

Codificação de Huffman



Base teórica e seu funcionamento

Alunos:

- Luis Vinicius Capelletto - Eng. da Computação
- Gabriel Henrique - Eng. da Computação

Conteúdo



- O problema
- O que é a codificação de Huffman
- Implementação.
- Como realizar a codificação

Tabela de frequência

Árvore de Huffman

Tabela de códigos

Codificação

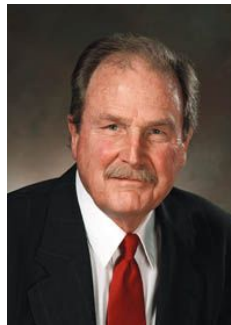
O problema



- . Problema de otimizar o uso do espaço de armazenamento.
- . Devemos nos certificar que os dados sejam armazenados da maneira mais econômica possível.
- . Textos e outros tipos de informação apresentam muita redundância (letras/pixels repetidos).
- . Exemplo: “Estrutura de dados”.

O que é a codificação de Huffman

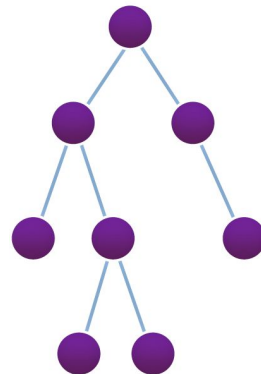
David A.
Huffman



- . A codificação de Huffman é um algoritmo de compressão sem perda de dados;
- . Utilizado principalmente para arquivos de texto ;
- . Código fundamental para o entendimento dos métodos de compressão atuais;
- . Parte do princípio de criar um formato binário onde são atribuídos menos bits para os símbolos mais frequentes e mais bits para os menos frequentes, em geral a saída no formato binário possui menos bits que o original;
- . Desenvolvido por David Huffman, enquanto aluno de doutorado do MIT por volta de 1950;
- . Os melhores desempenhos da codificação de Huffman acontecem quando as probabilidades de ocorrência dos símbolos são potências negativas de dois. Outra vantagem da codificação Huffman é que a codificação gerada não é ambígua;

Implementação - Teoria

- . Forma-se uma tabela contendo os símbolos presentes no texto analisado e suas respectivas frequências;
- . Implementa um tipo abstrato de dados (TAD) do tipo árvore binária ordenada, construída de baixo para cima (bottom up) a partir da tabela gerada;
- . A árvore formada é binária e ordenada, onde o maior valor é armazenado na raiz;
- . Os nós internos armazenam a soma dos valores de frequência de seus filhos;
- . Nós folhas contém os símbolos e suas frequências;
- . Códigos são gerados ao percorrer os caminhos do nó raiz a suas folhas;



Funcionamento

-Tabela de Letras/Frequências

Letra	Frequência
H	1
u	1
f	2
m	1
a	1
n	1

Exemplo: Huffman



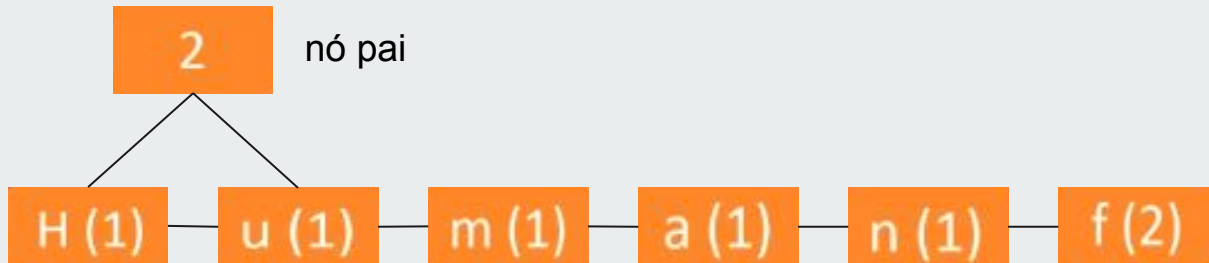
- . A princípio as letras e suas frequências são armazenadas e podem ser representadas por uma tabela;
- . A tabela pode ser armazenada através de uma lista encadeada, por heap, ou até em arrays e etc;
- . A forma implementada depende de sua utilização;

