# **Project Description**

Team Details :

* Rohan Sai Suhas Arava: S20210010028
* G Manohar : S20210010074

Topic : **News Documents Retrieval**

Project ID / Code : 44

Tech Stack:

* Rust
* NodeJS
* ReactJS

Tasks :

* Indexing components
* Searching components
* Refining searches based on specific filters
* Capturing Feedback
* Assessment Components

SubTasks:

* Parsing in Rust:

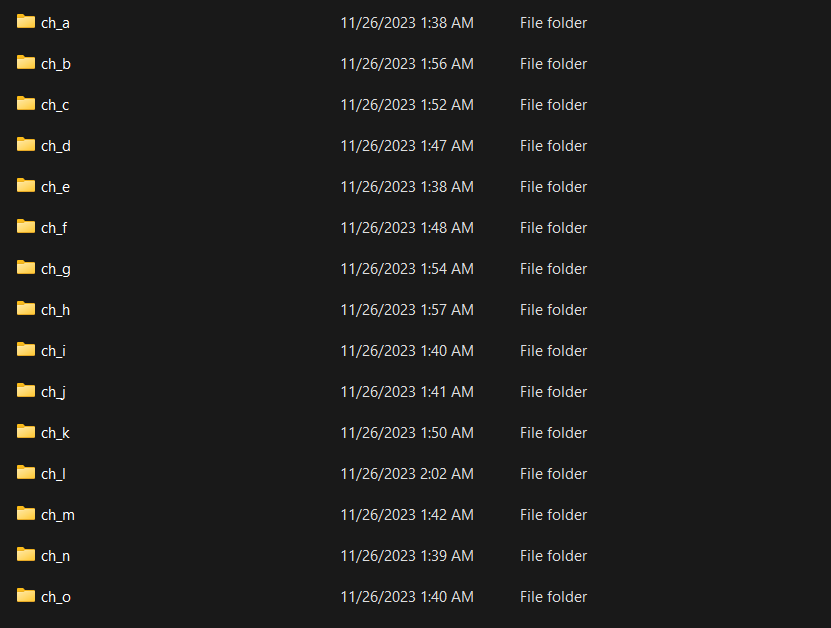
Parsing involves interpreting the structure of textual data to extract meaningful information. In Rust, libraries like nom or pest offer robust parsing capabilities. These libraries enable developers to define parsing rules and patterns, allowing the extraction of specific data from raw text. Leveraging Rust's strong type system and pattern matching capabilities, parsing tasks, such as extracting structured data from unstructured text or interpreting a specific syntax, can be efficiently implemented.

* Tokenizing in Rust:

Tokenizing involves dividing the text into smaller units called tokens, which could be words, sentences, or other defined units. In Rust, tokenization often starts with breaking down the input text into individual words or symbols, a process known as lexical analysis. The regex or tokenizer crates in Rust facilitate tokenization by defining rules or patterns to identify and extract tokens from the text. Developers can use iterators and pattern matching in Rust to tokenize text efficiently, enabling subsequent analysis or processing tasks.

* Construction of Index :

In the realm of retrieving world news from dataset, crafting an effective index involves organizing and structuring data for quick access. This process entails converting raw information into an optimized system that swiftly pinpoints relevant news articles. By creating a streamlined index that efficiently maps terms to documents and incorporates advanced techniques like embeddings, it enables rapid and accurate retrieval, ensuring timely access to pertinent global news updates.

