Group Name: Text Retrievers
Ajit Singh - ajitps2
Nila Krishnan - nilak2 (captain)

CS 410 Project Proposal

We decided to chose a topic under *Theme 1: Intelligent Browsing*, such that our project extends existing browser functionality with additional features. Within this theme, the specific type of extension-based project we want to create includes indexing the current page to allow users to search and query the page. Since default browser search functionality only allows for exact keyword or phrase match and is case-insensitive, it is not very convenient for complicated searches and does not allow for customization. This topic is very pertinent to our course material since our project would involve indexing the text within a page to retrieve relevant information related to the user's query. We would also need to implement a text retrieval function such as Okapi BM25, which we discussed heavily in class.

Since this project involves indexing the text found on a webpage, we first need a web scraping tool or crawler to gather all the text information off of the chosen webpage. There are many built-in libraries and modules we are going to research and explore before deciding which tool to use. Additionally, we need a retrieval function, such as Okapi BM25 to rank relevant text on the webpage in relation to the user-supplied query.

To demonstrate that our approach will work as expected, we will demonstrate our extension on multiple types of webpages to show that we highlight the most relevant text found on the webpage in respect to the user's query. In our demo, we will walk through our project from the perspective of a new user walking through multiple webpages utilizing our extension. As of now, the programming languages we plan to use are Javascript and Python.

In terms of workload, we plan to have at least 40 hours to put towards our project. These are the main tasks we foresee and the estimated time cost for each task:

Research on how to create a web extension	5
Research on how to integrate the BM25 as the backend of our extension	5
Designing the backend of our extension	9
Designing the frontend for our extension	9
Testing our extension on multiple webpages and making modifications	9
Planning and creating the demo	3