

Daily Log

Tuesday October 15

Felt pretty unmotivated but did small things like annotating my code and re indexing my main dictionary's list values so it makes more sense.

Thursday October 17

Day after the PSAT and this is also when I started to become more stressed by college stuff. Anyways, I started organizing structures around the information I want to extract using my updated main dictionary.

Monday October 21

I hate anchor days. But I made it work. Yay. By work, I mean I stopped being stupid and realized that its really easy to identify words by type if I sorted them into sets first.

Tuesday October 22

The numbers look right. I added a new dictionary to keep track of gender of characters (I'll probably automate that later). I made a few simple counters to check different numbers, but I guess I'll really know for sure when I graph the jaunt.

Thursday October 24

I did a lot of checks to make sure it works... I think it does but I can never be sure. This is also the perfect time for me to finish this, since I have been getting bored looking at the same wall of numbers every day. I really don't like numbers anymore. Harper's index is the closest thing to math I'm willing to do now.

Timeline

Date	Goal	Met
Two weeks	Bucket words by character and make it so you can look up types of words easily	yep!
One week ago	Finish last week's thing because that was a mess	correct :)
This week	Graph the data points generated from the previous week and see what it looks like. (This seems simple but I'm also really bad at math so this will probably take a week)	N/A
Today plus 1 week	Figure out how to run other books with the pipeline	N/A
Today plus 2 weeks	Run it on many other books and see what happens	N/A

Reflection

I have been having a bad couple of weeks, but that doesn't matter since I have done something with this project.

Essentially what I have now is a dictionary with character names as the value and an index value as the key. I did it this way since I think it's faster to use lists instead of dictionaries when I'm going through every word in the book. I also have a list (bigJaunt) with all the important info. Each index of bigJaunt corresponds to a character. The actual vales in the list is a list of integers that counts different part of speech and supersense tag. I also have a separate list with a dictionary. Again, each index again corresponds with characters. The list's dictionaries have keys that are words associated with a character and the number of times it's used. I'm using the lemma of each word so different forms of words are still counted together.