

Two character string may have many common substring. Substrings are required to be contiguous in the original string. For example, *photograph* and *tomography* have several common substring of length one (i.e. single letters) and common substring *ph*, *to* and *ograph* (as well as all substring of *ograph*). The maximum common substring length is 6. Let $X = \langle x_1, x_2, \dots, x_n \rangle$, $Y = \langle y_1, y_2, \dots, y_m \rangle$. Given an $O(mn)$ algorithm to find max common substring length for X and Y .

【92 年台大資工所】

Ans.

Let $c[i, j]$ 為 X_i 和 Y_j 的最長子字串的長度。

Input: X, Y

X_i : X 有 i 個元素時。

Y_j : Y 有 j 個元素時。

Output: length

1. $length \leftarrow 0$
2. for $i \leftarrow 1$ to n
3. $c[i, 0] \leftarrow 0$
4. for $j \leftarrow 1$ to m
5. $c[0, j] \leftarrow 0$
6. for $i \leftarrow 1$ to n
7. for $j \leftarrow 1$ to m
8. if $X[i] == Y[j]$
9. then $c[i, j] \leftarrow c[i-1, j-1] + 1$
10. $length \leftarrow \max(length, c[i, j])$
11. else
12. $c[i, j] \leftarrow 0$ // 不相同表示有斷點, 長度就重算。
13. return length