Assignment 1 Word Stats

Name: Arvy Salazar

Username: Ags531

Program Description

The program reads the file word by word. Whenever a word is read it is then pre-processed for comparing. Using a binary search that checks if the word exists, If the word already exists in the pool the word struct count is incremented. Otherwise, the word is added to the word pool and its start position and length is recorded. Using a binary search tree method, every word struct is added to a binary search tree method. If the count is higher than the struct word is put at the right position. If the count is lower than the struct word is put on the left position. If the count is equal then the struct word is compared alphabetically. Now that the word structs are in order print the top 10 and the bottom 10 using a in order traversal.

Data structures

Compact string storage – is used to store all the unique words. The reason why it is used is because compact string storage is efficient in storing strings. Than storing it in a string array.

Binary search tree – is used to organize the words in counts and alphabetically.

Algorithms

Isalpha() – is alpha is only used in processWord() to process the word for comparison by removing hyphens, commas and apostrophes.

Strcmp() – strcmp is used to compare two strings. This is mainly used in addWord() binary search and insertion sort as well to determine the word's alphabetical order, this function is also used in insert() to do compare of strings that determines where they go down the tree.

Strlen() -strlen is used to get the length of the string. This is mainly used in addWord()

Memset() – this is used at getWord() to make sure the char pointer is empty.