CS211 ALGORITHMS & DATA STRUCTURES II

LAB 2

Dr. Phil Maguire

HUFFMAN ENCODING

Pen and Paper Exercise

Take the following string, generate Huffman codes for it (don't forget the spaces), and show how it would be encoded:

apples and bananas

If we assume that each character requires 7 bits to be encoded in ASCII, then what level of compression will be achieved by using Huffman encoding? (i.e. what percentage reduction in size does Huffman encoding achieve relative to ASCII?)

Programming Exercise

(see HackerRank invitation)

Problem Statement

The goal is to convert a character String into a binary string Huffman encoding. As we've seen, the way a Huffman tree is built is not deterministic, because when you have multiple trees with the same lowest weighting, you can choose any two of those to connect together, and you can put either one on the left or right. So let's make it

deterministic by saying - where multiple trees have equal frequency, always connect the two trees containing the two lowest value ASCII characters, and always put the lowest of those two on the left. For instance, every ASCII character is associated with a number, so every tree will have a lowest value ASCII character in it - (e.g. capital A is the lowest ASCII letter). Always connect the two trees that have the lowest characters anywhere inside them, with the lowest on the left. Now everybody should end up with the same trees and the same coding. This as simple as saying: take the Priority Queue and as well as sorting on frequency, sort on lowest character as well. When two are popped from the Priority Queue, put the first to be popped on the left.

Input Format
A String like "hello".

Output Format
The Huffman encoding in binary like "1111100010"

Sample Input
hello

Sample Output
1111100010 (try it on paper for yourself)