

# Exercício Programa 2 - Busca de palavras

Tópicos de Programação

25 de janeiro de 2010

## 1 Descrição do problema

*Esse EP foi utilizado pela disciplina MAC 122 do curso de Bacharelado em Ciência da Computação em 2008 (<http://www.ime.usp.br/~am/122/eps/EP3.html>). Algumas adaptações foram feitas para viabilizar a sua aplicação no curso de verão Tópicos de Programação.*

Suponha que você está começando uma empresa de busca. Seu primeiro software vai receber sites (arquivos de texto) e fazer uma pré-análise, montando um histograma que conte quantas palavras ocorrem exatamente  $k$  vezes no site.

Uma palavra é definida baseada na função isalpha: é uma sequência maximal de caracteres para os quais essa função retorna 1. Para contagem de ocorrências, não se quer diferenciar entre maiúsculas e minúsculas; assim, "palavra", "Palavra" e "PALAVRA" são ocorrências da mesma palavra.

Para mais informação sobre a função isalpha e outras que serão úteis nesse EP, acesse: <http://www.cplusplus.com/reference/clibrary/>

## 2 A Tarefa

### 2.1 Entrada

A entrada será um texto, que deve ser lido do stdin, tendo seu final definido pelo caractere EOF (End-of-file). O arquivo terá no máximo 10.000.000 de caracteres.

### 2.2 Saída

A saída deve ser uma tabela impressa na stdout, tendo para cada  $k > 0$ , quantas palavras ocorrem exatamente  $k$  vezes, desde que esse número seja

positivo.

## 2.3 Restrições

Seu programa não deve usar funções de biblioteca para ordenação, você deve implementá-las quando necessário. O resto da biblioteca `libc` está liberado, incluindo funções sobre strings.

Seu programa deverá ser eficiente.

## 2.4 Exemplo de entrada e saída

Entrada:

```
De tudo ao meu amor serei atento
Antes, e com tal zelo, e sempre, e tanto
Que mesmo em face do maior encanto
Dele se encante mais meu pensamento.
```

```
Quero vivê-lo em cada vão momento
E em seu louvor hei de espalhar meu canto
E rir meu riso e derramar meu pranto
Ao seu pesar ou seu contentamento
```

```
E assim, quando mais tarde me procure
Quem sabe a morte, angústia de quem vive
Quem sabe a solidão, fim de quem ama
```

```
Eu possa me dizer do amor (que tive):
Que não seja imortal, posto que é chama
Mas que seja infinito enquanto dure.
```

Saida:

```
k ---> count
1 ---> 57
2 ---> 8
3 ---> 3
4 ---> 2
5 ---> 2
7 ---> 1
```

## 2.5 Bonus

As seguintes ações serão contadas como bonus:

- Escrever o programa com listas ligadas;
- Ter a opção, na linha de comando, de imprimir as palavras do texto em ordem lexicografica

## 3 Exemplos

Alguns exemplos com suas respostas podem ser encontrados em: <http://www.linux.ime.usp.br/~shakavp/Testes.zip>. O tempo de processamento para comparação encontra-se abaixo. Os testes foram em um processador *AMD Sempron 2800+*.

Teste	Tempo gasto
exemplo1.in	0.136s
exemplo2.in	0.687s
exemplo3.in	1.246s
exemplo4.in	2.121s
exemplo5.in	2.789s
exemplo6.in	2.138s
exemplo7.in	2.592s