

Abstract

Recommendation systems are one of the most used tools on e-commerce website. In United States, there are more than 300,000 books written every year and about 500+ movies released every year in the past decade. Various movies are often motivated by the books written. The books and movies share a lot in common such as genres and plots and thus our design performs a quick cross book-movie recommendation based on the analysis of description and other common features such as user-history, genres, plots, description, ratings and the number of awards. We also believe this can create a social impact to develop the culture of reading.

Data

We have used two datasets for Movies – IMDB and MovieLens dataset. The movie dataset was made sure to have the following basic features – Name, Director, Ratings, Description.

We used three datasets for Books – GoodReads, OpenSource Library and Kaggle having the following basic features – Name, Author, Rating, Description.

Both the data sets were combined and cleaned using MySQL, OpenRefine, DataWrangler and R. Post Data-Cleaning, we had up to 27,000 books and more than 6,000 movies.

Algorithm

For extracting similarity out of texts (description and plots), we used two approaches – the Naïve Bayes “Bag-of-words” and Cosine Similarity Measure. The Bag of Words approach analyzes the description of the media chosen by the user and performs a similarity search based on the number of similar words. The cosine approach looks for the similarity between two vectors created based on the title and the description of the movies. Moreover, we used user history to extract genres that the user prefers the most. The user history and text similarity are given equal weightage and then the final results are filtered on the basis of ratings of the book/movie and the number of awards won.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Evaluation

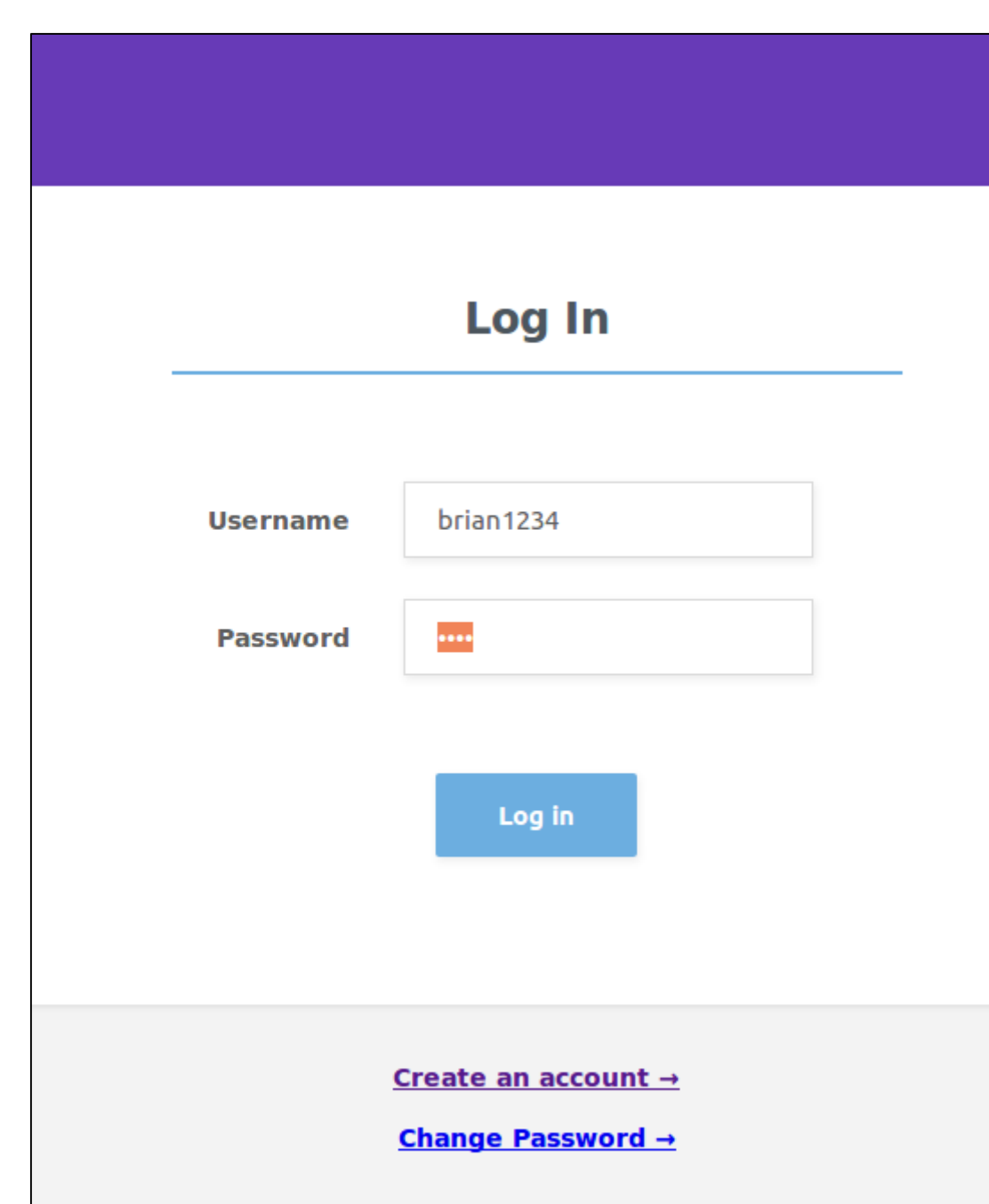
We evaluated our approach by performing a simple blind experiment. We asked volunteers to choose media (books and movies) that they have watched, and/or liked and asked few basic questions on whether they would read the recommended movies and/or books. We helped them by showing the trailers of movies and description of the books that were recommended. Among the recommendations, the volunteers in average liked 83% of them and would like to go ahead and read or watch the media.

