

CS 410 Project Progress Report

Members

Shi Yao Liu (shiyao3) (Captain)
Xinrui Zhu (xinruiz4)
Ryan Cedzo (cedzo2)

Summary

Up to this point, the initial UI and setup of the extension have been completed. The project is being completed on schedule so far and is on track to be finished in time before the end of the semester. The prerequisites for meeting all the objectives outlined in the proposal have all been met, the next step is to create the implementation and evaluate its performance.

Challenges

Currently, there have not been any challenges preventing the completion of the project up to this point. In addition, the team has discussed the project roadmap and clarified the requirements. In the following steps, we do not foresee any major blocking issues in the near-future.

Design

The following figure is a screenshot of the current design of the extension.

Enter Query to Find Relevant Sections and Links:

A screenshot of a web interface showing a search bar. The search bar has a dark background with the placeholder text "Enter Query..." in a light gray font. To the right of the search bar is a button labeled "Search" in a light gray font. The search bar is outlined with a thin red border.

The figure was generated from the source code. It is an initial draft and is subject to change in the future as more work is done as the project progresses.

Feedback and constructive criticism from peer-review is welcomed and greatly appreciated.

Code

Link to source code: <https://github.com/cedzo2/BrowserExtension>

The code contains the initial layout of the extension. It currently has no functional code, its purpose is to give us a starting point to discuss implementation. Instead of implementing algorithms from scratch, we agreed that it would be much wiser to use existing libraries. All necessary files are compiled into the extension folder except for the javascript file, which will be

used to perform the necessary tasks for the extension. This file includes the code used for scraping relevant links and making the necessary calculations. Once completed, the javascript file can then be linked to the manifest.json file.

Data Set

Nine Wikipedia pages have been chosen to evaluate the performance of this extension. For each of the nine pages, three query terms have been picked for testing. The team manually scraped the pages for links relevant to the corresponding query terms. Links not in the list are assumed to be irrelevant.

Evaluation

A method of evaluating the effectiveness of the system has been proposed and reviewed by the team. As with the design, this may change as more work is added and requirements are redefined when necessary. The method for evaluation is as follows:

- For every term, manually find the relevant links in the page they were sourced from.
- Compute the F1 measure (for each page) for the default search tool in the browser.
- Compute the F1 measure (for each page) for the intelligent browsing extension.
- With the aggregated F1 measures, perform a statistical significance test to see if the intelligent browsing extension makes a difference.
- If the data does indicate a difference, decide if it was more effective than the exact text match. This can only be done empirically (information retrieval is meant to serve humans, thus it is up to humans to make the final call). Ideally, it would be the users who decide but given the timeline and resource constraints, the decision must be made by the developers.

Feedback and constructive criticism from peer-review is welcomed and greatly appreciated.

Next Steps

The next step is to pick a BM25 algorithm for implementation followed by testing and evaluation. Tuning may be necessary and debugging the extension itself is definitely needed. The interface may need to be improved depending on peer feedback and the opinions of the team. The next step in UI design is to figure out how to present the results. Additional features to enhance usage and performance may be required as well. The prerequisites for meeting all the objectives outlined in the proposal have all been defined and set up, the next step is to create the implementation and evaluate its performance.