Bill Blunk

Programming Assignment 4

AssignmentFour

November 1, 2015 – 11:59PM

A task that we as systems software developers will commonly be assigned is to verify a file we receive in one format against a known good data file and return the results the end user. Automated spell checkers are used to analyze documents and locate words that might be misspelled. These programs work by comparing each word in the document to a large dictionary of words. If the word is not found in the dictionary, it is flagged as potentially incorrect. Specifics for this program are to receive a text file version of a novel and check the words in that file against a random dictionary file. The words to be checked against the dictionary are derived from the input novel text file. The methodology for handling the input text file to be spell checked was to create an algorithm that could handle the characteristics that varied from word to word in the input text file. The algorithm must handle unneeded characters such as punctuation by removing them from the word prior to checking that word against the dictionary. Another aspect of the algorithm was to handle the difference in capitalization that some words would have by changing all words to be checked to lower case prior to checking against the dictionary. As part of the learning for this programs development was to calculate average for the number of comparisons done for fond and not found words.

In conclusion, it was observed that the comparison averages were approximately one half from found to not found words. The algorithm for processing the input text data for this version of the program demonstrated the increased efficiencies of a linked list data structure.

Output:

run:

Number of misspelled words: 54648

Number of correct spelled words: 937492

Number of misspelled words comparisons: 403377564

Number of correct spelled words comparisons: 3049431067

The average number of comparisons for words found: 3252

The average number of comparisons for words notfound: 7381

BUILD SUCCESSFUL (total time: 39 seconds)