

Longest common substring

The longest common substring problem is the problem of finding the longest string (or strings) that is a substring of two strings

eg: $S_1 = "A \underline{B A B C}"$ $S_2 = "\underline{B A B C} A"$

method 1

Consider all substrings of 2nd string and find the longest substring that is also a substring of first string

$$O((m+n) * n^2)$$

method 2

Dynamic programming

$$LCS[i][j] = 1 + LCS[i-1][j-1]$$

$$\text{if } x[i-1] = y[j-1]$$

else

0

		A	B	A	B
	0	0	0	0	0
B	0	0	0+1=1	0	0+1=1
A	0	0+1=1	0	0+1=2	0
B	0	0	1+1=2	0	2+1=3
A	0	0+1=1	0	2+1=3	0

To backtrack

Find max number in matrix
and diagonally to back word
fill we get 0

		A	B	A	B
	0	0	0	0	0
B	0	0	1	0	1
A	0	1	0	2	0
B	0	0	2	0	3
A	0	1	0	3	0

ABA , BAB

These 2 are longest common substring.

memo = $\begin{bmatrix} m+1 \end{bmatrix} \begin{bmatrix} n+1 \end{bmatrix}$ initialise all with 0 ^{boxes}

For i in range(1, m+1):

for j in range(1, n+1):

if $x[i-1] = y[j-1]$:

memo[i][j] = 1 + memo[i-1][j-1]

else

memo[i][j] = 0

$O(m)$

$O(n)$

$O(1)$

$O(m \cdot n)$

OPTIMAL ANS

temp = max(memo)

FOR OPTIMAL SOLUTION

pos = position of temp = (i, j)

ans = $\begin{bmatrix} \end{bmatrix}$

ans.append(memo[i][j])

For k in range(temp):

i = i-1

j = j-1

ans.append(memo[i-1][j-1])

Time complexity $O(m \cdot n)$

space complexity $O(m \cdot n)$