What are the names and NetIDs of all your team members? Who is the captain?

- Timothy Cheung tcheung5 (captain)

What system have you chosen? Which subtopic(s) under the system?

- Topic: Free topic
- Sub-topic: Sentiment analysis

I have chosen to base my project on sentiment analysis. I will train a model based on Yelp reviews, such that different words and phrases are associated with a rating. The outcome is to develop a model that can predict a rating that a user would give based on the comment they made. In order to obtain data from yelp, I will use a web crawler.

Briefly describe the datasets, algorithms or techniques you plan to use.

- Web crawler: I will develop my own web crawler that can obtain the comment and ratings on any given Yelp page. I will also add an extension that will allow the crawler to identify multiple pages at a time to crawl from. For example, I may find a list of Chicago restaurants and the crawler will start crawling the pages linked to the first page of results.
- Sentiment analysis model: With training, the model calculates Expected Overlap of Words in Context (EOWC) to determine a rank by comparing similarities between the comment to be predicted and the training data.

If you are adding a function, how will you demonstrate that it works as expected? If you are improving a function, how will you show your implementation actually works better?

- N/A, not adding a function

How will your code communicate with or utilize the system? It is also fine to build your own systems, just please state your plan clearly.

- Two systems: crawler, predictor
- Crawler provides data for training and is only called upon when training is needed
- Predictor uses stored training data

Which programming language do you plan to use?

- Python

Please justify that the workload of your topic is at least 20*N hours, N being the total

number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

- Research: 5 hours. In order to utilize the model for the given topic, I will need to do research and small test samples to examine whether and how to fit the model.
- Implementation: 8 hours. Need to implement web crawler and predictor
- Tuning: 2 hours. In order to maximize accuracy of prediction, I will need to do some tuning.
- Documentation and miscellaneous: 5 hours. In order to prepare reports and other miscellaneous items.