Project Title: FindMyPaper

Project Summary:
This project is a web application to help users find academic papers by allowing users to search based on specific terms and recommending papers based on their previous likes. Users will create an account to save their liked papers.

Additionally, the platform will allow users to like papers they receive, enabling the application to recommend relevant papers in the future.The application will also feature a leaderboard displaying the top papers of the day and of all time, highlighting trending research and helping users discover the most popular literature within the specified timeframes. This combination of suggestion and dynamic content presentation enhances the overall user experience, making it easier for researchers, students, and professionals to find and engage with relevant academic papers.

Description of the application:
The application we are trying to develop is a comprehensive economic research paper search and recommendation system that focuses on simplifying the process of finding relevant economic publications. Users will be able to search for specific research papers using any keywords or phrases they want. Our goal with this is to make it simple for people to find papers based on their abstracts and crowd-sourced opinions.

In addition to the search feature, we will also integrate a recommendation system, which will recommend users relevant papers based on their previous likes. Our application will also allow users to see the most liked papers of all time, ensuring the presence of the most high-quality and impactful publications. To further encourage community engagement, we will showcase a leaderboard that displays the most upvoted papers of the day and of all time, allowing users to seek trending research topics. Our goal here is to make reading papers a more social activity, where people can directly see paper popularity based solely on their community's opinions, rather than mixing in several other factors like politics and awards.

Technically challenging features (creativity):

Our recommendation system uses a clustering algorithm to analyze paper abstracts and assign them to clusters based on content similarity. This process involves multiple steps. We begin by processing the abstracts, extracting meaningful features through techniques such as TF-IDF or word embeddings. This step is crucial for capturing the semantic content of each abstract. We then apply a clustering algorithm, K-means, to group papers into distinct clusters. Once papers are clustered, the system tracks user likes to recommend new papers from clusters that align with their interests. This clustering-based recommendation system introduces significant

complexity, as it requires sophisticated algorithm design, data preprocessing, and efficient handling of a large dataset.

Usefulness:
Our application helps users find relevant published economic-related research papers within a short time. It is more useful because our database is more focused on the economic field, and the application only considers the abstracts instead of the whole paper so that the search and recommendation process won't be distracted by other useless and irrelevant information.

This application also provides users with personal designs. Keeping track of what papers they liked allows users to easily come back to their favorite papers without the hassle of having to go back through their file history. The clustering-based recommendation system not only creates a personalized and engaging environment for users but also enhances the discovery of relevant research. By analyzing user preferences and clustering papers with similar content, our application can suggest papers that align closely with each user's interests. This system enables users to explore trending topics and novel ideas within their preferred clusters, making it easier to find content that resonates with them. As users interact with recommended papers, the system continuously refines its clustering and recommendation accuracy, providing a more efficient and user-friendly experience over time.

One of the most successful similar applications is Google Scholar. Our main differences are as follows:

1. Our application features a unique recommendation system, which sets it apart from Google Scholar. This system clusters papers based on content and suggests relevant papers to users based on their preferences and past interactions. By tailoring recommendations, our application offers a personalized and engaging experience that allows users to discover pertinent research even before they begin a search.

2. We only focus on economic papers and abstracts, whereas Google Scholar takes everything into account, which makes it possible to include some irrelevant information.

Realness
Our data is web scraped economic papers abstracts across multiple different publishing companies(Springer, Elsevier, Wiley, … etc). It is currently in csv format. It contains the Journal_Website, Journal_Name, Volume_Issue, Title, Authors, Abstract. It has a cardinality of 148,745 and a degree of 6. Having the abstracts and titles allows us to create our search algorithm while the Journal_Name and Journal_Website gives the user more insight as to where the paper is coming from.

Functionality

FindMyPaper has several different features. The user can log in, and manage their account information appropriately in a similar manner to other websites. Once logged in, the user can interact with the website in several ways. Our website allows a user to create a query that can include keywords, and the platform will utilize regex-based matching to search through paper abstracts for relevant results. In addition to search, FindMyPaper offers a personalized recommendation system powered by a clustering algorithm. Based on the content of the papers and user preferences, the system suggests new papers aligned with the user's interests. This recommendation feature enhances discovery by continuously refining suggestions as users interact with the platform. FindMyPaper also stores a page which contains the most popular papers of the day, and of all time. As a result, curious users can peruse the popular papers and satisfy their thirst for knowledge. Finally, users can query the papers that they have liked in the past, so they can easily come back to their favorite papers.

PAPER RANKINGS   USER PROFILE

TOP X PAPERS OF THE DAY   SEARCH BAR   NUMBER OF PAPERS

#1

#2   PAPER RESULT #1   NUMBER OF LIKES AND LIKE CHECKBOX

#3   PAPER RESULT #2   NUMBER OF LIKES AND LIKE CHECKBOX

#4   PAPER RESULT #3   NUMBER OF LIKES AND LIKE CHECKBOX

#5
                            PAPER RESULT #4   NUMBER OF LIKES AND LIKE CHECKBOX
#...

TOP X PAPERS OF ALL TIME

#1

#2

#3

#4   PAPER RESULT #...   NUMBER OF LIKES AND LIKE CHECKBOX

#5

#...

Division of Labor:

We aim to divide into groups of 2 to split work appropriately, where each group pair programs to complete each task.

Hao and Jason will work on the front end. This includes the User Login section, as well as rendering everything on the website, making sure the user experience is flawless. They'll display the information delivered by the backend appropriately and make the end to end website functional so that it is usable.

Prajwal and Alex will work on the backend. This includes the search algorithm, updating and maintaining our database tables appropriately (managing paper popularity, paper querying, etc), and determining the most popular papers of the day and all time (this includes some database operations, as well). They'll deliver information for the frontend to display appropriately.
t