Austin James
CSCE 578
Assignment 2 Proposal
2/11/2019

I propose to do a comparison of words and their synonyms from a thesaurus I generate from the GNU dictionary to the corresponding entry in the Moby thesaurus. I will then calculate what percentages of the synonyms were in both entries, and find the words that had the highest percentage of similarity.

First, I will use similarity of significant words in definitions in the GNU dictionary in order to identify synonyms for my thesaurus. My thesaurus will be outputted to a text file that will be organized so that I can read the key word from the file and search it in the Moby thesaurus. I can then use a comparison with the Moby thesaurus in order to determine how the accuracy my thesaurus.

I have found a command line interface that will allow me to query the Moby thesaurus from a locally cached copy on Github. This will allow me to access words individually using a bash script. I can then compare the two entries and determine which percentage of synonyms the two entries had in common. This will allow me to determine how accurate my thesaurus generator was. I will also find which entries in my thesaurus were most and least similar to their entries in the Moby thesaurus.