

## CS 410: Text Information Systems Project Proposal

### 1. Team

I will be working independently on this project. My netid is nehaam2.

### 2. Introduction

I plan to build a sentiment analysis tool on reddit comments to determine how students feel about specific courses at UIUC. This project is an implementation with some improvements of the CS 410 [Reddit Class Sentiment Analysis](#) project done in Spring 2019. I think this is an interesting application as it helps provide students with a quick and clear way to learn about students' opinions on courses. It is not always easy to find out and summarize opinions on courses as they are so varied and searching on reddit often requires sorting through a lot of noise.

#### Approach, Tools and Data

- Use the Reddit API to scrape for comments and posts about a user inputted course.
- Use Course Explorer API to link course code to course name and get other useful data for querying and final display. (Improvement 1)
- Query UIUC subreddit for comments about the course using Reddit API.
- Clean the data for sentiment analysis.
- Use NLTK, VADER or Textblob in Python to conduct sentiment analysis.
- Build a simple website that allows users to input a course code and displays results of sentiment analysis and general course information in a user friendly manner. No account creation or additional data input required.
- Build visual(s) that do one or more of - interactive charts to show statistics like percentage of positive and negative comments, compare sentiments associated with different courses compares (Improvement 2)

#### Outcome and Evaluation

- The outcome/deliverable of the project will be a simple website with functionality as described earlier.
- An important metric for evaluation is precision, recall and accuracy of the results of sentiment analysis and how well it classifies users' comments.
- Another good evaluation metric is whether the query given to the Reddit API ensures that all data related opinions on the course is picked up.

### 3. Programming Language

Neha Mathew  
Netid: nehaam2

I plan to use Python for the parts of the project that involve using apis to get data and conducting sentiment analysis as I am comfortable with its natural language toolkit. I don't have any experience with hosting a website so I plan to read more on the resources available before deciding what to use.

#### 4. **Workload**

Task	Hours
Querying data to get posts using Reddit API	2
Getting additional data about course using Course Explorer API	2
Data preprocessing	3
Sentiment Analysis baseline and improvements	3
Building and hosting website	7
Building Interactive charts/graphs	3
Report and Presentation	2
Total:	22

#### References:

Goering, K. Girard B. Moreno B. Ha D. Spring 2019. Reddit Class Sentiment Analysis.  
[https://mediaspace.illinois.edu/media/t/1\\_milkwa65/112201961](https://mediaspace.illinois.edu/media/t/1_milkwa65/112201961)