MC102 - Algoritmos e Programação de Computadores
Turmas QRSTWY
Instituto de Computação - Unicamp
Professores: *Hélio Pedrini e Zanoni Dias*
Monitores: *Andre Rodrigues Oliveira, Gustavo Rodrigues Galvão, Javier Alvaro Vargas Muñoz e Thierry Pinheiro Moreira*

Lab 05b - Distância de Tchonsky

Prazo de entrega: 22/04/2015 às 13h59m59s
Peso: 4

Um problema bastante estudado em Ciência da Computação é o de encontrar um padrão de caracteres, geralmente definido por uma *string* curta, em uma grande sequência de caracteres ou em um texto, que pode ser visto como uma *string* longa. Tal problema é encontrado em várias aplicações, tais como busca de uma palavra em um texto, detecção de plágio, correção ortográfica, análise de sequências de DNA e RNA, entre outras. No entanto, nem sempre estamos interessados em encontrar apenas o padrão exato, mas também padrões similares a ele. Nesses casos, um método muito comum é o seguinte. Primeiro, define-se uma métrica para *strings*, isto é, define-se uma função que, dadas duas *strings*, devolve um número que indica qual a distância (ou similaridade) entre essas *strings*. Depois, encontram-se todos os padrões de caracteres que possuem uma distância menor do que um certo valor, em relação ao padrão original de busca.

Uma das métricas mais simples para *strings* é a chamada Distância de Hamming. Dadas duas *strings* de mesmo tamanho, a distância de Hamming entre elas é igual ao número de posições nas quais elas diferem entre si. Dito de outra forma, a distância de Hamming entre elas é igual ao menor número de substituições necessárias para transformar uma *string* na outra. Ela foi definida pela primeira vez por Richard Wesley Hamming (por isso o nome da métrica) e foi utilizada na detecção e correção de erros em transmissões de sinais de telecomunicação, que podem ser vistos como sequências de *bits*.

Um dos problemas da distância de Hamming é que ela só faz sentido quando aplicada sobre *strings* de mesmo tamanho. Por essa razão, outra métrica bastante popular e que generaliza a distância de Hamming é a chamada Distância de Levenshtein, também conhecida como distância de edição. Dadas duas *strings*, a distância de Levenshtein entre elas é igual ao número mínimo de operações necessárias para transformar uma *string* na outra. As operações possíveis são inserção, remoção ou substituição de um caractere. Esta métrica foi definida pela primeira vez por Vladimir Iosifovich Levenshtein.

Outras métricas existem e, de modo geral, uma métrica visa cobrir as fraquezas das outras em determinados contextos. Com o objetivo de eliminar a necessidade de se criar diversas métricas, Noel Tchonsky, um notável linguista, teve a seguinte ideia: criar uma língua que facilitasse a criação de uma métrica única para as palavras (ou *strings*) que as constituem. Basicamente, ele criou uma língua com a seguinte propriedade: todas as palavras (ou *strings*) que possuem exatamente as mesmas letras (ou caracteres) são equivalentes. Por exemplo, as palavras "aatt", "atat", "atta", "taat", "atat", "tata" e "ttaa" são todas equivalentes e querem dizer "Pindamonhangaba" em Português. Por outro lado, as palavras "ta" e "at" são equivalentes entre si, mas não são equivalentes às citadas anteriormente, pois aquelas sete possuem dois "t" e dois "a", enquanto estas duas possuem apenas uma letra de cada. Aliás, "ta" e "at" querem dizer "atenção" em Português.

Considerando essa nova língua, Tchonsky propôs a criação de uma nova métrica, chamada Distância

de Tchonsky. Dadas duas *strings* `A` e `B`, seja `C(A,B)` o número de caracteres em comum entre elas, considerando que: (i) os caracteres não são "sensíveis ao caso", ou seja, letras maiúsculas e minúsculas são consideradas equivalentes, e (ii) os caracteres repetidos devem ser contados repetidamente. Por exemplo, supondo que `A = AatT` e `B = tTaa`, nós temos que `C(A,B) = 4`. Por outro lado, supondo que `A = aAtt` e `B = DaT`, nós temos que `C(A,B) = 2`. A distância de Tchonsky entre `A` e `B` é igual a `d(A,B) = |A| + |B| - 2 × C(A,B)`, sendo que `|A|` denota o tamanho de `A` e `|B|` denota o tamanho de `B`.

Além de linguista, Noel Tchonsky é músico e atualmente encontra-se em turnê internacional com seu quarteto de cordas chamado "É o Tchonsky". Por essa razão, ele lhe incumbiu a tarefa de implementar um programa em C que, dadas duas *strings*, calcula a distância de Tchonsky entre elas.

Entrada

A entrada é constituída de duas linhas:

- A primeira linha contém a *string* `A`, `1 ≤ |A| ≤ 100`;
- A segunda linha contém a *string* `B`, `1 ≤ |B| ≤ 100`.

Considere que as *strings* são constituídas apenas por letras.

Saída

Seu programa deve imprimir a resposta no seguinte formato: "`Distancia de Tchonsky = X`", onde `X` é igual à distância de Tchonsky entre `A` e `B`.

Exemplos

#	Entrada	Saída
1	`AatT` `tTaa`	`Distancia de Tchonsky = 0`
2	`aAtt` `DaT`	`Distancia de Tchonsky = 3`
3	`A` `z`	`Distancia de Tchonsky = 2`
4	`HipOPOtomonStroSESquipEDAliOFObIA` `NaMAStE`	`Distancia de Tchonsky = 26`
5	`ChA` `pnEUmouLTraMicrolAscOuscopicosSilicOVulcAnoconIOTiCo`	`Distancia de Tchonsky = 51`