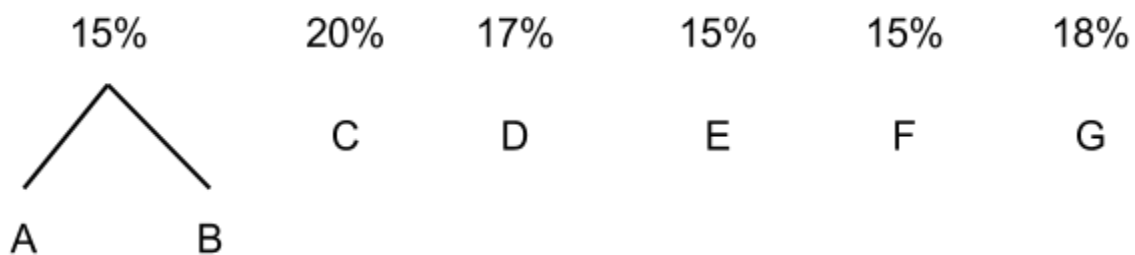CS 240 Tutorial - Huffman Encodings

Given the following letter frequencies give the Huffman encodings for each letter.

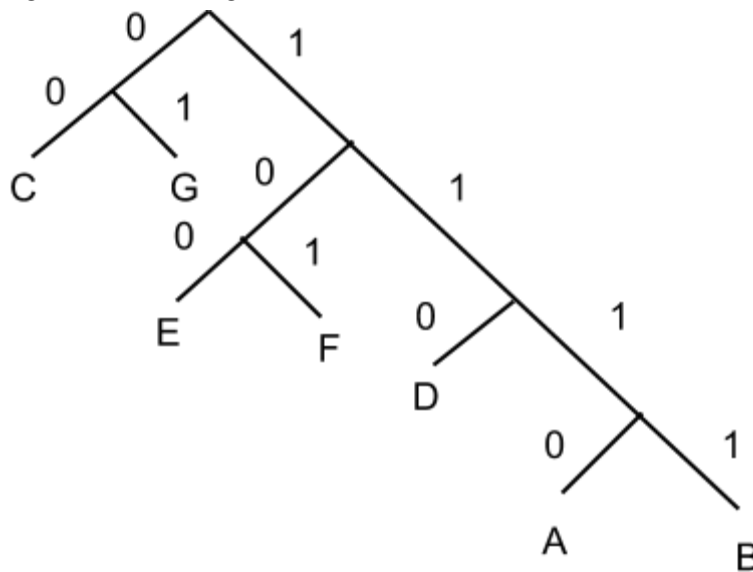A = 10%, B = 5%, C = 20%, D = 17%, E = 15%, F = 15%, G = 18%

10%	5%	20%	17%	15%	15%	18%
A	B	C	D	E	F	G

We take the two smallest frequencies (A and B) and combine them in a tree structure as follows.

15%	20%	17%	15%	15%	18%
A B	C	D	E	F	G

We continue doing the same things by taking the two smallest values and combining them in a tree. Eventually we get the following:

Therefore, the encodings are: A = 1110, B = 1111, C = 00, D = 110, E = 100, F = 101, G = 01

Using the encoding above, by what factor is the following string compressed?

GEEEBEFCGCBCDACECGCGAFDCG

2+3+3+3+4+3+3+2+2+2+4+2+3+4+2+3+2+2+2+2+4+3+3+2+2 = 67 bits

Normally for 7 characters you would need 3 bits per character to represent each character.

3bits * 25 characters = 75 bits

67/75 = 89% Therefore we saved 11%.