Huffman Coding

Huffman coding assigns shorter codes to symbols that occur more frequently and longer codes to those that occur less frequently.

It is used when the probability of elements in source file is known.

For example, imagine we have a text file that uses only five characters (A, B, C, D, E). Before we can assign bit patterns to each character, we assign each character a weight based on its frequency of use. In this example, assume that the frequency of the characters is as shown in table

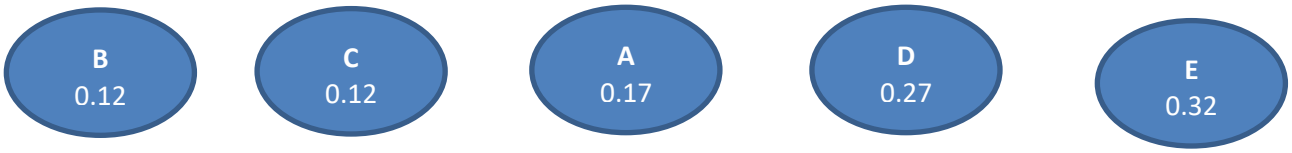
Character	A	B	C	D	E
Frequency	17	12	12	27	32
Probability	0.17	0.12	0.12	0.27	0.32
Binary Coding	000	001	010	011	100

Total bits = (Total Character =5 then 3 bit) * (Total no of Character =100)

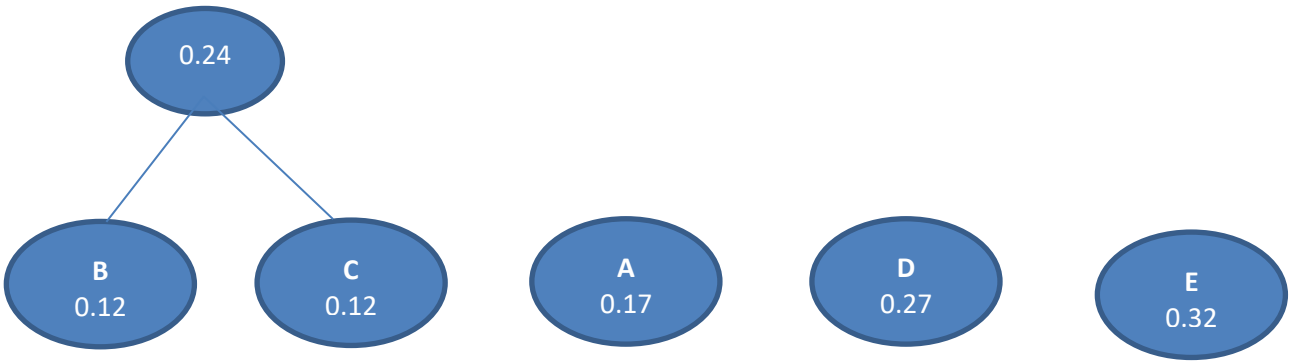
Total Bites = 3*100= 300 bits

Step1:

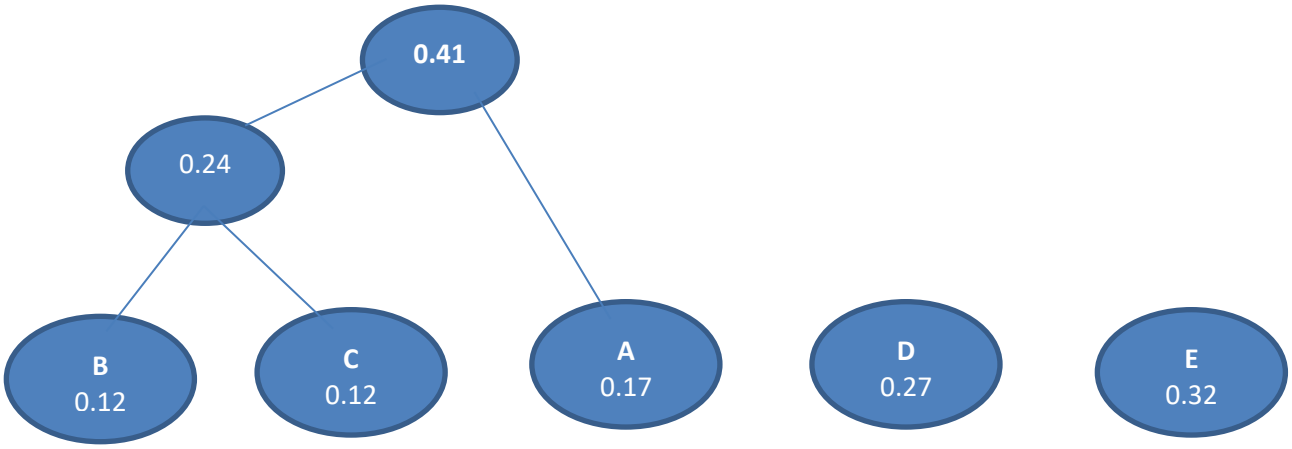
Arrange the Character according to the probability



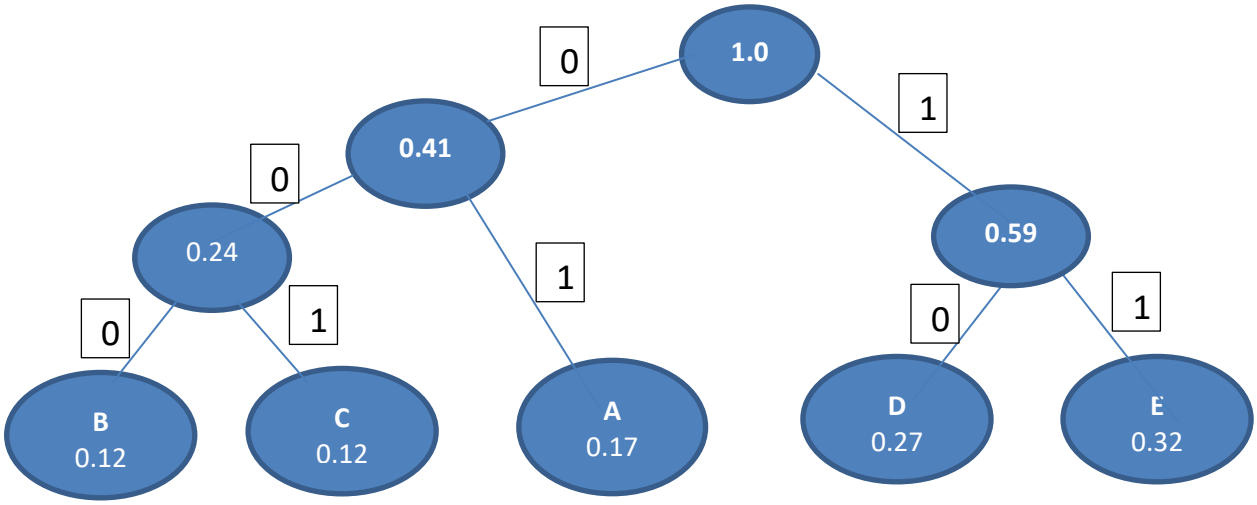
Step2: Combine two least probability



Again combine two least probability



Again combine two least probability and so on..



A character's code is found by starting at the root and following the branches that lead to that character. The code itself is the bit value of each branch on the path, taken in sequence.

Character	A	B	C	D	E
New Code	01	000	001	10	11
	2 bits used	3 bits used	3 bits used	2 bits used	2 bits used

Total bits used = 2 (17+27+32) + 3(12+12) = **224Bits**

Compressed massage: 224 Bits
Uncompressed massage: 300 Bits