Back to Study plan

Work on project. Stage 5/6: Inverted Index search

Project: Simple Search Engine

Hard 0/ 6 minutes

Description

Now your program can successfully search for all matching lines, ignoring cases and extra spaces. There is one problem though: you need to check whether each line contains a query string or not.

To optimize your program, you can use a data structure called **Inverted Index**. It maps words in a set of sentences/lines/documents in which the word occurs. As a result, when we receive a query, we can immediately find the answer without any comparisons.

At this stage, build an inverted index when the program starts, and then use the index for searching operations. You can implement it using **Map**. Classes. It connects an item with a list (or set) of indexes belonging to the lines that contain the item.

Suppose you have the following array or list of lines:

```
0: Katie Jackson
1: Jordan Harrington harrington@aol.com
2: Martha Medina
3: Erick Burgess

For these lines, the inverted index will look like this:

Katie -> [0]
Jacobs -> [0]
Erick -> [3, 3]
Harrington -> [1]
harrington@aol.com -> [1]
Jordan -> [1]
Medina -> [2]
Burgess -> [3]
```

The order of pairs is not important. If you are searching for Erick, you can immediately get the target fields using this mapping.

Note that the inverted index is not intended to search for parts of words, it is only used to search for full words.

Example

The lines that start with `>` represent the user input. Note that these symbols are not part of the input.

```
>>> Menu >>>
1. Find a person
2. Print all people
0. Exit
> 1

Enter a name or email to search all suitable people.
> ERICK
2 persons found:
Erick Harrington harrington@gmail.com
Erick Burgess

>>> Menu >>>
1. Find a person
2. Print all people
0. Exit
> 1

Enter a name or email to search all suitable people.
> M@Gmail.com
1 persons found:
Richard  May  rm@gmail.com

>>> Menu >>>
1. Find a person
2. Print all people
0. Exit
> 1

Enter a name or email to search all suitable people.
> jobs
No matching people found.

>>> Menu >>>
1. Find a person
2. Print all people
0. Exit
> 0

Bye!
```

Report bug

Start hard

Write a program

Code Editor | IDE

This widget does not work in Safari due to security settings. Please open this problem from within your IDE or use another browser.

Switch to IDE

src/search/Main.java

```
package search;
import java.io.File;
import java.io.FileNotFoundException;
import java.io.IOException;
import java.nio.file.Files;
import java.nio.file.Paths;
import java.util.*;

public class Main {
    private static Map<String,ArrayList<Integer>> findMapString, List<Integer> mp, String tofind) {
        Map<String, List<Integer>> map2map = new HashMap<>();
        ArrayList<Integer> arrList = new ArrayList<>();
        for (String s : mp) {
            if (mp.contains(s.toLowerCase())) {
                arrList = mp.get(tofind.toLowerCase());
                if (map2map.containsKey(s.toLowerCase())) {
                    arrList.addAll(temp);
                }
            }
        }
        if(arrList != 0) {
            map2map.put(tofind.toLowerCase(), arrList);
        }
        return map2map;
    }
    public static void printAll(Map<String> all) {
        for(int i = 0; i < all.length; i++) {
            System.out.println(all[i]);
        }
    }
    public static void printMenu() {
        System.out.print("\nSee Menu >>>\n 1. Find a person\n 2. Print all people\n 0. Exit\"\n);
    }

    public static String readFileAsString(String fileName) throws IOException {
        return new String(Files.readAllBytes(Paths.get("./".toRun())));
    }

    public static void searchStep1(int[] range) throws IOException {
        System.out.print("Enter a name or email to search all suitable people.\n");
        String tempFileAsStringFile = Files.readAllLines(Paths.get("./".));
        System.out.print("Enter a name or email to search all suitable people.\n");
//      System.out.print("Enter a name or email to search all suitable people.\n");
        ArrayList<String> lst = new ArrayList<String>(Arrays.asList(tempFile));
        int[] tmp = new int[range.length];
        for (int i = 0; i < lst.size(); i++) {
            String[] str = lst.get(i).split(" ");
            for (int j = 0; j < str.length; j++) {
                if(map.containsKey(str[j]) {
                    map.put(str[j],i.add(i);
                } else {
                    List<Integer> temp = map.get(str[j]);
                    temp.add(i);
                    map.put(str[j], temp);
                }
            }
        }
        System.out.print(Arrays.toString());
        int[] tmpCopy = tmp;
        int countFlag = scanner.nextInt();
        while (countFlag != 0) {
            int[] temp = new int[range];
            switch (countFlag) {
                case 1:
                    int[] tmpCopy = tmp;
                    break;
                case 2:
                    System.out.print("Enter a name or email to search all suitable people.\n");
                    int[] tmpCopy = tmp;
                    String scan = scanner.nextLine();
                    if(findMap(scan).size() != 0) {
                        ArrayList<Integer> arrList = new ArrayList<Integer>(map, scan);
                        for (int i0 = 0; arrList.size(); i0++) {
                            System.out.print(arrList[i0]);
                        }
                    }
                    printMenu();
                    countFlag = scanner.nextInt();
                    break;
                case 0:
                    System.out.print("\n\n\n\n List of people <<<\n");
                    printAll(Map<String>[0]);
                    printMenu();
                    countFlag = scanner.nextInt();
                    break;
                default:
                    System.out.print("\nUnsupported option! Try again.\n");
                    printMenu();
                    countFlag = scanner.nextInt();
                    break;
            }
        }
        scanner.close();
        System.out.print("\nBye!");
    }
}
```