Motivation & Innovation

Netflix, Amazon, IMDB, GoodReads are few among many of the examples that use advanced machine learning algorithms to recommend a choice of a movie, a book, or a product based on your history of choice and critique reviews. However, Netflix and IMDB would only recommend movies, GoodReads would only recommend books and so on and so forth. There are no current applications in the market that would perform cross recommendation between movies and books.

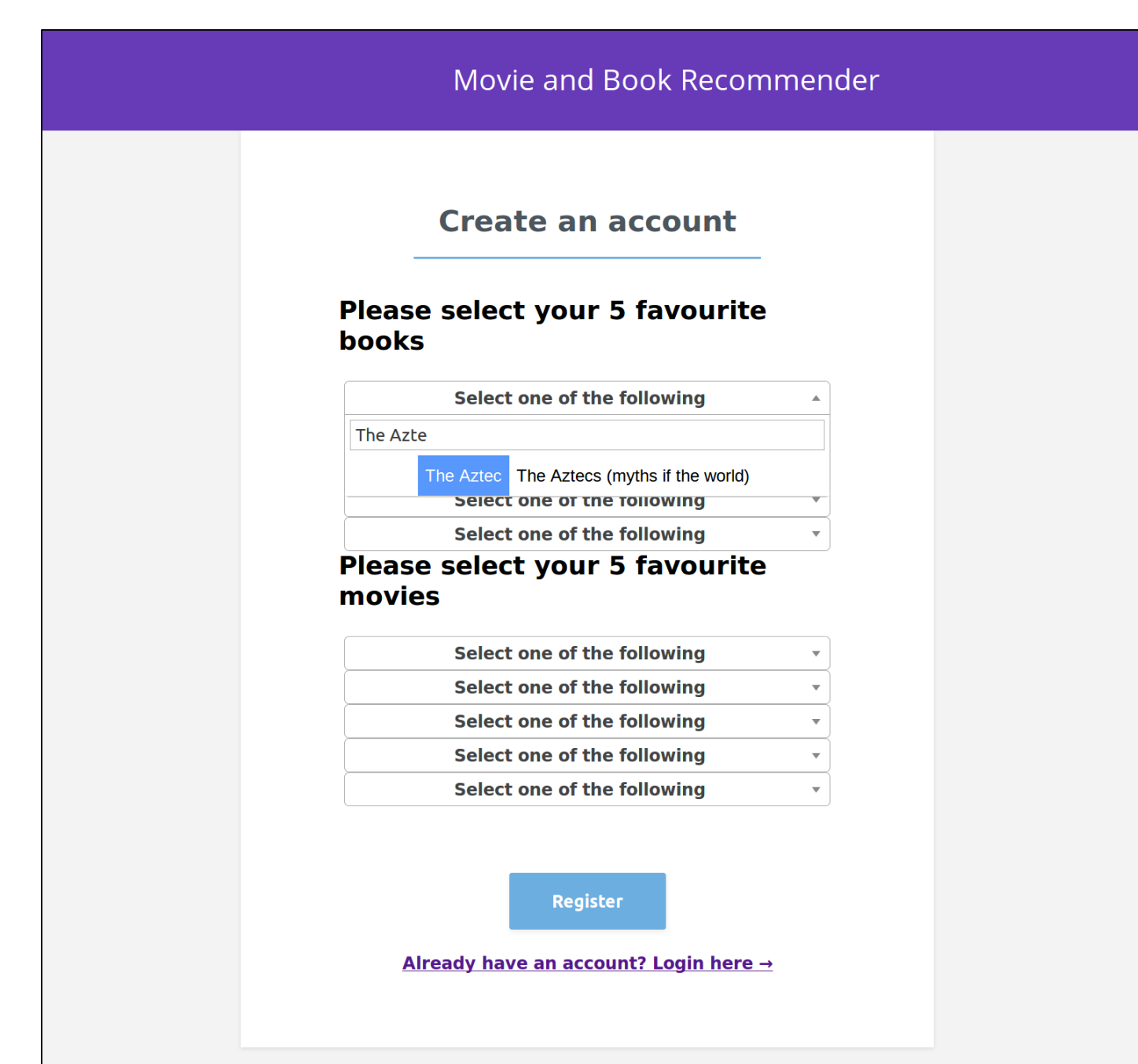
Movies and Books often seem to have similar themes and this occurs often because a lot of movies are inspired by books. This will help the user (1) to watch movies based on their choice of books and (2) to read books similar to their choice of genres to help inculcate the habit of reading.

