Team 17

Project Title: Mining and Evaluating Verb tags and Other Important POS tags inside Software Documentation

Date: 10/11/21

Members:

-William Sengstock – Team Leader

-Kelly Jacobson -

-Zach Witte -

-Sam Moore -

-Dan Vasudevan -

-Austin Buller -

-Jacob Kinser-

What we've accomplished in the past week/what we've been researching

-William Sengstock – Continued to work on NLP models and experimenting with different word embeddings. This week some of the other group members and I worked on modeling some data using spaCy and analyzing how it compares to others.

-Kelly Jacobson - Worked on a spaCy model that included various word vectorization techniques and POS tagging. Continued to research NLP stuff.

-Zach Witte - This week I experimented with spaCy and how it compares to NLTK. I also looked into detecting phrases using gensim's phrases model as well as different word embedding techniques.

-Sam Moore - This week I conducted experiments with spaCy in Python and also incorporated the Gensim library into my research as well. I took software documentation data and used that to test on to get a better understanding of what we will be working with the remainder of the semester and next.

-Dan Vasudevan -This week I worked on creating an NLP model that uses spaCy and Word2Vec. This was a valuable experience for me because I was able to train a model that takes in a word, vectorizes it, and spits out a similar word. This type of concept is what our entire project is about and it's great I was able to get experience with building an NLP model of that sort.

-Austin Buller - This week we started to learn and use spaCy to implement what we have been researching a few weeks ago, similar to what we did last week but with a different library. We used spaCy to process and create a model for software documentation from an HTML file that was taken from LeetCode.

-Jacob Kinser- For the past week we have been continuing to implement different libraries we learned about during our research phase. This week I mainly focused on StanfordNLP, spaCy, and word2vec. I tested these on software documentation provided to us by our client.

What we're planning to do in the coming week

-William Sengstock – This upcoming week will be dedicated to more model training revolving around the different types of word embeddings. As for the data being used, I will try to adapt my models to take in the software documentation that Andrew gave the group.

-Kelly Jacobson - Continue to build on the spaCy model from last week but also perform a manual review to make sure it is working right. I plan on focusing on evaluating the results of the model and how that model compares to manual review and other models.

-Zach Witte - This week I plan to improve the spaCy model with my group from last week. We will get the output to be consistent throughout the model and manually evaluate the data to see if our model is working and analyzing data how we want it to.

-Sam Moore - This week, we will continue to find ways to improve model performance with techniques discussed in our weekly meetings. I am planning on taking different software documentation data and testing that as well, and perhaps incorporating a different model and writing down the differences.

-Dan Vasudevan - We will continue to work with the model we built last week by creating updates to it. Some updates we can include are tuning the model's parameters to increase accuracy and using other libraries like StanfordNLP to enhance the performance of the model.

-Austin Buller - For this upcoming week, we are planning to build off and improve what we did this past week. Word2Vec will be used to create a model for the data from the HTML file. We also want to evaluate how removing stop words will affect the word relationship in the model.

-Jacob Kinser- For the coming week, I will be continuing to build upon my previous week's work. I plan on diving more into spaCy and word2vec. While looking into doc2vec. I plan on documenting how different techniques affect software documentation.

Issues we had in the previous week

-William Sengstock – I ran into some errors in regards to Word2Vec and trying to get my code to compile. With all of these various libraries, as well as learning Python, sometimes I face some obstacles and running code.

-Kelly Jacobson - I didn't have too many issues this week, just wasting time with bad code and then taking a lot of time to write the correct code.

-Zach Witte - I did not have any major problems this week. I had some trouble getting spaCy to properly install, otherwise, my only problems would just be coding errors that I had to fix.

-Sam Moore - There are not any major problems I ran into this past week. I did hit a few obstacles when trying to find ways to train my model due to lack of experience, but after I reached out to my team and did research, those obstacles were quickly overcome.

-Dan Vasudevan -There weren't any significant issues that I ran across this week. I did have roadblocks while writing the code for the NLP model. However, this was expected and I was able to overcome these roadblocks through the use of my resources and with the help of my peers.

-Austin Buller - Most of my issues this week came from trying to use Stanford NLP before switching to spaCy. I had issues running almost everything when using Stanford NLP and there aren't many good online resources, so my partner and I switched to spaCy instead. I also had an initial issue with spaCy which eventually was fixed by completely reinstalling Python and Anaconda.

-Jacob Kinser- The main issues I ran into this week involved StanfordNLP and word2vec. I was not able to get them entirely working and only achieved partial working

models. My goal is to work with those more this week to resolve any issues I still have left.