Project Title: **Research Paper Database**

Project Summary

We plan to build a web application to serve as a research paper database similar to Google Scholar and JSTOR. It allows users to search for papers and save papers to read later. Each paper will have attributes such as an ID, title, abstract, and published year. Each author will have attributes such as an ID, name, papers, and affiliations. The project aims to improve the search experience for research papers and facilitate academic research by incorporating new features such as advanced searching and note taking.

Description

Our application aims to improve the user experience that current academic research databases lack. We will provide a convenient query operation for research papers based on keywords, title, or author. Ideally, we would be able to provide both advanced querying options similar to advanced google search (e.g., exact match, exclude words, before or after a date) and the option for note taking or annotations on papers. Incorporating these two functionalities will create a more efficient research experience tailored toward users.

Creative component

We plan to add a note taking/editing functionality to our application for the research papers. For example, once a user finds and selects a paper, they can take in-app notes, make comments and annotations, and save their work locally. We can implement the annotation tool to allow users to highlight or underline text, as well as add notes to certain sections of the paper. This can be done in a web browser by using a database to store and retrieve a user's notes.

Usefulness

The basic (simple) functionalities of the web application are to search papers, save papers, add notes to papers, and user authentication and profile.

The complex functionalities of the web application are to calculate the similarity score of papers to a query, a recommendation system to suggest relevant papers to users, and a rating system for users' feedback.

There are similar applications such as Google Scholar which is a very common search engine for researchers to search for academic papers. We plan to create a web application with better personalized user experience by providing them with the functionalities to provide feedback on the papers, adding personal notes for later reference. We also plan to calculate the similarity score so that the users know how relevant the papers are compared to their search queries.

Realness

The datasets we plan to use include:

1. .json file from arXiv Dataset , which contains metadata of 1.7 million scholarly papers across STEM. Each entry contains an arXiv ID, authors, title, doi, abstract, and categories the paper belongs to.

2. .csv file from Papers by Subject on Kaggle, which contains more than 50,000 papers and their metadata in 155 subject areas. Each entry contains a title, abstract, main author, published date, and categories the paper belongs to.

Functionality

Expanding on the basic functions mentioned in the section Usefulness, the detailed descriptions of the functionalities the the website plans to offer are as follows:

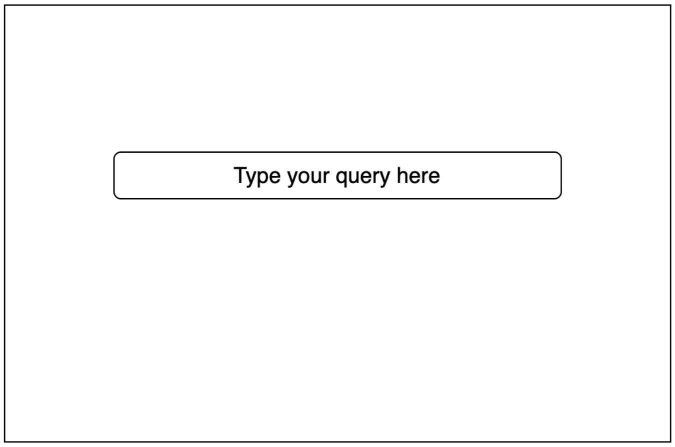
Simple/Basic functionalities:

1. Search papers: Users can search for research papers by multiple keywords and the results will display relevant papers with information like title, authors, abstract, and published year.
2. Save papers and add notes: Users can save papers on their dashboard and make notes on the papers for future reference.
3. User authentication and profile: Users can create accounts and login to the web application for functions such as saving papers.

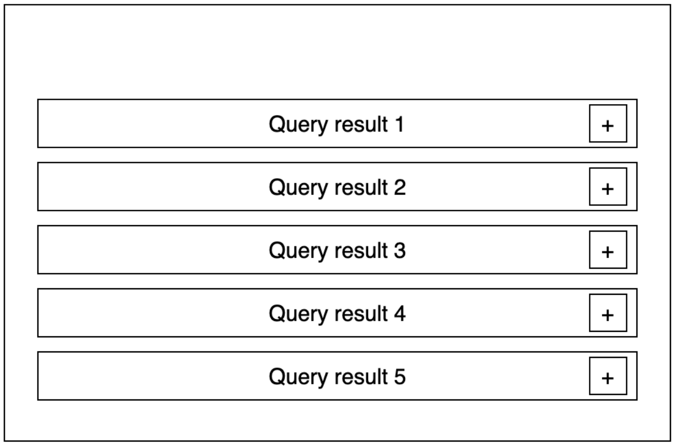
Complex/Extended functionalities:

