Name: KOENG Gana

ID: e20180441

Group: I5-GIC-B

Discussion_Lesson06

1. What is Huffman coding?
   Huffman coding is a compression technique used to reduce the number of bits needed to send or store a message.

2. Explain Huffman coding algorithm?

   The (real) basic algorithm of Huffman coding:
   1) Scan text to be compressed and tally occurrence of all characters.
   2) Sort or prioritize characters based on number of occurrences in text.
   3) Build Huffman code tree based on prioritized list.
   4) Perform a traversal of tree to determine all codewords.
   5) Scan text again and create new file using the Huffman codes.

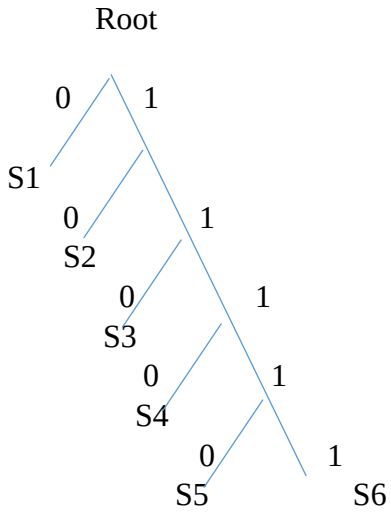
3. Why do we use Huffman tree?

   Huffman coding provides codes to characters such that the length of the code depends on the relative frequency or weight of the corresponding character.

   - Easy to find codeword
   - Easy to find average codeword length

4. Assume that we have 6 symbols, draw Huffman tree and find codeword of each symbol? You can choose your own values.

   Assume that we have 6 symbols: S1, S2, S3, S4, S5, S6. We know that:
   ➢ P(S1) = 0.9
   ➢ P(S2) = 0.8
   ➢ P(S3) = 0.7
   ➢ P(S4) = 0.6
   ➢ P(S5) = 0.5
   ➢ P(S6) = 0.4
   We put the biggest symbol on the top so:
   Now we start to build the Huffman tree.



Find codeword of each symbol:
S1->0
S2->10
S3->110
S4->1110
S5->11110
S6->11111

5. Find entropy and average codeword length from the word "lessonseemscool"?

P(l)= 2/15 = 0.133
P(e)= 3/15 = 0.2
P(s)= 4/15 = 0.266
P(o)= 3/15 = 0.2
P(n)= 1/15 = 0.066
P(m)= 1/15 =0.066
P(c)= 1/15 =0.066

$H = 0.133\log_2(1/0.133) + 0.2\log_2(1/0.2) + 0.266\log_2(1/0.266)$
$+ 0.2\log_2(1/0.2) + 0.066\log_2(1/0.066) + 0.066\log_2(1/0.066) + 0.066\log_2(1/0.066)$
H = 0.133x2.91+0.2x2.322+0.266x1.91+0.2x2.322+0.066x3.921
+0.066x3.921+0.066x3.921
H = 0.387+0.4644+0.5081+0.4644+0.7763
H = 2.6 bit
E = 0.266x1+0.2x2+0.2x3+0.133x4+0.066x5+0.066x6+0.066x6
E = 0.266+0.4+0.6+0.532+0.33+0.792
E = 2.92 bit