

**Etapa 1 – Projeto Final (Escrita/leitura de arquivos e contagem de termos em um documento)**

1 – Escreva um programa, na linguagem de sua escolha, para executar as seguintes tarefas:

- a) Dado um arquivo de texto qualquer como entrada, calcular o número de ocorrências de cada termo no arquivo. Observação: somente palavras com mais de três letras (>3) serão consideradas como termos de indexação. Desconsidere também as pontuações.
- b) Criar um arquivo texto de saída com a lista de termos em pares (termo, frequência de ocorrência do termo) e em ordem decrescente de ocorrências.

```
brasil 3  
liberdade 2  
podeis 1  
contente 1
```

2 - Com o algoritmo implementado, faça o seguinte:

- a) Verifique o funcionamento do algoritmo desenvolvido usando como caso de teste um (1) arquivo de texto em que o conteúdo seja a primeira estrofe do hino nacional brasileiro.

**Hino da Independência**

1

Já podeis da Pátria filhos  
Ver contente a Mãe gentil;  
Já raiou a Liberdade  
No Horizonte do Brasil  
Já raiou a Liberdade  
Já raiou a Liberdade  
No Horizonte do Brasil

**Refrão**

Brava Gente Brasileira  
Longe vá, temor servil;  
Ou ficar a Pátria livre,  
Ou morrer pelo Brasil.  
Ou ficar a Pátria livre,  
Ou morrer pelo Brasil.

2

Os grilhões que nos forjava  
Da perfídia astuto ardil,  
Houve Mão mais poderosa,  
Zombou deles o Brasil.  
Houve Mão mais poderosa  
Houve Mão mais poderosa  
Zombou deles o Brasil.

- b) Crie uma pequena base de dados de documentos usando os três principais hinos brasileiros (Hino da Independência, Hino à Bandeira e Hino Nacional). Cada documento será formado

por uma estrofe de cada hino, com o refrão aparecendo somente uma vez. Por exemplo, o documento 1 será a primeira estrofe do Hino da Independência, o documento 2 será o refrão, o documento 3 será a segunda estrofe e assim sucessivamente para cada um dos hinos. *Importante:* ficará a critério do aluno definir se cada documento será um arquivo de texto ou se todos os documentos estarão dentro de um mesmo arquivo de texto.

- c) Processe cada um dos documentos criados na letra 2-b). A partir do arquivo com os pares gerados (termo, frequência), trace um gráfico, onde o eixo X é a ordem da palavra na lista classificada por frequência e o eixo Y é a frequência da palavra. Explique o comportamento da curva gerada.