

**Universidade Federal de Ouro Preto**  
Instituto de Ciências Exatas e Biológicas – Departamento de Computação  
BCC202 – Estrutura de Dados  
Trabalho Prático 03

Aluno: Júlia Eduarda Miranda de Sousa 18.1.4084

### **Codificação de Huffman**

A codificação de Huffman é um método de compressão sem perdas, de modo que a informação obtida após a descompressão é idêntica à informação original. Esse algoritmo utiliza a frequência em que caracteres ocorrem no conjunto de dados a ser compactado determinando um número variável de bits para cada símbolo. Sua ideia básica é utilizar um menor número de bits para caracteres que aparecem com maior frequência e utilizar mais bits para caracteres menos frequentes.

A implementação do algoritmo pode ser feita através de uma árvore binária que armazena os caracteres em cada uma de suas folhas, de modo que o caminho da raiz até as folhas determina a sequência de bits que será usada para codificar cada caractere. Por convenção, os arcos (caminhos entre nós da árvore) da esquerda são numerados como 0 e os arcos da direita com 1 e, ao seguir o caminho até a folha desejada, os zeros e uns são concatenados formando o código do caractere.

A construção da árvore de codificação se dá a partir de um grupo de árvores. Em um primeiro momento, essas árvores são compostas por apenas um nó, que contém um caractere e seu peso. A cada iteração, as árvores são unidas criando uma nova árvore, de modo que a quantidade de árvores diminui a cada passo. Desse modo, temos uma lista de árvores, que deve estar na ordem crescente, fazendo com que o caractere de menor peso seja a raiz da primeira árvore e o caractere de maior peso seja o nó raiz da última árvore.

Uma nova árvore é criada, de modo que seu peso é a soma dos pesos das duas primeiras árvores da lista e elas se tornam subárvores dessa nova árvore. Assim, a árvore nova é inserida de forma ordenada na lista e são retiradas as árvores que agora compõem a árvore nova. O processo é repetido até que se obtenha uma árvore final com todas as árvores iniciais concatenadas, de modo que o nó raiz será o caractere de maior peso. Após a construção dessa árvore, quanto mais alta for a frequência de um termo, mais alto ele aparecerá na árvore

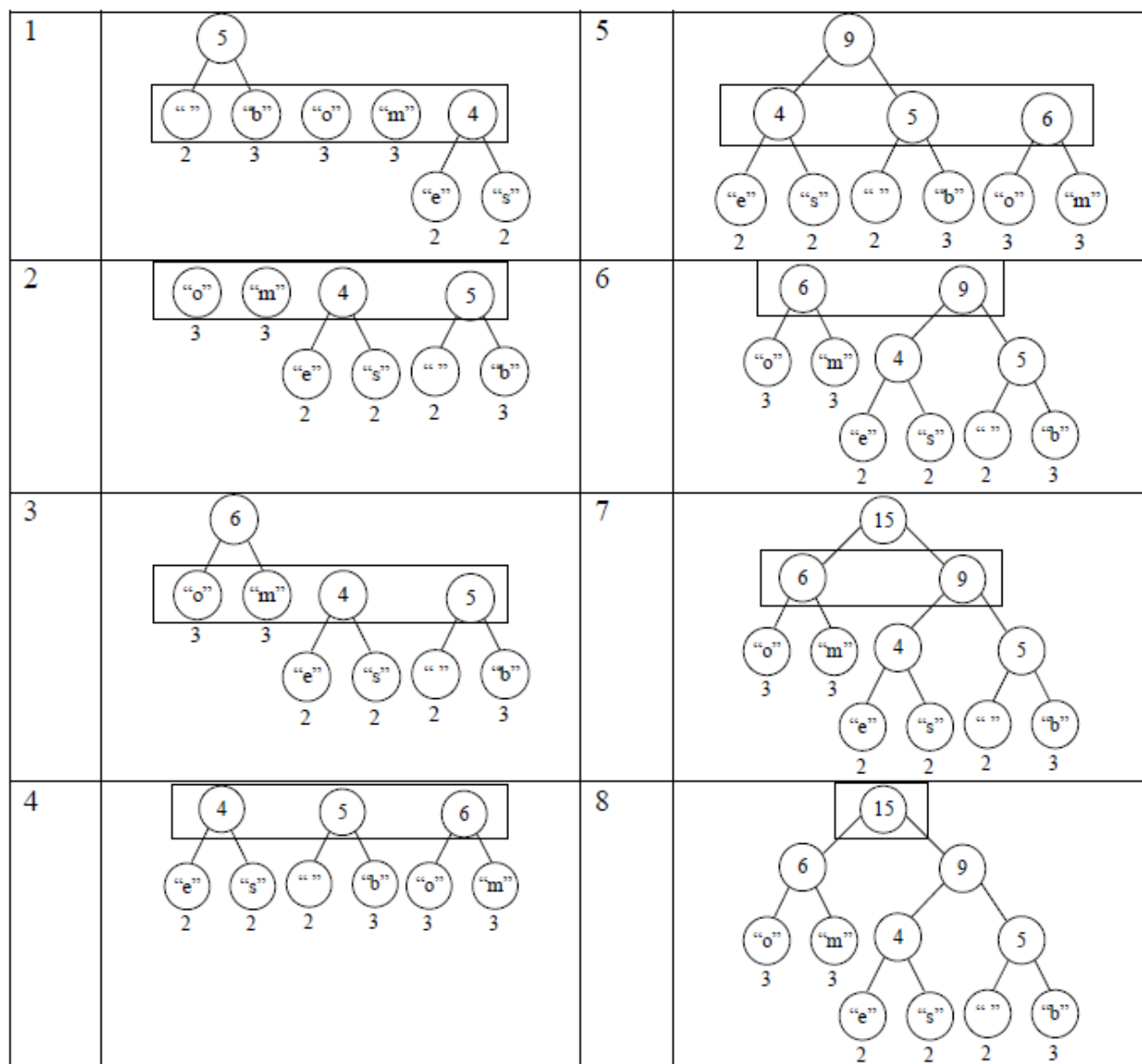


Figura: Exemplo de formação de uma árvore da codificação de Huffman < Retirado de: <https://inf.ufes.br/~pdcosta/ensino/2018-2-estruturas-de-dados/material/CodificacaoHuffman.pdf> >

## Tabela Hash

A tabela Hash é uma estrutura de dados do tipo dicionário, de modo que em seu caso médio é possível encontrar a chave desejada em tempo constante. Os registros são endereçados na tabela a partir de uma função de transformação aritmética na chave de pesquisa, sendo possível que alguns dados sejam adereçados a uma mesma posição, havendo colisões. Para resolver esse problema, cada posição da tabela pode conter uma lista encadeada, sendo composta por elementos que tenham o mesmo adereçamento. Essa técnica é conhecida como endereçamento fechado.

O cenário problema proposto é o cadastro de alunos de uma disciplina em um diário de classe, sendo a chave de pesquisa o número de matrícula. A tabela será construída por endereçamento fechado, uma vez que dificulta a remoção de dados e aumenta a complexidade para busca nos casos de colisão.