Free Topics

1. **What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.**
   Team members:
   - Zhekun Zhang, NetID: zhekunz2;
   - Ethan Ma (captain), NetID: ethanlm2;
   - Michael Xu, NetID: yhxu2;

2. **What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?**
   We plan to design a movie search system based on game input. The task is to output movie results that are related to the game (keywords/genres). This is interesting because there's no such movie search engines based on games, and people tend to look for movies related to the games they play. Current search engine would only match on the keywords on the game name, but we want to build a system that not only considers keywords, but the genre of the game as well.
   The ranking algorithm of movies will consist of two parts: keywords scoring and genre scoring. We plan to use kaggle datasets for the steam games and movies, https://www.kaggle.com/datasets/trolukovich/steam-games-complete-dataset, https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata?select=tmdb_5000_movies.csv. For the algorithm part, we plan to use the MeTA toolkit for implementation.
   We expect the output results would be a combination of movies that are related to the game name, and in the same genre. For example, searching for "Batman: Arkham Origins" would result in "the Batman", "The dark knight rises", "Batman begins", etc. We are going to evaluate the results in two categories: games have associated movies, and games don't have associated movies. For games that do have associated movies, e.g. Batman, top search results would be those associated movies. For games that don't have associated movies, e.g. shooting games, top search results should be action movies that contain shooting.

3. **Which programming language do you plan to use?**
   Python

4. **Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.**
   Tasks:
   Download, parse and clean up datasets (movie and games): 10 hrs
   Ranking system based on keywords + testing: 15 hrs
   Ranking system based on genre + testing: 15 hrs

Combine two ranking systems and refine parameters: 10 hrs
Build search engine infrastructure: 10 hrs
Testing and debugging: 10 hrs
Total engineering hours: 70 hrs