Web App



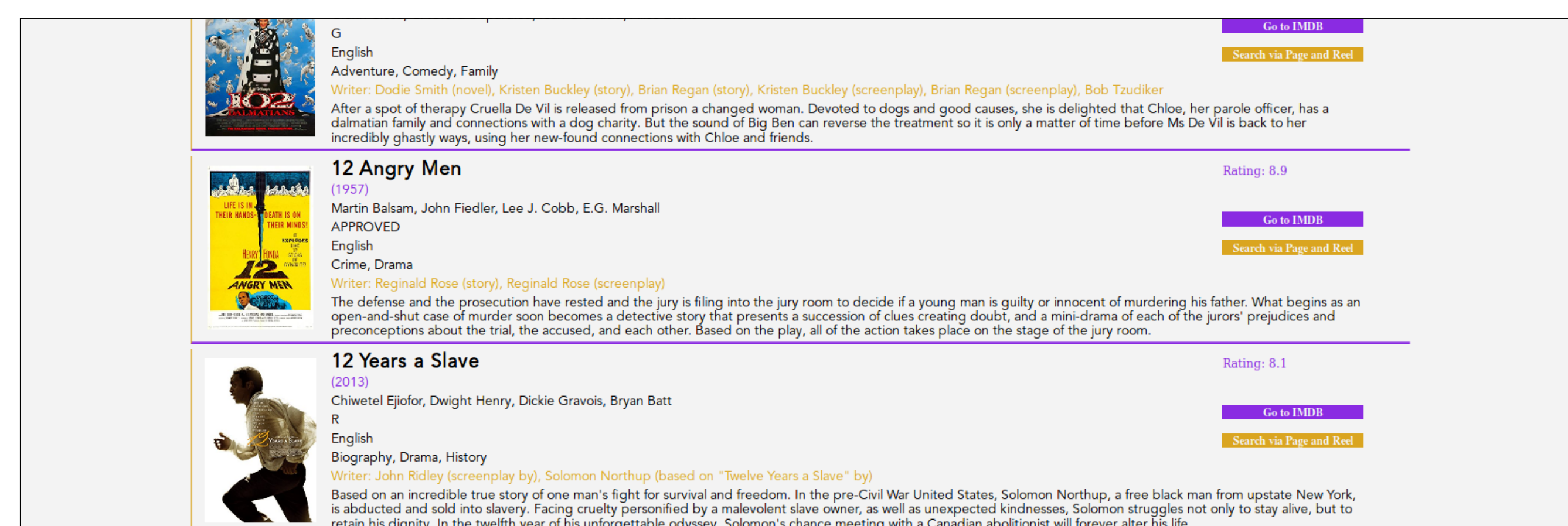
The Sign-Up/Log-In page features a purple header. Below it, a white box contains the 'Log In' section with fields for 'Username' (pre-filled with 'brian1234') and 'Password' (with a red 'X' icon), and a 'Log In' button. At the bottom, there are links for 'Create an account →' and 'Change Password →'.

Sign-Up/Log-In Page



The Selection Page has a purple header. Below it, a white box contains the 'Create an account' section. It prompts the user to 'Please select your 5 favourite books' and 'Please select your 5 favourite movies'. Each section has a dropdown menu for selection and a 'Register' button at the bottom. A link 'Already have an account? Login here →' is also present.

Selection Page



The Results/Prediction page displays three movie recommendations. Each entry includes a movie poster, title, year, genre, and a brief description. The first entry is '12 Angry Men' (1957), an English Crime Drama with a rating of 8.9. The second is '12 Years a Slave' (2013), an English Biography/Drama/History with a rating of 8.1. Each entry has buttons for 'Go to IMDB' and 'Search via Page and Reel'.

Results/Prediction Page

References

- Gediminas Adomavicius, Tuzhilin A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 2005, 17(6):734-749.2.
- Gong S: A Collaborative Filtering Recommendation Algorithm Based on User Clustering and Item Clustering. JSW 2010, 5(7):745-752
- Greg Linden BS, Jeremy York: Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Computing 2003, 7(1):76-80
- Nadav Golbandi YK, Ronny Lempel: Adaptive bootstrapping of recommender systems using decision trees. WSDM'11 2011
- Olivier Chapelle TJ, Filip Radlinski, Yisong Yue: Large-scale validation and analysis of interleaved search evaluation. ACM Transactions on Information Systems 2012, 30(1)