1. Similarity score: For each search result, a similarity score will be calculated to indicate how closely the paper matches the user's query. The research papers will be ranked in a decreasing order when displayed in the search results.
2. Recommendation system: Based on the user's recent search, personalized recommendations for papers with similar keywords will be generated.
3. Rating system: Users can rate the quality of the research papers and make comments for other users' reference.
4. Annotation tool: Allow users to highlight or underline text in the research papers.

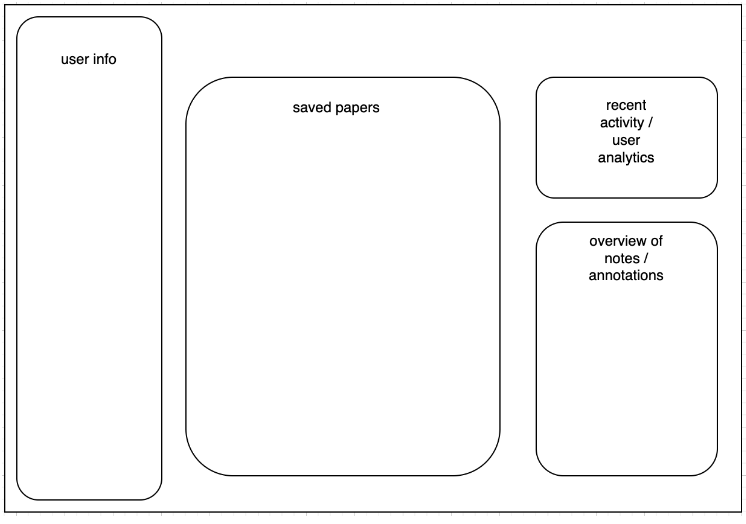
UI mockup



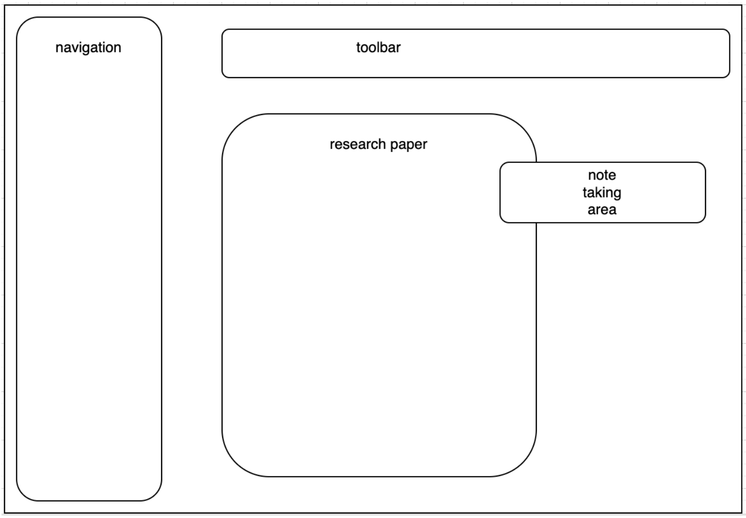
Users can search for papers by typing in their query keywords.



Query results with papers' information are displayed. Users can save the papers.


user info
saved papers
recent activity / user analytics
overview of notes / annotations

A user's dashboard will consist of an overview of their saved papers, recent activity, and notes.



Users will have the option to annotate papers and take notes.

Work distribution

Frontend:

- User login/registration: Claire, Yu Jun
- Search bar (intuitive autocomplete/suggestions) interface: Daniel, Ryan
- Query results interface: Daniel, Ryan
- User dashboard layout: Claire, Daniel
- Note taking tools/document viewer: Yu Jun, Ryan

Backend:

- User login/registration (authentication): Claire, Yu Jun
- Data processing/cleaning and database creation: Yu Jun, Claire
- Embedding generation and paper keywords assignment: Yu Jun, Daniel
- Similarity score calculation and search functionality: Daniel, Claire
- Update user’s dashboard (add/delete papers, handle note-taking): Daniel, Ryan
- Recommendation system: Claire, Yu Jun
- Rating system: Daniel, Ryan