

HUFFMAN ALGORITHM

Huffman coding is a lossless data compression algorithm that assigns shorter codes to more frequent characters and longer codes to less frequent ones.

It is based on constructing a binary tree (Huffman Tree) from character frequencies.

Steps of the Huffman Algorithm.

1. Calculate Frequency

- determine the frequency of each character in the string.

2. Build a Min-Heap

- Create a priority queue (min-heap) of leaf nodes where each node contains a character and its frequency.

Nodes are sorted by frequency.

3. Construct the Huffman Tree

- while the priority queue contains more than one node:
 - Remove two nodes with smallest frequencies
 - Create a new internal node with these two nodes as children and assign it a frequency equal to the sum of their frequencies.
 - Add new node back to the priority queue.
- The remaining node in the priority queue is the root of the Huffman Tree.

4. Generate Huffman Codes:

- Assign Binary Codes to characters by traversing the Huffman Tree
- Traverse left for 0 and right for 1.
- Leaf nodes represent characters with their codes.

5. Encode

- Replace characters in the input string with their corresponding Binary Codes.

Ex CONSTRUCT HUFFMAN TREE FOR MAHARASHTRA

Step 1 Calculate Character Frequencies

Character	Frequency
M	1
A	5
H	2
R	2
S	1
T	1

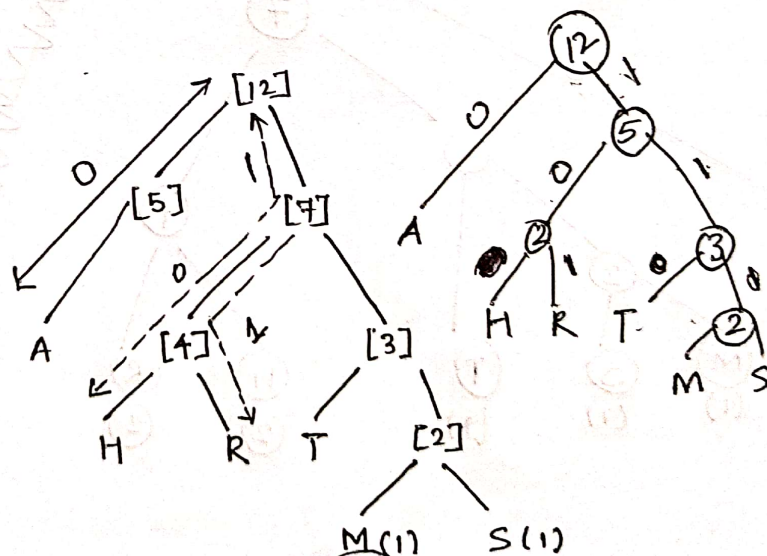
Step 2: Build Min-Heap

Initially, insert all characters into a min-heap sorted by frequency

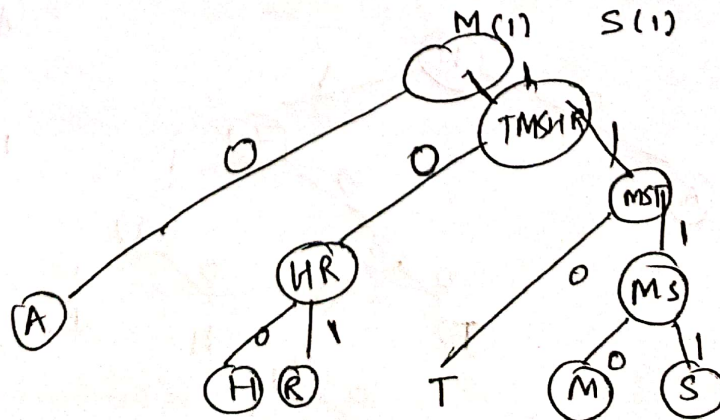
- Nodes: $[M(1), S(1), T(1), H(2), R(2), A(5)]$.

