Project proposal for Team Alpha – Johanan Isaac and Jiyao Zou

Team Members		
Name	**NetID**	**ROLE**
Johanan Isaac	jnisaac2	Captain
Jiyao Zou	jiyaoz3	Teammate

The topic we have chosen for our course project is a search engine for YouTube comments which displays ranked results for user queries. Currently, YouTube does not have any features which allow for searching the comment section of a video. This is a problem because it is inconvenient for users who wish to find specific comments, and restricts the browsing experience of the YouTube site. Our project aims to extend the functionality of the Chrome browser to offer users a way to query the comment section of any YouTube video, directly relating to the theme of Intelligent Browsing. To complete this, we will use some of the information retrieval concepts and techniques we have learned in this course, including term frequency and inverse document frequency weighting, the inverted index, ranking functions, and smoothing.

The format of our search engine will be a Chrome browser extension. When users are on a YouTube video, they will simply open the extension and input their queries into a search bar. The results will be displayed immediately and will include all information associated with each comment, including poster username, number of likes, and an option to view replies. We will make use of the Google YouTube API to extract comment data for our dataset, use the imported comment data as documents, and create some random queries to implement our ranking functions. In order to score each document, we will use metapy as our toolkit and apply pre-built rankers such as Okapi BM25, Dirichlet Prior Smoothing, or JelinekMercer. We will most likely choose Okapi BM25 as our ranking algorithm since it is more popular, but this may be changed as needed.

The programming languages we will use in this project are HTML, JavaScript, CSS, and Python. In the final submission, we will demonstrate that our project works as expected by uploading a video of the extension in use. We will test a variety of queries on YouTube videos with large and small comment sections and explain how and why our algorithm chooses which comments to display.

We expect to spend at least 45 hours in total on this project which fulfills the minimum requirement (2*20 = 40 hours) for a group of two. However, the hours required may increase as we progress. This may be due to time spent learning how to use metapy, challenges encountered when connecting to the Youtube API or extracting comment data, or other technical difficulties. We are planning to meet at least once a week so we can discuss the project's progress and perform some pair programming, if needed. Below is a breakdown of the main tasks of this project and an estimate of the time it will take to complete them.

Tasks	Time Required (In Hours)
Build user interface	5
Process data from YouTube API	5
Implement ranking function	10
Build search feature	10
Query and comment preprocessing	5
Testing results and bug fixing	5
Improving visual design	3
Documentation and demo	2