

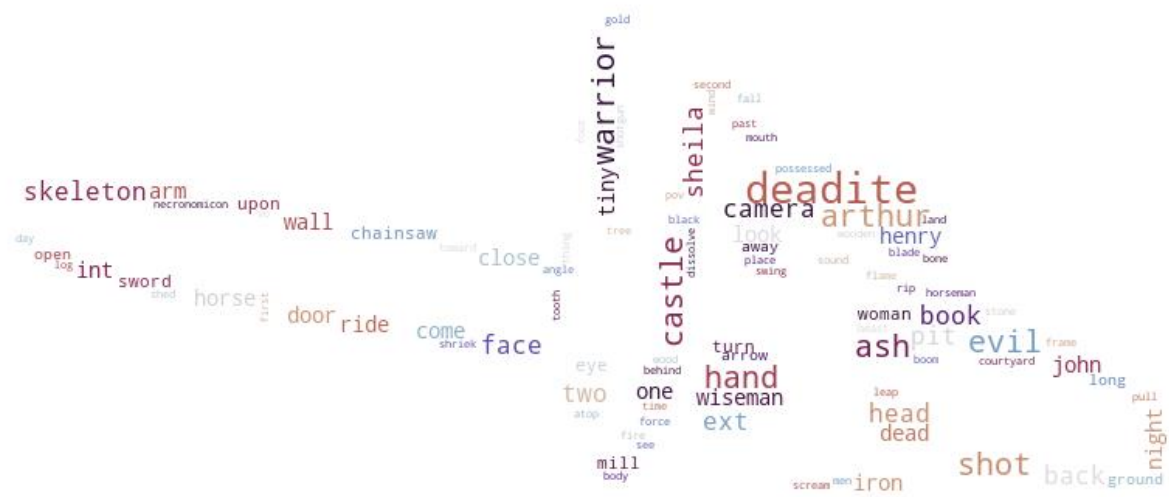
Caleb Klenda

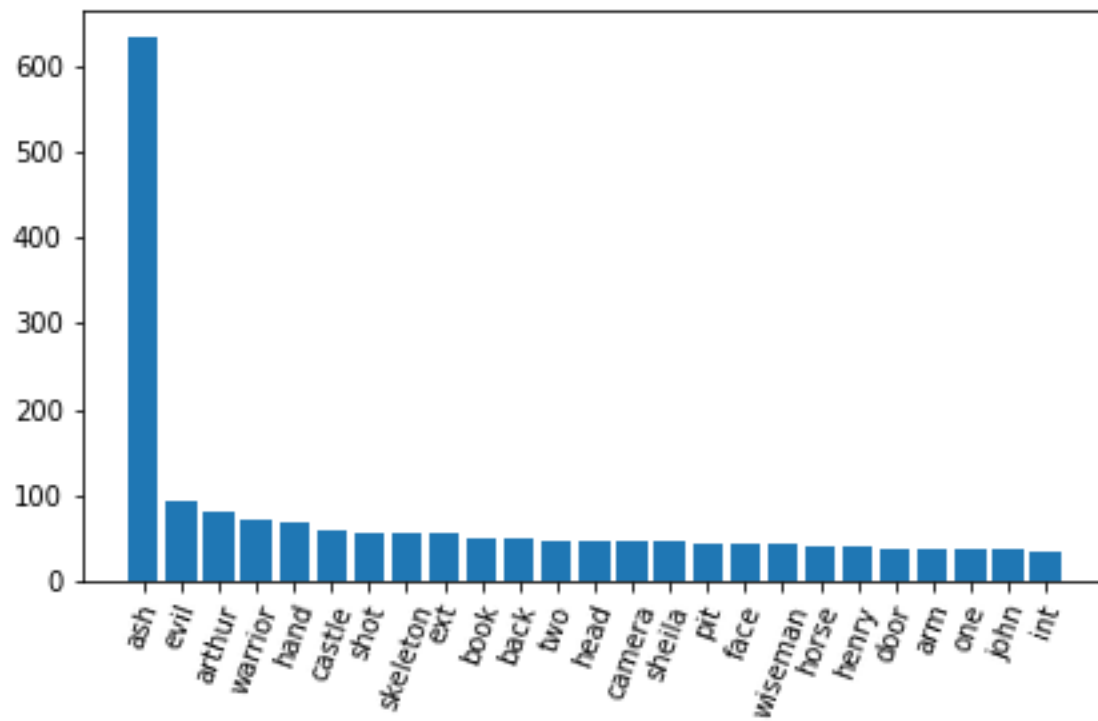
CSE 590

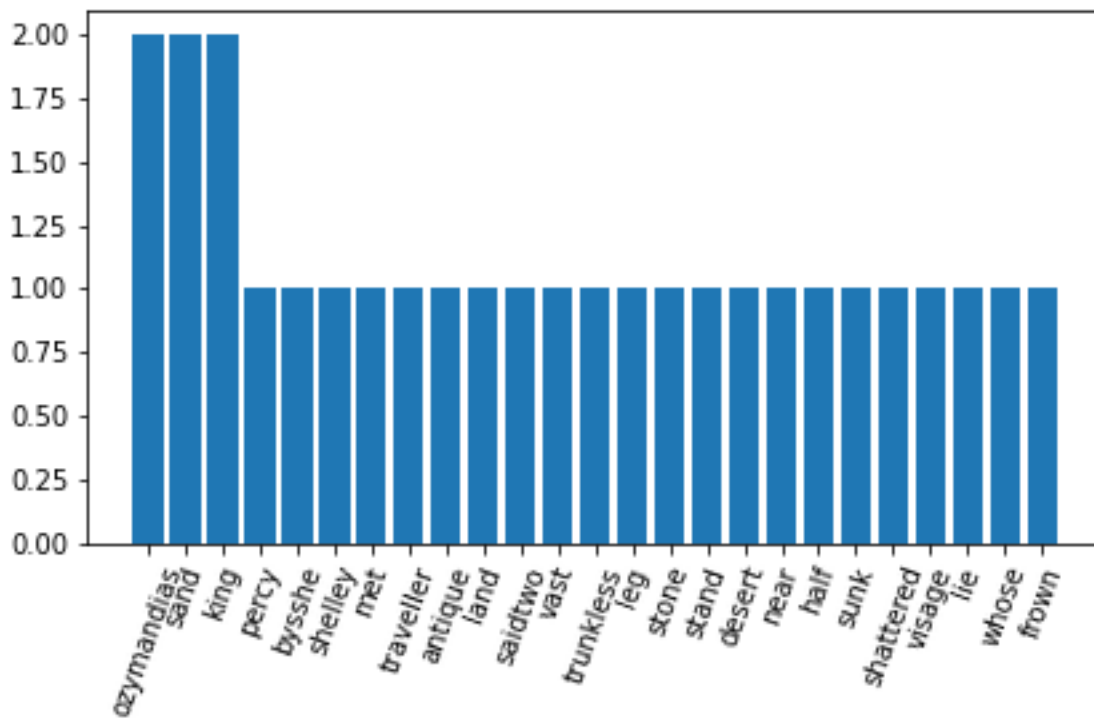
HW 3

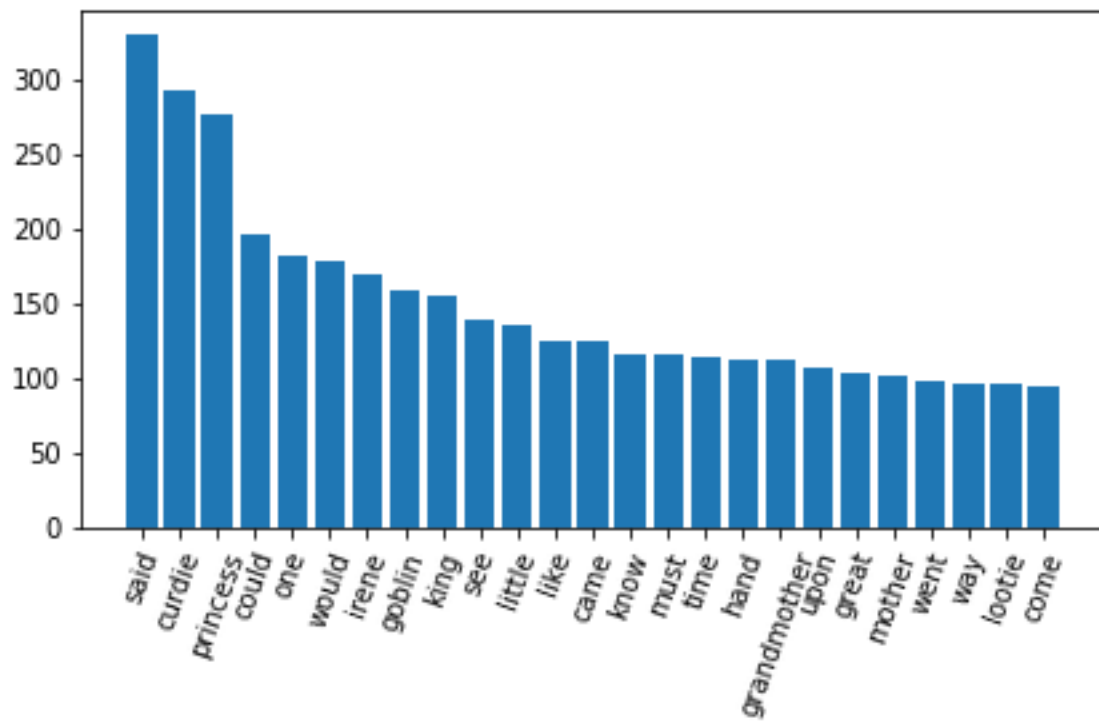
Part 1

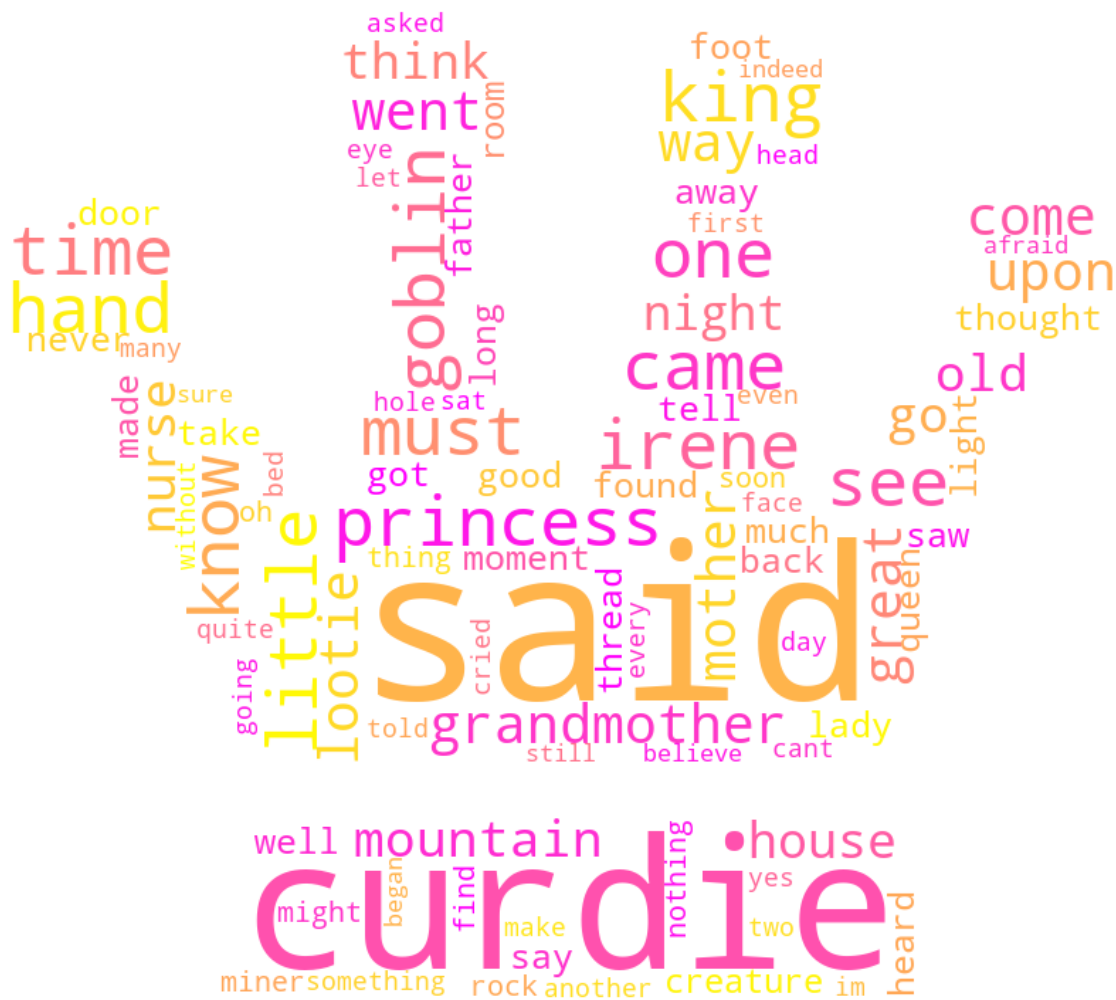
A movie script, a novel, and a poem were chosen for this assignment. "Ozymandius" was the chosen poem; it is short but involves a lot of varied vocabulary and intense tone. The movie script is that of the 1992 film *Army of Darkness*. This script has a lot of clutter in the text, but the actual words in the script are simple and often dealing with the main character Ash. The last text is the *Princess and the Goblin*, a children's novel of medium length. Though long, it consists of relatively easy reading level and a more light hearted topic area than the other two texts.











