

ALGORITMOS AVANÇADOS - Relatório Projeto 1

Problema escolhido: Código Huffman

Nomes: Gabriel Van Loon Bodê da Costa Dourado Fuentes Rojas - 10391607

Alberto Campos Neves - 10819033

O Código de Huffman

O Código Huffman se trata de um algoritmo que tem como objetivo comprimir cadeias de caracteres da melhor forma possível partindo do princípio de que os símbolos que aparecem com maior frequência em um determinado texto devem ocupar um espaço menor do que aqueles que aparecem raramente.

Para isso, os símbolos são representados na forma de códigos binários de tal forma que o código de um nunca seja prefixo do código de outro.

O Algoritmo de Huffman

Para decidir quais códigos serão atribuídos para cada símbolo, o Algoritmo de Huffman se utiliza de uma estratégia gulosa em que cada símbolo é representado inicialmente como um Nó Folha de uma árvore binária sendo eles totalmente independentes entre si.

Esses Nós são inseridos em uma Min-Heap que ordena os símbolos de acordo com o valor das suas frequências.

Nesse caso, o objetivo é gerar uma única árvore tal que os símbolos de alta frequência estejam o mais próximo possível da raiz.

O passo ótimo consiste de retirar os dois Nós com menor frequência e somá-los à um novo nó raiz que irá apontar para ambos. Em seguida esse nó será reinserido na Min-Heap.

O passo ótimo é repetido até que haja apenas um único nó da Min-Heap.

Entrada do Programa

O programa recebe como entrada uma sequência de caracteres ASCII que serão utilizados para a contagem de frequência de cada um dos símbolos. O sinal de fim de linha indica para o programa que o texto chegou ao fim.

Exemplo:

“Ao longo da sua vida, Huffman fez contribuicoes significativas para o estudo das maquinas de estados finitos, comutacao de circuitos, processos de sintese, e projetos de sinais.\n”

“aaaabbcd\n”

Saída do Programa

O programa primeiramente exibe os Nós da Árvore a partir de sua raiz até seus nós folhas. Cada nó é representado da forma “{A,B,C}” sendo **A** o valor de frequência daquele nó, **B** o bit indicando se o nó é folha ou não e **C** o símbolo (sendo representando pelo vazio caso o nó não se trate de um nó folha).

Em seguida o programa exibe uma tabela associando cada símbolo com a sua frequência e código binário.

Exemplo:

“Código Huffman

Árvore Gerada:

{8,0,' }

{4,1,'a'} {4,0,' }

{2,1,'b'} {2,0,' }

{1,1,'c'} {1,1,'d'}

<i>Símbolo</i>	<i>Frequencia</i>	<i>Código</i>
<i>'a'</i>	<i>4</i>	<i>0</i>
<i>'b'</i>	<i>2</i>	<i>10</i>
<i>'c'</i>	<i>1</i>	<i>110</i>
<i>'d'</i>	<i>1</i>	<i>111”</i>