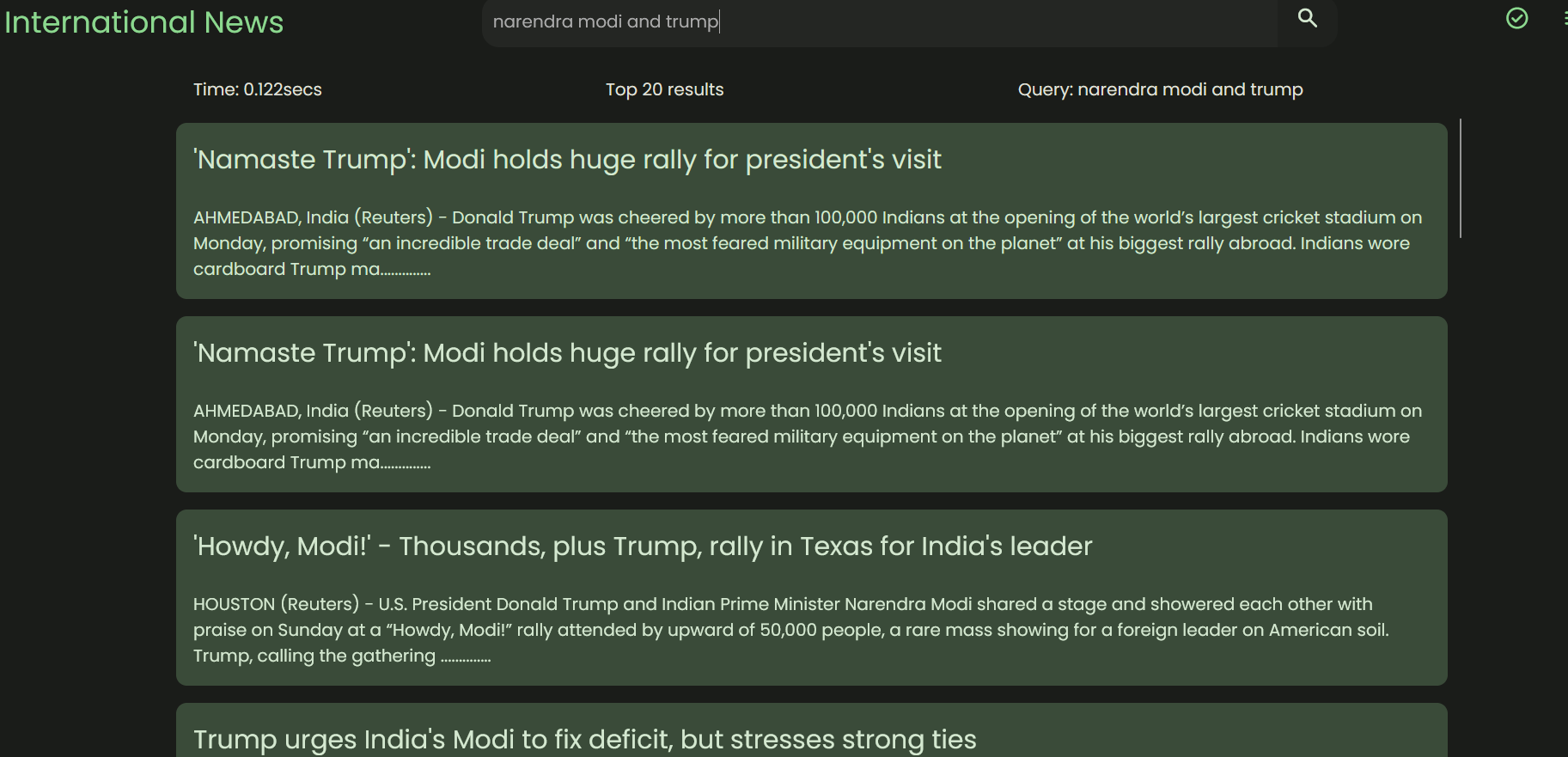


* Refining Searches :

Refining the search for retrieving world news involves optimizing the retrieval process to swiftly access global news documents. This enhancement includes advanced query techniques, metadata enrichmentuser personalization,ensuring efficient access to pertinent global news.

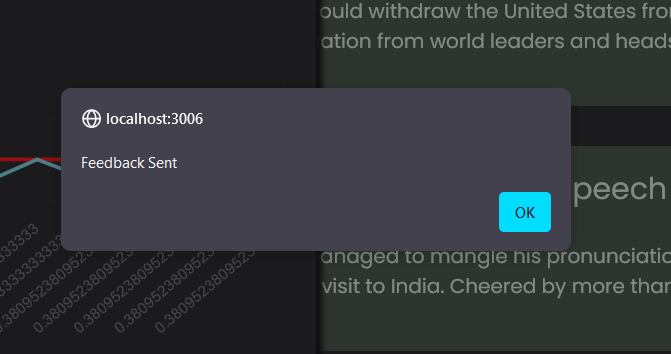
* Search Algorithm :

In our search algorithm, we've implemented cosine similarity by pre-calculating document weights and query weights beforehand, optimizing the retrieval process. This approach saves time by computing document relevance based on pre-prepared weights and quickly assessing query relevance at runtime. By leveraging cosine similarity, we efficiently match queries to documents, providing rapid and accurate search results in real-time.



