

# DNA突變

葛麗絲是在新加坡一家生物資訊公司工作的生物學家。她工作的一部份是分析各種生物的 DNA (脫氧核糖核酸) 序列。一個 DNA 序列是由 "A", "T", 和 "C" 字元所組成的字串。請注意, 在這個任務裏, DNA 序列**不會包含 "G" 字元**。

我們將突變定義為對 DNA 序列進行一次如下操作: 將序列中的兩個元素交換位置。例如, 透過一次突變操作我們可以將 "ACTA" 轉換成 "AATC", 其中被交換的是 "A" 和 "C" 這兩個用粗黑體標示的字元。

我們定義兩個 DNA 序列的突變距離為: 將一個序列轉換成另一個所需要的最小突變操作數目。如果不可能用突變操作來完成這個轉換, 他們的突變距離為  $-1$ 。

葛麗絲正在分析兩個 DNA 序列  $a$  和  $b$ , 兩個序列都包含  $n$  個元素, 元素的索引值 (indices) 為  $0$  到  $n - 1$ 。你的工作是幫忙葛麗絲回答  $q$  個詢問, 格式如下: 子字串  $a[x..y]$  和  $b[x..y]$  的突變距離是多少? 這裏一個 DNA 序列  $s$  的子字串  $s[x..y]$  指的是由  $s$  中索引值為  $x$  到  $y$  的連續字元所構成的序列。換句話說,  $s[x..y]$  就是序列  $s[x]s[x + 1] \dots s[y]$ 。

## 實作細節

你必須實作下列程序:

```
void init(string a, string b)
```

- $a, b$ : 兩個長度為  $n$  的字串, 描述兩個分析中的 DNA 序列。
- 在呼叫以下 `get_distance` 前這個程序會被呼叫剛好一次。

```
int get_distance(int x, int y)
```

- $x, y$ : 需要被分析之子字串的起始和終止索引值。
- 這個程序必須回傳子字串  $a[x..y]$  和  $b[x..y]$  的突變距離。
- 這個程序會被呼叫剛好  $q$  次。

## 範例

考慮下列呼叫:

```
init("ATACAT", "ACTATA")
```

假設評分程式呼叫 `get_distance(1, 3)`。這個呼叫必須回傳  $a[1..3]$  和  $b[1..3]$  (也就是 "TAC" 和 "CTA") 的突變距離。"TAC" 可以用兩個突變操作轉換成 "CTA": 首先  $TAC \rightarrow CAT$ , 然後  $CAT \rightarrow CTA$ ,

且可看出不可能用少於 2 個突變操作來完成這個轉換。

因此，這個呼叫應回傳 2。

假設評分程式呼叫 `get_distance(4, 5)`。這個呼叫必須回傳 "AT" 和 "TA" 的突變距離。"AT" 可以用一個突變操作轉換成 "TA"，而且可看出至少要 1 個突變操作來完成這個轉換。

因此，這個呼叫應回傳 1。

最後假設評分程式呼叫 `get_distance(3, 5)`。因為不可能用突變操作將 "CAT" 轉換成 "ATA"，這個呼叫應回傳 -1。

## 條件限制

- $1 \leq n, q \leq 100\,000$
- $0 \leq x \leq y \leq n - 1$
- $a$  和  $b$  中的每一個字元都是 "A", "T", 和 "C" 中的其中之一。

## 子任務

1. (21 points)  $y - x \leq 2$
2. (22 points)  $q \leq 500$ ,  $y - x \leq 1000$ ,  $a$  和  $b$  中的每一個字元都是 "A" 和 "T" 中的一個。
3. (13 points)  $a$  和  $b$  中的每一個字元都是 "A" 和 "T" 中的一個。
4. (28 points)  $q \leq 500$ ,  $y - x \leq 1000$
5. (16 points) 無其他限制。

## 範例評分程式

範例評分程式讀取輸入的格式如下：

- line 1:  $n\ q$
- line 2:  $a$
- line 3:  $b$
- line  $4 + i$  ( $0 \leq i \leq q - 1$ ): 第  $i$ -th 次呼叫 `get_distance` 的  $x\ y$  值。

範例評分程式以下列格式輸出你的答案：

- line  $1 + i$  ( $0 \leq i \leq q - 1$ ): 第  $i$ -th 次呼叫 `get_distance` 的回傳值。