

1º Trabalho de Implementação de Algoritmos e Estruturas de Dados (2023-2):

Referência cruzada de palavras

Descrição do problema:

Uma referência cruzada é um processamento que permite criar referências entre partes de um texto. Nesse trabalho, seu programa deve ler um texto (ou um conjunto de palavras) e criar uma lista com as palavras que fazem parte dele, e para cada palavra diferente incluir a quantidade de vezes que cada palavra aparece. O programa deverá poder exibir as palavras ordenadas por ordem de frequência das palavras, e por ordem alfabética.

Organização: trabalho individual ou em duplas

Apresentação:

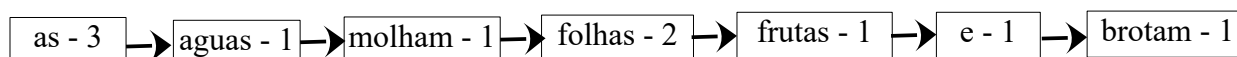
O trabalho será apresentado para a professora, em aula. Para a apresentação, se o trabalho for feito em dupla ambos os participantes devem estar presentes e participar da explicação sobre a implementação realizada.

Até o dia da apresentação, fazer *upload* do trabalho na tarefa correspondente no AVA do trabalho: código-fonte da implementação.

Representação:

1) uma lista encadeada de palavras (pode ser simples ou dupla), com as palavras inseridas por ordem de aparição ou por ordem alfabética:

“as aguas molham as folhas, as frutas, e brotam folhas.”



2) listas auxiliares para responder as consultas solicitadas, como:

- lista de letras (pode ser uma tabela *hash*), para acessar as palavras por ordem alfabética, ou iniciadas por uma letra (pode ser implementada como vetor ou como lista encadeada). Se for uma lista encadeada, pode conter apenas as letras que tem palavras associadas.
- lista de frequências: lista que mantém as referências das palavras por ordem de frequência

O texto a ser submetido à referência cruzada pode estar originalmente num arquivo (arquivo *texto.txt*), ou pode ser digitado pelo usuário (texto inteiro ou cada palavra).

Implementação:

Pode ser implementado em C, C++, C#, Java ou Python, desde que as referências entre os nodos sejam manipuladas explicitamente pelo programa (não pode usar classe, método ou função pronta da linguagem nem para o *hash* nem para listas encadeadas).

Cada nodo deverá conter as informações acima enumeradas (a palavra, nro de ocorrências, referências), mas pode conter outras informações necessárias para a realização das operações.

Pode utilizar header.

A interface pode ser bem simples (textual), o importante é que as operações funcionem.

Algumas operações básicas a serem oferecidas pelo programa:

- INSERE: permite a inserção de uma palavra do texto na lista. Palavras não podem ser duplicadas. Ao inserir uma palavra, seu número de ocorrências é inserido também.
- CONSULTA_PALAVRA: consulta uma palavra na lista, informando seu número de ocorrências.
- REMOVE: remove uma palavra da lista (com suas ocorrências)
- CONTA: retorna o número total de palavras (total de nodos da lista)
- CONTA_OCORRÊNCIAS: retorna o número total de ocorrências das palavras da lista (soma os contadores das palavras)
- EXIBE_PALAVRAS: exibe a lista das palavras (só as palavras, ou com os respectivos contadores, em **ordem alfabética (de A a Z)** ou em **ordem alfabética inversa (de Z a A)** de palavras (deve ter as duas opções).
- EXIBE_PALAVRAS_LETRA: exibe a lista das palavras da lista iniciadas por uma determinada letra (só as palavras ou com contadores), **em ordem alfabética (de A a Z)** ou **em ordem alfabética inversa (de Z a A)** de palavras (deve ter as duas opções).
- EXIBE_PALAVRAS_NRO_OCORRENCIAS: exibe a lista das palavras (as palavras com os respectivos contadores) **em ordem decrescente de número de ocorrências**, ou **em ordem crescente de número de ocorrências** (deve ter as duas opções)
- EXIBE_PALAVRAS_NRO: exibe a lista das palavras que possuem um determinado número de ocorrências informado.