Team 17

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

Date: 10/18/21

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 – For this week I used the StanfordNLP word embedding to see how it trains models for software documentation. Specifically, I looked at how it used POS tagging for words and punctuation, along with how it tokenizes the document.

-Kelly Jacobson - I worked on making word clusters using CBOW and Skip-Gram models and comparing how they performed. I also researched supervised word clustering and what kind of data you need for that type of model.

-Zach Witte - This week I worked on analyzing the pros and cons of NLTK. I used what I found and compared them with Will and Jack to come to a better conclusion about what toolkit our group should use for our project. Our whole group then compiled what we did this week into one paper.

-Sam Moore - This week I worked on analyzing the accuracy of the CBOW model. I focused on studying what different factors could change it through research on previous experiences. We then used that research to conduct experiments and present findings to our client and team.

-Dan Vasudevan - This week I worked on analyzing the accuracy of the CBOW model and how it changes when I add and remove certain factors such as numbers, punctuation, and tokenization. Then I compiled all the results and displayed them in the document we created as a group.

-Austin Buller - This week I worked on evaluating the differences between supervised and unsupervised learning models. I did this evaluation because when we decide what our final model will be for this project, we will have to choose either of these methods, each having pros and cons.

-Jacob Kinser- This week I worked on analyzing the pros and cons of SpaCy. I did this to determine if we should use any parts of this model going forward in our project. As a group, we compiled our different research into one report to present to our client.

What we're planning to do in the coming week

-William Sengstock – Zach, Jack, and I will continue to work with StanfordNLP to work with the data given by our client. Continuing on, we will then present our findings along with the other group members for our client, and our instructor who will be joining our weekly meeting.

-Kelly Jacobson - I will be working with Austin again to make a complete model using word2vec, CBOW, and Skip-Gram based on the guidelines the client gave us. We have a pretty important presentation this Thursday so I will try to spend extra time on the project this week.

-Zach Witte - I will be working with William and Jack to create a Stanford NLP model. The other two small groups will be using exclusively different models, so we can get a better idea about the differences between them.

-Sam Moore - I will be working with Dan to make a spacy model with data given to us by our client. We will present our findings/results and propose any suggestions/questions we may have for the following week.

-Dan Vasudevan - I will be working with Sam to make a spacy model with some new pseudocode data our client gave us. Then our group is going to present all of our various results to the client this Thursday.

-Austin Buller - I will be working with Kelly this next week to evaluate models based on the accuracy of POS and AST tags, and then will try to cluster data based on these tags.

-Jacob Kinser- I will be working with William and Zach to make a Stanford NLP model to present to our client on Thursday. Our team will take our results for our experiments and put together a presentation for our faculty advisor.

Issues we had in the previous week

-William Sengstock – Learning about StanfordNLP proved to be a bit difficult because it is not documented very well. On their website, they only provided basic examples for the functions provided, so there was a lot of experiment done on my end to have it work for our models. Starting to work on requirements earlier will prove to be beneficial so I have more time to mess around with the data.

-Kelly Jacobson - I got started late and did not spend as much time on the project as I should have. I need to get better about starting early.

-Zach Witte - I started working on our deliverables a little late and didn't get as much done as I would have liked to. I still made good findings and had good information to present, but I could have gathered more data.

-Sam Moore - I did not have as much time as I would have liked to conduct meaningful experiments on Jupyter Notebook, so I talked with our client and got everything resolved. I also had connection issues during meetings because the wifi was being upgraded at my apartment complex. Everything was resolved through communication with my team and our client.

-Dan Vasudevan - There weren't any significant issues, however, our group did start this week's work a little late so I felt a little rushed while building the model. Overall, our Thursday meeting went well and we have a clear plan for next week.

-Austin Buller - The biggest issue this week was trying to evaluate the difference between unsupervised and supervised models because the data was in a raw form and not labeled at all which is needed for supervised learning.

-Jacob Kinser- Our group struggled this week in determining what our client wanted us to present. This ended up causing us to start late, and have to slightly rush our work. Although our client said he was pleased with our presentation.