Part 2

There are some similarities between *Army of Darkness* (AoD) and the *Princess and the Goblins* (PaG). Both have very highly rated the main character, in the case of AoD Ash, and in the case of PaG Curdie. A large difference though is the drop off after the main character's mention. For PaG the next 10 or so words are only 40% less than that of the most common word. For AoD, the dropoff is nearly 80%. This difference is expected and makes some sense. A movie script will refer very frequently to the main character either in the dialogue or in the instructions. A novel will tend to refer to the main character indirectly, and there is more diverse discourse since other character's perspectives can be explored. *Ozymandias* is an outlier here. Due to its shorter nature, there is almost no repeat words within the poem after the processing. This is a unique and interesting quality of the poem, but it makes further comparison with it difficult since every word is used with the same frequency: once. This is expected as many poems are very intentional with their words and only repeat key phrase (though there were no such phrases in the chosen poem).

Part 3

The scores were as follows 6.94 for Ozymandius, 6.11 for PaG, and 4.19 for AoD. These differences are minor, especially with the near identical nature of the poem and novel. However, the poem's brevity combined with the fact the PaG is written for children makes such a similarity not all that surprising. AoD scoring so low is likely due to the fact that movie scripts are written very simply as they are not intended to be read, but acted, and thus a lot of extra complexity comes along outside of the text itself.

Part 4

Pairwise similarity was as follows:

