CS410 Course Project – Improved Search on Current Web Page

Name: Yibo Li
NetID: yibol2
Email: yibol2@illinois.edu
Team Leader: Yibo Li

What topic have you chosen? Why is it a problem? How does it relate to the theme and to the class?

The topic is to create a Chrome extension to improve search experience for current web page. The user can have the similar search experience in the current page by typing in words, phrases and sentences and the extension will index and retrieve ranked paragraphs in the current web page. The ranking will be based on the relevance if the paragraphs.

The motivation is influenced by the suboptimal search functionality that offered by Google Chrome on current web pages (the current search capabilities are limited to exact keyword match). The lack of intelligence issue will be addressed by the extension.

The topic is directly related to the theme: Intelligent Browsing by using search techniques to improve use experience and is highly related to the class material: Text Retrieval.

Briefly describe any datasets, algorithms or techniques you plan to use.

The team will be using Google Developer to the build the extension. The main algorithm used in the text retrieval and ranking function is Vector Space Model (BM25).

How will you demonstrate that your approach will work as expected?

The team will record a video using the Chrome extension on some text rich web pages by typing in the key words and show the ranked list of paragraphs, and the audience can assess the relevance of the list to decide if the approach works.

Which programming language do you plan to use?

JavaScript.

Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

Consider this is a one-person team who lacks coding experience, the majority of the time will be spent on learning how to use Google developer to build the extension and how to code in JavaScript (~10 hours). Seven hours will be spent on the building the functionalities and embedding the search ranking algorithm (BM 25). The remaining three hours will be spent on testing the final Chrome extension for real world search, de-bugging and improve the functionalities.