

MOVIE MANIA

MAJOR UPDATES:

Apr 3: To write into files, you would need the FileSystem API(fs.js) which is a built-in module in NodeJS. Because this cannot be done without NodeJS, we move the User Registration component into Customisation and replace it with Recommender System. **References to NodeJS added.**

Apr 4: Marks Distribution and Problem Statement Finalised.

Apr 6: You are allowed to use **anything** which satisfies the project requirements (For eg: using database programs instead of reading/writing json files). But ensure you add a proper section in your report for **“Project Setup”** mentioning all requirements to be installed and how to run your project on someone else’s laptop. This is to make the grader’s life easier and hence yours!

Problem Statement:

Your goal in this project is to design a website interface which serves as a one-stop movie repository. This project involves

1. Creating your own simple database by scraping from online repositories.
2. Developing a web interface where users can search information about movies.
3. Algorithm ideation and implementation for Movie Recommendation.
4. Additional features under customisation.

Possible Learning Outcomes:

Python requests/Selenium :Web Scraping

Creative Ideation: Algorithm Design (Lowkey intro to ML)

(For people with great interest – Ofcourse beyond the course project limits):

For database storage: Can use MySQL | PostgreSQL | MongoDB | SQLite

For Web Frameworks: Can use ExpressJS | Django | Flask

Basic Requirements:

The following lists the layout and description of basic tasks to be done, above this students are free to add customisations of the proposed choices which is up to the creativity of the student.

Curation of the movie database: Collect information about at least 100 movies covering a good range of genres(Top 100 IMDB is a potential choice). Some of the good online sources include IMDB, Metacritic and Rotten Tomatoes. Information that must be extracted(but not limited to, include the following):

- 1) Movie Title, Year of Release, Duration Length
- 2) Rating from a well known website
- 3) Genres, Language(Optional)
- 4) Directors and/or Main Cast
- 5) Short Summary or the movie plot
- 6) Atleast 1 image related to the movie (Can be image path/URL)

Website Interface: Design a simple web interface with a search box for movie titles which displays all information related to the movie when clicked on submit. Good UI design is not expected for this part. You can attempt it under customisation.

Recommender System: Design a recommender interface, where you ask the user for rating some movies. Based on the ratings provided, the user is suggested 3-5 more movies. You are free to decide your own algorithm based on different criteria. One simple example is demonstrated below, which uses the genres and language as a criteria for recommendations. Algorithm design need not be completely correct/accurate. Marks will be provided for creative ideation.

Example:

Input: Ratings (out of 10):

All of Us Are Dead - 9 (Horror/Action/Thriller/Drama)

Moving - 8 (Action/Sci-Fi/Fantasy/Thriller)

Be With You - 2 (Romance)

Titanic - 3 (Romance/Drama)

The Silent Sea - 10 (Sci-Fi/Thriller/Mystery)

Good suggestions :

Squid Game (Thriller/Horror/Drama/Action)

Alice in Borderland (Sci-Fi/Thriller/Drama)

Night has Come (Mystery/Thriller/Horror)

Rationale: Input suggests that user doesn't like romantic drama, but is interested in action thrillers. Also, most of the movies that user has rated are Asian dramas.

Customisation:

To obtain full marks in this section, students can do some/all of the following to a satisfactorily good extent.

- 1) **User registration and rating:** The website should include a user registration feature, allowing users to create accounts to rate movies. When a user intends to rate a movie, they must first validate their login credentials. Only registered users are permitted to login and rate movies. Students are required to create a registration page where users can sign up for new accounts. The system must ensure that each username is unique: If a username is already taken, the website displays an error message, prompting the user to choose a different username.
- 2) **Rating Meter:** Implement a rating meter unique to your website, showcasing ratings from registered users. If a movie has received any ratings, the average rating should be displayed alongside other scraped information.
- 3) **Recommender for registered users:** When a user wishes to see movie recommendations, he/she must first enter valid credentials. The user can then see the movies they have rated for, and based on the ratings provided, the user is suggested movies.
- 4) **Spell Check on Search Query:** There may be certain misspelt/case insensitively written/missed out words when a user searches for a movie. Design some algorithm/mechanism to handle just this.
 - i) When searched for e.g. **“all of us is ded”**. The spell checker must provided suggestions saying: **Do you mean “All of Us are Dead”?**

- ii) You may also provide multiple suggestions to the user to choose from and when clicked on either of these, it must redirect to the details page for that movie.
- 5) **More Scraping:** Scrape positive and negative user reviews from Metacritic as well as critic reviews from Rotten Tomatoes and display it alongside other scraped information for a movie.
- 6) **Enhanced UI:** Enhance the overall appearance(UI) of the website. The website can be made responsive, ensuring compatibility across different screen sizes. You are allowed to use frontend toolkits like Bootstrap, Tailwind etc.
- 7) Other interesting customisation ideas are allowed, but you need to discuss them with the TA in charge if they are worthy of marks.

