



Universidade de São Paulo – ICMC  
Bacharelado em Ciência da Computação  
SCC0221 – Introdução à Ciência da Computação I  
Prof. Rodrigo Fernandes de Mello – [mello@icmc.usp.br](mailto:mello@icmc.usp.br)  
Monitores: Victor Forbes – [victor.forbes@usp.br](mailto:victor.forbes@usp.br),  
Yule Vaz – [yule.vaz@usp.br](mailto:yule.vaz@usp.br)

## Exercício: Autocomplemento

### 1 Descrição

As interfaces modernas de desenvolvimento de programas e de busca de palavras-chave, bem como aplicativos de digitação de dispositivos móveis, possuem o sistema de autocomplemento para auxiliar o usuário a escrever uma mensagem ou código com maior rapidez. Esse sistema consiste em apresentar ao usuário, durante a escrita de uma palavra, uma série de prováveis candidatos para o termo em questão. Isso pode ser feito armazenando-se um dicionário com as frequências com que cada palavra foi escrita no código-fonte, aplicativo ou texto e, a partir dessa informação, calcular as probabilidades de cada palavra ocorrer a medida que o usuário digita um caractere.

Por exemplo, dada uma entrada “ca”, calcula-se

$$P(\dot{x}_{ca}) = \frac{\dot{x}_{ca}}{\sum_{x_{ca} \in X_{ca}} x_{ca}} \quad (1)$$

em que  $X_{ca}$  é o conjunto de frequências de todas as palavras que começam com “ca”,  $x_{ca}$  é a frequência de uma determinada palavra que pertence a esse conjunto e, finalmente,  $\dot{x}_{ca}$  é o termo sobre o qual se quer calcular a probabilidade. Com isso é possível organizar os termos a partir da chance que esse tem de ocorrer e apresentar ao usuário.

### 2 Objetivo

**Você deverá implementar um sistema de autocomplemento que apresente as cinco palavras mais prováveis de ocorrer.**

### 3 Proposta

Seu programa deverá ler como entrada, tal como apresentado abaixo, dois nomes de arquivos: um contendo o dicionário e as frequências de seus termos e o outro contendo a “query”.

```
dic1.dat  
query1.dat
```

O arquivo de dicionário, formatado da seguinte maneira,

```
ABACATEIRO 95  
ABACAXI 52  
ABACO 33  
ABADE 73  
ABADESSA 73
```

ABADIA 85  
ABAF 93  
ABAFADO 49  
ABAFAMENTO 23  
ABAFAR 21  
ABAFO 90  
ABAIANADO 95

será utilizado para se calcular as probabilidades dos termos. A partir da “query”, disposta com uma letra em cada linha do arquivo tal como apresentado abaixo,

O  
T  
I  
T  
E

você deve calcular, para cada sequência de letra  $\{\{c_1\}, \{c_1, c_2\}, \{c_1, c_2, c_3\}, \dots\}$ , as probabilidades dos termos que se iniciam com tais sequências ocorrerem e imprimir os cinco que apresentam maiores chances. Por exemplo, nesse caso deve ser calculado, primeiramente, a probabilidade de todas as palavras que começam com 'O' e selecionadas as cinco mais prováveis. Logo após, efetua-se esse procedimento novamente sobre todas as palavras que começam com 'OT', depois sobre as que começam com 'OTI' e assim por diante. Note que, eventualmente, menos de cinco termos podem ser selecionados, nesse caso imprima todos. Além disso, o seu programa deve parar ou quando sobrar apenas uma palavra candidata ou quando toda a “query” for avaliada.

Para cada sequência da “query”, deve-se imprimir as cinco (ou menos) palavras candidatas em cada linha tal como apresentado abaixo:

### Entrada

C  
O  
M  
P  
E  
N  
S  
A  
R

### Saída

CADASTRAR CAIXEIRO-VIAJANTE CAMARONES CAMINHAO CAMUFLAGEM  
COCHEIRA COLONIZADO COMANDOS COMPETIR COMPILAR  
COMANDOS COMPETIR COMPILAR COMPUTADO COMA  
COMPETIR COMPILAR COMPUTADO COMPENSADOR COMPLICAR  
COMPETIR COMPENSADOR COMPENDIO COMPENSACAO COMPETIDOR  
COMPENSADOR COMPENDIO COMPENSACAO COMPENSA COMPENSATORIO  
COMPENSADOR COMPENSACAO COMPENSA COMPENSATORIO COMPENSAR  
COMPENSADOR COMPENSACAO COMPENSA COMPENSATORIO COMPENSAR  
COMPENSAR

## 4 Observação

- Considere apenas caracteres maiúsculo;
- Podem ser utilizadas expressões regulares (regex) para encontrar as palavras que se iniciam com a cadeia de caracteres desejada.