Funcionamento

. Ordena-se por ordem de frequência os nós símbolos;

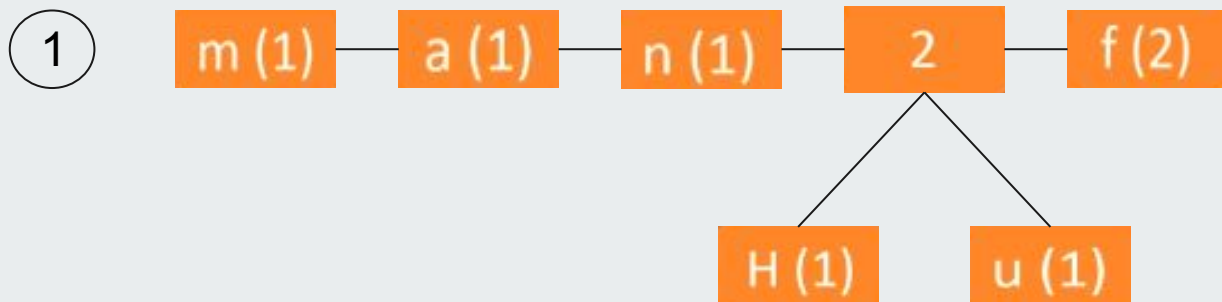


. Define-se um nó pai que liga os 2 símbolos de menor frequência, esse nó contém a soma das frequências dos 2 nós filhos e nenhum símbolo (ou nulo '\0');



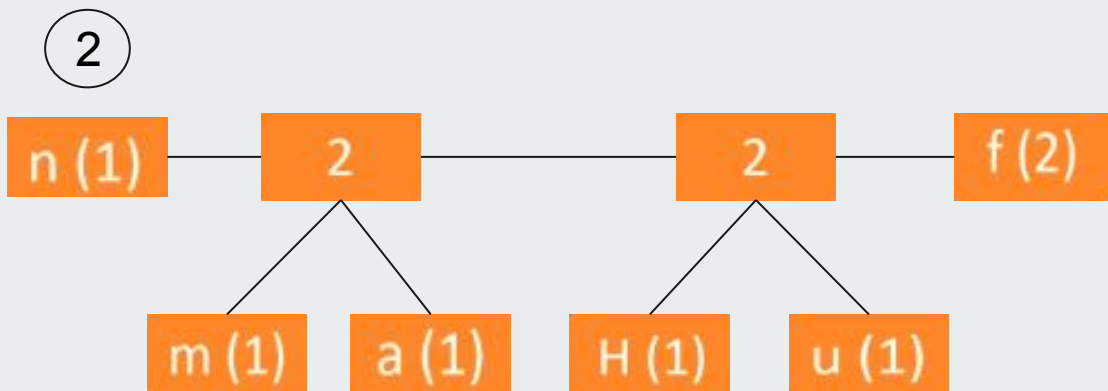
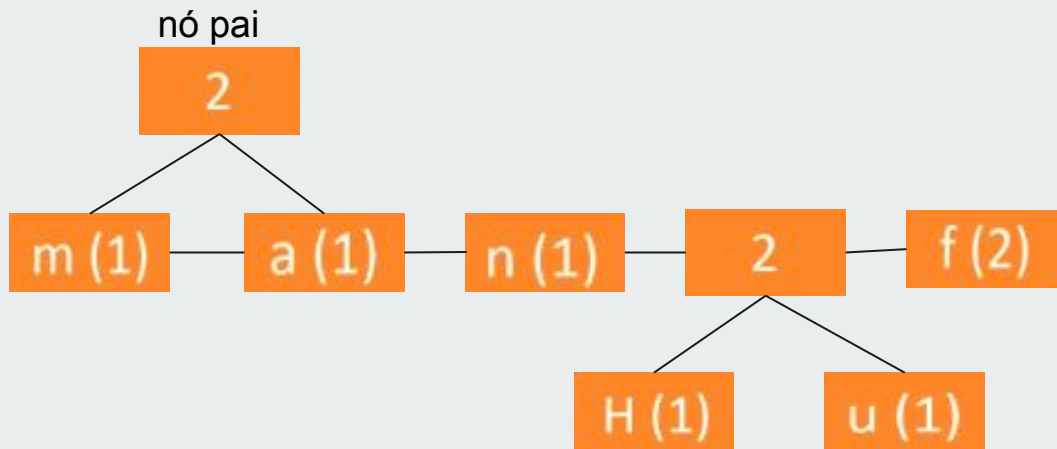
Funcionamento