References/Help:

For Web Scraping – Python requests module and BeautifulSoup [[Reference](#)]. If you see a Forbidden Error, that may be because the website blocks scrapers. Check [this](#) for some workarounds. To generate fake user agents, you can use [this](#) python module

Usage:

```
from fake_useragent import UserAgent
user_agent = UserAgent.random()
```

Sometimes, websites generate content dynamically after initial HTML response, in such cases simple requests cannot capture it. In such cases, you need a headless browser automation tool like Selenium [[Reference](#)], [[Reference2](#)], which can render the Javascript and capture the dynamically loaded content.

Dataset is supposed to be stored as a json file. Your javascript code needs to parse this json file and show the details of the searched movie. [[Reference](#)]. Similarly you can also store registered user information in another json file(s).

Some servers cause issues when displaying fetched data citing security concerns. If you are not using NodeJS, there is an alternate solution using python's http.server. Check this [doc](#).

File System in NodeJS [[Reference](#)]

Marks Distribution(15 marks):

These marks will be awarded via **viva(s)**, where you demonstrate and explain the implementation aspects of the project.

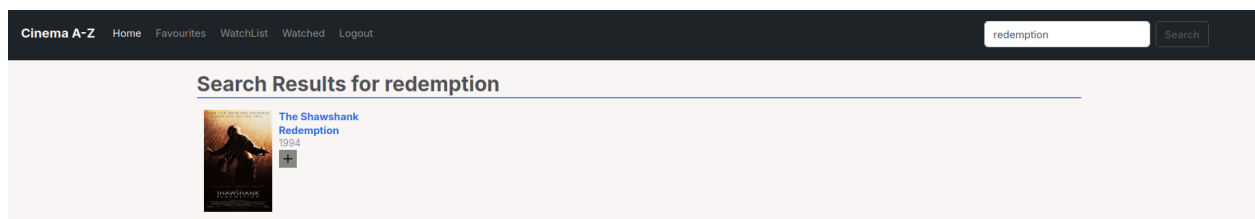
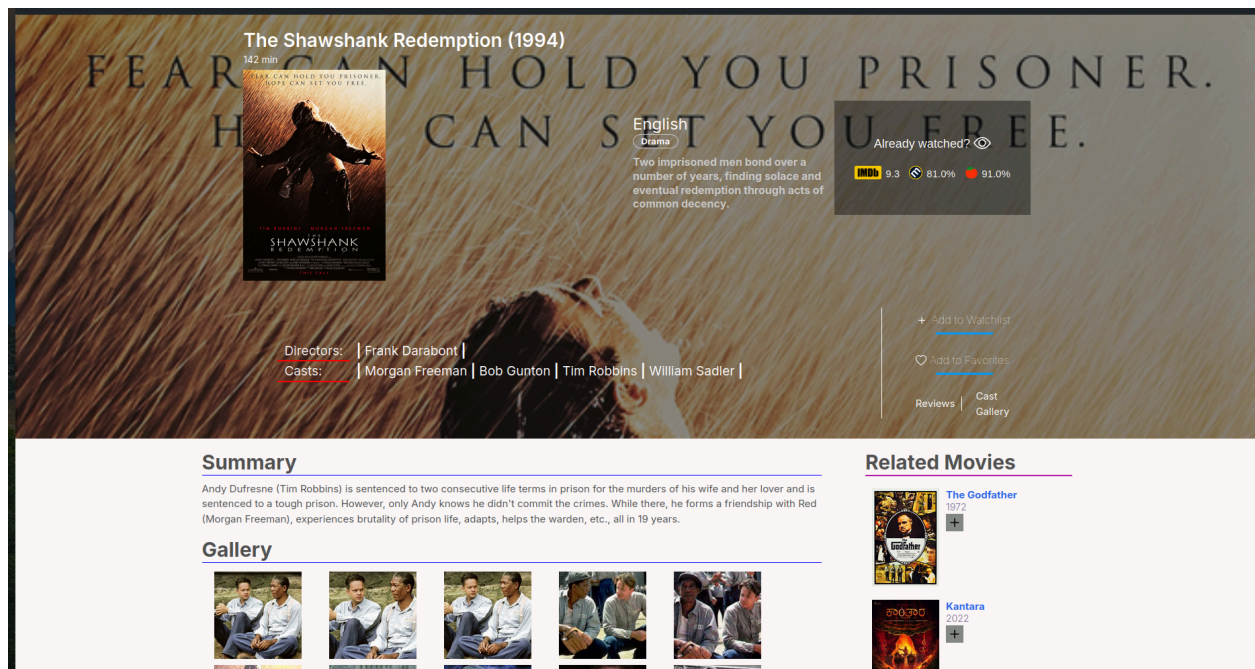
- Basic Implementation : **8 marks**
 - Curation of movie information dataset : **3 marks**
 - Scraping all required information
 - Interactivity of the website: **2 marks**
 - Search Bar works as intended
 - All information scraped is displayed
 - Movie Recommendation: **3 marks**
 - Recommender interface works as intended
 - Algorithm Design
- Customisation : **5 marks**
- Overall code quality, database design, file organisation, LaTeX based report : **2 marks**

