

LCS (Longest Common Subsequence)

최장 공통 부분자열

1. 해를 분석 \rightarrow 부분재귀 분할

2. 부분재귀 해 \rightarrow 큰 문제의 해를 표현하기 (포함성)

D.P. table

3. 작성된 순서로 채워 (DP table)

4. table에서 해를 계산 알고리즘의 정확성 증명.

$X = ABCBDAB$

부분자열: 문자열에서 몇개를 지우고 남은 문자열

공통 부분자열: 두 집합의 공통으로 속한 부분자열.

$Y = BDCABA$

ex) A, B, C, D $\Rightarrow 1$

AB, DA $\Rightarrow 2$

BCAB, BDAB $\Rightarrow 4$

LCS \rightarrow 가장 긴 길이의 공통 부분자열 찾는 문제

$X_m = x_1, x_2, \dots, x_m$

$X_3 = x_1, x_2, x_3$

\rightarrow 3개의 문자열

$Y_m = y_1, y_2, \dots, y_m$

* $LCS(X_m, Y_m) = X_m$ 과 Y_m 의 최장길이 부분자열의 길이.

\Downarrow

$LCS(i, j) = X_i$ 와 Y_j 의 LCS의 길이

$X_i = x_1, x_2, \dots, x_i$

$Y_j = y_1, y_2, \dots, y_j$

① 마지막 문자열이 서로 같은 경우 \Rightarrow 공통 부분자열로 땀은 것의 정도 순서 \times

if $x_i = y_j$:

\Rightarrow 길이 + 1 증가

이후 $X_i = x_1, x_2, \dots, x_i$

$Y_j = y_1, y_2, \dots, y_j$

" y_{j-1} 까지 공통 부분자열 계산.

$\Rightarrow LCS(i-1, j-1) + 1$

② 마지막 문자열이 서로 다른 경우.

\rightarrow 둘 중이 한개만 땀린다. 두개 모두 땀어지는 않는다.

if $x_i \neq y_j$:

$X_i = x_1, x_2, \dots, A$ \checkmark x_i 는

(둘 중 x_i 가 선택 되었을때)

$Y_j = y_1, y_2, \dots, B$

\rightarrow 두개의 LCS를 구한다.

$\Rightarrow \max(LCS(i, j-1), LCS(i-1, j))$

둘중의 값.

$LCS(\bar{i}, \bar{j}) = x_i (x_1, x_2, x_3 \dots x_i)$ 와 $y_j (= y_1 \dots y_j)$ 이 LCS 길이

$$LCS(\bar{i}, \bar{j}) = LCS \begin{cases} LCS(\bar{i}-1, \bar{j}-1) + 1 & \text{if } x_i = y_j \\ \max \left(\begin{array}{l} LCS(\bar{i}-1, \bar{j}) + 1 \\ LCS(\bar{i}, \bar{j}-1) + 1 \end{array} \right) & \text{if } x_i \neq y_j \end{cases}$$

각 글자에 대한 *반복*
이전 글자까지의 LCS 길이
이 글자와 이전 글자 중 하나와 비교하여
이 글자와 이전 글자 중 하나와 비교하여

X \ Y	O ₀	B ₁	D ₂	C ₃	A ₄	B ₅	A ₆
O ₀	0	0	0	0	0	0	0
A ₁	0	0	0	0	1	1	1
B ₂	0	1	1	1	2	2	
C ₃	0			2	2	2	
B ₄	0			2	3	3	
D ₅	0			2	3	4	
A ₆	0			2	3	4	4
B ₇	0			2	3	4	4

$LCS(0, 5) = LCS(2, 0)$

$\uparrow \leftarrow Y_j \Rightarrow X = [], Y = [\dots]$

None

$m \times m$ 칸 $\times O(1)$

$= O(m \times m) = O(m^2)$

($\therefore m = m$ 갯수)

$LCS(7, 6) = LCS(X_7, Y_6)$



X \ Y	O ₀	B ₁	D ₂	C ₃	A ₄	B ₅	A ₆
O ₀	0	0	0	0	0	0	0
A ₁	0	0	0	0	1	1	1
B ₂	0	1	1	1	2	2	
C ₃	0	1	1	2	2	2	
B ₄	0	1	1	2	3	3	
D ₅	0	1	2	2	3	4	
A ₆	0	1	2	2	3	4	4
B ₇	0	1	2	2	3	4	4

대각선으로 따라 올라간 것을 바로 결과는.

\Rightarrow **BCBA** \Rightarrow LCS의 길이

이러한 과정은 모든 배열 쌍에 대해 반복함

\uparrow (여기까지 완료함 후)