Group G Project Proposal: WeeeRead

One Sentence Summary:

This site generates recommended reading for book lovers, teachers, and information-seekers based on previously enjoyed books, using Goodreads' reader data to identify other books loved by users with similar taste.

Audience and Goals:

A WeeeRead user is anyone in search of a good book. This website is intended for people like a teacher or librarian looking for new literature to recommend their students, a book aficionado of any age seeking new titles, a professional hunting for further reading on particular topic. WeeeRead users may not know the type of book they're searching for, but they know books they've enjoyed before, and our site can use that information to draw inferences about other books they may appreciate. Our users are not necessarily technological savants, so they're looking for a site that's simple to navigate, something that will take care of the mass data analysis necessary to find a statistically likely good book on the back end and produce basic recommendation information as a result.

Requirements (with acceptance criteria as hollow bullet points)

- View recommendations based off user input (regardless of genre)
    - Displays 3 book recommendations
- Clearly delineate which books are "input" and which are "output" (recommendations) in a novice-friendly, accessible and assistive-technology-friendly way
    - Users can tab through options and enter input typing
    - Instructions and information is laid out clearly and in a large text size
- View recommendations within genre
    - If this option is selected, only books within genres one of the input books will be recommended
- Generate age-appropriate content
    - If option is selected, filter explicit
- See more information about recommended books
    - Shows popularity (number of ratings) of recommended books
    - Shows average rating of book
    - Shows how strong a recommendation it was
    - Shows (or links to) description of book
- Show suggestions for how to find the book
    - Provide working links to library or bookstore recommendations
    - Never return 0 results

Summary and metadata of the dataset we are using:

The dataset we are using for this project is entitled the "goodbooks-10k" dataset. This is a set of information pulled from the "Goodreads" website and uploaded to GitHub by user zygmuntz. The dataset was uploaded to GitHub under the Creative Commons

Attribution-ShareAlike 4.0 International License, which means that anyone is allowed to use the data in essentially any way they please. Goodbooks-10k was uploaded on September 13, 2017, and has never been uploaded. There are 10,000 books referenced in this dataset, and 53,424 Goodreads users referenced. It appears that the ratings are “overall” Goodreads ratings rather than goodbooks-10k specific, so if there were a book that had over 53,424 reviews or ratings, that would not be illogical. Our particular version of the data was downloaded on October 4, 2019.

The full dataset consists of five files: book_tags.csv, books.csv, ratings.csv, tags.csv, and to_read.csv. Each file contains different, but related information. All of the data types of the same names have the same meaning, some information just crosses over between files as necessary. Below are two tables: one is a table of what data types are in each file and the other is a table of what each data type label means.

File	Data Types
book_tags	goodreads_book_id, tag_id, count
books	book_id, goodreads_book_id, best_book_id, work_id, books_count, isbn, isbn13, authors, original_publication_year, original_title, title, language_code, average_rating, ratings_count, work_ratings_count, work_text_reviews_count, ratings_1, ratings_2, ratings_3, ratings_4, ratings_5, image_url, small_image_url
ratings	user_id, book_id, rating
tags	tag_id, tag_name
to_read	user_id, book_id

Data Type	Explanation
authors	The author of a particular book
average_rating	The average rating on a scale of 1-5 of a particular book
best_book_id	Goodreads ID of most popular edition of a given book
book_id	An arbitrary number assigned to each book by the dataset
books_count	Number of Editions
count	The quantity of books with a particular tag
goodreads_book_id	A number assigned to each book by Goodreads (if multiple editions, refers to most popular edition)

image_url	The url for a particular book's book cover
isbn	The 10 digit isbn of a particular book
isbn13	The 13 digit isbn of a particular book
language_code	The language in which a particular book was written
original_publication_year	The original year in which a particular book was published
original_title	The original title of a particular book
ratings_1	The number of times a particular book has been rated one star
ratings_2	The number of times a particular book has been rated two stars
ratings_3	The number of times a particular book has been rated three stars
ratings_4	The number of times a particular book has been rated four stars
ratings_5	The number of times a particular book has been rated five stars
ratings_count	The number of users who have rated a particular book
small_image_url	The url for a thumbnail of a particular book's book cover
tag_id	The number assigned to a particular tag
tag_name	The name of a particular tag
title	The title of a particular book
user_id	The id number of a particular Goodreads user
work_id	Refers generally to a book (all editions)
work_ratings_count	Number of users who have rated any edition of a particular book
work_text_reviews_count	Number of users who have left text reviews on any edition of a particular book

Team Collaboration Plan

We intend for this project to be wholly collaborative, such that we all have a hand in all parts of the project. Our intent, though, is to have point people for specific tasks. Emily is particularly interested in graphic design, so she will be taking charge on the front end work for the website. Evan is excited about understanding and formatting the data, so they will focus their energy on the five data sets the project requires. Elizabeth is particularly interested in the

back-end functionality of the website, and thus will be the coordinator for this part of the project. Evan will be joining her after the data formatting is finished.

In order to keep one another on track, we intend to have brief check-ins at the beginning of class to assess our progress, and bi-weekly, outside-of-class meetings focused on making progress on the website itself.