Similarity for PaG to Ozymandias is 0.955

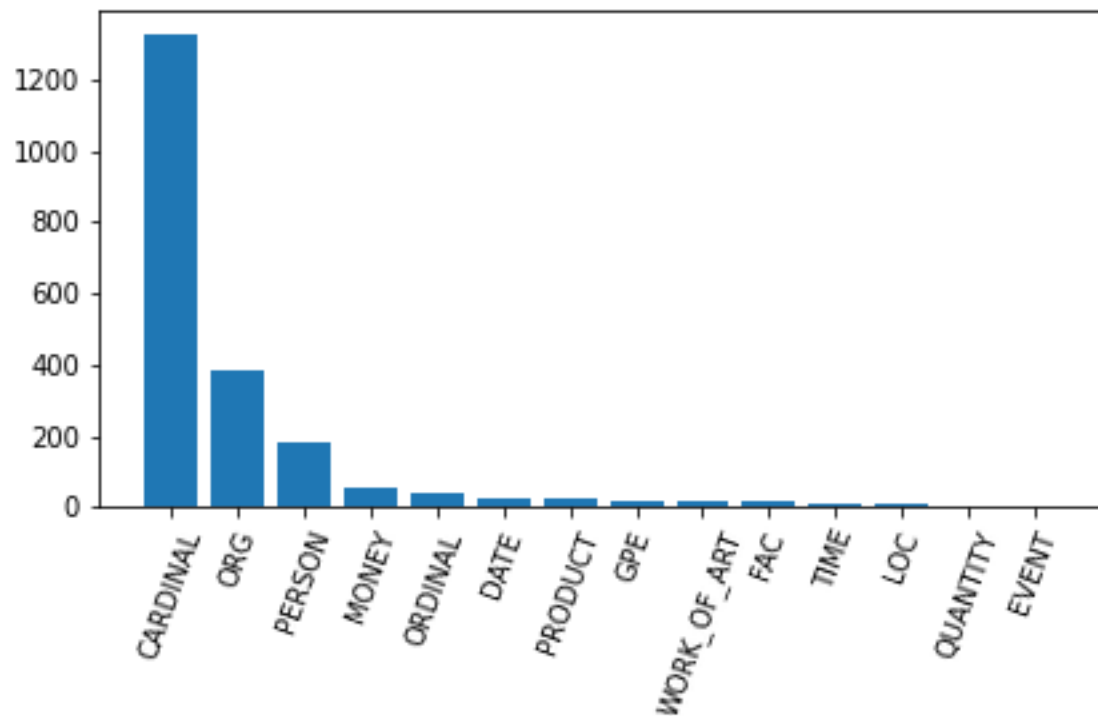
Similarity for Ozymandias to AoD is 0.948

Similarity for aAoD to PaG is 0.958

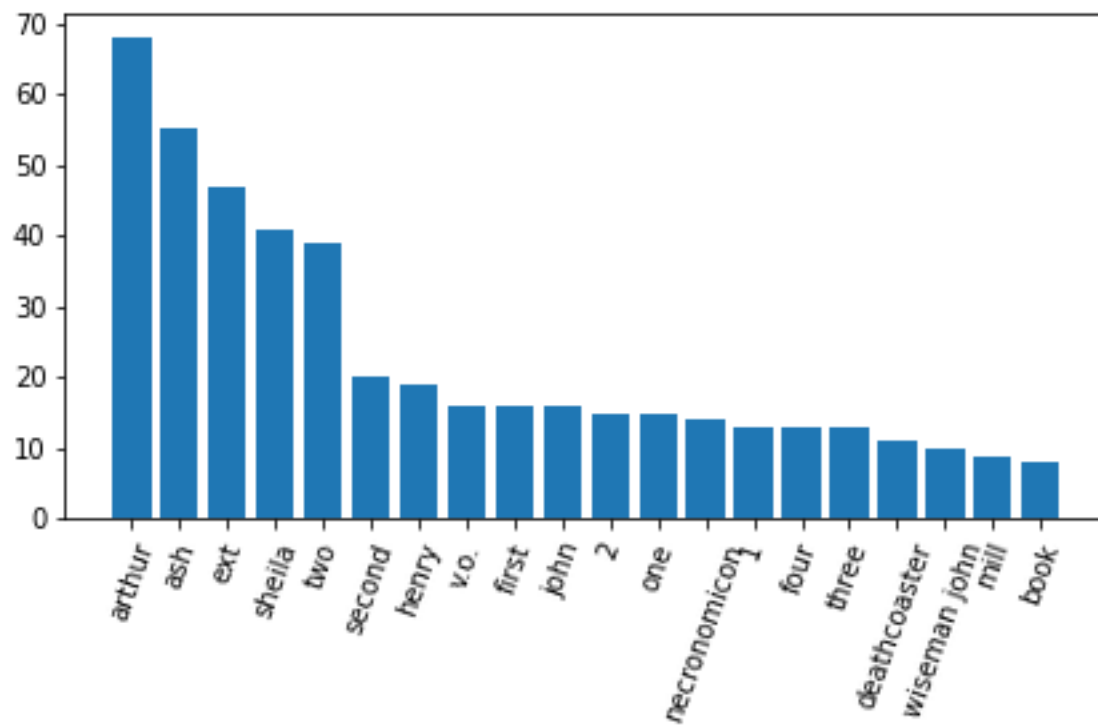
These all seem way higher than expected. These texts are pretty far removed in terms of content, tone, and length. To see such similar scores makes me question if further processing is need before comparison, or if they texts were overprocessed. Given the difference in reading scores and word counts, I would expect a larger difference in these numbers.

