

Group: Douglas Rudolph, Nandan Thakkar

Professor: Dr. Fransisco

Course: Systems Programming - CS:214

Due Date: March 29th, 2017

Assignment: Assignment 2

Design:

The goal for my indexer was modularize the code base. The code base for my project consists of 5 files - each of which handles their own part of the project

Files and Descriptions:

1. Filesystem.c:

- i. Filesystem.c's purpose is to handle all the functions pertaining to the file system. Going into a little more detail, the main purpose of this file is to recursively walk through the file system and get all the files and call a function to then extract all the tokens.

2. File_IO.c:

- i. File_IO.c's purpose is to take a file found in the file system and read the file character by character. From which, it manages to pull out all the tokens and adds them to a *char***.
- ii. File_IO.c's other purpose is to write the organized information back to a file specified by the user.

3. File_Hashtable.c:

- i. File_Hashtable.c's purpose is to take the tokens from a file and store them in a hashtable in alphanumeric order based on the filename.

4. Token_Hashtable.c

- i. Token_Hashtable's.c's purpose is to take the token data hashed within File_HashTable.c, sort and count the unique tokens from a file

5. Driver.c

- i. Driver.c's purpose is run all four of the other files and then write it back to the file the user specified.

Time Complexity:

Reading in a file takes $O(3N)$. 1 iteration to find the size of the file, 1 iteration to get *rewind* back to beginning of the file, and then 1 iteration to then read the data to the end of the file.

Generalized to $O(N)$.

Writing the data to the File_HashTable Takes $O(N)$ due to the fact that each file has its tokens stored during the process of reading a file.

Sorting the tokens from a file takes $O(N^2)$ due to the fact that I am using insertion sort.

Counting the amount of unique tokens takes $O(N)$ because the data is already sorted, thus finding the groups of unique tokens can be done in one iteration.

Writing to a file takes $O(N)$, because we have to go through all tokens that were read in and write the XML tags.