**SI507 Final Project Proposal**

Name: Yuru Gong

**Summary:** This project will implement a Binary Search Tree or a Balanced Binary Search Tree to store movie information from IMDB. Users will be able to search a rating number, or a rating range to see a movie list. The program will provide an interactive command line prompt for user to choose data/visualization options. Display graphs options using matplotlib.

**Dataset:** <https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset?select=IMDb+names.csv>

**Steps:** download three datasets (IMDb movies.csv, IMDb names.csv, IMDb ratings.csv) from this repository and merge them by `imdb\_title\_id` and randomly select top 2000 records.

**User interaction:** prompt user to search a number in range of (1-9.9) or enter a range (n, n) to return a list of movies starting with the highest rating. If user entered range (n, n), prompt user graphs options list, after user enter a choice number, display graphs using matplotlib.

**Graphs options:** only available when the user search with a rating range, and will use the returned movie list to generate:

1. Bar chart of genre/country/language
2. Line chart of years vs rating