Lab Report

Michael Kim

010572235

4/17/2017

**Program Statement**

The goal of the programming assignment is to make a simple searching program. To search documents, first the program needs to read file and extract the words from the reading file. After extracting words, those words need to be sorted in alphabetical order. Finally, the program should output how many unique words in document, and how many times those unique words come out for each. If user input N integer, the top N numbers of unique most used words should be printed out. The inputs to the program is the book.txt file which is given in website.

**Design**

For reading file, reading string method is used. after read string that reads until the space,

Remove the non-alphabetical letter from string. For example, “Sam-I-am.”, ‘Sam’, ‘I’ and ‘Am’ should be extracted. After removing all non-alphabetical letters, those words should be stored in a string array. For next, the string array should be sorted in alphabetical order.

For sorting string array, selection sort is used. after selection sort, new function for analyzing array is made. This function makes two array which are integer array for count words, and string array for string words corresponding to count number. This function will count how many identical words in the string array for each word. After counting the number of corresponding word, it stores its count number to integer array. Corresponding word also is saved to word array. For example, if letter ‘a’ is used 33 times, at integer array[0] is 33, for string array[0] = ‘a.’ for last user will be asked for numbers of how many top often used word to check. For sorting two arrays for ranking often used word, bubble sort was used.

**Implementation**

For reading file, while loop was used with din >> string. It stops when it is end of file. for loop was set for checking each character of string. Tolower function was used for change all capital letter to lower case. For next, !salpha function was used to find non-alphabetical character in the string. Then substr function was used for making string until non-alphabetical letter is encountered. For example, “Sam-I-am,” first variable i, for finding non-alphabetical is 3. Second variable k was used for starting variable. So for extract ing ‘Sam,’ substr(k,i-k) was used so it can be substr(0,3) when k = 0. If k >0, substr(k+1, i-k-1) was used. After extracting word, it saved it to the array. For sorting extracted word array to alphabetical order, selection sort was used. Alphabet has its own number, so it can be compared like integer. For last, analyzing array function was made. For the first step, it makes two arrays of count array and word array. Count array will store counted numbers of each word. Word array will be sorted corresponding to count array. For to count words, while loop was set. It loops until the identical words are found. after finishing counting words, it saves that counted number to count array, and the word that is used for counted is saved to word array. Both number and word which are correspond to each other should be saved at same location of each array. For sorting both array, bubble sort was used. when bubble sort found two values to swap, word array also swaps for same location. Before finishes the function, cin was used for user input for top number of most used word to be printed.

**Testing**

First I started with book.txt. it come out correctly with 50 unique words. Most used word was ‘i’ and ‘not.’ Each word used 83 times. Least used word was ‘if’ with 1 time. For other testing, in project5 description, implementation part was used for testing. There were 91 unique words. Word ‘a’ and ‘you’ was used most with 9 times each. For last, first paragraph of the article from <http://www.bbc.com/sport/football/39623727> was used. there were 93 unique words. Most commonly used word was ‘the’ with 7 times.

**Conclusion**

The overall result of the assignment came out with correct output with few statements. Project worked successful. I want to try to make this assignment with coding file for checking error for “ ” or ‘ ’ sign.

**Citation**

- John Terry: Chelsea captain to leave club at end of season - BBC Sport. (n.d.). Retrieved April 17, 2017, from http://www.bbc.com/sport/football/39623727