//CSCI203 Assignment 1

//Name: Kazi Swad Abdullah

//UOW ID: ksa255

A high-level description of the overall solution strategy:

- The program prompts the user for a filename

-The application opens the file and reads it one word at a time and store word in array of characters of length 20.

- Every word is passed through the processWord function that uses the standard library functions isalpha to only keep the alphabetical characters and tolower to convert it to lowercase.

- The processed word is searched using the search function. The word is searched in the AVL tree to find the instance of the word. If there is an existing occurrence than the key is returned, and count is incremented for that word.

-It the word read in does not exist in the tree, then a new word structure is created with the start index set to counter1. The word is then added to the array and length of the structure is incremented. The word structure is then added to the WORDS array and added to the AVL tree.

- After all the words have been added to the WORDS[] array and sorted with quicksort. The quicksort partitions the array based on a pivot chosen as the last element.

A list of all of the data structures used, where they are used and the reasons for their choice:

- Word structure has been used to store information about each word in the list of words that have been read and the number of times it has occurred. The information stored includes the index where the word begins, length of the word and the number of occurrences of the word. It was used to keep track of the unique processed words.

- **Node** structure has been used to store information about the AVL tree. Each node contains the height of the node in the tree and a pointer to the left and right child of the node. A key has been used as an index for the array of words structures called Words[] and is used to do comparisons for insertion operations.

A list of any standard algorithms used, where they are used and why they are used:

-**AVL Tree** has been used to store the words read, it was used because searching the tree has a complexity O(logn) also because insertion has a complexity of O(logn)

-**Quicksort** has been used to sort the array in descending order, it was used because it has a time complexity of O(n.logn)