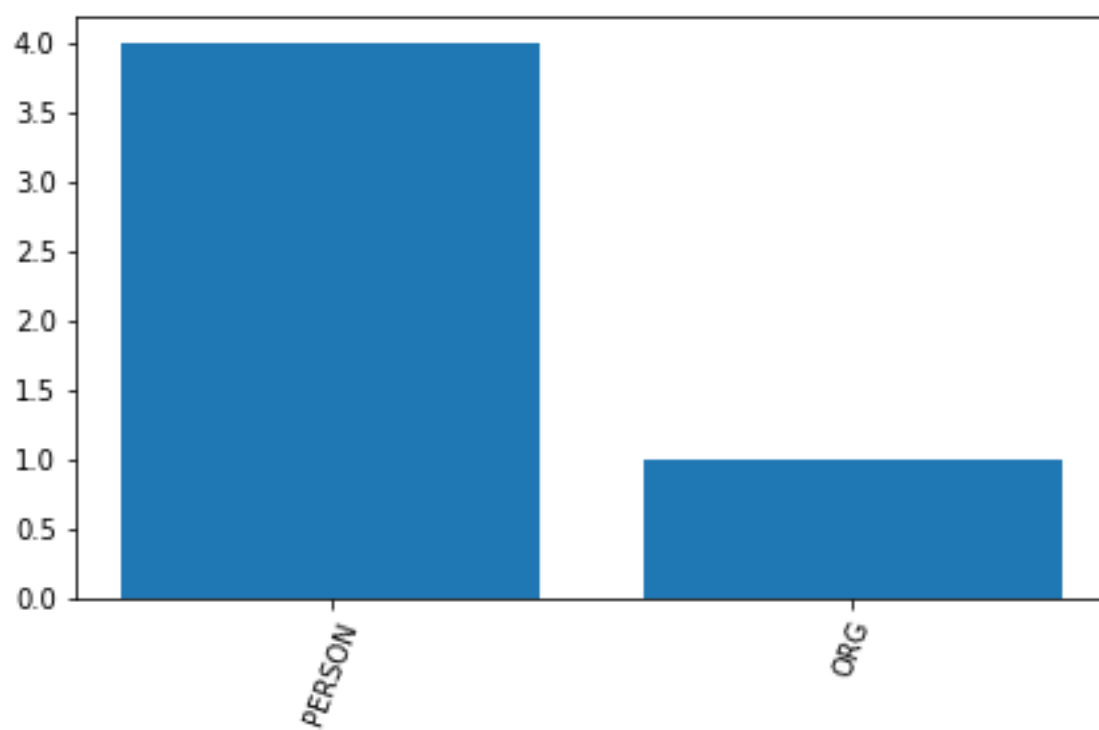
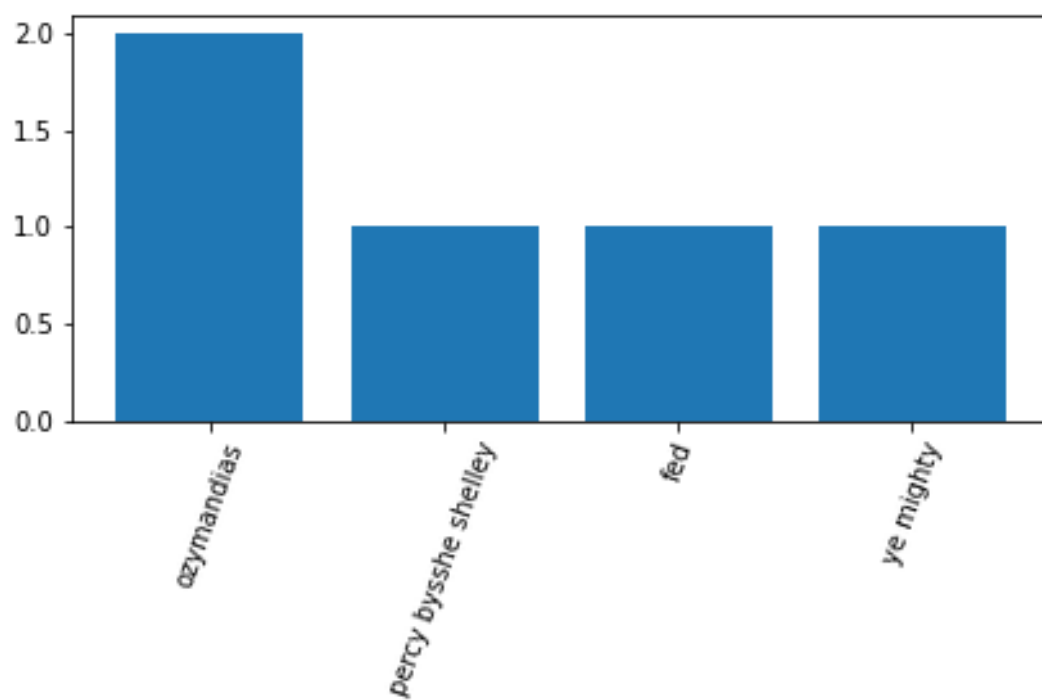
Part 5

