

Abhishek Prajapati

Assignment 4

Indexer Read Me

This assignment reads the input from a single file or all the files in a directory. It will open the file, read the text and store it in a different file in JSON format. There are several different data types I used to make this work. I used Data type to represent pair :<file_name> : number_of_occurrences to keep track of the number of occurrence of the text in the file. I used Utility data type to store the index_paired array, and another data type to represent word in a list of pair index, and to store all the words with its occurrences. All the data types are declared in the .h files. All the parsed data from the text file is stored in simple dynamic arrays.

The running time of my algorithm is following:

- Searching file name through files takes linear time $O(n)$
- searching word through all words takes linear time $O(n)$
- all the data is sorted using quick sort before writing it in the JSON file.

The Algorithm I used is

- read file line by line
- split lines by separators and use tokenizer to tokenize the lines
- Update the information related to each word
- Sort the data using sorted_list
- write the data to JSON file
- free all the created memory