PaperPlex

Summary

Our PaperPlex research paper managing application focuses on creating a system for managing academic research papers. It will allow users to create and manage personalized reading lists by adding or deleting papers. This application will also incorporate a search functionality, allowing users to find papers based on keywords and filters. Furthermore, users can privately take and save notes linked to specific papers.

To further enhance our application, we plan on integrating a vector database that aims to improve the accuracy and efficiency of search queries when finding a related paper. Additionally, our paper manager will track the number of views for each paper, which provides a popularity rating for said paper and its usefulness/relevancy to the user. Overall, the application creates a user-friendly platform for managing academic research papers for students and researchers.

Description of Application

While you're researching machine learning, you find an article that would be valuable to said research, then another, then another! You write them down on a document, but when you go to check it later, you can no longer remember which was which, what was so important in them, or who to cite on your paper. While doing research, it's important to not only find and keep track of your sources, but to notate your thoughts on the subject material as you're reading and understanding it. With our paper manager, you could instead save all of these papers and keep annotations on them as you're working, easily find the DOI or author, and more.

Creative Component

An interesting feature of our application is a vector database extension that would support queries that can match the semantic meaning of similar papers. We could also extend that to support recommendations of similar papers. We would need to first create these embeddings using an open-source machine learning model and insert those into a vector database (PineconeDB, pgvector, Marqo). We can then create queries that search for items that are semantically similar to the search text as well as finding similar items given a paper.

Usefulness

Our tool tailored to researchers and students who need to manage vast amounts of academic literature. This will streamline the research process by providing a platform to store, search, and take notes on academic papers. The common challenge of tracking relevant papers, organizing notes, and finding related papers efficiently will be addressed.

On our website, users will be able to create an account in order to:
1) Add and delete papers from their reading list.
2) Search for papers with specific topics (using keywords and filters like author and publication date)
3) Take and save notes on individual papers

Global features will include:
1) View tracking to see the number of times each paper is viewed by different users. This will be the main way papers are sorted when no filters are applied.
2) Integration of vector database

There are similar websites that allow you to keep a reading list of academic papers such as Google Scholar but it doesn't keep track of paper popularity/views or allow you to type in notes. There are also websites that have other social features, but users can't search papers or recommend similar papers by semantic meaning which we aim to solve.

Realness

We obtained a dataset from Cornell called "ArXiv" on Kaggle, which describes metadata for academic research papers, and an id to link to each paper. This dataset is stored as a json, and contains the following information (as per its description on Kaggle):

id: ArXiv ID [https://arxiv.org/pdf/{id} can be visited to find the paper itself]
submitter: Who submitted the paper
authors: Authors of the paper
title: Title of the paper
comments: Additional info, such as number of pages and figures
journal-ref: Information about the journal the paper was published in
doi: [https://www.doi.org](Digital Object Identifier)
abstract: The abstract of the paper
categories: Categories / tags in the ArXiv system
versions: A version history

Therefore, it has a degree of 10 and a cardinality of 2,501,827. We feel confident that we can fulfill the requirements with a single database by implementing the other features outlined above.

For example, we'd have additional, separate tables for user log-in information, user-taken notes, a reading list, and recommended papers.

Functionality

Our website would offer an account system to create user accounts and log them in. There is a taskbar where the user can view their account information, their saved articles, and the notes they've taken. On the homepage, there is a search query section where the user can type in the name of a paper, author, etc. Once queried, the website will return a list of relevant papers, including whether the user has saved the paper with the star icon (or not saved with a star cut-out), and how many views that paper has. Clicking on a paper will show the user the PDF of the paper, their notes, and (again) the saved status and the current number of views.