Simplified Search Engine

A Web search engine allows users to retrieve relevant information from source files, thereby identifying relevant pages on the Web containing given keywords. Here is a simplified model of a search engine.

1.Inverted Files

The core information stored by a search engine is a dictionary, called an inverted index or inverted file, storing key-value pairs (w,L), where w is a word and L is a collection of references to pages containing word w. The keys (words) in this dictionary are called index terms and should be a set of vocabulary entries and proper nouns as large as possible. The elements in this dictionary are called occurrence lists and should cover as many Web pages as possible.
We can efficiently implement an inverted index with a data structure consisting
of the following:
An array storing the occurrence lists of the terms (in no particular order)
A compressed trie for the set of index terms, where each external node stores the index of the occurrence list of the associated term.

2.Retrieve

With our data structure, a query for a single keyword is similar to a word matching query. Namely, we find the key word in the trie and we return the associated occurrence list.
When multiple keywords are given and the desired output is the pages containing all the given keywords, we retrieve the occurrence list of each keyword using the trie and return their intersection.

3.Input

The input is the keywords by which users want to search.
Note: the keywords should be separated by space
eg: Please input your keywords(separated by space): "computer course"

4.Output

The output of every search will return the files which contains all the keywords.

eg:

----------Here is search result----------

Find 4 files which contains your keyword(s)!

file name: 1.txt file path: C:simplified-search-engine\input\1.txt

file name: 5.txt file path: C:simplified-search-engine\input\5.txt

file name: 7.txt file path: C:simplified-search-engine\input\7.txt

file name: 8.txt file path: C:simplified-search-engine\input\8.txt