Step 3: Construct Huffman Tree

1. Extract two smallest nodes: $M(1), S(1)$
• (create a new node $[MS(2)]$)
2. Extract two smallest nodes: $T(1), MS(2)$
• (create a new node $[TMS(3)]$)
3. Extract two smallest nodes: $H(2), R(2)$
• (create a new node $[HR(4)]$)
4. Extract two smallest nodes: $TMS(3), HR(4)$
• (create a new node $[TMSHR(7)]$)
5. Extract two smallest nodes: $A(5), TMSHR(7)$
• (create the root node $[ATMSHR(12)]$)



$S = 1111$
 $M = 1110$
 $A = 0$
 $H = 100$
 $R = 101$
 $T = 110$

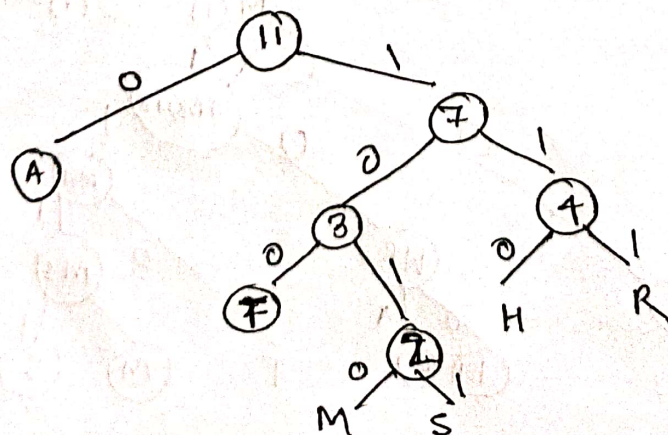
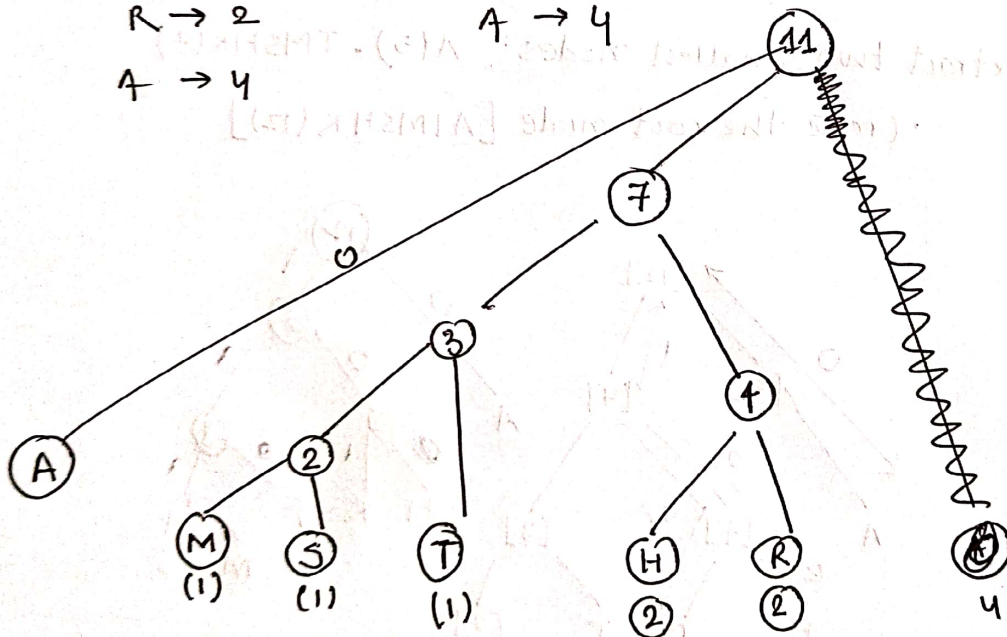


MAHARASHTRA \rightarrow MAHARASHTRA

$M \rightarrow 1 \checkmark$
 $A \rightarrow 4 \checkmark$
 $H \rightarrow 2 \checkmark$
 $R \rightarrow 2 \checkmark$
 $S \rightarrow 1 \checkmark$
 $T \rightarrow 1 \checkmark$

