CS 410 Project Proposal

Team: TBD
Member: Hao Mou (haomou2)

Note: I have drafted this new proposal for two reasons: (1) The originally proposed project was too technically challenging to implement. (2) I got feedback that the originally proposed project had too little relevance to the topic of this course. This newly proposed project should ideally be within my technical capabilities and have greater relevance to text information systems.

Q: What topic have you chosen? Why is it a problem? How does it relate to the theme and to the class?

I plan to implement a Chrome extension that enhances Reddit browsing by analyzing sentiments of Reddit comments on the fly and displaying the "positivity score" next to each comment. A higher "positivity score" means there is a higher likelihood that the comment is positive. This should help users stay away from the toxic comments and be mindful of their own mental welfare while browsing Reddit. This closely relates to text categorization discussed in the class.

Q: Briefly describe any datasets, algorithms or techniques you plan to use.

I will be using the VADER sentiment analyzer (https://github.com/cjhutto/vaderSentiment). I will be using this pre-trained model as it is difficult to train and then predict on the fly. VADER is a python tool, so I will need to run it on the server end and the Chrome extension can then send text to it for analysis. If I can find a JavaScript port of VADER, then it should be possible to run VADER in the browser and no server end would be required.

Q: How will you demonstrate that your approach will work as expected?

I will demonstrate the work by browsing Reddit with and without the extension. When the extension is on, we should see a "positivity score" next to each comment and, ideally, they should reflect the positivity of the comment.

Q: Which programming language do you plan to use?

JavaScipt for the extension and Python for the server end. Only JavaScript if a VADER port could run in the browser.

Q: 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.

I have not developed a Chrome extension before, so learning how to develop an extension and how to use an extension to extract content from a website and then inject content back into it would take a considerable amount of time. The sentiment analysis part should not be too difficult with a pre-trained model.