

Indexing Text

Daniel Yee
CIS 2168

Purpose

In this assignment, you will create an index for the complete works of William Shakespeare (shakespeare.txt). This will help you practice using binary trees and other data structures.

The Problem

An IndexTree is a special kind of Binary Search Tree that we'll use to build our text index, like the one you find at the end of a textbook. Since text files don't have traditional pages (without extra work to keep track of how many lines can fit on a particular print copy), we will use line numbers like those in your IDE.

The IndexTree contains a new node type that stores the word, the number of occurrences, and a list of all the lines that the word appears on. Nodes in the tree will be sorted by comparing the word data.

You must 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. You must also provide functionality to allow someone to search the index for a keyword, displaying the keyword and a list of the lines that it is found on.

Challenge

10 points extra credit

Implement the IndexTree class by implementing and extending the BinaryTree class.