

MAC 121 - Algoritmos e Estruturas de Dados I**Segundo semestre de 2020****Quarto Exercício-Programa – Entrega: 14 de dezembro****Índice remissivo de palavras em texto**

O objetivo deste exercício-programa é implementar a estrutura de lista ligada e usá-la para construir um índice remissivo de um texto.

Um índice remissivo é uma lista de palavras de um texto, em que cada palavra ocorre no máximo uma vez, e, depois da palavra, listamos em quais linhas do texto a palavra ocorre. Ainda, quando uma palavra ocorre mais de uma vez em uma mesma linha, isso também é mostrado.

Exemplo: Para o texto¹

```
Joao amava Teresa que amava Raimundo
que amava Maria que amava Joaquim que amava Lili
que nao amava ninguem.
Joao foi para os Estados Unidos, Teresa para o convento,
Raimundo morreu de desastre, Maria ficou para tia,
Joaquim suicidou-se e Lili casou com J. Pinto Fernandes
que nao tinha entrado na historia
```

O índice remissivo seria

```
amava: 1(2) 2(3) 3
casou: 6
com: 6
convento: 4
de: 5
desastre: 5
e: 6
...
```

Você pode considerar que o texto está em ASCII (sem acentos). Uma palavra é uma sequência de letras separada por branco ou algum sinal de pontuação (vírgula, ponto, etc), e palavras com letras maiúsculas e minúsculas são diferentes.

Sua tarefa, neste EP será ler o texto, linha por linha, e para cada palavra da linha, guardar na tabela as ocorrências destas palavras. No fim do texto, você deverá apresentar o índice remissivo, ou seja, as palavras em ordem alfabética com suas ocorrências.

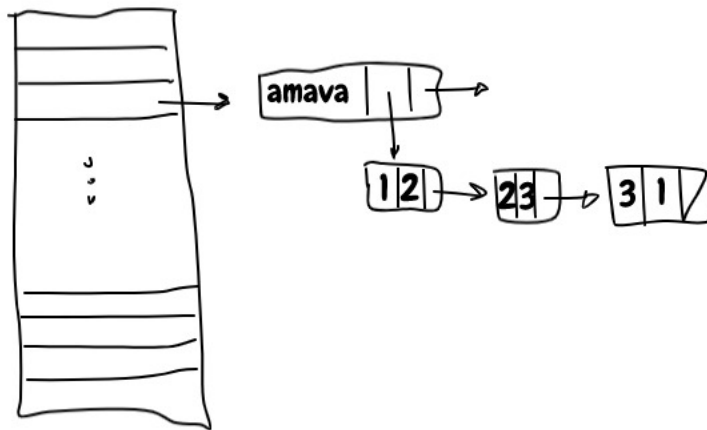
¹Quadrilha, Carlos Drummond de Andrade

A estrutura de dados

A estrutura de dados que vocês deverão usar é a seguinte. Teremos duas estruturas para guardar as palavras:

- uma tabela de hashing;
- um vetor com ponteiros para células da tabela de Hashing.

A tabela de hashing permite saber se a palavra já ocorreu ou não. O vetor poderá ser ordenado no final, e poderemos apresentar o índice remissivo em ordem alfabética, como mostrado no exemplo.



Temos, como pode ser visto na figura acima, para cada entrada na tabela de hashing, uma lista ligada das palavras que têm o mesmo valor na função de hashing. Para cada célula desta lista ligada armazenamos a palavra e uma lista ligada de ocorrências desta palavra no texto. Cada célula desta lista ligada de ocorrências armazena a linha em que ocorreu, e quantas vezes a palavra ocorreu naquela linha.

Para testar seu programa use textos grandes. Um site em que você pode buscar textos para seus testes é Projeto Gutenberg que guarda uma grande coleção de ebooks. A única entrada do seu programa será o nome do arquivo com o texto, e a saída deverá ser o índice remissivo, como mostrado no exemplo.