Movie Search with Vague Description

Mengjia (Zoe) Ding

ding.me@northeastern.edu

Github repository: https://github.com/zoeding12/SearchMovie-WithVagueDescription

I.   Problem Identified

People love movies. There are countless movie-related websites, posts, lists for audience to choose which movie they would love to watch. Audience can browse popular movies by genre, country, time, actors, or even by recommendations based on their watching history. However, how could a user randomly find a movie different from his/her past taste? The user might just have an abstract feeling about what movie he/she wants to watch and don't want to scan through all different recommendations on the home page. How could he/she search with no specific keywords?
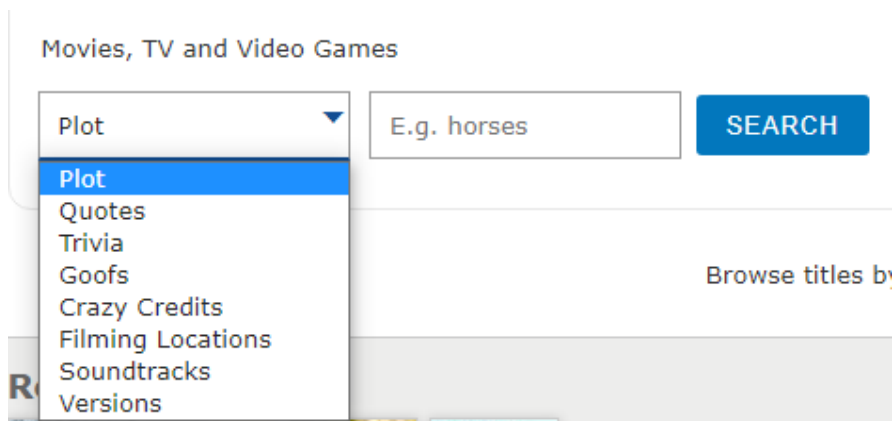
a)   Available movie search services
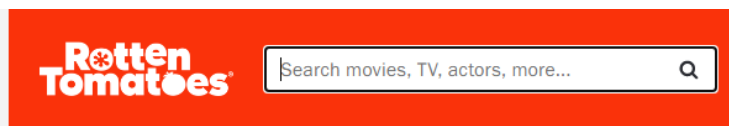
1. IMDB

keyword searching on IMDB:
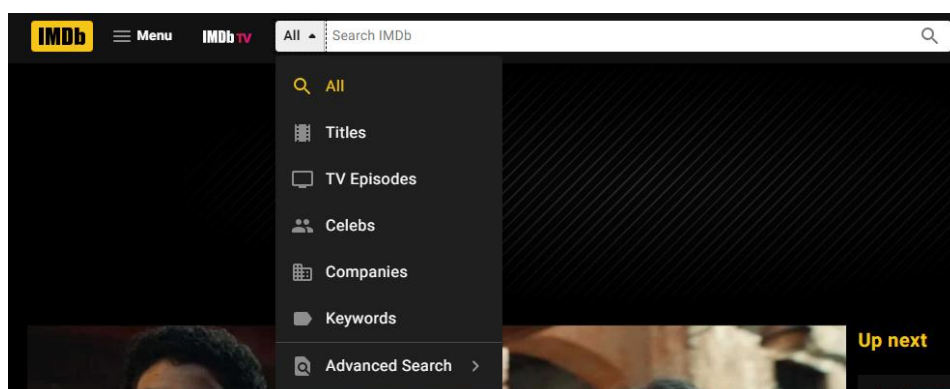
advanced search with more specific information:

The search feature on IMDB is quite accurate with the exact keyword and has lots of different classifications to choose from as filters. Items in the advanced search focuses more on "helping people find the exact movie with known information".
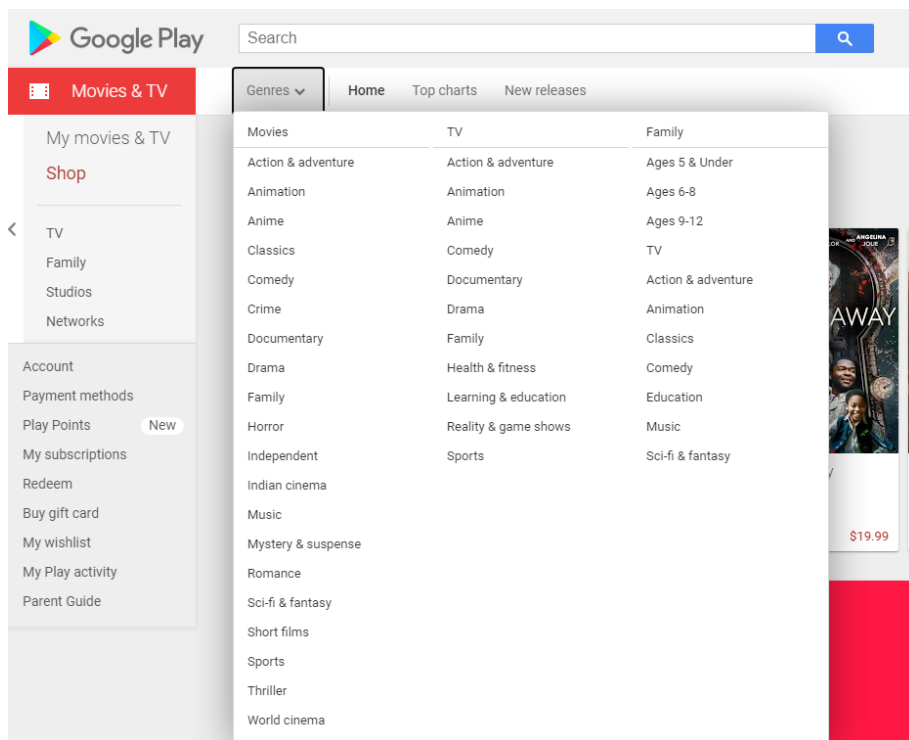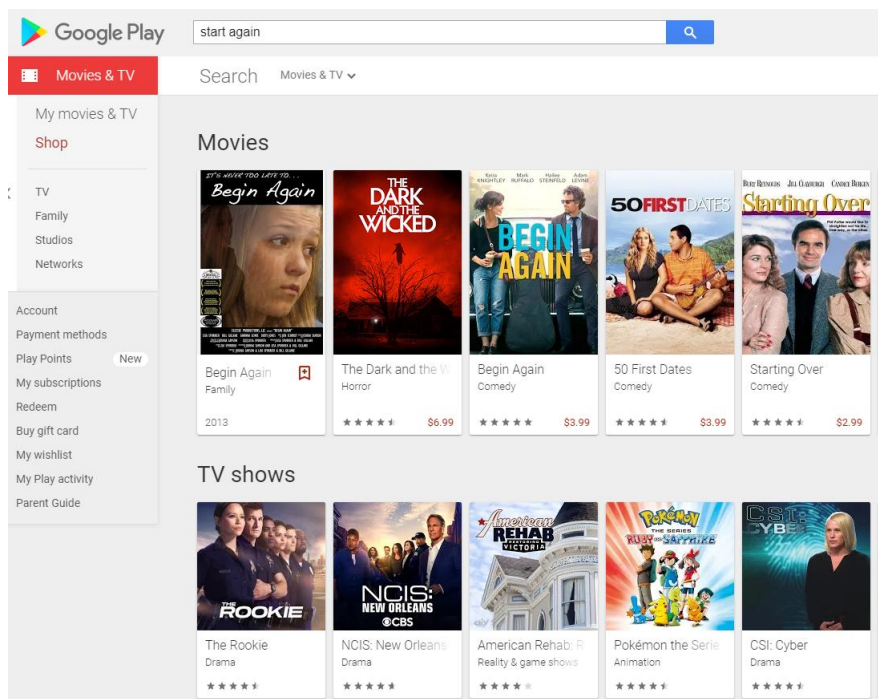
2. Rotten Tomatoes

Rotten Tomatoes doesn't have advanced search or filters.