- Report must contain the following:
 - Highlight the challenges faced during scraping and what did you do to fix them. For eg: IMDB blocks web scrapers(if using python requests)
 - Mention what information is scraped from which online repository. How do you automate scraping?
 - Explain and put images to your web interface and functionalities, including user registration and movie ratings.
 - Explain the structure of databases created.
 - Include explanation of your algorithm, and a testcase demonstration.
 - Add references to any sources you have copied code from or taken inspiration from in the bibliography, otherwise it may be considered as plagiarism.

Note: All submitted code/report will be checked for plagiarism and if caught, your case will be forwarded to DDAC.

Demo:

The following is a demo to get an idea of how the details page can be designed. You are not supposed to replicate it, but are encouraged to be creative in design.



Search Results for god



The Godfather

1972



The Godfather Part II

1974



OMG: Oh My God!

2012



Reviews for Kantara:

Reviews

User Reviews

10

itguyvoice

Epitome of acting! Chilling experience! It's a mystical thriller and slice of life..... Top class direction, BGM, cinematography, choreography & touch of spirituality. I watched yesterday night, and the movie's experience is still with me like an hang over and not able to get out of it!!!!!! ...

[Read more](#)

10

shyamtheju

No words to explain.... Outstanding Cinema. All departments hardwork shows in the screen. Aravind Kashyap cinematography is next level. BGM and Song from Ajaneesh gives goosebumps. The Director and Actor Rishabh Shetty is fire. Saptami Gowda did excellent and romantic part is very raw and goes with ...

[Read more](#)

1

smartarpit-91770

I agree with the other reviewer this movie is a complete waste of your time. I would have given 5 out of 10 stars for the concept, direction and acting but recent allegations of plagiarism deserves 1 star. Apparently they have stolen the "varah roopam" song that was the only highlight of the entire...

[Read more](#)

1

YourBossHere-00

Omg. It's definitely not a masterpiece which everyone was hyping about. I get it yes it's from South and u need to hype and it and praise just any product coming from there but atleast use the butter in limit not overdo it like u guys did here by making it the highest rated movie right now. Coming ...

[Read more](#)

Critic Reviews

☆

Abhimanyu Mathur

Rishabh Shetty's mesmeric blend of action and mythology ranks among the best of Indian cinema.

☆

Times of India Staff

B Ajaneesh Loknath's music is soothing. With plenty to savour for both mass and class audience, Kantara undoubtedly makes for a wonderful theatrical experience.

Submission: (Tentative)

Submit all your files, including the dataset scraped as submission.tar.gz.

Name the html page which contains the movie search box as index.html. The TA will open the index.html file on a live server and will check if all the above criteria are met. TA will navigate to and fro different pages starting from index.html, so ensure you have linked the pages properly. We expect good coding practices, which includes modularisation, file organisation and documentation.

More details TBA

Contact:

For any discussion or queries, contact the TA incharge at 7995308887

TA Incharge: Sabyasachi Samantaray