails.
Fin: Ask me more about the above topics or just Fincher in general or type 'exit' to end the conversation.
Jaideep: food
Fin: Sorry, I don't have information on that.
Jaideep: Tell me about The Social Network
Fin: It stars Jesse Eisenberg as Zuckerberg with a supporting cast of Andrew Garfield, Justin Timberlake, Armie Hammer, and Max Minghella.
```

NLP techniques used in the chatbot:

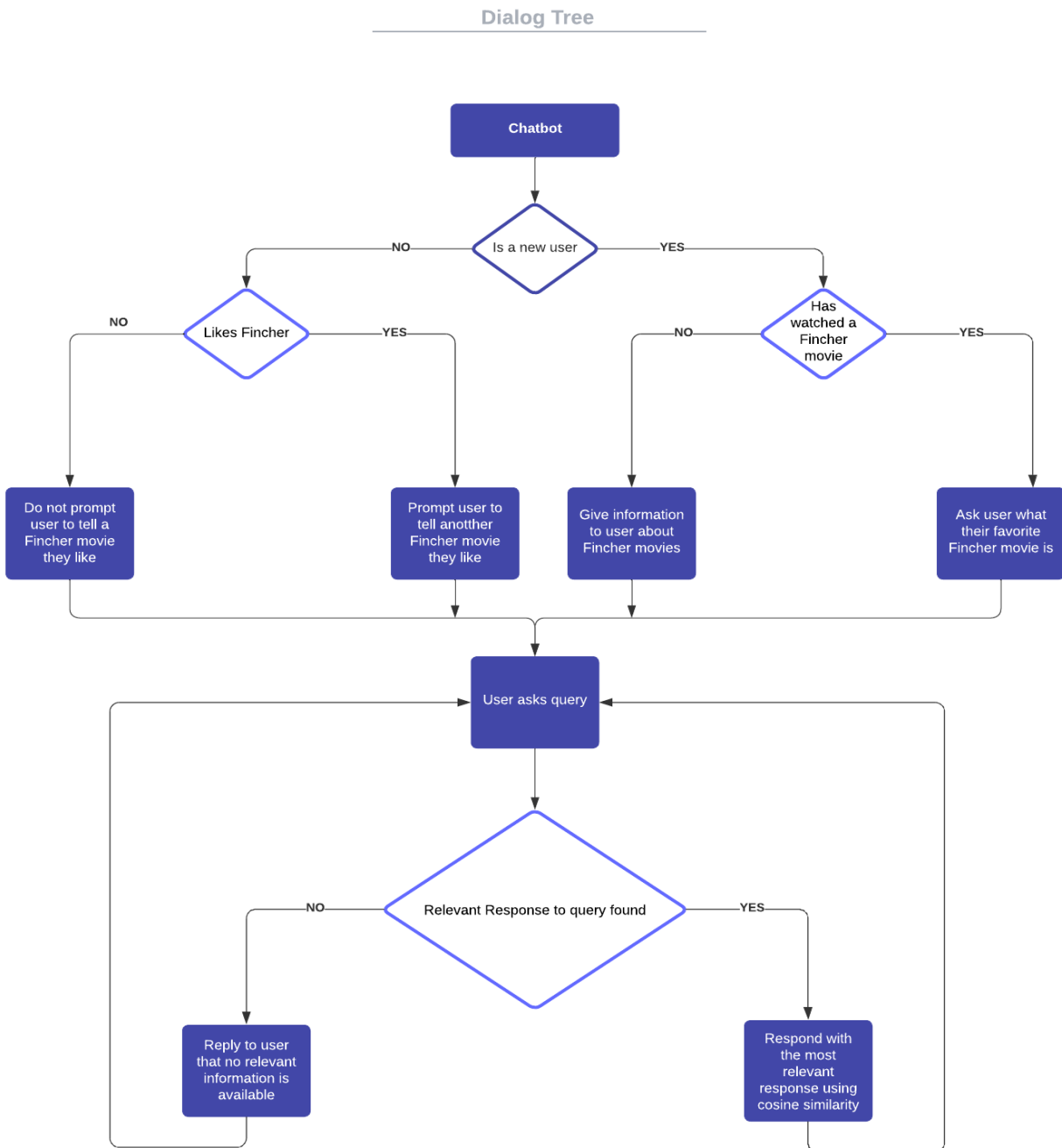
- Tokenization and lemmatization: The user input is processed by spaCy's NLP pipeline to extract tokens and lemmatize words. This helps in matching the user's input to predefined sentences.
- NER: Named Entity Recognition is used to extract named entities like person, organization names from the user input. This provides more context to match the user input.
- Cosine Similarity: Cosine Similarity is used to find the best matching sentence in the knowledge base to the user input. This has been important in retrieving sentences despite the term not appearing in the sentence. For example, mentioning Facebook would give responses related to 'The Social Network', a movie based on the founding of Facebook.

Chatbot Logic:

The chatbot checks whether the user is new or returning based on their name.

- If it's a new user who does not like Fincher or does not know about him, they are asked if they watched any Fincher films. If they did, their favorite film of his is asked. If not, the user is given the names of some Fincher films.
- If the user is new and they like Fincher, their favorite movie of his is asked and the information is stored before the conversation proceeds.
- If the user is returning and likes Fincher, the chatbot remembers them and remarks that the user likes a movie which they previously mentioned that they liked.

- If returning and does not like Fincher or does not know about him, the chatbot remembers them and they are just simply asked to ask any questions to learn about Fincher.



The dialog continues until the user enters 'exit'.

Examples of the above cases can be seen in the sample dialog interactions.

Sample Dialog Interactions:

The below two interactions are for new users.

