

Asst2: Inverted Index

Coding by Taner Doganay & Liam O'Brien

Our program firstly creates a node to hold the strings we will later tokenize, the name of files, and the number of occurrences of these words. Our first major function is `indexDir`. This function creates a pointer that traverses through our file system and opens objects in order. If the pointer finds a directory it then opens that directory and read the files. If the pointer finds a file it then opens the file. Our next function `removeBadCharAndTokenize` simply traverses through the strings and removes any non-alphabetical or non-numerical characters with spaces. The function then tokenizes the strings to be sorted. Our next step is to create a hash table to store our strings. Our next major function is `addWord` which looks at our files and adds the words as tokens. If the word has not already occurred, it is added as a new entry to our indexer.