

Especificação do Trabalho 1 – Índice de palavras e sugestão ortográfica

Cada aluno deve desenvolver um programa na linguagem C que receba dois nomes de arquivo como parâmetro. O primeiro arquivo terá um dicionário de palavras, e o segundo arquivo terá um texto qualquer. Todas as palavras contidas no texto e no dicionário possuem apenas os caracteres minúsculos [a-z] e maiúsculos [A-Z] sem acento. O texto ainda pode conter espaços, traços e pontuações, que devem ser considerados separadores (a sentença “pode-se” deve ser considerada duas palavras: “pode” e “se”). O exemplo abaixo mostra como o programa será testado.

Ex:

```
./a.out dicionario.txt texto.txt
```

O trabalho deverá ser enviado por e-mail (mpamplona2.ufba@gmail.com), com o assunto “TRABALHO 1 - EDAIL.2014.2”. Este e-mail deve conter um único anexo de nome “trabalho1.tgz”. Dentro deste arquivo compactado deve haver uma pasta com o nome “trabalho1” que contém o código fonte e um arquivo Makefile que gere um executável chamado “a.out” no diretório “trabalho1”. Cada aluno receberá um arquivo de identificação que deve ser incluso inalterado no diretório “trabalho1”.

Para cada palavra do dicionário existente no texto (esta análise não é case-sensitive), o seu programa deve imprimir em quais de linhas no texto ela aparece (a mesma linha não deve ser listada mais de uma vez). Deve-se utilizar o seguinte padrão para a impressão:

word x

word x, y

word x, y, z

onde **word** é a palavra contida no dicionário (que deve ser impressa exatamente como aparece no dicionário) e **x**, **y**, e **z** são as linhas em que a palavra **word** aparece. No exemplo acima, temos casos em que a palavra ocorre em apenas uma linha, em duas linhas, ou em três linhas. No entanto, não há um limite para a quantidade de ocorrências. Se mais de uma palavra do dicionário aparecer no texto, as palavras devem ser impressas em ordem alfabética. Cada linha da saída representa uma palavra do dicionário e suas respectivas linhas. Palavras que não aparecem no texto não devem ser impressas.

Ex:

```
word 1
```

```
zebra 1, 2
```

Em seguida, para cada palavra do texto que não está no dicionário, o seu programa deve sugerir a palavra mais próxima, usando o padrão abaixo:

world word

sendo **world** a palavra no texto não contida no dicionário (que deve ser impressa exatamente como aparece no texto) e **word** a palavra mais próxima no dicionário (que deve ser impressa exatamente como aparece no dicionário).

A distância entre duas palavras é dada pela quantidade mínima de modificações em uma das palavras para que esta fique igual a outra. São consideradas modificações a troca de uma letra por outra letra, a remoção de uma letra ou a inserção de uma letra. Se duas palavras no dicionário estiverem igualmente

distantes da palavra do texto, deve-se imprimir a alfabeticamente menor, como no exemplo abaixo:

Dicionário	Texto	Saída
abacate abacaxi	abc	abc abacate

Segue abaixo um outro exemplo de entrada e saída esperada:

Dicionário	Texto	Saída
a arremessada bola chutada foi seria	A bola foi chutada! A bola SERIA ARREMETIDA.	a 1, 2 bola 1, 2 chutada 1 foi 1 seria 2 ARREMETIDA arremessada

O dicionário não necessariamente será fornecido de forma ordenada.