

# DNA Storage?

Por Maratona de Programação 2003, IME-USP  Brasil**Timelimit: 1**

A Universidade Charles, situada em Praga, a exemplo de diversas outras universidades de renome ao redor do mundo, instituiu recentemente um programa interdepartamental de pós-graduação na área de biologia computacional. Integrante do corpo docente, Ms. Dolejšková está atualmente interessada no problema das árvores filogenéticas, e trabalhando, portanto, com  $n$  cadeias de DNA. Para simplificar o trabalho, Ms. Dolejšková resolveu trabalhar apenas com cadeias gênicas de comprimento  $m$  (isto é, todas as cadeias possuem exatamente  $m$  bases nitrogenadas).

Um subproblema interessante envolve o armazenamento das  $n$  cadeias em disco. Até o momento, Ms. Dolejšková está utilizando um esquema ingênuo que requer  $n \times m$  caracteres, além dos delimitadores. Isto é, todas as sequências são gravadas dentro de um arquivo texto, sequencialmente. Mr. Chuchle, um colega de departamento e especialista em técnicas de armazenamento, sugeriu uma alternativa que pode ser mais econômica.

Segundo Mr. Chuchle, é possível armazenar uma cadeia juntamente com informações que permitam transformá-la em outras. Mais especificamente, considere duas cadeias de DNA  $D_1 = \text{ACTA}$  e  $D_2 = \text{AGTC}$ , onde A, C, G, T representam as bases nitrogenadas adenina, citosina, guanina e timina, nesta ordem. Observe que é possível transformar  $D_1$  em  $D_2$  trocando-se as bases nitrogenadas C e A das posições 2 e 4 de  $D_1$  para G e C, respectivamente. Considere agora uma terceira cadeia  $D_3 = \text{CGTC}$ . É necessária apenas uma modificação para transformar  $D_2$  em  $D_3$  e são necessárias três modificações para transformar  $D_1$  em  $D_3$ . Logo, é vantajoso permitir a transitividade das modificações entre as cadeias.

Ms. Dolejšková observou rapidamente que, se as cadeias envolvidas forem muito diferentes entre si, este esquema de armazenamento alternativo não oferece ganhos. Assim, em vez de adotá-lo prontamente, ela solicitou a você que construa um programa que recebe as  $n$  cadeias, e determina o número mínimo de transformações que devem ser gravadas (além de uma cadeia) para que seja possível, no futuro, obter-se novamente as  $n$  cadeias originais. Baseado no resultado fornecido por seu programa, Ms. Dolejšková vai decidir qual dos esquemas deve utilizar em cada instância de dados que tiver.

## Entrada

Seu programa deve estar preparado para trabalhar com diversas instâncias. Cada instância tem a estrutura que segue. Na primeira linha são fornecidos dois inteiros  $n$  e  $m$  ( $0 \leq n \leq 100$  e  $1 \leq m \leq 1000$ ) que representam, nesta ordem, o número de cadeias de DNA e o comprimento delas. Nas próximas  $n$  linhas são fornecidas as  $n$  cadeias, uma por linha, sem espaços adicionais. Cada cadeia é uma sequência de caracteres tomada sobre o alfabeto  $\Sigma = \{A, C, G, T\}$ . Um valor  $n = 0$  indica o final das instâncias e não deve ser processado.

## Saída

Para cada instância solucionada, você deverá imprimir um identificador Instancia  $h$  em que  $h$  é um número inteiro, sequencial e crescente a partir de 1. Na linha seguinte, você deve imprimir o número mínimo de transformações que devem ser gravadas para esta instância. Uma linha em branco deve ser impressa após cada instância

### Exemplo de Entrada

### Exemplo de Saída

3 4	Exemplo de Entrada	Instancia 1	Exemplo de Saída
ACTA		3	
AGTC			
CGTC		Instancia 2	
4 5		4	
AAAAA			
ATATA			
ATCTA			
AACAA			
0 0			

VII Maratona de Programação IME-USP 2003.