Systems Programming
Assignment 3: Indexer
Jon Lee (jl1424) and Niharika Nagar (nn162)

Our indexer uses a hashtable to store the tokens in multiple files. The files in a directory are processed within the program indexer.c. First, indexer.c determines whether the given argument is a file or directory. If the file/directory is not found or the item found is not a file/directory, an error is returned. Otherwise, if it is a directory the directories within the the files within the directory are processed and the subdirectories are processed recursively. Within the processFile() method, for each line given to a buffer, the method extractWords() is called, which stores the tokens in each line into a hashtable, from the method InsertTable() in hashtable.c. InsertTable() converts the token to lowercase and then stores it using a key which is calculated for alphabet by subtracting the ASCII values by 96 and modding it by 27 and for numbers by subtracting the ASCII values by 47 and adding 26; therefore, the range for the key is 1 - 36. As InsertTable() stores the tokens, it sorts them using insertion sort. For any tokens that appear in multiple files, the files are sorted by the count. We do this by

Overall, the Big-O time of our program is $O(n^2)$. Each token is received from a file which is done in time O(n). Each token is then inserted into the table which is done in time O(n) as well.
The Big-O space is O(n) since we are using a hashtable.