

## MAC 323 – Algoritmos e Estruturas de Dados II

Primeiro semestre de 2023

### EP-Substitutivo – Entrega: 15 de julho

Este exercício-programa é de entrega opcional, para os alunos que não puderam entregar um dos EPs durante o semestre, ou que desejem melhorar sua média.

#### Códigos de Huffman

Como visto em sala de aula, os códigos de Huffman dão um algoritmo de compactação de arquivos, baseado na frequência de uso destes caracteres. A ideia é associar números binários com menos bits aos caracteres mais usados nos textos. Se você quiser saber mais sobre códigos de Huffman, consulte o ótimo livro [1].

O objetivo deste (último???) exercício-programa é construir dois programas:

- um compactador de arquivos baseado nos códigos de Huffman, como visto em sala de aula. Seu programa deverá ler um arquivo de texto, contabilizar a frequência de uso de cada letra do alfabeto, montar a árvore ótima e gravar o arquivo compactado, que deverá conter a árvore e o arquivo compactado, usando a codificação obtida;
- um descompactador de arquivos, que deve ler um arquivo binário que contém a árvore usada na compactação e o texto compactado, e o descompacta, obtendo o arquivo original.

Faça testes com diversos tipos de arquivos (textos em português, outras línguas, códigos, imagens, etc) e compare a compactação obtida pelo código de Huffman com pacotes que você costuma usar. Nesse exercício-programa, você também deve entregar um pequeno relatório com seus testes e resultados.

Para a compactação, você deve executar o programa em linha de comando, tendo um arquivo “.txt” como entrada, e deve gerar um arquivo “.bin”. Já na descompactação, o arquivo “.bin” é dado na linha de comando, e você gera o arquivo, de mesmo nome, “.txt”.

Exemplo:

- `./compacta bibliadeGutenberg.txt`
- `./descompacta receitas.bin`

## Referências

- [1] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, *Introduction to Algorithms*, 4th. edition, McGraw Hill.
- [2] D. A. Huffman, “A Method for the Construction of Minimum Redundancy Codes”, *Proceedings of the IRE* 10, 40, 1098 – 1101 (1952).
- [3] D. A. Lelewer, D. S. Hirschberg, “Data Compression”, *ACM Computer Surveys* 19, 3, 261–296 (1987).