1) Consider the following code for the letters A through H:

A 0

C 1010

E 1100

G 1110

B 100

D 1011

F 1101

H 1111

What is the result of decoding the string: 100010100101101100011010100100000111001111

100 0 1010 0 1011 0 1100 0 1101 0 100 100 0 0 0 1110 0 1111

B A C A D A G A F A B B A A A G A H

2) Given the character frequencies character B C D E F G frequency 32% 28% 16% 6% 10% 8%

a. Using Huffman encoding, what is the code for character F? (Suppose that when constructing a sub tree from 2 nodes we always place node with higher frequency on the left; and the left branch of a node gets value 0, the right one gets value 1).

F = 11

b. Encode the text AABCDEFEEFG and calculate the average code length of this string.

f(A) = 2, f(B) = 1, f(C) = 1, f(D) = 1, f(E) = 3, f(F) = 2, f(G) = 1

A = 000, B = 0010, C = 0011, D = 110, E = 01, F = 10, G = 111

Average Code Length = ((3x2) + (4x1) + (4x1) + (3x1) + (2x3) + (2x2) + (3x1)) / 11 = 2.73

c. Check whether the string 10010101011011001101010111 is a valid encoded string of the above codes or not.

10 01 01 01 01 10 110 01 10 10 10 111

F E E E E F D E F F F G

NO

3) What is the maximal length of a codeword possible in a Huffman encoding of an alphabet of n characters?

The maximal length of a codeword possible in a Huffman encoding of an alphabet of n characters is n-1

4) Show that a Huffman tree can be constructed in linear time if the alphabet’s characters are given in a sorted order of their frequencies.

A Huffman tree can be constructed in linear time if the alphabet’s characters are given in a sorted order of their frequencies. This is because the algorithm for constructing a Huffman tree involves repeatedly selecting the two characters with the lowest frequencies and merging them into a new character with a frequency equal to the sum of their frequencies. If the characters are already sorted by frequency, this selection process can be done in constant time, allowing the entire tree to be constructed in linear time.

5) Given a raw message BBBBBUUUUXXXUUPPPPPPPUUKKKKKKKK. What is the compressed output if you apply the run-length encoding algorithm for that message?

5B4U3X2U7P2U8K

6) Suppose you are using the LZW algorithm to encode the message ABBABABACCDABCCDB and contents of the dictionary at the beginning of encoding are: (1) A (2) B (3) C (4) D. After the encoding process completed:

a. What strings are contained in the dictionary?

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Pos | Dictionary | Output |
| 1 | 1 | (5) AB | (1) |
| 2 | 2 | (6) BB | (2) |
| 3 | 3 | (7) BA | (2) |
| 4 | 4 | (8) ABA | (5) |
| 5 | 6 | (9) ABAC | (8) |
| 6 | 9 | (10) CC | (3) |
| 7 | 10 | (11) CD | (3) |
| 8 | 11 | (12) DA | (4) |
| 9 | 12 | (13) ABC | (5) |
| 10 | 14 | (14) CCD | (10) |
| 11 | 16 | (15) DB | (4) |
| 12 | … | … | (2) |

b. What is compressed string?

(1)(2)(2)(5)(8)(3)(3)(4)(5)(10)(4)(2)