first, we have to compute the number of words, and then we can calculate the number of distinct words.

we have to delete punctuations except for the apostrophes. we also have to delete the titles. (normal text cleaning)

after that, we can create the verses based on the words we've found.

a sliding window-like method has been used to create verses with lengths of 4 and 8.

we can find the size of these verses and words using the "asizeof " function.

after finding the verses, we can create the fingerprints using the hashing function introduced in the lab description and store them in a set, just like the verses and their sizes are given.

in this .py file, two plots are created. one is how the memory changes when the probability of having false positives(epsilon) in the output increases and the other one is the plot showing how the value for “b” changes the probability of having false positives in the output(epsilon).

In the end, both the verses and the fingerprints are stored in a dictionary and the size of the dictionary is also given.

Note: the .zip file only contains the txt file and .py file. No pdf files are included