

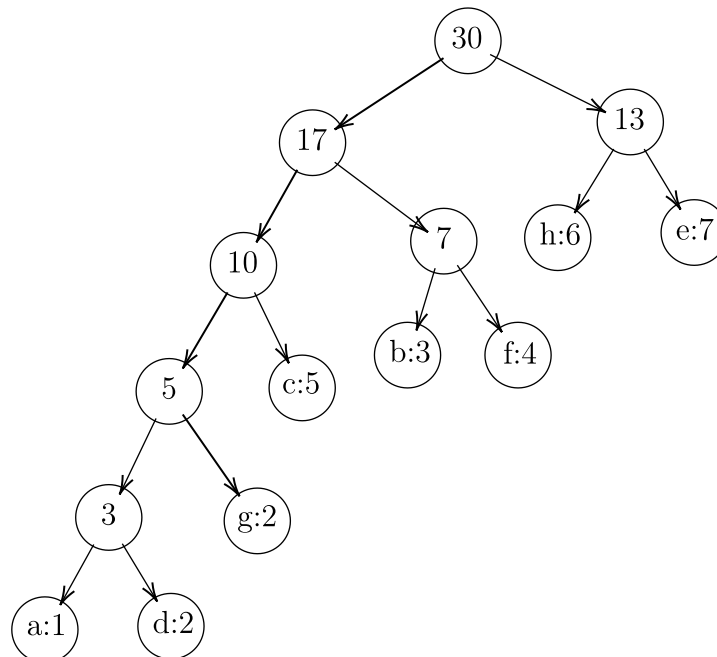
HUFFMAN CODING

Algorithms Analysis

github.com/erngv

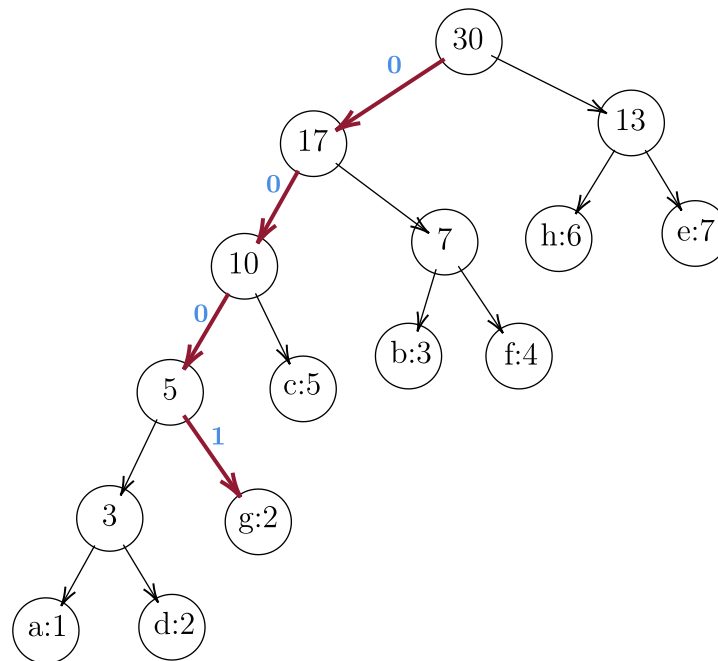
1. What is an optimal Huffman code for the following set of frequencies?
a:1, b:3, c:5, d:2, e:7, f:4, g:2, h:6

To answer this question, let us construct a Huffman tree by making a leaf node for every unique character and sorting them in increasing frequency order. Afterwards, we will connect the two nodes with minimum frequency, creating a new internal node with frequency corresponding to the sum of the two frequencies of the node children. Then, we will repeat the previous step until we reach a root node.



Since each path from a node parent to a left child node will have a value of **0**, and each path from a node parent to a right child node will have a value of **1**, now we can derive optimal variable-length codes for each character by following a path from the root node of the tree to each node that contains a character with its frequency.

For instance, an optimal code for character **g** is **0001**, following the path shown below:



Therefore, an optimal Huffman code for this problem would be:

a \longrightarrow 00000
 b \longrightarrow 010
 c \longrightarrow 001
 d \longrightarrow 00001
 e \longrightarrow 11
 f \longrightarrow 011
 g \longrightarrow 0001
 h \longrightarrow 10