

Resolução – Lista 7 (Projeto e Análise de Algoritmos)

Março - 2023 / Leticia Bossatto Marchezi – 791003

Questão 1

O que é o código de Huffman? Explique para que ele serve e qual sua diferença para um esquema de codificação padrão

Resolução:

O algoritmo de Huffman é uma estratégia utilizada para compressão de dados sem perda de informação. Ele utiliza uma codificação em forma de dicionário com base na probabilidade de ocorrência dos símbolos. Dessa forma, os símbolos mais frequentes possuem representação de menor tamanho, e os de menor frequência com tamanhos maiores.

A diferença para outros esquemas de codificação é que o algoritmo de Huffman apresenta uma economia de espaço pois os símbolos com maior taxa de ocorrência são codificados de tal forma a serem representados com menor custo, por exemplo com a necessidade de uso de menos bits, enquanto outros algoritmos codificadores apenas atribuem uma quantidade fixa de bits para todos os símbolos, desprezando assim a frequência de ocorrência.

Questão 2

Explique o funcionamento do algoritmo de Huffman. Porque ele é um algoritmo guloso?

Resolução:

O algoritmo de Huffman é guloso pois utiliza das escolhas ótimas locais para construir uma melhor solução, tomando 2 símbolos de menor probabilidade por vez e atribuindo códigos maiores à eles, até alcançar os símbolos de maior probabilidade. A abordagem aplicada é bottom-up, ou seja, inicia-se resolvendo os subproblemas até alcançar o caso mais complexo.

Questão 3

Supondo que as frequências de ocorrência dos símbolos seja dada pela tabela a seguir, crie os códigos de Huffman para cada símbolo:

Resolução:

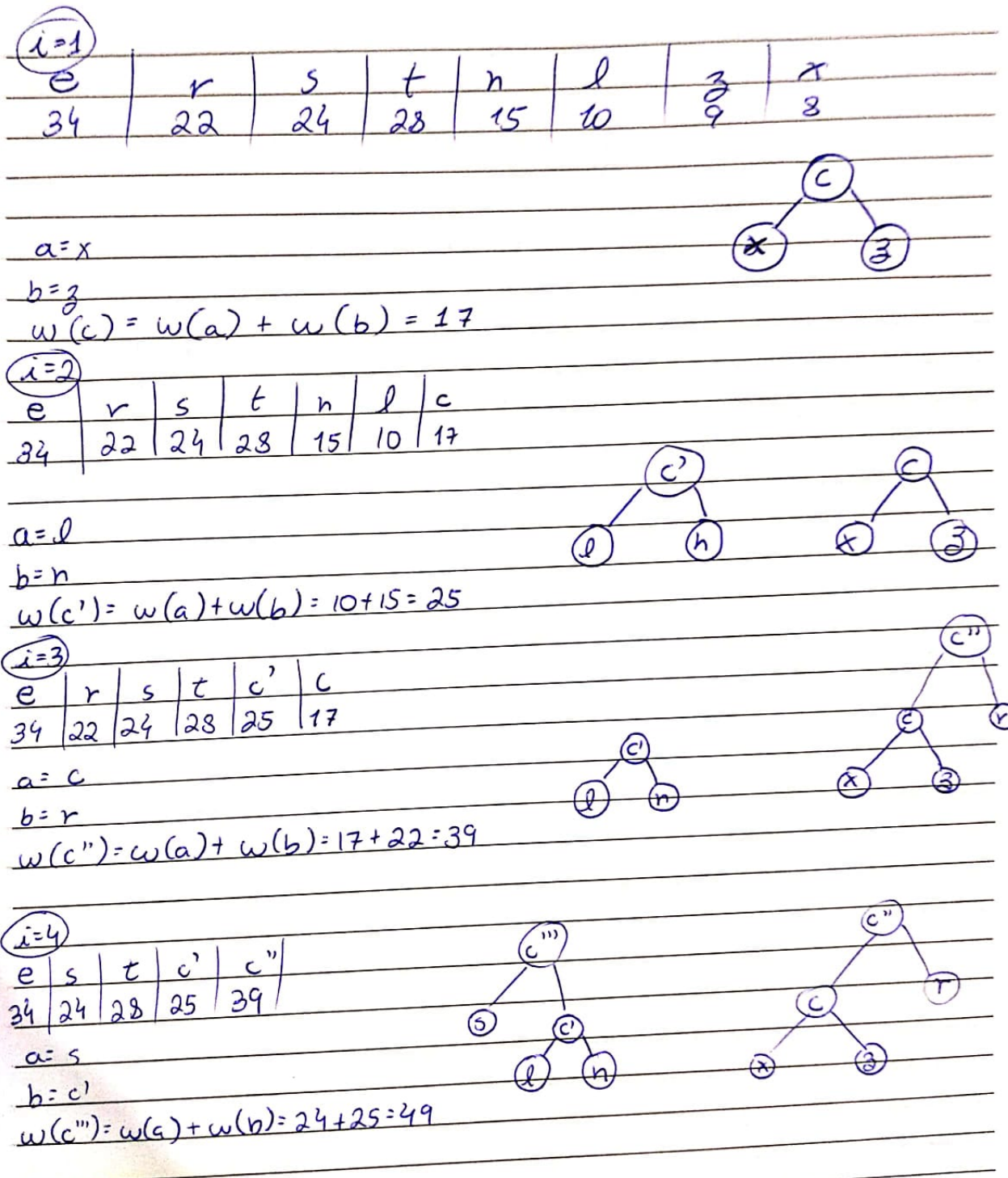


Figure 1: Resolução

Resolução:

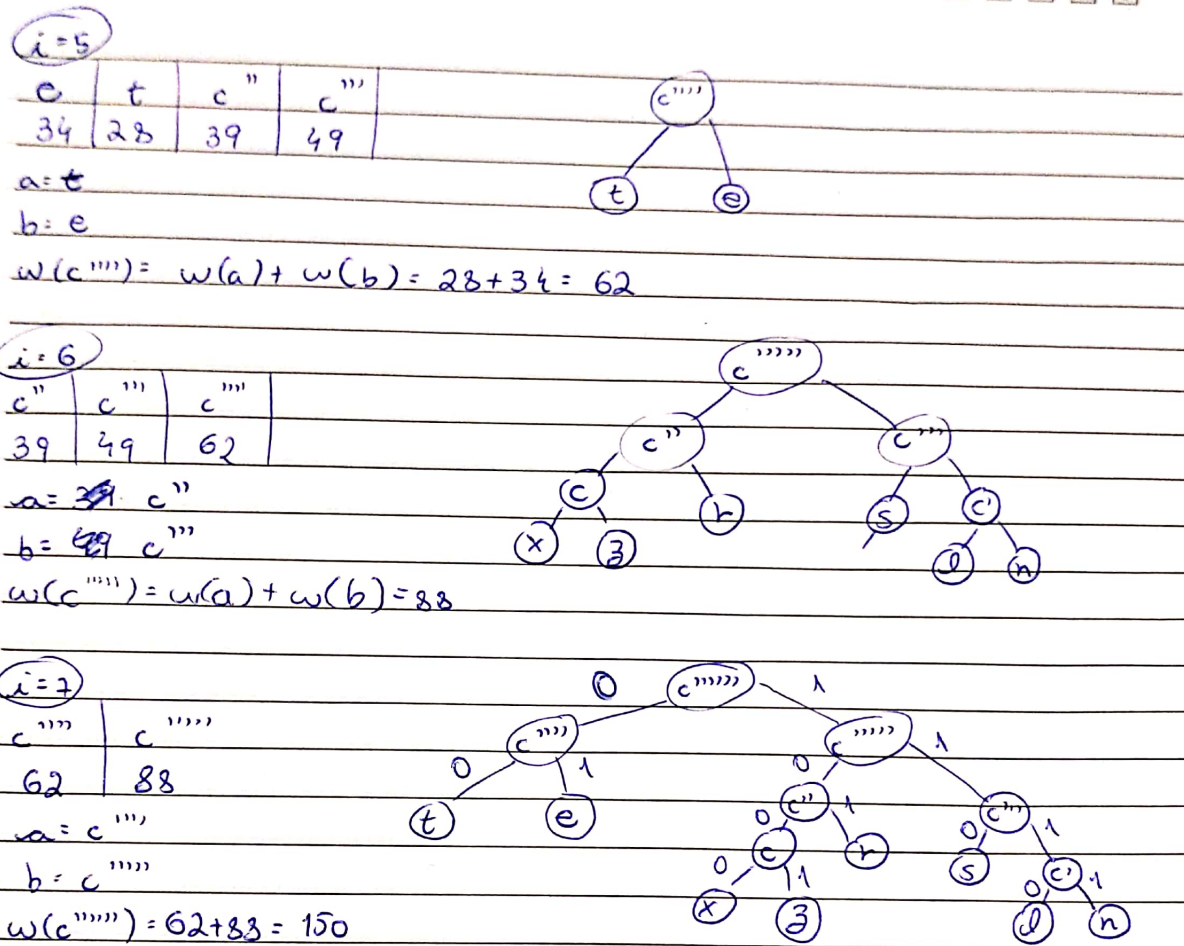


Figure 2: Resolução

Resultado da codificação:

