Team members: Morgan Aloia (mta47), Evan Chen (xc374), Caleb Chiam (cc982), Alena Hutchinson

PM1 Piazza Post: https://piazza.com/class/jqnqfmmybnj1l9?cid=478

Github Repo: https://github.com/tectonic8/cs4300sp2019-hs786-amh345-xc374-mta47-cc986

Heroku URL: https://the-book-was-better.herokuapp.com/

Abstract: The user will input titles of movies that they have enjoyed, and our system will output book recommendations based on these movies. We plan to combine general information about the movie and book characteristics from databases such as TVTropes and ISBNdb, with social information from book and movie reviews on sites like RottenTomatoes and Goodreads, to make these recommendations.

Application breakdown

Input: A set of movie titles (e.g. Fight Club, Your Name, The Hunger Games)

Output: A list of 10 book recommendations with easy-to-skim summaries, ranked by similarity and popularity.

Use cases: This project was inspired by the decline of print sales in conjunction with an ever-increasing boom in film. As movies have begun to overtake books as a more accessible means of entertainment, many avid readers or movie-goers curious about taking a foray into print are left without adequate means of retrieving information about what books they might enjoy. Thus, our project.

To give an example: maybe you've watched a film like Fight Club and thought to yourself, "Dang, I really wish I could find a book that would compare to this dissociative-identity, counter-culture, Stick-It-To-The-Man movie!" What about a body-swap romantic fantasy like Your Name (Kimi no Na Wa) or a dystopian battle royale-esque thriller like The Hunger Games.?

Some of these have obvious answers, which our system will identify as the most relevant, e.g. Fight Club and The Hunger Games are based on books of the same name, but the real interest of the project would be to recommend books with similar themes and genres. If you liked Fight Club, you might enjoy reading American Psycho. If you liked The Hunger Games, something like Battle Royale or the Divergent series might be up your alley. Our system will identify such books and rank them according to similarity and popularity, so that you can trust that these reads will be excellent.

Data sources:

- Movie/book characteristics data:

    - TVTropes (https://github.com/raiben/tropes_open_data)
    - ISBNdb (book database with API: https://isbndb.com/apidocs)
- Movie review data:
    - RottenTomatoes (https://github.com/ninetwenty-one/rotten-reviews)
    - https://data.world/datasets/rottentomatoes
- Book review data:
    - Goodreads (https://github.com/OmarEinea/GoodReadsScraper)

Information retrieval component: Internally, we represent each book / movie as part of a document-(trope/category) matrix. A query must be part of our database, i.e. matches one of the entries in our database, and can thus be parsed as a query vector of trope/category features. We can then use the query vector to retrieve similar documents using a TF-IDF similarity metric.

Social component: Our system will utilise movie/book rating information to create a popularity score (weighted according to ratings and number of ratings) that will be part of the ranking metric.

Machine learning: We aren't sure how to incorporate ML yet, we plan to design this after ML is covered in the course.

Possible avenues for expansion: Depending on whether we implement the above successfully, we might expand this in terms of search directionality (e.g. book title query -> movie suggestions instead), or by expanding the media included (e.g. add anime data.) In other words, The Book Was Better might be generalizable to The Book/Movie/Anime Was Better, taking in a Book/Movie/Anime title input.