Project Proposal – CS 410 Fall 2023

Team Member Information:

Team Name: EMKW
Captain: Emily Meyer

Team Members	NetID
Emily Meyer	emilynm4
Mark Falcone	markf4
Kin Yik Lam	kinyikl2
Warren Dietz	wadietz2

Objective:

Free Topic: Book Recommender System: Your top ten! Best Sellers List Ranked based on a favorite book of yours.

Detailed Description: Best Selling Lists tell you the most popular books that are available currently, but they don't let you know which new book is most likely going to be your favorite. This project aims to rank the best selling list based on similarity to a book the user has recently enjoyed, to let the user know what their top choice for their next book to read should be.

Task: Take the current list of the New York Times best sellers and rank it based on a book that the user inputs.

Relevance to the class: This class explores ranking algorithms and evaluating different methods to present the best documents to the user.

Measurements of Success:

Expected Outcome: Application that takes input of a book that the user likes and outputs a ranked New York Times Best Seller list.

Evaluation Criteria: Relevance Judgements from the users on the ranked list, and user interviews focused on the user interface.

Methodology & Project Scope:

Planned Approach:

Constrain a corpus on the New York Times Best Seller List book descriptions. Create an inverted index. Create a mapping between a book id and its description document. Take in a book that the user likes, run BM25 on the corpus with the inputted book's description as a query. Output a ranked New York Times Best Seller List.

Datasets/Algorithms/Techniques:
BM25 - algorithm discussed in class.
Rank books in the New York Times best sellers list.
.
Google Books API - https://developers.google.com/books
Get book descriptions.

New York Times Books API - https://developer.nytimes.com/docs/books-product/1/overview
Get best sellers books from the New York Times website.

Programming Language: Javascript, Python

Workload: 80 hours (4 Team Members * 20 hours each)

Tasks	Estimated Time Cost (hours)
Get information from Best Seller List	10
Get description information from Google Books	10
Creation of Inverted Index	10
Ranking Algorithm Implementation	25
User Interface	25
Total	80