

## Project proposal for Team Alpha – Johanan Isaac and Jiyao Zou

### 1. About our team

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

### 2. Topic information

#### a. What topic have you chosen?

A search engine for YouTube comments in a Chrome browser extension

#### b. Why is it a problem?

Currently, YouTube does not offer an option to search the comment section of a video. This is inconvenient for users who wish to find specific comments about the video.

#### c. How does it relate to the theme and to the class?

This topic is related to Theme 1: Intelligent Browsing in project instruction because it builds onto the Chrome browser to provide a better way to use YouTube. It is also related to the topics of information retrieval, ranking functions and search engine techniques we have learned in class.

### 3. Datasets, algorithms, or techniques

We will develop a chrome extension that can be used to search Youtube video comments based on the specific query, and return relevant comments according to their ranking scores. We will use the Google YouTube API to collect the dataset. We will use some ranking functions we learned from class such as the Okapi BM25 or JelinekMercer to score each document.

### 4. How will you demonstrate that your approach works as expected?

We will test several different queries on YouTube videos with large and small comment sections and explain how our algorithm chooses which comments to display.

### 5. Programming languages chosen

HTML, JavaScript, CSS, and Python

### 6. Hour distribution and planning - $20 \times (2 \text{ team members}) = 40$ hours minimum

Build user interface for the extension - 5 hours

Process data from the YouTube API - 5 hours

Implement a ranking function - 10 hours

Build search engine - 10 hours

Query and comment preprocessing - 5 hours

Create a display for retrieved comment chains - 5 hours

Improve visual design - 3 hours

Documentation and demo - 2 hours

**Total hours expected to finish the project - 45 hours**