pa4

Analysis

Iterating over n number of files – O(n) Writing tokens into the file – O(1) Tokenizing the file into m number of tokens – O(m) Overall performance – O(mn)

Brief Overview

We tweaked the “Tokenizer” from the first assignment to conform to this particular assignment because our older version of tokenizer failed at some tests but this is seems to be working fine plus it has to take in files instead of strings. Another way we had to tweak tokenizer from first assignment is that it uses all characters that aren't letters or numbers as tokens. In order to read in a file OR a directory we made indexer.c which calles the sorted list that puts one word at a time from the files provided. Files get tokenized and inserted into a list if it is valid and then it is outputted into the destination file that the user names. Algorithms that were used were linked lists and hashmaps. We also used sorted list program from second assignment which we also tweaked.