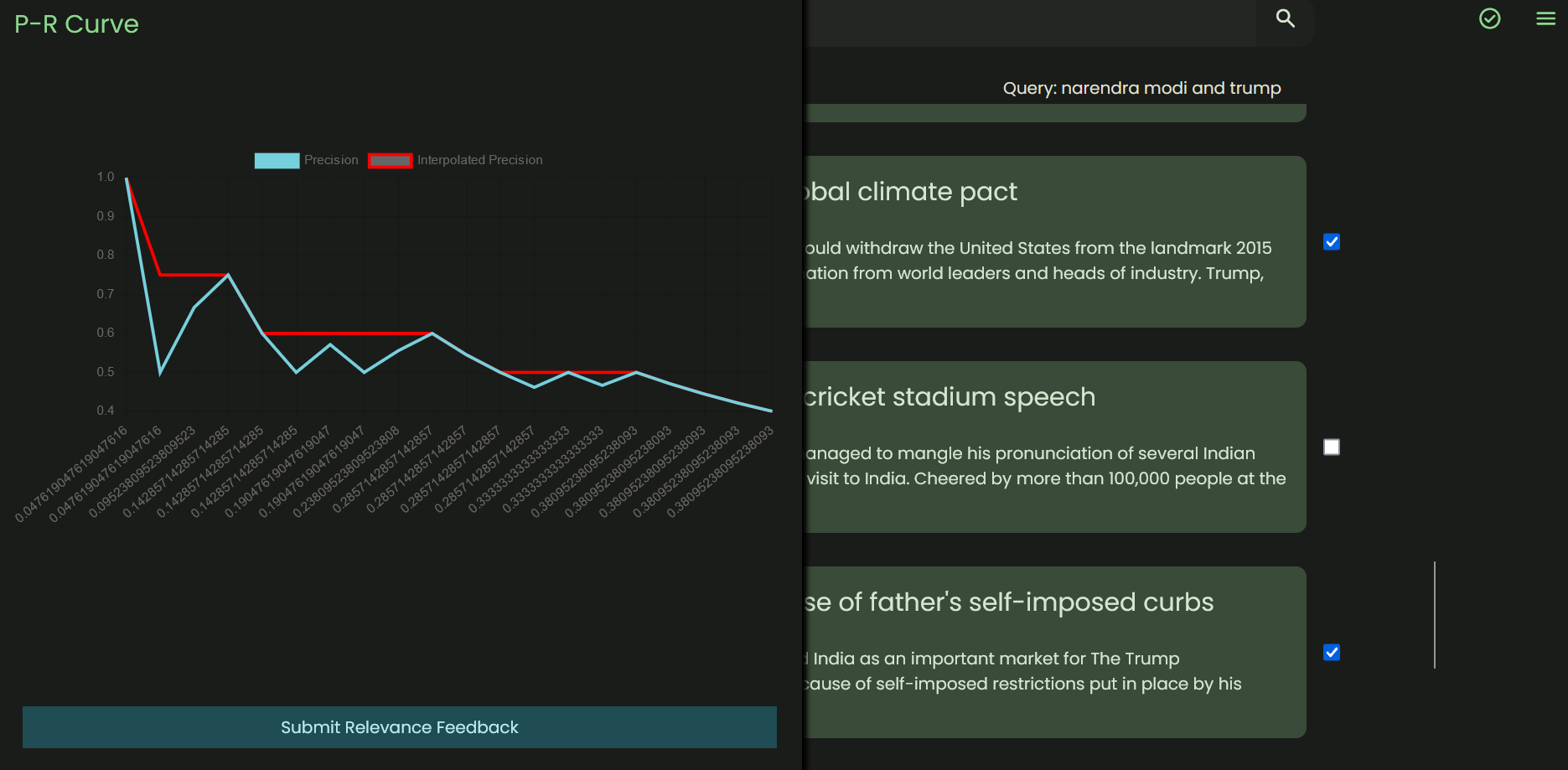
* Relevance Feedback :

Our system employs relevance feedback, utilizing user-provided document relevancy to adjust weights dynamically. When a user marks a document as relevant, we augment its weights by adding a bias, enhancing its significance for future searches. Conversely, if a document is marked as irrelevant, we diminish its weights by subtracting the same bias. This adaptive mechanism refines the search algorithm, improving result accuracy based on user feedback.



