

Indexing With Trees

Andrew Rosen

Abstract

In this lab you create an index for the complete works of William Shakespeare (`pg100.txt`), although you can use it on other files. It will test your ability to put together multiple data structures and objects.

Check each file to see what you must complete.

1 Your Assignment

We will create an `IndexTree`, a special type of Binary Search Tree. The `IndexTree` does a bit more than your standard tree, as we will use it to build an index of a file. This will be much like the index you find at the end of a textbook, where each topic is listed in alphabetical order with the pages it is found on. Since files don't have traditional pages to work off of, we will instead use line numbers. Furthermore, rather than building an index of only a few select topics, we will build an index over all the words in the file.

To do this, we use a special type of node. The `IndexTree` is made up of special `IndexNodes`. Rather than using generics, each `IndexNode` stores a word, the count of occurrences of that word, and a list of all lines that word appeared on (this means that each `IndexNode` will hold their own list). Nodes in the tree will be sorted by the `String`.

Use an `IndexTree` object to store an index of all the words that are in the provided text file, then display the index by performing an inorder traversal of the tree.

2 How to Read a File

Good news, everyone! You already know how to do most of it. I've attached a file showing how to use `Scanner` to read every line of text from the file, split that line into individual words, and print out each word. Be sure to scrub each word of those pesky punctuations, like commas

3 Advice

Get the file reading working first, then work on the `IndexTree`. **Do not attempt to work on the tree without figuring out how to get the file**

reading working. How to do it depends on your IDE and your setup. Resources and tutorials are available online. Generally to be able to read the file in an IDE, you need to put it in the top most level of your project folder.

My other advice is start early.