- . Os dois elementos símbolos que estavam presentes no início da lista são retirados;
- . Insere-se de forma ordenada o nó pai na estrutura contendo os demais nós símbolos;



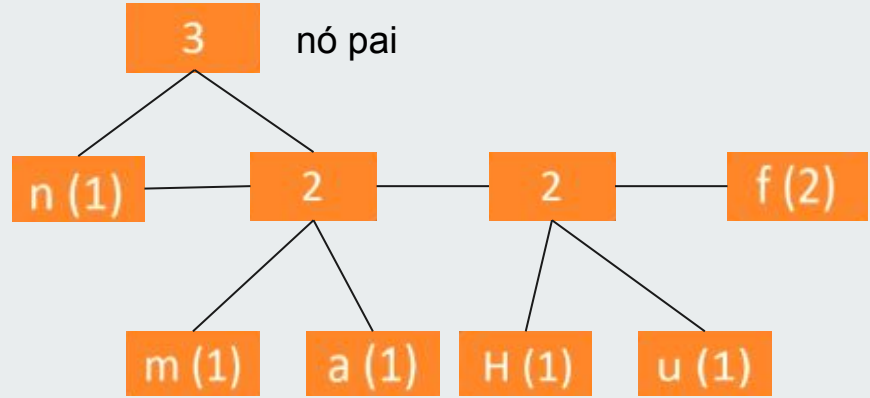
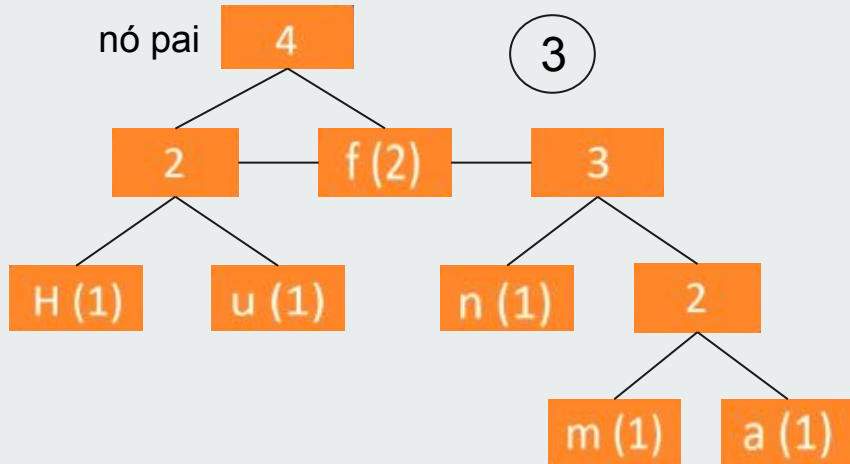
Funcionamento

. Repete-se a operação;



Funcionamento

. A operação é repetida até que a estrutura que representava a tabela inicial contenha apenas 1 elemento;



Funcionamento

. Para gerar os códigos de cada símbolo, percorre-se a árvore gerada da raiz até cada nó folha, atribuindo-se 0 caso o percurso seja pelo ponteiro esquerdo e 1 caso percurso passe pelo caminho do ponteiro da direita;