Correct.

It was a tricky task, but you nailed it!

49 users solved this problem. Latest completion was **1 minute ago**.

Hint users solved this problem. Latest completion was **1 minute ago**.

- Hard 0/6 Problems Solved
- Operating system basics 0/1 Problems Solved
- Data structures 0/1 Problems Solved
- Working with files 0/1 Problems Solved
- Mock Solution 0/1 Problems Solved
- Flash talks 0/1 Problems Solved

Show all

Join a study group for the project Simple Search Engine

Discuss your current project with fellow learners and help each other.

Comments (40) | Hints (5) | Useful links (1) | Solutions (37) | Hide discussion

Share something User 18442759

Place a snippet Please do not post solutions here

Sort by: Last posted

Jose b wrote: 21 days ago

My program is fully working & cannot pass tests.

Reply · Report

Charlie Sror 21 days ago Topmost Comment

need to post some information, code, error, something.

Reply · Report

Jose b wrote: 20 days ago

already works fine. I failed at test 5.

Reply · Report

Nick Martarison 6 months ago

"Looks like you're printing unknown people when you enter option 2."
>>> Menu >>>
1. Find a person
2. Print all people
0. Exit

Reply · Report

Ilya Markarov 6 months ago

The problem was in the print, but it was in 1. "Find a person" option, because with the input "bob" I was prints a double search instead of 1, all keys which contain "bob" because I did a key contains("bob") instead of "hashmap.containsKey("Bob")" "Bob" "lower Case"".

Reply · Report

Kazaar 7 months ago Topmost

Can anyone help me?
in case when I click the check button, I get the error:

Wrong answer in test #1

Reply · Report

S V R T R 7 months ago Topmost

Hi, please contact support https://support.hyperskill.org/hc/en-us/requests/new if you didn't solve this problem

Reply · Report

Kazaar 7 months ago

It turned out that during the test, the nameslist file is created in one directory and looks for it in another.
I found workaround for this problem:
Just created a link to the file in the directory in which the test system is looking for it.

ln -s "C:\Users\username\IdeaProjects\Simple Search Engine\names.txt" "C:\Users\username\IdeaProjects\Simple Search

Reply · Report

Ivan KORNIEV 7 months ago Topmost

I get an Error in test #1 despite closing both file reader and scanner. My code all works locally but I can't get it to pass this test due to the FileSystemException. Any idea how I can solve this?
java.nio.file.FileSystemException
The file names.txt can't be deleted after the end of the test. Probably you didn't close File or Scanner.

Reply · Report

Charlie Sror 6 months ago

try using Files.newBufferedReader(Paths.get("path"));) and see if you still have this issue. Per documentation for readString:

Reads all content from a file into a string, decoding from bytes to characters using the UTF-8 charset. The method ensures that the file is closed when all content have been read or an I/O error, or other runtime exception, is thrown.

Reply · Report

Ciao 6/4/2022 8 months ago

Hum where I can get the data?

Reply · Report

Dorodnitsyna Sleeper 7 months ago

I don't understand what I need to do. Where can I get the data? Please help me

Reply · Report

Dorodnitsyna Sleeper 7 months ago

Login 4

Reply · Report

Ciao 6/4/2022 8 months ago

Feedback from the tests is useless.

