Project Description: News Document Retrieval System

Project Overview:
The News Document Retrieval System aims to provide users with an efficient platform to retrieve state and national news articles. The system will leverage web scraping, text processing, indexing, searching, filtering, feedback capture, and assessment components to deliver relevant news articles to users based on their queries.

Tasks:
Web Scraping (**Arun KoushikV**)
Implement web scraping to extract state and national news articles from reputable news websites.
Ensure compliance with terms of use and legal considerations.
Text Processing and Indexing
Preprocess news articles to extract relevant information.
Build an inverted index for efficient retrieval.
Searching and Ranking
Implement a search mechanism to retrieve news articles based on user queries.
Rank search results based on relevance.
Refining Searches with Filters
Add filters for refining searches, such as by state, category, or publication date.
Enhance user experience in navigating and filtering news articles.
Feedback Capture Mechanism (Lead Member)
Develop a system to capture user feedback on retrieved news articles.
Use feedback to continuously improve the relevance of recommendations.
Timeline:
[Include a timeline for each task and subtask.]
Dependencies:
Python 3.x
Flask (for web interface, if applicable)
Relevant libraries for web scraping, text processing, and natural language processing.
Evaluation Criteria:
The success of the News Document Retrieval System will be evaluated based on the following criteria:
Relevance of Search Results: Precision, Recall, F1 Score.
User Satisfaction: Captured through user feedback.
Efficiency: Time taken to retrieve and display search results.
Scalability: Ability to handle a growing corpus of news articles.
Legal Compliance: Adherence to terms of use and legal considerations in web scraping and data usage.
Notes:
Customize the project structure and components based on the specific requirements and constraints