

# CS410 Project Progress Report

Author: Zixuan Ge, zge7@

Name: Zixuan Ge, zge7

Captain: Zixuan Ge

Group Name: dontYellAtMe

Github Repository: <https://github.com/marcuzge/CS410CourseProject>

I have made progress on getting familiar with YouTube's data API v3 and performed experiments to retrieve raw data from the API. I read through the API documents and was able to create a YoutubeService instance to retrieve all comments for a video with the generated API key. (8 hours) I was also able to clean the comments with Panda, by removing whitespaces, stop words, special characters, emojis etc. (3 hours)

I have yet to start reading documentation on NLTK and analyzing comments with NLTK's Vader sentiment analysis tools. I will need to review the lecture videos again on sentiment analysis to convert scores into categories "Positive", "Negative" or "Neutral" smoothly. I have also yet to start on the pie chart creation using Panda to showcase the percentage of each category for a Youtube video. To create a command-line tool, I will need to start on tasks around setting up corresponding flags and parsing user inputs to finish the project end-to-end. Last but not the least, I plan to collect data from human evaluators using the cranfield evaluation methodology we have learned from lectures.

I am not facing any challenges. However getting the Youtube API working to retrieve Youtube comments for a video was trial and error. Since I have never used the Youtube API before, I faced some authentication issues when setting up the API keys. In the end, I was able to bypass the OAuth setup as creating a web application was a stretch goal in my original project proposal, let alone with user account authentication. A lot of the details on the API were also not clearly specified. For example, the API only allows a maximum of a hundred queries on comments at once, so we have to keep querying if the "nextPageToken" key is in the response. I spent around two hours trying to debug this issue.