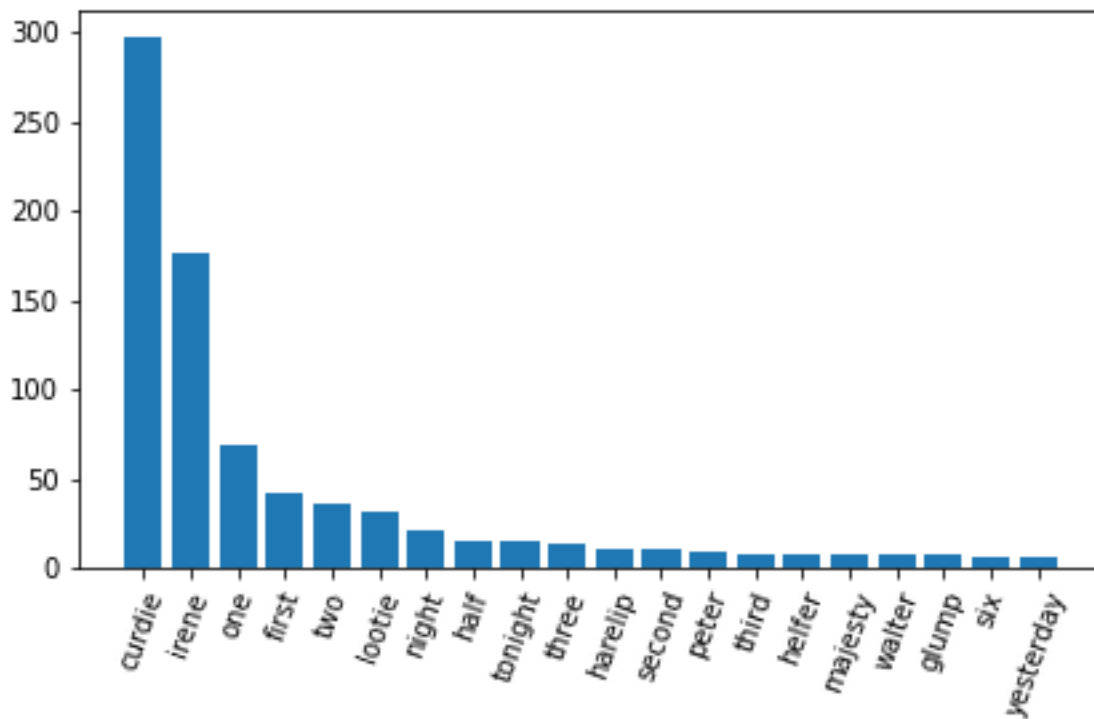
Both PaG and Ozymandias had PERSON as their top entity type, and also the main character as their top entity name. This tracks with the data found in the previous parts. Surprisingly, AoD was quite different, listing cardinal as the top type and Author over Ash for the name.



Aod Figures







Pag Figures