3. Google Play Movies & TV

Users can use different predefined categories to browse movies:

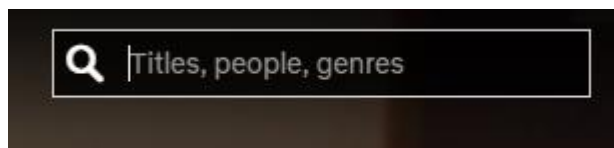For search, it will be general search without filters:

4. Netflix

User can use genres to browse popular movies (the ranking of movies might

be affected by the companies own business concern):

For search, it will be general search without filters:

b) Scenario where available services fail

If a user has an abstract sense of preference instead of having a specific movie or category/genre to watch, and wants to pick a good movie without wasting much time browsing, it would be very difficult to realize on all available services since all these services focus on exact searches.

c) Summary of the problem to solve

The proposal aims to help users search for good movies (with high ratings) by a simple query with abstract description about what movie they want to watch. Queries would be like "family, adventure", "sad, science fiction", "happy, horror"

II.   Implementation Details

a)   User Interface

Query page:
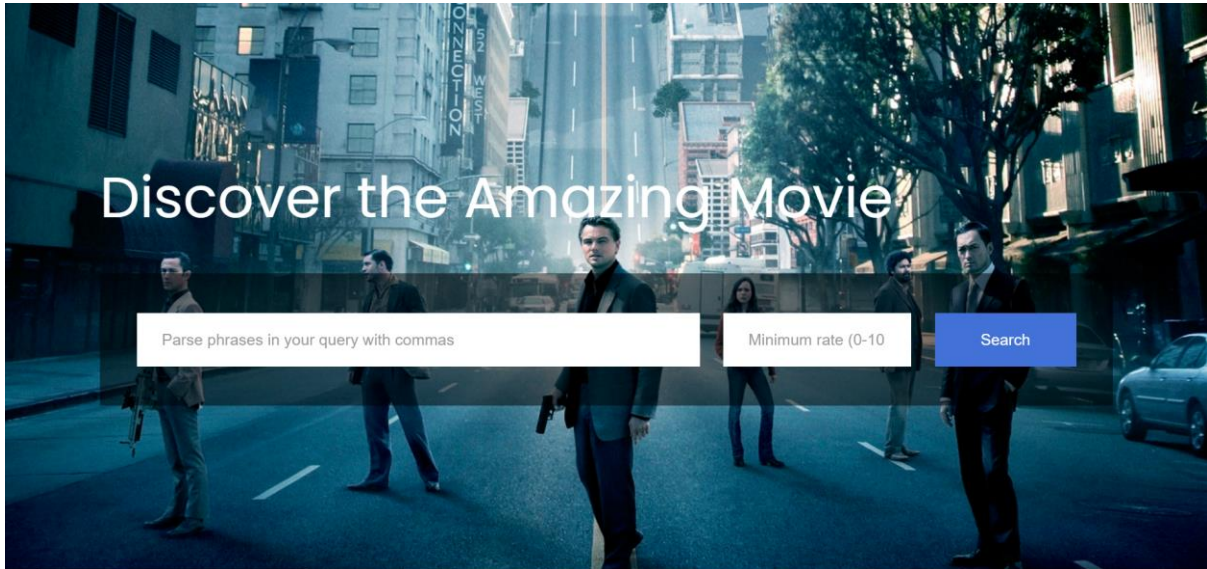
User can enter their query and a minimum rate score they want to search for.

Result page:

User can see the first 10 movies' title and a short description.

b)   Tools and libraries used

- Tomcat with Java (web server)

- Elasticsearch with Java (indexing and searching)

- fastJson (parse json)

- openCSV (read dataset from csv file)

c) Dataset

Basic information(85854 records):

https://www.kaggle.com/stefanoleone992/imdb-extensive-

dataset?select=IMDb+title_principals.csv

Supplement information for movie plot(34990 records):

https://www.kaggle.com/jrobischon/wikipedia-movie-plots

d) Mapping Structure in Elasticsearch

There are 9 main fields for each movie. Among them, Title, Description, Plot, Actors,

Director, Writer are used for full text searching; Genre will first be used as a filter if

there exist specific genre keywords in the query; Average Vote Score and Number

of Votes are two metrics which will both help rank and filter out movies that has

the rate lower than the user's requirement.

e) Searching and Ranking

For the query, user can help improve the accuracy by including commas between

different phrases. The application will parse the query with commas so genre

keywords can be recognized more easily.

