

Olivia Duong, Ashwin Anand, Fabio Gunderson, Skandan Venkatraman

Professor Maria Striki

Cloud Computing and Big Data 16:332:494-516

23 October 2023

Final Project Proposal - Group 7/13

Goal/Proposed Solution

Potential Title: Personalized Movie & Show Recommendation Engine Powered by Apache Spark: A Web-Based Entertainment Recommendation System

Introduction/Background

This project was selected to develop a customized movie/show recommendation system using Apache Spark software technology. The resulting movie/show recommendation service will be deployed on a web application.

We selected Apache Spark as the software for our project because of its open-source nature. It's specifically designed for large-scale data processing and analytics, making it well-suited to handle our extensive movie datasets.

Goal

The main objective is develop the movie recommendation service with Apache Spark and host the service on a webpage. The client will be able to receive a list of movie/show recommendations based on an unlimited number of words in their search.

Related Work

Our selected project was executed in Fall 2018-2019 by another colleague. Thus, it is not a recent undertaking. However, our team will take it a step further and deploy the movie

recommendation service to a web application. Another difference would be the implementation of show recommendations for our service. Hence, A Web-Based Movie & Show Recommendation System.

Timeline

Timeframe: Five weeks (Due in December)

1. Week 1: Literature Review of References
2. Week 1: Research on Data Sets (IMdb or MovieLens)
3. Week 1: Set Up and Implement Required Software on Personal Devices
4. Week 2: Coding Time
5. Week 2-3: Coding & Testing
6. Week 3-4: Determine Type of Web Application to Use & Deploy Webpage
7. Week 4-5: Conclusion of Results and Limitations
8. Week 5: Final Submission & Future Work (if time permits)

Design & Methodology

- Movie Recommendation Service
 - Software: Spark
 - Dataset: IMDB Non-Commercial Datasets (Free)
 - <https://developer.imdb.com/non-commercial-datasets/>
 - Number of words in search is unlimited
 - The service will go through every movie in dataset and rank ones that contain the most matches with keywords in the user's search
 - Example Search: Christmas Movies/Shows with John Candy
- Web Page/Web Application

- Language: HTML/CSS
- Open Source Code Editor: Brackets
- Host through Github, free
- Purchase Google Domain, \$12/year to have custom website name
- Or Django

Future Work

If time permits, implement searches to be more complex such as recommending an episode within a said show that is happy/sad/etc. This would be dependent on the data we have and if we can retrieve it episode specific. Moreover, the group could organize the list of results of movies & show recommendations based on word search in the form of images. This would require another data set of images/covers for each movie.

Limitation for this project currently involves datasets of movies that are not fully updated/not real-time. Newer released movies past a certain date will not be incorporated/available in our search.

References

“IMDb Non-Commercial Datasets.” IMDb, IMDb.com,

developer.imdb.com/non-commercial-datasets/. Accessed 19 Oct. 2023.

“Spark Programming Guide.” Spark Programming Guide - Spark 2.2.0 Documentation,

spark.apache.org/docs/2.2.0/rdd-programming-guide.html. Accessed 19 Oct. 2023.