

# CS410 Text Information Systems - Fall 2022

## Project Progress Report

### Team SALE

- Eunbi Go - eunbigo2 (Team leader)
- Akarsh Bhagavath - akarshb2
- Lingfei Yang - lyang26
- Sruthi Jayanti - sjayan3

### Progress as of 11/13

#### Completed

- Frontend
  - We came up with a basic user interface for the Chrome extension. The interface includes a text input where the user can search for book titles as well as a search button. The current user interface is pictured below.

## Book Recommender System

Search for a book title to retrieve similar book recommendations!

Search

- We have also inserted an icon for the extension.



- Backend
  - We started to look at the initial raw data (5 core), and noticed that it is still too large for project purposes especially for both sentiment analysis and recommendations. Instead, we further explored alternatives, and found the Amazon Kindle review dataset from kaggle<sup>1</sup>. This data is extracted from our initial

---

<sup>1</sup> <https://www.kaggle.com/datasets/meetnagadia/amazon-kindle-book-review-for-sentiment-analysis>

raw data, but contains fewer reviews, i.e. 12,000 reviews. It is ready for sentiment analysis.

- We lowered cases for all words, analyzed the data and verified that the languages of all review texts are English except one, which is Spanish. Thus, we removed that review data. Further, we observed that some review texts are as short as one word or two. For example, “enjoyable”, “very hot”, “love it”. They can convey sentiment and thus are considered valid. Thus, we did not remove short reviews.

## Pending

- Frontend
  - We must still connect the dataset and backend components to the frontend elements. We have to retrieve the results in a readable format. This is what we will focus on in the coming weeks.
- Backend
  - **Dataset:** The dataset mentioned in the completion section contains Asin, which is Amazon’s book ID locator, review texts and ratings. However, the data does not contain book names, authors, book descriptions, etc. which are required for our recommendation steps. After investigation, we found that this missing information can be available through [https://www.amazon.com/dp/\[Asin\]](https://www.amazon.com/dp/[Asin]) where Asin is known. If we are not able to find a direct mapping table, we are going to use web crawler skills to grab the book information.
  - **Process:** In the current stage, we put more emphasis on the data preparation. In the next step, we will implement sentiment analysis first, and create recommendations after that.

## Challenges

So far, we have not faced many challenges on the frontend components.