Programming Assignment 4

Garrett Mestel
C202
Jul 7 2017

The overall problem is simple. We have a file and we want to see how many words are spelled correctly and how many are not. We also want to know the average of how many times the words are checked against the words in the dictionary. To do this though, we will need to read in both a dictionary and the file we want to check.

There are two main steps in the algorithm. The first is reading in the dictionary. The dictionary the problem uses contains just the words and not any definitions of those words. All we need to do for this one is have a list for each letter and read in the next word and add it to the correct list. We do not need to worry about getting each list in alphabetical order. The second half is reading in the file we want to check. To do this we need to go line by line. First, seeing if the line is empty and if it is not removing any special characters other than apostrophes, because some of the words do have them but we do need to check that it is not at the beginning because no word starts with an apostrophe. The last part is to take each word and see if it is in the list. To make this work we will need four counters; one to count the word found, one for words not found, and two counting the amount of string comparisons for both other counters. These would each be changed by the result of comparing the final word to the dictionary. What we want to program to output is the first two counters and the average or the second two counters divided by its respective first counter.

The program itself is designed using an array of custom linkedlist class for the dictionary. One thing that this linkedlist has is in its method to see if it has a value in it, it also has the option that will return the number of times it went through. This gives us the time for each of the words that can then be added together to give us the second two counters.

The result of using the test file is: (test file in github page)
914750 The Avg of the word found is 3549.9994971303636
63841 The Avg of the word not found is 7442.381525978603

One observation that we noticed was that the average of words not found is around double that of words found. This makes sense as they would end up being the worst case and average case times and most search algorithms' times are like that.