

Introduction

Query is the third and final section of the Tiny Search Engine. The goal of query is to search for user input words in the indexed documents and list the URLs with the query words based on descending frequency of occurrence.

Design Choices

Argument: Check if the input indexed file and html folder to the query are valid.

usage: ./query [indexed file] [html files folder]

indexed file: the output of the indexer, frequently named indexer.dat

html files folder: output of the crawler, frequently named data

Invalid indexer.data or html folder will terminate the query.

Query: After the argument validity is established, process the user input query.

A query could be a single word or multiple words. A user can input logical queries by using the logical words OR and AND. A query cannot just consist of a single logical word or two consecutive logical words as shown below.

Query:> AND (invalid)

Query:>OR (invalid)

Query:> queryWord AND OR queryWord2 (invalid)

Query:> queryWord OR OR queryWord2 (invalid)

Query:>queryWord1 AND queryWord2 (valid)

Query:>queryWord1 OR queryWord2 (valid)

Sort: Recursively sort the query based on the frequency of occurrence of the word.

Print: Print the URLs in the descending order.

Note: most methods calls used in query.c are based on queryHelper.c. The latter includes methods like querySort, ANDMethod, ORMethod, getDocumentList, printResult, etc.

Test

Thorough testing conditions are implemented in QBATS.sh. Some of them are as follows.

Test Conditions

Input Argument Size: 1 (smaller)

Input Argument Size: 5 (bigger)

Incorrect directory path

Incorrect filename

Correct Input

The result of the QBATS.sh is saved in a log file named after the date stamp. I have, however, not implemented unit testing cases.