



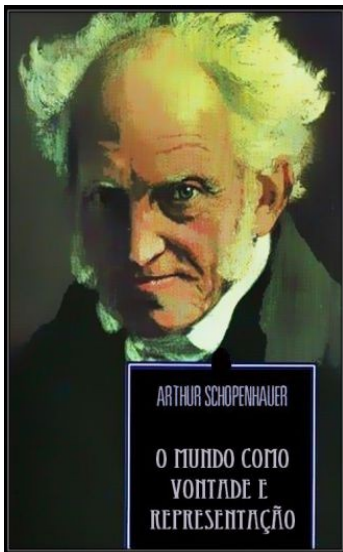
Estrutura, Pesquisa e Ordenação de Dados (TSI32B)

Professor: Ivan Luiz Salvadori

Trabalho 3

Atenção: Esta atividade será avaliada e corresponderá nota de acordo com o plano de ensino da disciplina.

“



O Mundo como Vontade e Representação é a obra fundamental de Schopenhauer. Admitindo a distinção kantiana entre fenômeno e númeno, o autor considera que o mundo é ao mesmo tempo vontade e representação. A vontade é a coisa em si, que se esconde por trás do fenômeno, que é a representação, ilusão, aparência; ela existe em todos os seres da natureza, desde os mais ínfimos até ao homem, no qual atinge o seu «mais alto grau de objetivação». O conhecimento, em geral, é um meio de preservar o indivíduo e a espécie, e está, portanto, ao serviço da vontade. Mas pode ultrapassar essa sujeição e «existir para si mesmo como um claro espelho do mundo», como acontece na arte. A liberdade total, no entanto, só é conseguida quando a vontade se reconhece a si mesma como dor e se resigna. Assim, a vontade, que é «esforço constante, sem objetivo e sem descanso», é abolida, e, com ela, todo o mundo: é o nada - **«Sem vontade não há representação nem mundo.»**

Desenvolva um programa em Java capaz de ler um arquivo de texto (.txt) que representa o livro supracitado e fazer a contagem de palavras com maior ocorrência no texto. As palavras devem ser ordenadamente organizadas de forma decrescente, apresentando as palavras de maior ocorrência no início da listagem. Palavras escritas de forma maiúscula ou minúscula devem ser consideradas uma mesma palavra. Ou seja, não deve haver diferenciação quanto ao uso de maiúscula ou minúscula. Além disso, para que o programa seja mais efetivo na contagem de palavras, deve-se ignorar os seguintes caracteres:

- “,” (vírgula)
- “.” (ponto)
- “-” (traço)
- “:” (dois pontos)
- “;” (ponto e vírgula)
- “?” (interrogação)
- “\” (aspas)

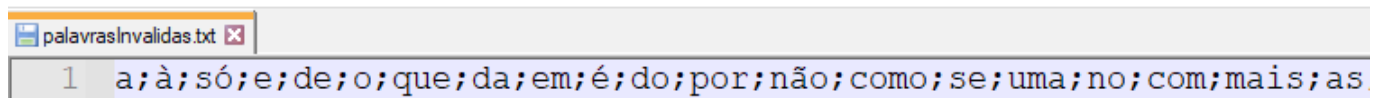
O programa deverá ser implementado com o uso da estrutura de dados correspondente a tabela de espalhamento. O programa deverá apresentar o número de palavras únicas, seguido da listagem das 20 palavras de maior ocorrência, conforme exemplificado pela imagem a seguir:

Saída esperada:

```
Contando palavras com Hashmap
Total de palavras únicas: 17740
As palavras com maior ocorrência são:
=====
1) vontade (1264 ocorrências)
2) conhecimento (945 ocorrências)
3) apenas (831 ocorrências)
4) mesmo (813 ocorrências)
5) também (765 ocorrências)
6) assim (716 ocorrências)
7) são (714 ocorrências)
8) razão (706 ocorrências)
9) isso (664 ocorrências)
10) qual (641 ocorrências)
11) pode (623 ocorrências)
12) mundo (593 ocorrências)
13) vida (568 ocorrências)
14) sem (555 ocorrências)
15) tempo (544 ocorrências)
16) sobre (527 ocorrências)
17) sempre (519 ocorrências)
18) todo (498 ocorrências)
19) toda (488 ocorrências)
20) portanto (476 ocorrências)
Tempo de processamento: 0.18 segundos
```

O Programa deverá desconsiderar as ocorrências de palavras especiais, como artigos e palavras de ligação, como, por exemplo, o, a, as, do, das, etc...

As palavras que devem ser ignoradas estão no arquivo **palavrasInvalidas.txt**. A imagem a seguir ilustra algumas palavras do conteúdo do arquivo:



### Requisitos específicos:

- O programa deve processar o arquivo TXT disponibilizado para a atividade.
- O caminho do arquivo TXT deve ser o mesmo do arquivo JAR executável.
- O arquivo TXT deve ser nomeado como “livro.txt”.
- O programa deve seguir os exemplos da versão do professor.
- O tempo de processamento é opcional e não será considerado para fins de avaliação.

**Instruções gerais:**

- Esta atividade pode ser desenvolvida em grupo de até 4 alunos.
- A entrega deve ser feita no Moodle da disciplina através da submissão dos seguintes arquivos:
  - Arquivo de texto com os nomes, e-mails e RA dos autores.
  - Código-fonte do projeto (pasta compactada ZIP)
  - Arquivo JAR executável
- A entrega deve ser feita por apenas um membro do grupo.
- A entrega deve ser realizada através do Moodle até o dia 13/12/2022 23h59.

Toledo – PR, 29/11/2022