Team Members

Captain: Robert Leggat - NetID: rleggat2

Topic Choice

We have chosen to develop a browser extension that enables users to search within the current web page using the BM25 information retrieval ranking function.

Rationale

When dealing with large pages of text, users could more efficiently search for given information with additional tools. Search functions that rely on exact keyword matches limit the users' ability to find relevant information.

Datasets, Algorithms, and Techniques

We plan to use JavaScript for the extension's development.
For the core functionality, we will implement the BM25 ranking algorithm to calculate the relevance of different sections of the web page to a user's query. We will tune these parameters to ensure the best performance on our given webpages.
We will focus our development on webpages with long text sections, like wikipedia articles and news stories.

Demonstration

We will demonstrate the extension's functionality by providing a live demonstration where users can input search queries, and the extension will highlight or display relevant sections of the current web page.
We will also compare our approach to the existing keyword matching search function, showing that our BM25-based search delivers more accurate and context-aware results.

Programming Language

We plan to use JavaScript for the development of the browser extension. JavaScript is the primary language for browser extensions and web-related projects.

In addition, I will use Python for developing a backend Flask server to perform the bm25 indexing and searching.

Workload Justification

The estimated workload for our topic is at least 20*N hours, where N is the total number of students in our team. The main tasks include:

1. Development of the extension (est. 10 hours).
2. Implementation of the BM25 algorithm and parameter tuning (est. 5 hours).
3. User interface design and testing (est. 4 hours).
4. Documentation and user manual creation (est. 3 hours).

Deliverables at Final Stage

1. Documented source code of the browser extension.
2. A live demo showcasing the extension's functionality with user interaction.
3. A comparative analysis of our implementation against existing extensions that accomplish similar goals, demonstrating the improvements.