Name of Group: JSEN
Group Members: Sarah Zhang, Evan Xiang, Jingwen Zhang, Nancy Li

Description of Project: Scraping college subreddits on Reddit and creating lists of most popular keywords from different colleges. Users enter the college, time frame, and amount of keywords. The scope of the project will be limited to the ten top colleges according to USNews.

Goals of Project: The goal is to be able to scrape any given college subreddit and to return most popular keywords/phrases from the subreddit. This will give us an idea of the interests of the student body at a particular college during a particular time period.

Data Sources We Plan to Use:
HTML and API from subreddit links

Lists of Tasks to Complete and Timeline for Completing them:
Week 4

- Presentation

Week 5

- Collect data (all ten schools)

Week 6-7

- Clean data (for one school)
- Create corpus of insignificant/repetitive words (from one school, limited data)
- Set up database for one school
- Get database working for one school, so that the method can be applied to the rest of the data

Week 8-9

- Clean remaining data and add to database
- Write program to search database based on specified time/school parameters
- Design basic Python UI
- If all other tasks are completed, optionally create simple website interface