System Programming (Spring 2017)

Assignment2: Recursive Indexing

Contributors: Joseph Moussa & Enkai Ji

Goal:

Given a set of files, an indexer will parse the files and create an inverted index, which maps each token found in the files to the subset of files that contain that token.

Basic overview of the entire program:

Count.h

Export a function called "countFromSingleFile" which will find all the tokens in a single file and their occurrences. It will return a linked list in which tokens and their occurrences are stored one by one. When the function encounter a token when reading the file, it will compare it to all the tokens in the current linked list. If the token is existed, count++. Or, insert it into the linked list. So the time complexity is O(n^2)

Main.c

Basically, it including a function that can recursively traverse the directory, a linked list that stores all tokens and the corresponding filename and counts, and a function that can print the linked list out as XML format.