

MAC 323 – Algoritmos e Estruturas de Dados II

Primeiro semestre de 2023

Tabela de Símbolos – **Entrega: 15 de maio**

O objetivo deste exercício-programa é testar diferentes implementações para uma tabela de símbolos ordenada. As operações a serem implementadas são:

- void add (Key key, Item val);
- Item value (Key key);

Você deverá fazer vários testes com diferentes implementações para tabelas de símbolos. Você deverá, **pelo menos**, implementar a tabela de símbolos usando:

- vetor dinâmico ordenado (VO);
- árvore de busca binária (ABB);
- treaps (TR);
- árvores 2-3 (A23);
- árvores rubro-negras (ARN).

Para cada uma destas implementações vocês deverão fazer testes empíricos para analisar o tempo que leva para montar a Tabela de Símbolos usando a operação **add**, e consultas usando a operação **value**.

A tabela de símbolos vai conter como chave palavras de um texto lido. O campo **value** terá várias informações da palavra:

- número de ocorrências da palavra no texto;
- número de letras na palavra;
- número de vogais sem repetição na palavra;

Para fazer testes com textos grandes você pode usar os disponíveis no Projeto Gutenberg. Aí você encontra grandes livros escritos em várias línguas. Teste também com outros tipos de dados, como dicionários, códigos muito grandes e textos gerados aleatoriamente, no estilo *loren ipsum*.

Depois de ler o texto e montar uma tabela de símbolos com as palavras do texto, você vai responder consultas sobre estas palavras:

- quais as palavras mais frequentes no texto (F);
- dada uma palavra, quantas vezes ocorre no texto (O “palavra”);
- quais as palavras mais longas (L);
- quais as maiores palavras que não repetem letras (SR)
- quais as menores palavras com mais vogais sem repetição (VD).

Observe que se houver empates em alguma consulta queremos exibir **todas** as palavras que atingem o máximo (ou mínimo).

Você poderá pré-processar a Tabela de Símbolo a fim de responder as consultas acima de forma eficiente (tempo proporcional ao tamanho da resposta). Estruturas auxiliares utilizadas deverão ser eficientes e econômicas.

Neste exercício-programa você deverá fazer um relatório descrevendo seus testes, e o relatório é muito importante. Faça vários testes para mostrar o desempenho de cada implementação.

Descrição da entrada para os testes

Um caso teste do EP2 consiste de diversas linhas de entrada. A primeira linha indica a estrutura a ser testada (VO, ABB, TR, A23, ARN). Em seguida, na segunda temos um inteiro N com o número de palavras do texto. Então, temos N palavras que constituem o texto, que poderão ocupar diversas linhas do input. Logo após o texto, temos um inteiro Q , que representa o número de consultas a serem realizadas na tabela de símbolos. Em sequência, temos Q linhas, cada uma com uma consulta como definido acima. Você deve seguir o padrão para permitir que os monitores rodem testes com seus programas.

Caso teste 1: Exemplo abstrato	
1 E	// Estrutura a ser testada
2 N	// Número de Palavras do texto
3 Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis	
4 feugiat tellus vel dui varius mollis ac id lacus. Donec vitae	
5 mauris a tellus semper.	// Texto qualquer com N palavras
6 Q	// Número de consultas a serem feitas
7 Q linhas com uma das consultas descritas anteriormente	

Veja os exemplos da próxima seção para melhor entendimento.

Observação

Neste exercício programa o objetivo é a implementação e teste de estruturas de dados para tabelas de símbolos (árvores rubro-negras, ABBs, árvores 2-3, e treaps). Dessa forma, **não será permitido o uso de implementações prontas destas estruturas** disponíveis em bibliotecas das linguagens de programação.

Exemplos

Caso teste 2: Machado de Assis		
1	V0	
2	18	
3	ao verme que primeiro roeu as frias carnes do meu cadaver dedico	
4	como saudosa lembranca estas memorias postumas	
5	3	
6	O dedico	
7	VD	
8	L	
		// Fim da Entrada
		// Saída
9	1	// A palavra dedico ocorre 1 vez
10	memorias	// A palavra mais curta com mais vogais distintas (4)
11	lembranca	// A palavra mais longa

Caso teste 3: Carlos Drumond de Andrade

```
1 A23
2 53
3 João amava Teresa que amava Raimundo
4 que amava Maria que amava Joaquim que amava Lili
5 que não amava ninguém.
6 João foi para os Estados Unidos, Teresa para o convento,
7 Raimundo morreu de desastre, Maria ficou para tia,
8 Joaquim suicidou-se e Lili casou com J. Pinto Fernandes
9 que não tinha entrado na história
10 3
11 F
12 O Maria
13 SR
                                     // Fim da Entrada
                                     // Saída
14 amava                          // 'amava' é a palavra mais frequente
15 2                               // 'Maria' ocorre 2 vezes
16 Raimundo                       // 'Raimundo' é a maior palavra que não repete letras
```

Caso teste 4: Sequoia

```
1 ARN
2 51
3 Mother of the Forest was a giant sequoia in the Sierra Nevada that
4 was stripped of its bark and used as an exhibition tree. Discovered in
5 1852 by Augustus T. Dowd, it was the second largest tree in Calaveras
6 Grove, standing at 328 feet tall with a circumference of 93 feet.
7 2
8 VD
9 F
                                     // Fim da Entrada
                                     // Saída
10 sequoia                        // menor palavra com 5 vogais sem repetição
11 the was                        // palavras mais frequentes
```
