

## Project Progress Report CS410 Fall 2023

David Nieves

dnieves3@illinois.edu

Some of the main tasks needed to accomplish are

1. Obtain a reliable dataset of shows and movies 1-2 hours

**Progress made Updated:** I have found a database in excel that I can use that will provide the data for the application currently only at 10,000 titles but will keep looking and see if I can find a more updated and wider list of movies/shows database.

2. Organize dataset and begin planning front-end design and implementation 3 hours

**Progress made Updated:** I have created the general layout of the design of the webpage and how the user interface will look.

3. Design a front-end application and layout that applies a search and retrieval function for said database 10-20 hours

**Progress made Updated:** I have set the search bars and have the design of how the database will be converted back into a website UI. The user types into the search bar and that connects to the database and lets the user select movies in it that then will be compared using the BM25.

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

**Progress made Updated:** I have started the backend and the BM25 implantation and the rank scoring which will compare and output the top 5 similar movies based on the similarities in the descriptions. I still might edit this to try to get better results especially if I change the dataset. I have completed most of the applications basic functions calling on the database via the search bar and have connected them with some issues.

5. Test and debug and make improvements to the scoring algorithm 10 hours

**Progress made Updated:** I have started testing but the application is not fully functioning yet so still building the complete application and many edits will be made to both the database and the ranking function.

**Remaining tasks:** The remaining tasks are to finish editing step 2 and trying to make a nicer user interface with backend functionality. For Step 3 I need to continue making sure that as I make any edits that the search bar maintains compatibility and no errors pop up. Step 4 has the biggest amount of work to do at the moment as I need to add the BM25 ranking function and optimize it while again maintain connection to the front end that consistently returns relevant movie results to the one the user entered.

**Challenges:** The main challenge I expect and occasionally run into is sometimes the code runs but the output is the same as the previous one without checking new information from the user. At the moment I plan to focus on maintaining that stability without that error. I also need to find a way to implement a new or updated database and test it as it will be different from my current database. Overall, the code and idea seem to work and the only errors I seem to have had have been not optimized code or some issues with the front end. I think the code will need a lot more testing and changing some variables but should make the deadline. I could also work on cleaning up the front end as well.