Project Proposal – CS410 Fall 2023

Sole Member/Project lead

David Nieves

dnieves3@illinois.edu

1. The team consists of just me David Nieves. I am the captain so I will be the only one making edits to the program and any changes to the code or GitHub repository.

2. I chose a free topic to create a movie recommender program. The task is the user chooses and provides examples of movies or shows and the program will process these and output recommendations based off a excel database of other movies and shows. The program will search for the user's choice and read the title and description then search for related movies and shows with similar titles and descriptions.

   It's an important program because we often have push recommenders in our daily lives that search our queries and search for recommended content that's related. This program is good practice for designing a basic search and comparing application. There was much discussion about push applications in this course such as Spotify, and Netflix. This application works in a similar way with initial user input then outputting similar results.

   The expected outcome is that the program can successfully output relevant movies/show by analyzing the user input and returning similar yet different results. I expect to evaluate my work on how relevant the output is. Are the movies/shows that are being returned relevant? If I change the database size, does it hurt relevance? Can I improve the relevancy output?

3. I plan to use a python application that uses a excel worksheet with a list of names of movies and their descriptions. I may try to incorporate more datasets later. The program will prompt the user to "enter a movie/show" then will output a list of shows or movies that are relevant for the user to compare. This will be shown to the user using Dash to build the front-end interface which will prompt the user then output the results of the recommendations.  I plan to use BM25 as it's the most stable approach for this program and seems the most applicable.

4. Since I am the only person working on the program, I expect to allot at least 10hrs a week on this project if not less or more depending on flexibility of schedule.

Some of the main tasks needed to accomplish are

1. Obtain a reliable dataset of shows and movies 1-2 hours
2. Organize dataset and begin planning front-end design and implementation 3 hours
3. Design a front-end application and layout that applies a search and retrieval function for said database 10-20 hours
4. Design an algorithm using BM25 that implements the search function to the front end and searches the database outputting a similarity score for the 1 movie/show compared to the related outputted 5 movies/shows 10-20 hours
5. Test and debug and make improvements to the scoring algorithm 10 hours