SHANNON ENCODING

It is a top-down encoding stratigy.
You dont rearrange the occurance of alphabets.
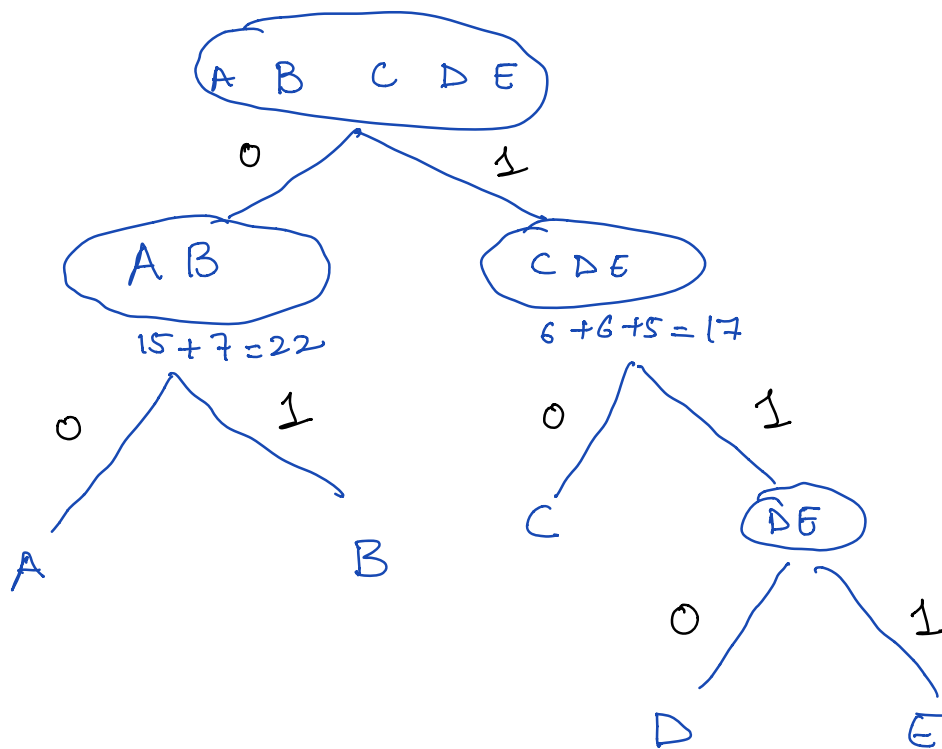We sort them on the basis of their frequencies.

ALGORITHM

- Get the list of probabilities of each character based on their occurances.

- Sort the list of symbols according to their frequences.

- Divide the list into 2 parts with prob of left part as close to the right part.

- Assign left part 0 & right part as 1.

- Recursively divide 2 part till each gets unique code.

Example

A	B	C	D	E
15	7	6	6	5	= 39

Count

Occurance $15/39$ $7/39$ $6/39$ $6/39$ $5/39$

A B C D E

0   1

A B   C D E

$15 + 7 = 22$ $6 + 6 + 5 = 17$

0   1   0   1

A   B   C   D E

0   1

D   E

$A = 00$ $C = 10$ $E = 111$

$B = 01$ $D = 110$

ENTROPY

$$\text{Entropy} = -\sum_{i=1}^{n} p_i \log_2 p_i$$

$$p_i = \frac{\# \text{ of } i}{\text{Total count}}$$