```
C:\Users\abhir\OneDrive\Desktop\Chatbot\Chatbot>python chatbot.py
Fin: Hello! I'm Fin, a David Fincher fan chatbot.
Fin: What's your name? Jaideep
Fin: Do you like David Fincher? Type 'yes' if you do, or 'no' if you do not like him or don't know much about him: no
Fin: Hello Jaideep! Nice to meet you. I see you do not like David Fincher or know much about him.
Fin: Have you watched any Fincher movies before? Type 'yes' or 'no'
Jaideep: no
Fin: Well, if you're interested, here are some popular David Fincher movies: Fight Club, The Social Network, Gone Girl, Seven, The Girl with the Dragon Tattoo, The Killer, Alien 3, Zodiac.
Fincher has won many awards, made several book adaptations, including one about Facebook, ventured into his own production company 'Propaganda', has made music videos before entering filmmaking, and has a longtime collaboration with composer Trent Reznor from the Nine Inch Nails.
Fin: Ask me more about the above topics or just Fincher in general or type 'exit' to end the conversation.
Jaideep: food
Fin: Sorry, I don't have information on that.
Jaideep: Tell me about The Social Network
Fin: It stars Jesse Eisenberg as Zuckerberg with a supporting cast of Andrew Garfield Justin Timberlake Armie Hammer and Max Minghella.
Jaideep: What about Alien
Fin: Alien3 was panned by critics and largely ignored by audiences the only way it could have done worse would have been if they actually called it AllMale Outer Space Penal Colony. Fincher hates talking about Alien3 which started filming without a finished script and plunged him into daily battles with 20th Century Fox over control in the end the studio recut Fincher's version.
Jaideep: Which awards did he win?
Fin: The film received awards for Best Motion Picture Drama Best Director Best Screenplay and Best Original Score at the 68th Golden Globe Awards.
Jaideep: Did he make music videos?
Fin: Set on a directing career Fincher cofounded production company Propaganda Films and started directing commercials and music videos.
Jaideep: Tell me about Propaganda Films
Fin: There's always been a deep humanism to David's work says director Greg Gold one of Propaganda's founders.
Jaideep: Who is Trent Reznor?
Fin: That's the foundation of my career. Collaborators edit frequent Fincher collaborators Trent Reznor and Atticus Ross over the course of his career Fincher has shown loyalty to many members of his cast and crew.
Jaideep: Which films of his are book adaptations?
Fin: The villain of the story is played at one point or another by almost all of its heroes and midway through the film the role of protagonist actually switches from Jesse Eisenberg's Mark Zuckerberg to Andrew Garfield's Eduardo Saverin.
Jaideep: exit
Fin: Goodbye! Have a great day.
```

