

3º laboratório de estruturas de dados, 16_05_2022

Implementar um compactador e um descompactador de arquivos texto, de acordo com o algoritmo de Huffman.

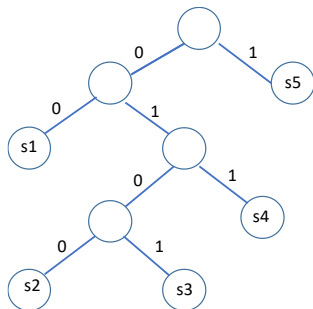
Ref: Estruturas de dados e seus algoritmos, Jaime Luiz Szwarcfiter e Lilin Markenzon, 3ª ed. LTC

No método de Huffman é codificado um texto para se obter a sua compactação. Supõe-se que o texto seja constituído de um conjunto de símbolos e que é conhecida a frequência de cada símbolo. O método consiste em atribuir um código a cada símbolo de forma a compactar o texto na íntegra. Os códigos procurados são dados por uma árvore binária de prefixos, ou seja, nenhum código é prefixo de algum outro. A árvore binária de prefixo é estritamente binária e rotulada. Para cada nó interno v , da árvore T , a aresta que conduz ao filho esquerdo é rotulada com “zero” e a que conduz ao filho direito é rotulada com “um”. Cada símbolo está associado a uma folha da árvore. Os códigos dos símbolos são sequências binárias. O código do símbolo “s” é igual à sequência dos códigos das arestas do caminho que vai da raiz até a folha “s”.

Ex: símbolo-código

s1	00
s2	0100
s3	0101
s4	011
s5	1

A árvore de prefixos:



O símbolo s5 é codificado como 1, s4 é codificado com 011 e assim sucessivamente. Uma vantagem de utilização de códigos de prefixo é a facilidade para executar as tarefas de codificação e decodificação. O símbolo de maior frequência é s5 cujo caminho é 1, em seguida s1 cujo caminho é 00, depois s4 cujo caminho é 011 e, em seguida, s2 ou s3.

Assim, a palavra s1s3s2s4s1s5 é codificada como 0001010100011001 (00-0101-0100-011-00-1), que, em binário, cabe em 2 bytes ou 16 bits que deve ser armazenada em bytes sucessivos. Com base na tabela de frequências a decodificação é tarefa trivial.

Teste o compactador com um arquivo formado pela letra do poema "Trem de Alagoas" de Ascenso Ferreira.

Gere um relatório indicando como construiu a árvore de prefixos, exiba a frequência dos símbolos em ordem decrescente (do mais frequente para o menos frequente) e o percentual de compactação obtido. Exiba o arquivo original e o arquivo descompactado, indicando se a compactação e descompactação funcionou a contento. Indique o que não funcionou, se for o caso.