Alexander Scharf
CS410 (Fall 2020)
Course Project

Background:
After learners complete an online business course, they are prompted to enter a "reflection" on how they can apply the knowledge from the course to their job or daily life. These reflections are shared with other learners so they can deepen their understanding, learning how others applied their learning.

This project aims to:
1.) to analyze "useful" and "not useful" reflections, finding syntactic elements that make up each
2.) gather user input for a user reflection and predict whether that reflection is "useful" or "not useful" via a web application

The project uses a real data set from an online learning service with reflections in both English and Japanese language.

1.) Progress made thus far

- The data set has been extracted for both English and Japanese text
- The labeling of the data set (“useful” or “not useful” is being done manually and took longer than expected just for English, but is nearly complete
- Decided to use spAcy as framework and researched its usage as part of tech review
- Ran preliminary analysis on data set (number of words, parts of speech, common words) for both “useful” and “not useful” reflections
- Coding for training model complete by referencing spAcy sample code
- Preliminary web app created and hosted on personal server (take a look at http://alexscharf.com/)

2.) Remaining tasks

- Complete labeling the English data set
- Begin labeling the Japanese data set
- Make adjustments to the code as needed for the Japanese data set
- Make the web app slightly more user friendly
- Conduct user tests
- Update documentation and clean up code to remove debugging

3.) Any challenges/issues being faced

- I thought I could outsource labeling the training data with Amazon Mechanical Turk, but labeling it required too much domain knowledge - taking more time than I thought
- Unrelated to the direct goal of this project, but fiddling with the public facing web server took longer than expected since I don’t have experience in this area. Ended up switching from Apache to Nginx
- Labeling the Japanese data set will take more time than expected. I plan labeling a train data size of 1000 reflections for English, I may do half of that for Japanese

- Still a bit unknown how much rewriting the original application will be necessary for Japanese or any other strange bugs like character encoding, especially with web app