The first thing when searching is to find if there exist specific genre keywords such

as "Drama", "Comedy". If found, then this genre will be deleted from the query and become a filter which allows the app only searches for movies from the specific genre. For this project, the genres used are all from the dataset.

Then, the application will do full-text search with the query according to a customized weight combination. Comparing to other available main stream services, I lower the weight of title and instead increase the weight of plot in order for the search results to be more content-related.

When ranking retrieved results, the app will sort with the Lucene score first, then average rate, then number of votes.

III. Testing

a) query: "start again, romance, music", minimum rate: 7

Your search result for: start again, romance, music

Movie Title	Description	Rate
His Butler's Sister	A young girl visits New York for two reasons: to see her half-brother, and to try to start a musical career.	7.0
Chihayafuru Part I	Chihaya is a high school student who learns that her childhood friend who went back to Fukui Prefecture, doesn't play Karuta anymore. Believing they will meet again, Chihaya and her close friend start a Karuta club at their high school.	7.1
Once Again	After a chance phone call leads to daily conversations, a widowed restaurant owner and a lonely film actor plan to finally meet in person.	7.0
Groundhog Day	A weatherman finds himself inexplicably living the same day over and over again.	8.0
Pure Country	Dusty gets tired of all the smoke and light effects, wonders what happened to doing it the way that made him famous and happy. Takes a walk to find what seems to be missing and finds himself and his music again!	7.0
Begin Again	A chance encounter between a down-and-out music-business executive and a young singer-songwriter, new to Manhattan, turns into a promising collaboration between the two talents.	7.4
The Commitments	When Jimmy Rabbitte wants to start a band, he has open auditions at his house.	7.6
Gün Keçdi	Schoolmates who loved each other, but didn't say it meet again years later.	8.3
I Love You Again	Boring businessman Larry Wilson recovers from amnesia and discovers he's really a con man...and loves his soon-to-be-ex wife.	7.5
Sujon Sokhi	Conflict between two brother causes separation of a family. Their children unite them once again.	7.6

b) query: "sad, science fiction"；minimum rate: 7

Your search result for: sad, science fiction

Movie Title	Description	Rate
Inception	A thief who steals corporate secrets through the use of dream-sharing technology is given the inverse task of planting an idea into the mind of a C.E.O.	8.8
The Matrix	A computer hacker learns from mysterious rebels about the true nature of his reality and his role in the war against its controllers.	8.7
National Theatre Live: Frankenstein	Created by Victor Frankenstein, the one known only as the creature sets out to discover the world and the meaning of life.	8.7
Interstellar	A team of explorers travel through a wormhole in space in an attempt to ensure humanity's survival.	8.6
The Prestige	After a tragic accident, two stage magicians engage in a battle to create the ultimate illusion while sacrificing everything they have to outwit each other.	8.5
Back to the Future	Marty McFly, a 17-year-old high school student, is accidentally sent thirty years into the past in a time-traveling DeLorean invented by his close friend, the eccentric scientist Doc Brown.	8.5
Terminator 2: Judgment Day	A cyborg, identical to the one who failed to kill Sarah Connor, must now protect her teenage son, John Connor, from a more advanced and powerful cyborg.	8.5
Avengers: Infinity War	The Avengers and their allies must be willing to sacrifice all in an attempt to defeat the powerful Thanos before his blitz of devastation and ruin puts an end to the universe.	8.4
Alien	After a space merchant vessel receives an unknown transmission as a distress call, one of the crew is attacked by a mysterious life form and they soon realize that its life cycle has merely begun.	8.4
Ivan Vasilevich menyaet professiyu	An ordinary Soviet building manager, living in the 20th century, is extremely similar to a Tsar of All Rus' - Ivan IV the Terrible (1530-1584). He would never have known about this, but one day his neighbor created a time machine.	8.4

