

## **ATIVIDADE – CODIFICAÇÃO DE HUFFMAN**

Os códigos de Huffman são códigos binários de comprimentos variados que são determinados a partir da frequência de uso de um determinado caractere. A principal ideia é associar números binários com menos bits aos caracteres mais usados num texto, possibilitando a sua compactação. O algoritmo de compactação usando os códigos de Huffman tem três fases:

- Cálculo da frequência de cada caractere no texto.
- Execução do algoritmo de Huffman para construção de uma árvore binária.
- Codificação propriamente dita. Uma árvore de Huffman é uma árvore binária em que cada nó interno tem dois filhos.

A árvore binária é montada obtendo sempre as menores frequências da lista e juntando-as numa sub-árvore em que os nós folha são organizados colocando-se às ocorrências com maior frequência à esquerda e os com menor frequência à direita. O nó raiz conterá a soma das frequências dos nós folha.

No início, cada um dos caracteres forma uma árvore que é composta apenas pela raiz e cujo conteúdo é a frequência com que esta letra ocorre no texto. Em seguida, são escolhidas as duas árvores com as menores frequências associadas a elas são transformadas em uma só árvore cujo valor é a soma do valor destas duas. Este processo é repetido até a existência de uma única árvore

## **COMPACTAÇÃO E ÁRVORES BALANCEADAS**

O fator mais crítico durante a compactação de um arquivo é o tamanho do buffer original. Caso o buffer seja muito grande provavelmente conseguiremos valores de frequência iguais para todos os bytes do buffer e, talvez, as tabelas de frequência ficarão extensas. Dessa forma, a árvore de Huffman estará não estará completamente balanceada e assim não será possível “resumir” os bits de itens com maior frequência, causando baixa compressão.

Logo é importante que os compactadores utilizem um buffer fixo e dividam os dados originais em blocos, tornando-se possível realizar uma melhor compactação e mantendo as árvores de Huffman balanceadas.