APCS :: Lab 21: Letter Counting

Counting letter frequencies in text documents is an important procedure used in many data compression, encoding, and encryption schemes. You’ll get a flavor for it here. **Parts 1 and 2 are required.**

**Part I:** Finish implementing the BST we started in class so that you can add Comparables to the tree and find objects within the tree.

**Part II:** Use your BST to count the occurrences of letters in a given text document.

To do this you need to:

- Create a tiny object (call it CharCount) that you can store as the value in your TreeNodes. The object should hold a char and an int. CharCount should implement the Comparable interface and its compareTo method should compare based on the char value. (In this way, a BST full of CharCount objects will be ordered by the ASCII value of the char).
- Write a separate class (could be done in main) that processes a text document and examines each character.
- For each character in the text document try to find it in the tree. If the character is already in the tree increment the integer associated with that char. If it’s not in the tree, construct a new CharCount object and add it to the tree making sure to set its integer to 1. Thus, each node of the BST will contain a CharCount object whose character is one of the chars from the text document and whose integer is number of times it occurred in the text.
- Once the document has been processed print out all of the chars and their associated counts (do an in-order printing).

**Advanced Options (in no particular order):**

- Print out the char counts in order of the count (look up Comparator in the java API.)
- Write an Iterator for the BST. Easy way: make an arraylist out of an in-order traversal of the tree and return an iterator of that. Harder Way: write a custom iterator that only maintains pointers into your tree – if you want to get good with trees, do this.
- Implement a remove method for the tree – this will also help you get comfortable with trees.

**NOTES:**

I’ve included a text document with the project distribution: decl.txt which is the Declaration of Independence. If you want more texts to process, there are thousands plain text version of classic texts on the machines in the computer lab under /UH202UserData/ClassicLiterature/