c) query: "new york, love, adventure"；minimum rate: 7.2

Your search result for: new york, love, adventure

Movie Title	Description	Rate
Escape from New York	In 1997, when the U.S. president crashes into Manhattan, now a giant maximum security prison, a convicted bank robber is sent in to rescue him.	7.2
King Kong	A film crew goes to a tropical island for an exotic location shoot and discovers a colossal ape who takes a shine to their female blonde star. He is then captured and brought back to New York City for public exhibition.	7.9
North by Northwest	A New York City advertising executive goes on the run after being mistaken for a government agent by a group of foreign spies.	8.3
The Tall Target	A New York City detective, traveling by train between New York and Baltimore, tries to foil an on-board plot to assassinate President-elect Abraham Lincoln before he reaches Baltimore to give a major pre-Inauguration speech in 1861.	7.2
Spider-Man: Homecoming	Peter Parker balances his life as an ordinary high school student in Queens with his superhero alter-ego Spider-Man, and finds himself on the trail of a new menace prowling the skies of New York City.	7.4
The Great Race	In the early 20th century, two rivals, the heroic Leslie and the despicable Professor Fate, engage in an epic automobile race from New York to Paris.	7.2
Die Hard: With a Vengeance	John McClane and a Harlem store owner are targeted by German terrorist Simon Gruber in New York City, where he plans to rob the Federal Reserve Building.	7.6
Transamerica	A preoperative transgender woman takes an unexpected journey when she learns that she fathered a son, now a teenage runaway hustling on the streets of New York.	7.4
I Am Legend	Years after a plague kills most of humanity and transforms the rest into monsters, the sole survivor in New York City struggles valiantly to find a cure in this post-apocalyptic action thriller.	7.2
Till We Meet Again	Till We Meet Again is a 1944 American drama film directed by Frank Borzage and written by Lenore J. Coffee. The film stars Ray Milland, Barbara Britton, Walter Slezak, Lucile Watson, ...	7.2

d)   query: "brad pitt, drama, war"；minimum rate: 7

Your search result for: brad pitt, drama, war

Movie Title	Description	Rate
At Close Range	In 1978 Pennsylvania, an absentee father is reacquainted with his estranged teenage sons who become intrigued with his romanticized life of crime.	7.0
Look Back in Anger	A disillusioned, angry university graduate comes to terms with his grudge against middle-class life and values.	7.0
Fury	A grizzled tank commander makes tough decisions as he and his crew fight their way across Germany in April, 1945.	7.6
Since You Went Away	With her husband away to fight in World War II, a housewife struggles to care for their two daughters - and a pair of lodgers who have moved in - alone.	7.5
Party	A planned social gathering is held to commemorate Divakar Barve for the award he receive pertaining his contributions to Indian Arts and films. The social event is greeted by some important...	7.2
Inglourious Basterds	In Nazi-occupied France during World War II, a plan to assassinate Nazi leaders by a group of Jewish U.S. soldiers coincides with a theatre owner's vengeful plans for the same.	8.3
Legends of the Fall	In the early 1900s, three brothers and their father living in the remote wilderness of Montana are affected by betrayal, history, love, nature, and war.	7.5
Sleepers	After a prank goes disastrously wrong, a group of boys are sent to a detention center where they are brutalized. Thirteen years later, an unexpected random encounter with a former guard gives them a chance for revenge.	7.6
Once Upon a Time... in Hollywood	A faded television actor and his stunt double strive to achieve fame and success in the final years of Hollywood's Golden Age in 1969 Los Angeles.	7.6
Babel	Tragedy strikes a married couple on vacation in the Moroccan desert, touching off an interlocking story involving four different families.	7.4

IV. Steps to start the application

1. download all the codes

2. go to "elasticsearch-7.10.0/bin" and start "elasticsearch"

3. go to "VagueMovieSearch_zoe/bin" and start "startup.bat"(windows) or "startup.sh"(mac)

4. use a browser and enter "http://localhost:8080/hw1ZoeDing/query.jsp" to search for movies

V. Possible Future Improvements Directions

a) Import more movie reviews written by real users so each movie can have more subjective texts for searching

b) If we can gather enough users' subjective comments towards movies such as words "comfortable" or "desperate". And then use machine learning approaches to expand all these tags to more movies. It would be more accurate to search for movies based on vague descriptions.