```
C:\Users\abhir\OneDrive\Desktop\Chatbot\Chatbot>python chatbot.py
Fin: Hello! I'm Fin, a David Fincher fan chatbot.
Fin: What's your name? Rohit
Fin: Do you like David Fincher? Type 'yes' if you do, or 'no' if you do not like him or don't know much about him: no
Fin: Hello Rohit! Nice to meet you. I see you do not like David Fincher or know much about him.
Fin: Have you watched any Fincher movies before? Type 'yes' or 'no'
Rohit: yes
Fin: That's good to know.
Fin: Which Fincher movie did you like the most? Please provide the name: Fight Club
Fin: Nice! Fight Club is a great choice.
Fincher has won many awards, made several book adaptations, including one about Facebook, ventured into his own production company 'Propaganda', has made music videos before entering filmmaking, and has a longtime collaboration with composer Trent Reznor from the Nine Inch Nails.
Fin: Ask me more about the above topics or just Fincher in general or type 'exit' to end the conversation.
Rohit: Did he make music videos?
Fin: Martin just gave hope for a Game of Thrones prequel CNET Tech Culture David Fincher returns to his Vogueing video days in new comedy David Fincher returns to his Vogueing video days in new comedy The acclaimed director follows House of Cards with an HBO show based on the music video industry that spanned the likes of Michael Bay and Zack Snyder. by Richard Trenholm richtrenholm 5 May 2015 12:21 pm BST By Richard Trenholm richtrenholm Directed by David Fincher the 1990 promo for Vogue by Madonna is hailed as one of the best music videos ever.
Rohit: tell me about his films
Fin: Before he became an established auteur Fincher directed dozens of music videos dating back to 1984 through his production company Propaganda Films.
Rohit: Has he won academy awards?
Fin: Perhaps he means for the awards voters or for any studio executive who take him on though right now few would dream of it. A year and a half since he agreed to turn the improbable story of a billionaire computer nerd into one of the most original pictures in years Fincher 48 a large authoritative man coolly commanding even in jeans and a black sweatshirt has become a superstar flooded with offers for projects like his upcoming The Girl With the Dragon Tattoo.
Rohit: What is Fincher's favorite film?
Fin: But just because we know that we are watching a film doesn't make us any less susceptible to the illusion.
Rohit: What is his movie with Mark Zuckerberg?
Fin: But watch the various kids Dustin and Billy making screwdrivers the stoned dude on the couch watching Shark Week the students in the lab Erica Albright's roommate in her flannel pajamas the kids in the pizza parlor all while beautiful people on E dance to Serato synced songs. Watch the subtle ways in which Saverin and Zuckerberg interact.
```

Then for returning users:

```
C:\Users\abhir\OneDrive\Desktop\Chatbot\Chatbot>python chatbot.py
Fin: Hello! I'm Fin, a David Fincher fan chatbot.
Fin: What's your name? Abhi
Fin: Hello Abhi! Welcome back. I remember you like David Fincher.
Fin: By the way, you mentioned that you like 'Zodiac'.
Fin: Please name another Fincher movie you like.
Abhi: Zodiac
Fin: That's great! Zodiac is indeed a fantastic film.
Fincher has won many awards, made several book adaptations, including one about Facebook, ventured into his own production company 'Propaganda', has made music videos before entering filmmaking, and has a longtime collaboration with composer Trent Reznor from the Nine Inch Nails.
Fin: Ask me more about the above topics or just Fincher in general or type 'exit' to end the conversation.
Abhi: What movie did he make about Facebook?
Fin: But you probably did have nights like the one in which Zuckerberg and his few friends hacked the Harvard network and came up with the germ of the idea that would eventually become Facebook.
Abhi: Is it the social network?
Fin: But watch the various kids Dustin and Billy making screwdrivers the stoned dude on the couch watching Shark Week the students in the lab Erica Albrigh ts roommate in her flannel pajamas the kids in the pizza parlor all while beautiful people on E dance to Seratosynced songs. Watch the subtle ways in which Saverin and Zuckerberg interact.
Abhi: awards won by Fincher
Fin: Its early morning on New Years Day and the directors latest film The Social Network is just embarking on a rollercoaster ride that will see it win the top critics awards then lose at the allimportant PGA DGA and SAG mainly to rival The Kings Speech.
Abhi: Did he make music videos ?
Fin: Before he became an established auteur Fincher directed dozens of music videos dating back to 1984 through his production company Propaganda Films.
Abhi: Does he own a production company?
Fin: He has certainly been well compensated by the corporate fat cats he claims to loathe at one point in the mid80s he says Propaganda was billing nearly 8 0 million a year.
Abhi: Tell me about Gone Girl
Fin: Its just about survival. Five years after Panic Room Fincher returned on March 2 2007 with Zodiac a thriller based on Robert Graysmiths books about the search for the Zodiac a real life serial murderer who terrorized communities between the late 1960s and early 1970s.
Abhi: Why does he like working with Trent Reznor?
Fin: By far their most lighthearted collaboration was a Heineken commercial that ran during the 2005 Super Bowl. 18of 14 Trent Reznor Finchers relationsh ip with the Nine Inch Nails frontman began with the opening credits sequence of Seven although their first proper collaboration was the music video for Only .
Abhi: exit
Fin: Goodbye! Have a great day.
```

