

#### Problem 4：其它

子題 1：最長共同子序列(Longest common subsequence)。(程式執行限制時間: 2 秒) 10 分

「最長共同子序列」 (Longest Common Subsequence, LCS)。LCS 是兩個序列(sequence)，在各自所有的子序列(subsequence)之中，一模一樣而且最長的那個子序列(subsequence)。

給 2 個字串，請你輸出他們的最長共同子序列 (Longest common subsequence) 的長度。

也就是說，在這兩個字串各自所有的子序列之中，內容相同而且長度最長的那個子序列。舉例來說有兩個字串 abcdghxy 和 aedfhrz，它們的最長共同子序列為 adh，長度為 3。

j		0	1	2	3	4	5	6	7
i		Y	a	e	d	f	h	r	z
0	X	0	0	0	0	0	0	0	0
1	a	0	1	1	1	1	1	1	1
2	b	0	1	1	1	1	1	1	1
3	c	0	1	1	1	1	1	1	1
4	d	0	1	1	2	2	2	2	2
5	g	0	1	1	2	2	2	2	2
6	h	0	1	1	2	2	3	3	3
7	x	0	1	1	2	2	3	3	3
8	y	0	1	1	2	2	3	3	3

<http://lcs-demo.sourceforge.net/>

定義：給定兩個序列(Sequence) $X = \langle x_1, x_2, \dots, x_m \rangle$ 和 $Y = \langle y_1, y_2, \dots, y_n \rangle$ ,求 X 和 Y 所構成的最長共同子序列  $LCS(X,Y) = Z = \langle z_1, z_2, \dots, z_k \rangle$ 為何？

[注意] 子字串(Substrings)和子序列(Subsequences)的差別

例如有一條字串(String)  $S = "a t g a t g c a a t"$ ，則

子字串(Substrings) of S : "g a t g c", "t g c a a t"

子序列(Subsequences) of S : "a g g t", "a a a a"

解法

$$C[i,j] = \begin{cases} 0, & \text{if } i = 0 \text{ or } j = 0 \\ C[i-1, j-1] + 1, & \text{if } i, j > 0 \text{ and } x_i = y_j \\ \text{Max}(C[i, j-1], C[i-1, j]), & \text{if } i, j > 0 \text{ and } x_i \neq y_j \end{cases}$$

若 $i$ 或 $j$ 為0，表示 $X$ 或 $Y$ 這兩條序列的其中一條為空序列。

若 $x_i = y_j$ ，則 $c[i, j]$ 所表示的序列長度，是由 $\langle x_1, x_2, \dots, x_{i-1} \rangle$ 和 $\langle y_1, y_2, \dots, y_{j-1} \rangle$ 兩序列所構成之最長共同子序列的長度，即 $c[i-1, j-1]$ 再加上1。

若 $x_i \neq y_j$ ，則 $c[i, j]$ 的序列長度是由下列兩個不同的最長共同子序列長度當中之最大值所構成：

1.  $\langle x_1, x_2, \dots, x_{i-1} \rangle$ 和 $\langle y_1, y_2, \dots, y_j \rangle$ 兩序列所構成之最長共同子序列的長度 $c[i-1, j]$

2.  $\langle x_1, x_2, \dots, x_i \rangle$ 和 $\langle y_1, y_2, \dots, y_{j-1} \rangle$ 兩序列所構成之最長共同子序列的長度 $c[i, j-1]$

```
function LCSLength(X[1..m], Y[1..n])
    C = array(0..m, 0..n)
    for i := 0..m
        C[i,0] = 0
    for j := 0..n
        C[0,j] = 0
    for i := 1..m
        for j := 1..n
            if X[i] = Y[j]
                C[i,j] := C[i-1,j-1] + 1
            else
                C[i,j] := max(C[i,j-1], C[i-1,j])
    return C[m,n]
```

[https://en.wikipedia.org/wiki/Longest\\_common\\_subsequence\\_problem](https://en.wikipedia.org/wiki/Longest_common_subsequence_problem)

**輸入說明：**

第一列的數字  $n$  代表有幾組資料要測試， $2 \leq n \leq 8$ ，第二列起為每組的測試資料，之後每二列為每組的測試資料。每組測試資料為二列字串，字串內容為英文字母或數字，大小寫的英文字母不同，每列最多有 80 個字元。

**輸出說明：**

對輸入的每組測試資料，輸出它們最長共同子序列的長度。

輸入檔案 1：【檔名：in1.txt】

2

ACCGATGCAGCGCTC

CCGATGA

abcdghxy

aedfhrz

輸入檔案 2：【檔名：in2.txt】

2

13567

24680

ab12

Abc1

輸出範例：【檔名：out1.txt】

7

3

輸出範例：【檔名：out2.txt】

1

2