In my case test #4 and #5 failed because I was using Machine and Pattern-class. If I searched for "SQ" I got 16 results, but this is wrong. This should result in no match. Only full addresses should be allowed.
In short: do not search for substrings, only the full word.

Ilya 1 · Reply · Report

Pseudo0100100 2 years ago

how to resolve/complain/report failed something wrong in test:TestExecution failed for task :Simple_Search_Engine:test:compileTestJava:
error: cannot find symbol
  symbol: variable SearchEngineTests
  location: class SearchEngineTests.NAMESE()

symbol: variable SearchEngineTests

Reply · Report

Ian Emsley 2 years ago

"Wrong answer in test #3

Reply · Report

Maria Krasavaev about 1 year ago

Hi Ian, were you able to solve this problem? I have the same one. Everything looks correctly.

Reply · Report

Roman Dexter 9 months ago

Red #3 is supposed to print three names. If yours is only printing two, then you are somehow missing one. Look at your code and try to figure out how you might only get two entries instead of three.

Reply · Report

Lukas 2/8/2022 1 year ago

Can someone help with test 5? Im struggling now for about an hour. When searching for "@" I'm getting 14 results (no duplicates, everything seems correct), when searching for "jacob" I'm getting 3 results (also everything fine). After checking input text from previous tests my results seems perfectly fine so im clueless whats wrong.

Reply · Report

Kaiser 1 year ago

That was a very informative and useful exercise

Reply · Report

Vitali Sauce over 1 year ago

> the example, you search for 901@gmail.com and find an entry, even though in the dataset you provided above, this mail is nowhere to be found.

Reply · Report

Dmitrii Mastov over 1 year ago

Error in test #1

java.nio.file.FileSystemException

The file names.txt can't be deleted after the end of the test. Probably you didn't close File or Scanner.

Reply · Report

Dmitrii Mastov over 1 year ago

But I've closed FileScanner
I've tried to check code from previous stage and it gail Error only in test #5.
Now code quite the same just added Map.

Reply · Report

Dmitrii Mastov over 1 year ago

Error in test #5

I'll pass just another bug.

Reply · Report

Maria Krasavaev over 1 year ago

Got the same issue. Just close FileReader.

Reply · Report

Alexander over 1 year ago

Wrong answer in test #4
Looks like you're printing unknown people when you enter option 2.
What does it mean?

Reply · Report

Misha Usmanov about 1 year ago

using equalsIgnoreCase instead of Matcher class helps me.

Reply · Report

Alexander about 2 years ago

Wrong answer in test #5

Search result is not equal to the expected search

How I can understand What exactly is wrong ?!

Reply · Report

Ivan STRAKHI about 2 years ago

I has the same problem :(

Reply · Report

Ivan STRAKHI about 2 years ago

actually with is an idea and plugin problems, but the instruction link didn't help me

Reply · Report

Ivan STRAKHI about 2 years ago

use the instructions and restart the computer and everything will be ok! :)

Reply · Report

S V R T R user 1 year ago Topmost

At this stage you shouldn't check substring, but the whole word. That can be a problem with test 5.

Ali 2 years ago

Bartosiewicz Rob over 1 year ago

Problem occurs only while searching '@' symbol. Maybe that will help:

Reply · Report

yyn over 1 year ago

Thank you, I wasted so much time, but I'm glad it finally works.

Reply · Report

Dimosthenis Stelio over 1 year ago

Thanks!

Reply · Report

Alex Cabaul almost 2 years ago

I got an error that has been previously fixed.

Failed to post submission to the Hyperskill. Please, try to check again

Reply · Report

Kyle almost 2 years ago

Hi, please check out this article on our Help Center:
https://support.hyperskill.org/hc/en-us/articles/360008790215-I-have-a-problem-with-IDE-or-Ide/tools-issues-what-should-i-do-

Reply · Report

Nicolas over 1 year ago

Rita Dzuchuk, I suppose, your mistake is the name of file "names.txt".
In tests of HyperSkill there is no this file.

Reply · Report

Anton Moskovski over 1 year ago

The description suggests that the key in the index should the whole word, as shown in the example. However, in the final tests, only part of the word is used as a search term("@gmail.com") and test expects that it watch all entries which contain that.

Reply · Report

Lucas Arnaud over 1 year ago

Wrong output in test case 5 Search result is not equal to the expected search.responseFromClient - what to do?

Reply · Report