```
C:\Users\abhir\OneDrive\Desktop\Chatbot\Chatbot>python chatbot.py
Fin: Hello! I'm Fin, a David Fincher fan chatbot.
Fin: What's your name? Vikas
Fin: Hello Vikas! Welcome back. I remember you do not like David Fincher or know much about him.
Fincher has won many awards, made several book adaptations, including one about Facebook, ventured into his own production company 'Propaganda', has made music videos before entering filmmaking, and has a longtime collaboration with composer Trent Reznor from the Nine Inch Nails.
Vikas: Which thriller films has he made?
Fin: It makes an untellable story clear and fascinating. 2011present Established auteureditRooney Mara Daniel Craig and Fincher at the premiere of The Girl with the Dragon TattooIn 2011 Fincher followed the success of The Social Network with The Girl with the Dragon Tattoo a psychological thriller based on the novel by Swedish writer Stieg Larsson.
Vikas: Which film is a book adaptation?
Fin: Sorkins fictional Harvard features moments possibly fashioned out of whole cloth painting what some have described as an inaccurate and unfair portrait of Zuckerberg and his motivations.
Vikas: Is the social network based on a true story?
Fin: Or The Social Network which contrasted Mark Zuckerbergs frictionless online communication with his fractured reallife friendships.
Vikas: Thoughts on Alien 3?
Fin: Alien 3.
Vikas: Thoughts on Alien 3?
Fin: From the beginning Alien 3 was hampered by studio intervention and several abandoned scripts.
```

Appendix for knowledge base:

I have not modified the knowledge base created in part 1 as there are way too many lines to even consider changing anything. It uses the same screenshots present there so I am avoiding putting them here again.

Appendix for sample user models:

The sample user models are pickled in the file ‘**user1_data.pkl**’. A readable version of the users can be seen below. It documents their names, their preference towards Fincher, and if they’ve liked any of his movies, it is saved here.


```
C:\Windows\System32\cmd.exe x + v
Microsoft Windows [Version 10.0.22631.3155]
(c) Microsoft Corporation. All rights reserved.

C:\Users\abhir\OneDrive\Desktop\Chatbot\Chatbot>python user.py
Unpickled Data:
{'Abhi': {'likes_fincher': 'yes', 'liked_movie': 'Zodiac'}, 'Vijay': {'likes_fincher': 'no', 'liked_movie': None}, 'Vikas': {'likes_fincher': 'no', 'liked_m
ovie': 'The Game'}, 'John': {'likes_fincher': 'yes', 'liked_movie': 'Zodiac'}, 'Jake': {'likes_fincher': 'yes', 'liked_movie': 'The girl with the dragon tat
too'}, 'Chandra': {'likes_fincher': 'no', 'liked_movie': 'The Social Network'}, 'Charan': {'likes_fincher': 'no', 'liked_movie': 'Fight Club'}, 'Rohit': {'l
ikes_fincher': 'no', 'liked_movie': 'Fight Club'}, 'Harsha': {'likes_fincher': 'no', 'liked_movie': None}, 'Srikanth': {'likes_fincher': 'no', 'liked_movie':
None}, 'Shashank': {'likes_fincher': 'no', 'liked_movie': None}, 'Sahith': {'likes_fincher': 'no', 'liked_movie': None}, 'Jaideep': {'likes_fincher': 'no',
'liked_movie': None}}

C:\Users\abhir\OneDrive\Desktop\Chatbot\Chatbot>
```

My evaluation of the chatbot:

I think the chatbot does well in identifying the relationship between words that are similar. When I mention an actor, it links them to the movie which they acted in most of the time. For example, when I mention Daniel Craig, it sometimes does fetch sentences related to The Girl with the Dragon Tattoo, a movie which he starred in. It does a good job in remembering users and their preferences. In most cases, it does identify names and organizations that appear a lot of times, like Fincher's own production company 'Propaganda', and award names like 'BAFTA' and 'The Oscars'.

A drawback is that sometimes it would fetch a sentence despite not having much relevance to the question asked. I have tried my best in fine tuning this aspect, but it is to be expected from an NLP based chatbot.