$M \rightarrow 1$
 $S \rightarrow 1$
 $T \rightarrow 1$
 $H \rightarrow 2$
 $R \rightarrow 2$
 $A \rightarrow 4$

$\Rightarrow T(1) = 1 \checkmark \rightarrow MST(3)$
 $MS \rightarrow 2 \checkmark$
 $H \rightarrow 2$
 $R \rightarrow 2$
 $A \rightarrow 4$



$A = 0$
 $T = 100$
 $M = 1010$
 $S = 1011$
 $H = 110$
 $R = 111$

Letter

freq

e

120

d

42

l

42

u

37

c

32

m

24

k

7

z

2

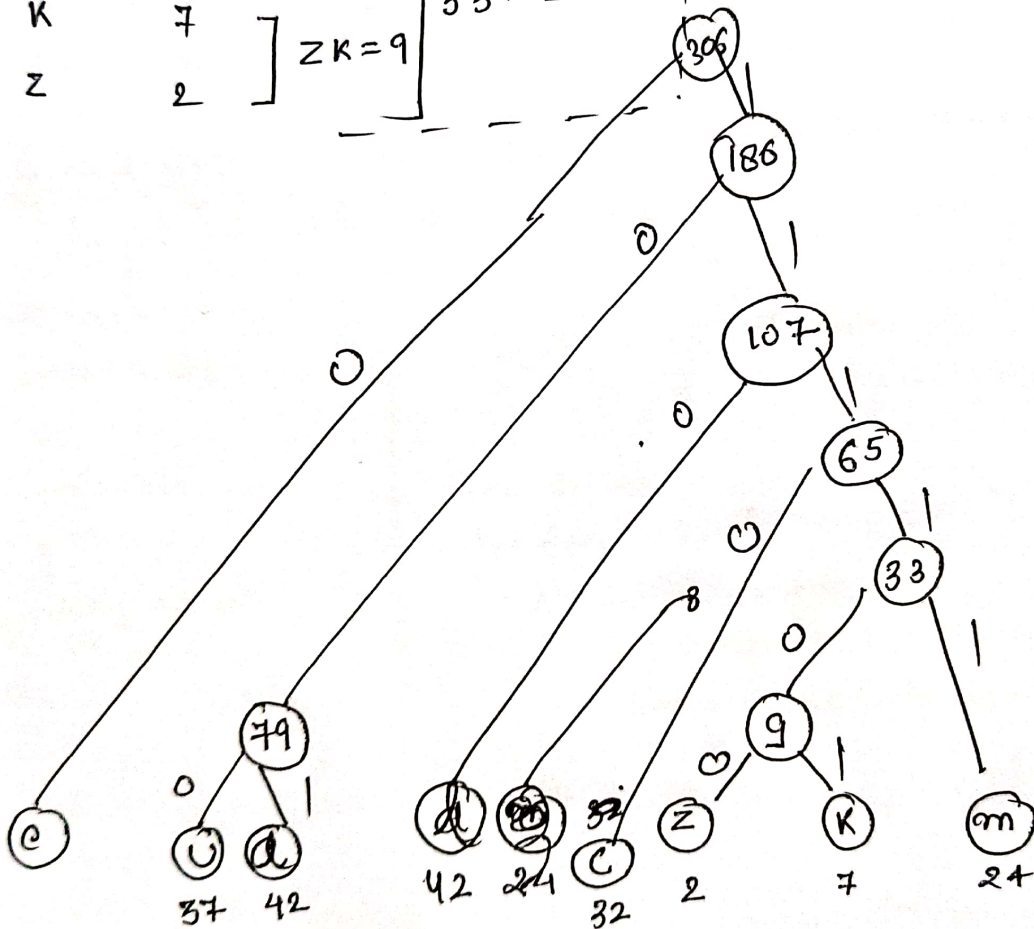
ul = 79

33 = zkm

65 = zkm c

zk = 9

186
120
306



e = 0 ✓

u = 100 ✓

d = 101 ✓

l = 110 ✓

c = 1110 ✓

z = 111100 ✓

k = 111101 ✓

m = 11111 ✓

