Algorithm Design

IIT Goa
Deadline:08-04-23 (23:00pm)

(Lab - 08)

– You may form groups of two each and stick to the submission guidelines as before.

Write a program that reads two long strings $Str1$ and $Str2$ and:

1. Compute the following sequences:
   (a) Prints the Longest Common Subsequence of both the strings.
   (b) Prints the Shortest Common Supersequence of both the strings.
2. We want to send $Str2$ to the receiver and we will do it in the following way. Assume that there is a cost of 1 unit for deleting a character, 2 for replacement and 3 for insertion:
   (a) Find the edit distance between $Str1$ and $Str2$. Rather than sending $Str2$, we will send $Str1$ together with the changes to make to convert $Str1$ to $Str2$. Say, $0a5$ may mean insert (indicated by 0) a before $Str1[5]$. Or 1@9 may mean Delete (indicated by 1) $Str1[9]$ (@ is just a separator), or even $2z7$ to express replacing (indicated by 2) $Str1[7]$ by z. Now, the whole message you want to share takes the form $Str1\#OP_1\#OP_2\#...\#$ where OPs (Operations) communicate the changes to make.
   (b) We will share the above information using Huffman encoding for which:
      i. : Write down a subroutine that computes the frequency of each character in a given string (the argument to be received) and returns the character array together with the associated frequency array sorted in the increasing order of frequency.
      ii. Construct the Huffman prefix tree recursively from the information obtained from 2(b)i (*Hint: Insertion sort maybe useful here*).
      iii. Construct the coding table (mapping each alphabet to the associated codeword) by traversing the tree obtained from 2(b)ii.
      iv. Encode the message from 2a using 2(b)iii.

You may print the outcome for each of the sub-problems from within the function itself.

*******END*******