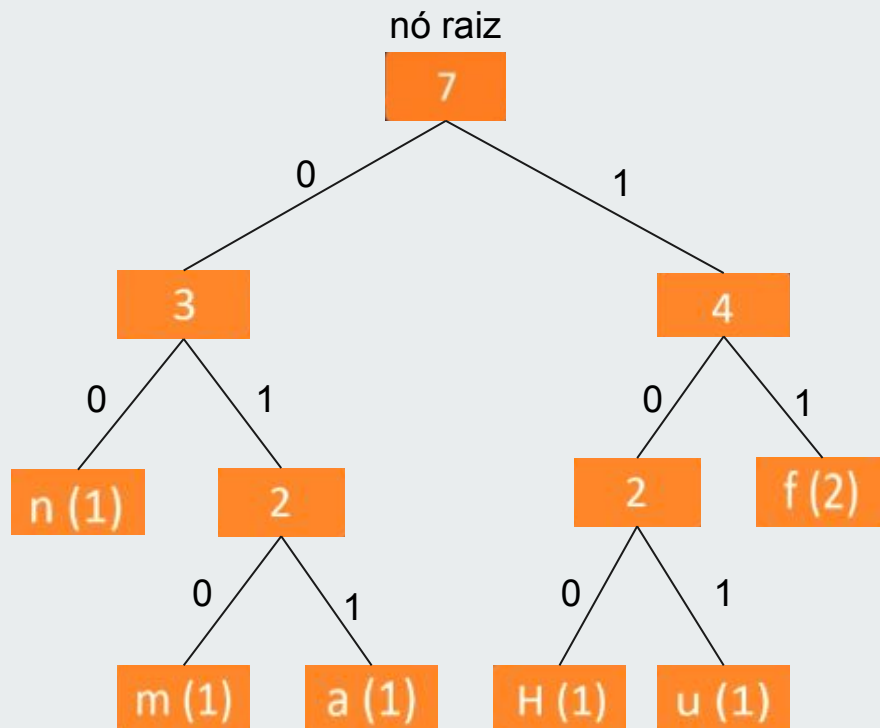


Tabela de códigos

Letra	Código
H	100
u	101
f	11
m	010
a	011
n	00

. Árvore de Huffman gerada;

