

InterFatecs. Você quis dizer InterFatecs?

Arquivo fonte: sugestao.{c | cpp | java}

Autor: Leandro Luque (Fatec Mogi das Cruzes)

Cezinho anda enfrentando problemas com a sua língua materna e também com a sua professora de português. Depois de ler em seus trabalhos frases como “Inglês não é meu forte, mas no português eu destróio” e “Eu odío isso”, a professora de Cezinho ficou muito preocupada e resolveu investigar o que estava acontecendo. Ela descobriu que o menino usava e abusava das correções automáticas de alguns programas de computador e não se preocupava em escrever corretamente. O sistema de correção automática preferido dele é o do serviço de busca *web* que utiliza. Ele pesquisa “Sésar soro acaba” e o sistema corrige: “Você quis dizer César Sorocaba?”, e por aí vai.

Procurando encaminhar o menino, a professora sugeriu que ele estudasse alguns materiais e também que fizesse um exercício relacionado à funcionalidade que ele gosta do sistema de busca. Para o exercício, a professora entregou a Cezinho duas listas, uma com palavras grafadas corretamente e outra com palavras grafadas incorretamente. A professora então escolhia uma palavra da lista de palavras com erro de grafia e pedia para Cezinho indicar qual palavra da lista de palavras grafadas corretamente ele iria sugerir. Ele tinha que escrever: “Você quis dizer <palavra sugerida da lista de palavras com grafia correta>”.

Para encontrar uma sugestão de palavra, a professora o ensinou a calcular uma métrica de distância entre dois textos chamada de **Distância de Levenshtein**. Utilizando esta métrica, ele deveria então calcular a distância da palavra grafada incorretamente em relação a todas as palavras grafadas corretamente e escolher como sugestão aquela que tivesse a menor distância. Caso houvesse empate de distância entre duas ou mais palavras, a professora sugeriu que ele escolhesse a primeira destas palavras considerando a ordem alfabética.

A **Distância de Levenshtein** corresponde ao menor número de alterações (inserções, substituições ou deleções) que devem ser feitas em um texto x para se obter outro texto y . Uma forma de calculá-la (existem várias outras) é por meio de uma matriz como a apresentada a seguir, criada para calcular a distância entre as palavras VAZA e CASA.

		C	A	S	A
	0	1	2	3	4
V	1	1	2	3	4
A	2	2	2	1	2
Z	3	3	3	2	2
A	4	4	4	3	2

Para criar esta matriz, deve-se inicialmente preencher as células da primeira linha (1, j) com o valor do seu índice j e as células da primeira coluna (i , 1) com o valor do seu índice i . Em seguida, deve-se preencher o restante das células (i , j) da matriz baseando-se na seguinte regra

$$\text{Matriz}[i][j] = \begin{cases} \text{Matriz}[i-1][j-1], & \text{se } x(i) = y(j) \\ \text{Mínimo}(\text{Matriz}[i-1, j] + 1, \text{Matriz}[i, j-1] + 1, \text{Matriz}[i-1, j-1] + 1), & \text{caso contrário} \end{cases}$$

onde $x(i)$ representa o i -ésimo caractere do texto x e $y(j)$ o j -ésimo caractere do texto y .

Após calcular o valor para todas as células, o valor encontrado no canto inferior direito da tabela corresponde ao menor número de alterações que devem ser feitas em x para se chegar à y . No caso de VAZA e CASA, o menor número de alterações é igual a dois (2).

Apesar de conhecer muito bem esta versão do algoritmo da **Distância de Levenshtein**, Cezinho é muito preguiçoso e resolveu enganar sua professora. Ele pediu para o seu irmão, Javinha, um “cabra” muito esperto, para desenvolver um programa que, dada uma lista de palavras grafadas corretamente e uma palavra com grafia incorreta, faz uma sugestão conforme a regra definida pela professora de Cezinho. Como Javinha está empregado, ele pediu para você ajudá-lo na escrita deste programa.

Entrada

Inicialmente um valor N é informado, $1 \leq N \leq 100$, indicando a quantidade de palavras com grafia correta que serão informadas. Seguem-se N linhas, cada uma com uma palavra composta por até 20 caracteres maiúsculos sem espaço. Em seguida, é informado um valor I , $1 \leq I \leq 1000$, indicando o número de palavras com grafia incorreta que deverão ser verificadas. Seguem-se I linhas, cada uma com uma palavra composta por até 20 caracteres maiúsculos sem espaço.

Saída

Para cada palavra com grafia incorreta informada, imprima uma linha com a frase “voce quis dizer” (sem acento e em minúsculas), a palavra da lista de palavras grafadas corretamente que tem a menor distância de Levenshtein em relação à palavra com grafia incorreta e a distância de Levenshtein entre esta palavra e a palavra com grafia incorreta. Estas informações devem ser separadas por um espaço em branco. Caso, para uma determinada palavra com grafia incorreta, haja empate de distância entre duas ou mais palavras da lista de palavras grafadas corretamente, utilize a regra especificada para a professora para selecionar uma das palavras.

Exemplos

Entrada:	Saída:
6	voce quis dizer CRUZEIRO 1
MOGI	voce quis dizer MOGI 1
SOROCABA	voce quis dizer SOROCABA 2
SAOJOSE	
CRUZEIRO	
PINDA	
PIRAMBIRA	
3	
CRUSEIRO	
LOGI	
ROSOCABA	