

Ethan Ma (ethanlm2)  
Zhekun Zhang (zhekunz2)  
Michael Xu (yhxu2)  
CS 410: Text Information Systems  
Prof. Chengxiang Zhai  
14 November 2022

## Progress Report 1

So far we've decided to begin on 3 aspects of the project:

1. Download, parse, and clean the datasets for movies and games
2. Begin creating a ranking system based on keywords
3. Begin creating a ranking system based on genre

For aspect 1, we've completed the cleaning of the data:

1. Movies: 4 Columns
  - a. Title
  - b. Keywords
  - c. Genres
  - d. Description
2. Games: 4 Columns
  - a. Name
  - b. Tags
  - c. Genres
  - d. Description

These are stored as `clean_movies.csv` and `clean_steam_games.csv` on github under the data folder. We also have the original csv files: `steam_games.csv`, `steam_top_100.csv`, and `tmdb_5000_movies.csv` included as well to show the source of our data.

For aspect 2 we have begun our research and experimentation process into how we should design the ranking system based on keywords.

For aspect 3 we have also begun our research and experimentation process into how we should design the ranking system based on genre.

For both functions we have solidified our decision to go with MeTA for implementation, with python as our main programming language.

For general purposes, we have also begun researching a remote Linux server to work on so that we can keep our environment decoupled to our personal workstations.

For our next steps, once we finish our implementation of the ranking functions, we need to test them both individually and together, so that we can further refine recommendations to the users. This combination of both keywords and genres should help users receive results from highly relevant movies based on the games that the user inputs.