Text Information Systems – Course Project – Proposal

This document is my proposal for the course project for Text Information Systems.

Team

The registered team name is "Wonder from Down Under".

This is a single person team with captain and sole member **Alan Anderson**, with netID **alansa3**

Topic

The topic I have chosen is to write a browser extension to allow the user to search the text in the page using a more sophisticated algorithm than simple keyword 'exact match'.

Often, a google search will yield pages with lengthy text, and there is merit in being able to quickly locate the points in the text relevant to the query. For instance, if we wanted to learn about the economic impacts of the War of the Roses, we might well find that useful information is buried in a long document on the War of the Roses. We could search for the keyword 'economic', but we might get a better ranking of relevant sections by performing a BM25 search for a larger set of relevant terms.

Algorithm, language and implementation

I intend to implement my project as a Chrome extension using Javascript. I intend to use the BM25 algorithm.

Testing

I will test the functionality by selecting some queries relating to complex topics that have lengthy text documents online explaining them (e.g. sub-topics relating to a major war or significant civilization). While these history-related use cases are the ones that sprang first to mind, based on personal experience, I will see if I can also find use cases in more technical fields.

I will compare the results of applying the BM25 algorithm with the results of a simple 'exact match' keyword search, to see if the top-ranked parts of the document are more relevant.

Workload

I have not previously written any browser extensions, and have limited Javascript experience. So I expect to spend more than 20 hours across the range of activities:

- Developing and documenting this proposal
- Learning how to build a browser extension
- Implementing B25 in Javascript
- Developing use cases for testing and running tests and comparisons
- Improving the usability of the extension (if time allows).