

ASSIGNMENT 4 INDEXER

Design:

argument 1 is the file in which the words are being written to in JSON format

argument 2 is what we are searching and if it doesn't exist , an error message appears

First I made error checks for the command line arguments , checked if files were files and if directories were directories, if not the program exited with an error message.

Interacted with the user if the file for argument 1 already existed and asked them if they wanted to overwrite it.

Second I made a directory recursive traversal function

whenever it reaches a file it will call tokenize function in indexertk.c , if argument 2 is a file and not a directory it will just tokenize that file

indexertk.c tokenizes the words in files, disregards non alphanumerical characters including white spaces.

then I call search function in indexer.c file

the search function creates a data structure of linked list in memory

a list of word nodes and each node has a list of file nodes

word nodes hold a pointer to the list of file nodes for that word node, the word and a pointer to the next word.

the file node hold the pathname, the count number of times the word appears in that file, and a pointer to the next file node.

then write to file function is called which writes the structure to the file indicated by argument 1 in JSON format

files used are closed and program exits after it outputs the file

analysis: $O(n^2)$