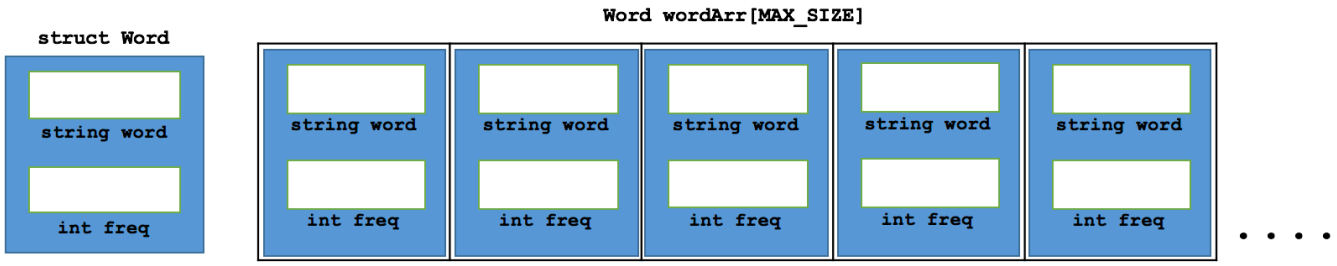
Eric Sehun Oh
CIS 22B

Homework 4 Report: ***Structures and Strings: Text Analyzer***

Data Structure



In order to keep track of words and their frequencies, a `struct Word` was defined containing a `string word` and an `int freq`. In addition, an array of `Word structs` were also defined with size `MAX_SIZE`. In addition, there are three `int` variables of `num_words, max_length, and highest_freq` which keeps track of the number of words encountered, the maximum length of words encountered, and the highest frequency, respectively. The previously mentioned variables were updated as the input file was parsed and processed.

Functions

The functions that are used in the program are as listed:

```
void parseFile(string file, int &num, int &m_length, int &h_freq, Word[]);
void updateArr(int &num, int &h_freq, int &max, string word, Word[]);
void displayHighestWords(int num, int freq, Word[]);
void writeToFile(string file, int num, int h_freq, int max, Word[]);
```

The `parseFile` function accepts a string variable of the filename, `int` variable of number of words, `int` variable of highest frequency of a word, and a Word array containing `Word structs` of words and frequencies. The function reads in each line and parses the line into words with only characters and calls the `updateArr` function to store words into arrays. It does this by reading in each block of characters separated by spaces. Afterwards, the word is "cleaned/processed" by removing any non-alpha characters from the end of the word and by converting uppercase characters to lowercase for matching strings later on.

The `displayHighestWords` function accepts an `int` variable of number of words, `int` variable of highest frequency of a word, a Word array containing `Word structs` of words and frequencies. The function prints out the word(s) that have the highest frequency.

`The updateArr` function accepts an `int` variable of number of words, `int` variable of highest frequency of a word, `int` variable of the longest length of a word, and a `Word` array containing `Word structs` of words and frequencies. The function will update the array that contain words and the frequencies. It also updates the number or words found, the highest frequency of a word, and the longest length found so far from the input file. Each time the `updateArr` function is called, it will sort the `Word` array in a similar manner to that of an insertion so that it is in alphabetical order,

The `writeToFile` function accepts a `string` variable containing name of input file, `int` variable of number of words, `int` variable of highest frequency of a word, `int` variable of the longest length of a word, and a `Word` array containing `Word structs` of words and frequencies. The function will write to an output file the processed information of the input file.