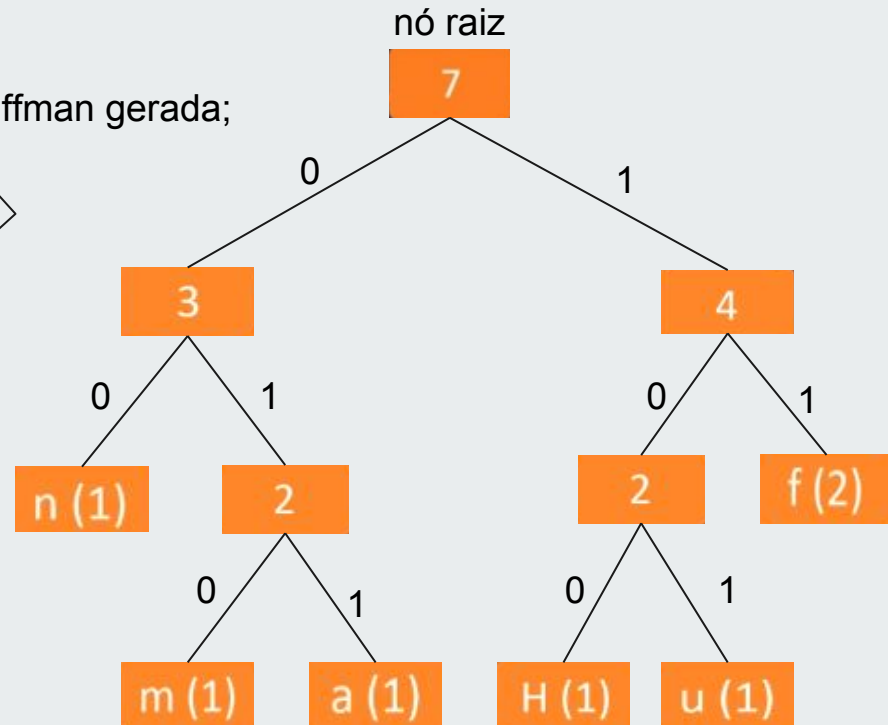
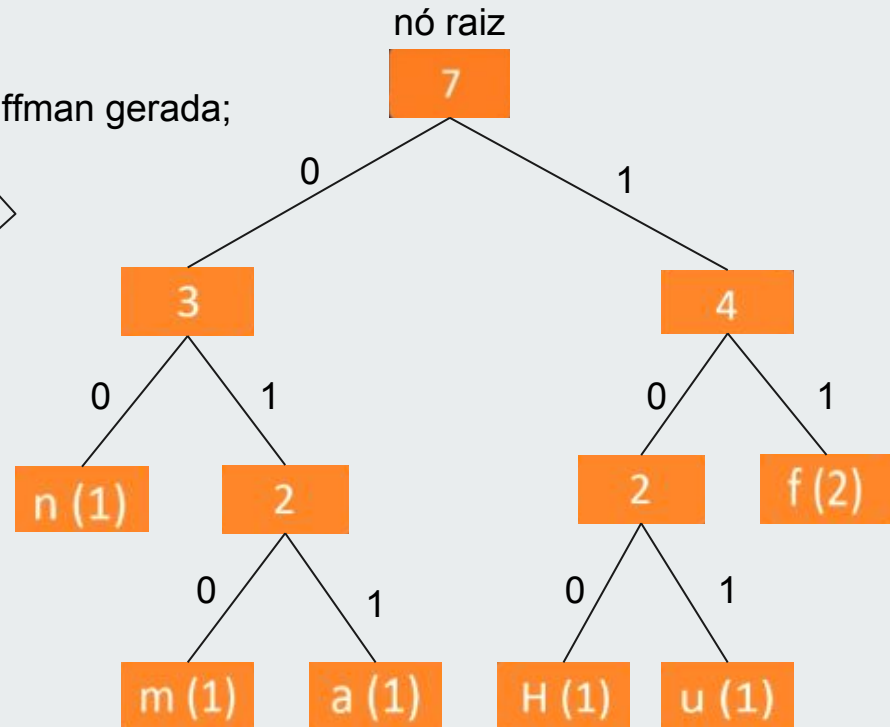


Tabela de códigos

Letra	Código
H	100
u	101
f	11
m	010
a	011
n	00

. Árvore de Huffman gerada;



Código gerado: 100101111101001100

Verificação

H u f f m a n

Código antes: 0100100001110101011001100110110011011010110000101101110
(56 bits).

Código gerado: 100101111101001100
(18 bits).

Economia de 38 bits;
Taxa de compressão: 67,86%

Aplicações



.Compressão de imagens (jpeg)

.DEFLATE (usada nos arquivos do padrão ZIP e PKZIP)

.gzip

.png

.bzip2

Considerações Finais

- . Para descomprimir ou expandir novamente o código comprimido deve-se armazenar juntamente com o arquivo comprimido a árvore produzida para codificação ou a tabela de códigos formada;

- . **Curiosidade:**

- . Aplicativo chamado **Huffman Encoder** na Play Store:

Codifica um texto de entrada qualquer e fornece informações sobre o cálculo da entropia de compressão, sua taxa de compressão e sua tabela de símbolos, frequências e conversão binária;

Fim

Canarinho pistola deseja uma ótima Copa a todos!!!



#VEMHEXA



Referências

<https://www.ime.usp.br/~pf/estruturas-de-dados/aulas/compress.html>

<https://www.ime.usp.br/~pf/estruturas-de-dados/aulas/huffman.html>

<https://www.youtube.com/watch?v=MaPGqrdjCo8>

<http://www.eletrica.ufpr.br/marcelo/TE072/012007/Nilson-JPEG.pdf>