e = 01

r = 101

s = 110

t = 0

n = 1111

l = 1110

z = 1001

x = 1000

Questão 4

Os caracteres de a até h possuem como frequências em um texto o padrão de uma sequência de Fibonacci, ou seja: a:1, b:1, c:2, d:3, e:5, f:8, g:13, h:21 Se utilizarmos o código de Huffman para representar esses símbolos, mostre qual será a sequência de bits da cadeia de caracteres fdheg Para isso, mostre o processo de construção da árvore e criação dos códigos de cada símbolo. Generalize sua resposta para definir os códigos dos n símbolos se as frequências são dadas pelos n primeiros números da sequência de Fibonacci.

Resolução:

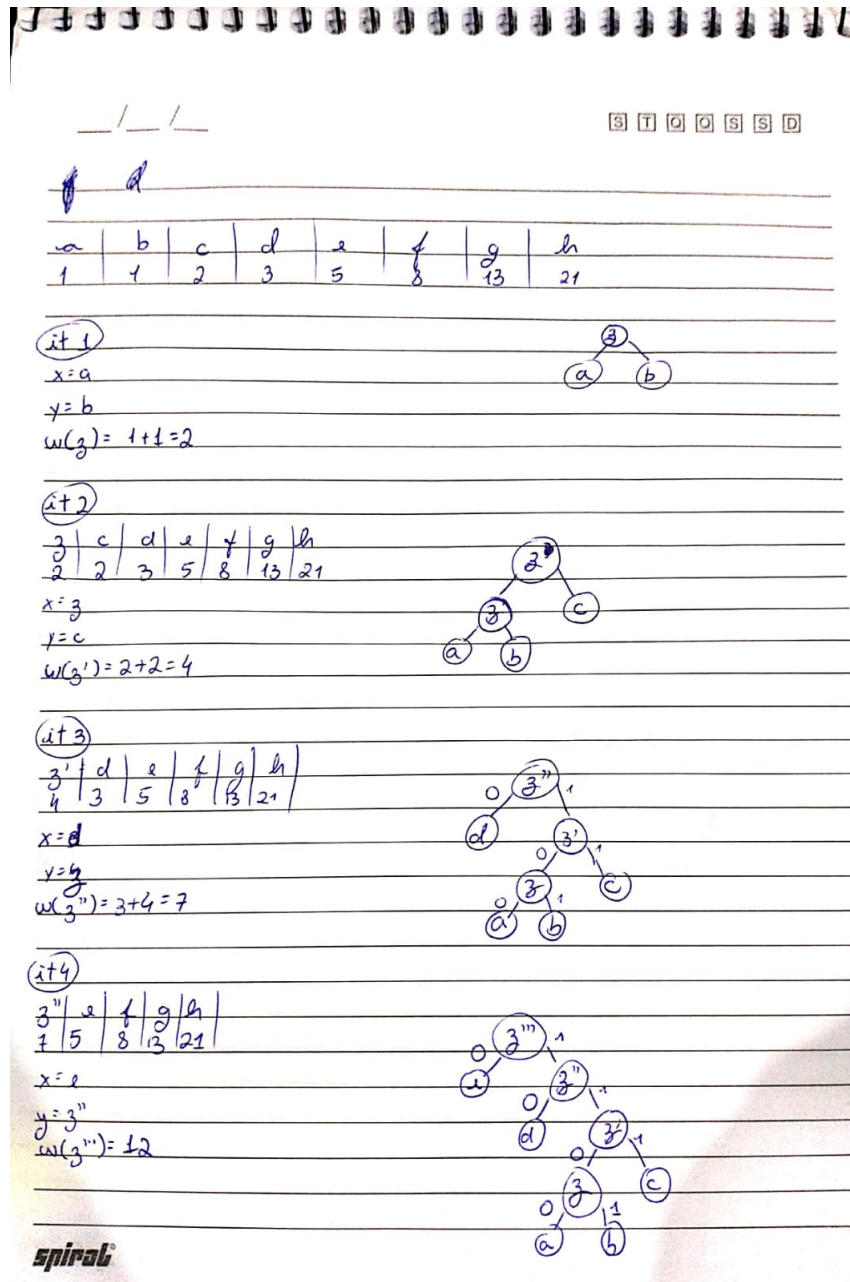


Figure 3: Resolução

Resolução:

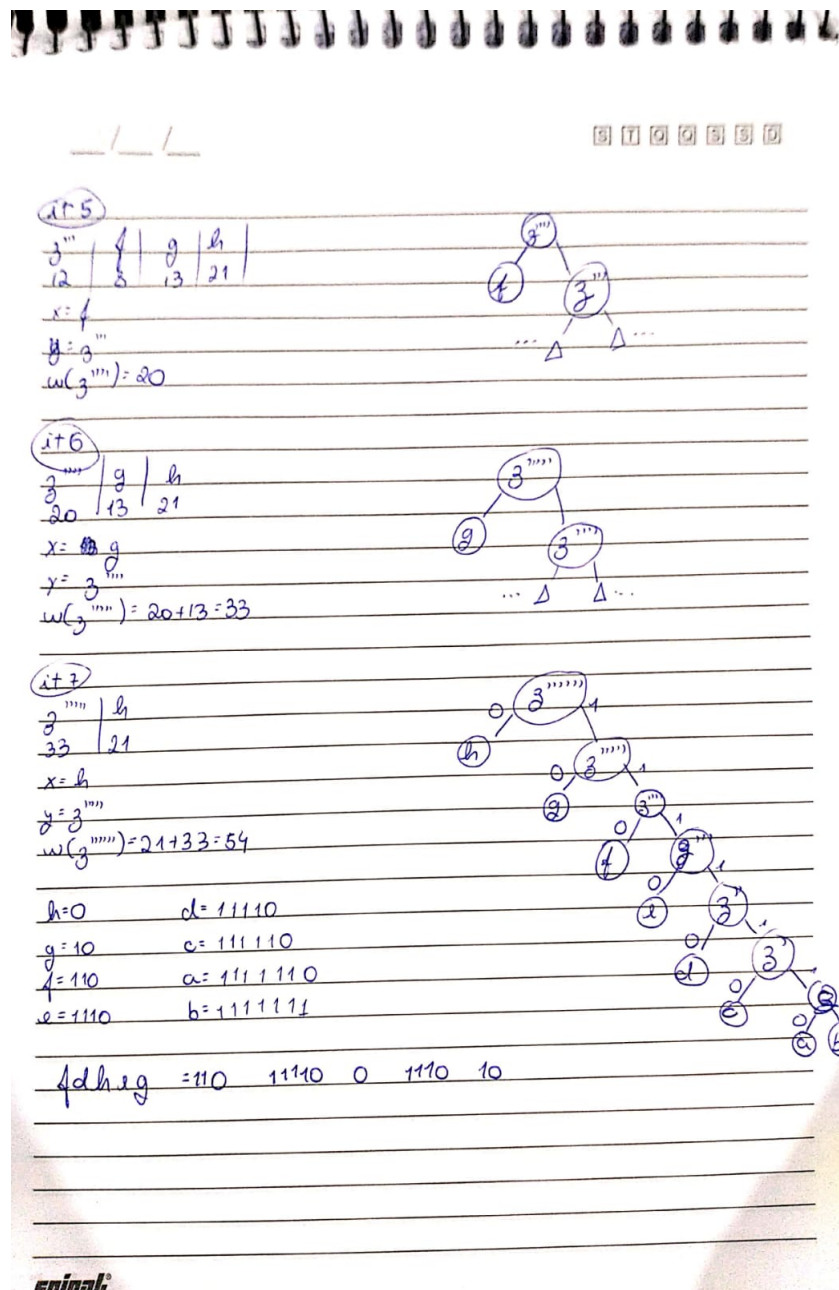


Figure 4: Resolução

Em resumo, devido à sequência de fibonacci, para um novo símbolo adicionado à sequência de caracteres este assumirá o código 0 e deve-se adicionar um bit 1 à esquerda da codificação de todos os outros símbolos anteriores.

Questão 5

Mostre que o código de Huffman é ótimo no sentido de minimizar o tamanho do código gerado.

Resolução:

De acordo com o teorema de Shannon, a quantidade de informação de um símbolo é logaritmo do inverso da sua probabilidade.

A distribuição de probabilidades é afetada pela entropia $H(w)$, que é um limitante inferior ao tamanho esperado de código C $l(C)$. Para provar isso, pode-se calcular a diferença entre $H(w)$ e $l(C)$.

Dado que o algoritmo de Huffman gera uma árvore binária, a quantidade de nós na última camada é sempre maior do que dos nós intermediários.

Em casos que a quantidade de informação é correspondente ao tamanho do código, os símbolos mais frequentes estão mais próximos da raiz e possuem portanto codificação menor, e os símbolos menos frequentes possuem maior quantidade de informação e maior código.