CS 410 Project Proposal
Personalized Email Search Engine (Intelligent Browsing)

Team Member: Ashley Yeah (ayeah2@illinois.edu)

In today's world, there is no shortage of advanced and fast search engines. In fact, we are probably all heavily reliant on using Google on a daily basis. However, there are often situations where we want an intelligent search engine for our own personal and private information. Thus, I have chosen the Intelligent Browsing theme and my goal with this project is to create a personalized and intelligent search engine chrome extension for my own Gmail account by indexing all of my personal emails and ranking relevant search results based on my search queries. This project will heavily relate to the text retrieval system topics covered in this course.

For datasets, I will need to download my personal email data. This can be done through Google's personal data export process. Then I will need to create an inverted index on the dataset and most likely utilize the BM25 retrieval function to return a ranking of relevant emails. I will ultimately compare my search engine to the built in search on Gmail, which only really works for exact word matches and see if my solution provides a better search result that doesn't necessarily rely on exact word matches.

For programming languages, I will utilize Javascript for the frontend and chrome extension functionalities. As for the text retrieval system, I will most likely use python along with the metapy library, although I am also considering the possibility of using elasticsearch, which is a more robust system for creating an inverted index on a large set of documents.

Since I am working on a project alone, it will definitely take at least 20 hours to complete the basics of this project. As an estimate, this would be the breakdown of work:

Download + Clean Dataset	3 hours
Create Index + Mapping of Documents	5 hours
Implement Ranker Function	5 hours
Connect Ranker to Interface + Frontend through Chrome Extension	7+ hours

This should be plenty of work to do, but in the case that I can finish, I would like to incorporate more data, including more personalized data, such as my docs, github, and notepad to create an ultimate personalized search engine for my private data.