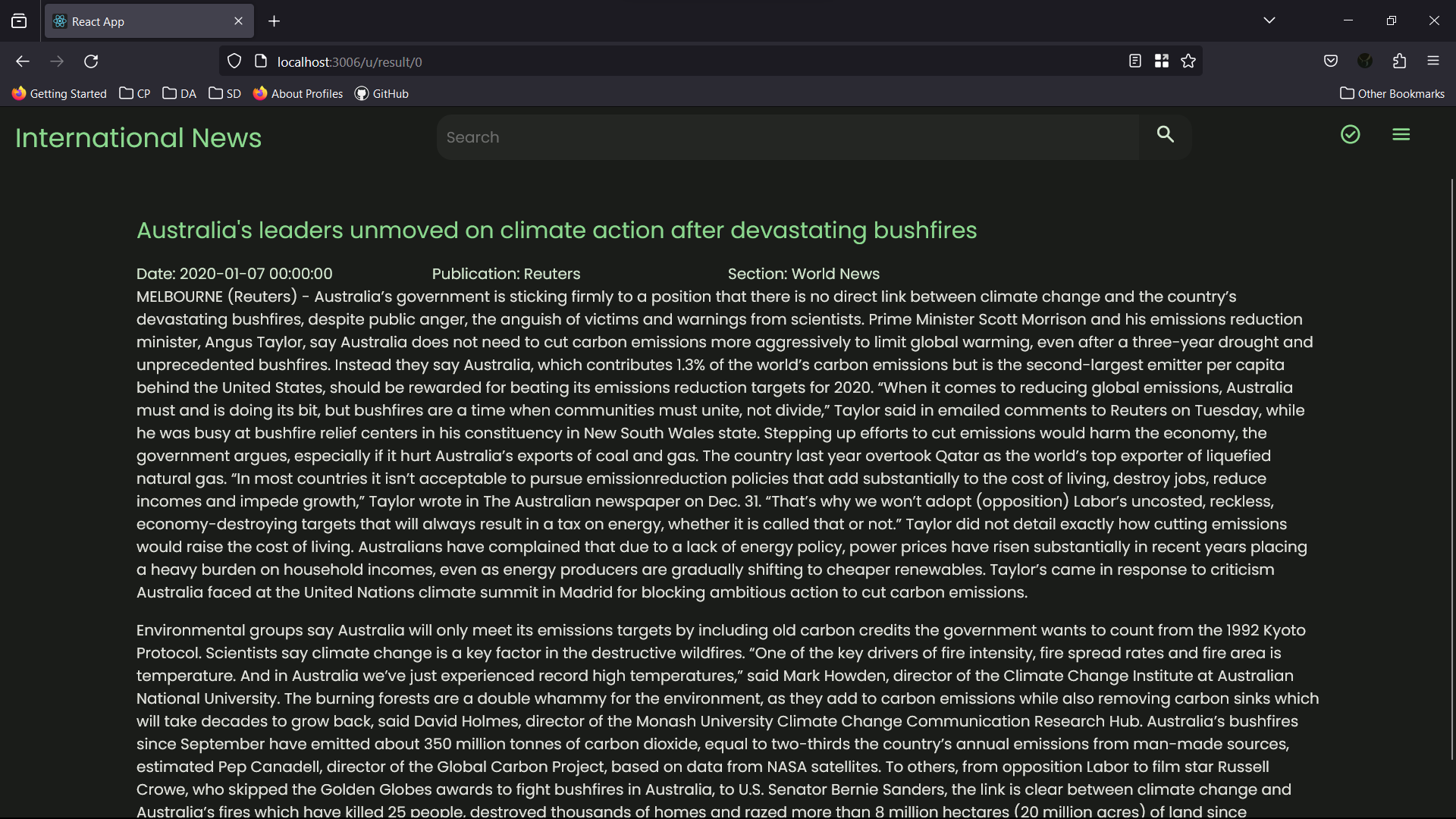
* Evaluation and Assessment Components :

We leverage user feedback to calculate Precision-Recall (PR) curves for evaluation and assessment purposes. By considering relevant and non-relevant documents indicated by users, we generate the PR curve, illustrating the trade-off between precision (the proportion of retrieved documents that are relevant) and recall (the proportion of relevant documents retrieved). Additionally, we employ interpolated PR curves, emphasizing precision at specific recall levels, aiding in the comprehensive evaluation of our system's retrieval performance based on user feedback.



* User Friendly Website :

Our user-friendly website features a search bar enabling users to retrieve the top 10 documents based on their queries. Additionally, we've incorporated sidebars allowing users to provide feedback on document relevance. Users can label documents as relevant or non-relevant, contributing to our relevance feedback system. Moreover, the sidebar contains evaluation metrics such as the Precision-Recall (PR) curve, facilitating an interactive and informative experience for users to assess and contribute to the system's performance based on their feedback.



Link To Demo Video: https://youtu.be/tsOl8vEKmCg