Fundamentos de análisis y diseño de algoritmos

LCS (Longest Common Subsequence)

Multiplicación de matrices

Propiedades generales de la programación dinamica

Dada una secuencia $X=\langle x_1,x_2,...x_m\rangle$, se dice que $Z=\langle z_1,z_2,...,z_n\rangle$ es una subsecuencia de X si los simbolos z_i se dan en X siguiendo el orden que se dan en Z (no necesariamente uno seguido de otro)

Por ejemplo, para $X=\langle A,B,C,B,D,A,B\rangle$, algunas subsecuencias son:

Por ejemplo, para $X=\langle A,B,C,B,D,A,B\rangle$, algunas subsecuencias son:

Sin embargo, no son subsecuencias:

Problema; Dadas dos secuencias $X=\langle x_1,x_2,...x_m\rangle$ y $Y=\langle y_1,y_2,...y_n\rangle$, encontrar la(s) subsecuencia(s) más larga(s) entre a X y Y

Indique algunas subsecuencias

Indique una solución al problema

Problema; Dadas dos secuencias $X=\langle x_1,x_2,...x_m\rangle$ y $Y=\langle y_1,y_2,...y_n\rangle$, encontrar la(s) subsecuencia(s) más larga(s) entre a X y Y

X=ABCBDAB

Y=BDCABA

Algunas subsecuencias son: A, CB, ABA, BDAB, CAB, DAB

El problema de la subsecuencia más larga tiene varias soluciones: BCBA, BCAB y BDAB

Problema; Dadas dos secuencias $X=\langle x_1,x_2,...x_m\rangle$ y $Y=\langle y_1,y_2,...y_n\rangle$, encontrar la(s) subsecuencia(s) más larga(s) entre a X y Y

¿Cuál podría ser una solución al problema?

Problema; Dadas dos secuencias $X=\langle x_1,x_2,...x_m\rangle$ y $Y=\langle y_1,y_2,...y_n\rangle$, encontrar la(s) subsecuencia(s) más larga(s) entre a X y Y

¿Cuál podría ser una solución al problema?

Enumerar todas las subsecuencias de X y examinar si es también una subsecencia de Y. Tomar al final la subsecuencia más larga

Existen 2^m posibles suconjuntos (subsecuencias) del conjunto de m elementos \rightarrow Tiempo exponencial

Problema; Dadas dos secuencias $X=\langle x_1,x_2,...x_m\rangle$ y $Y=\langle y_1,y_2,...y_n\rangle$, encontrar la(s) subsecuencia(s) más larga(s) entre a X y Y

La solución al problema presenta una característica especial

Problema; Dadas dos secuencias $X=\langle x_1,x_2,...x_m\rangle$ y $Y=\langle y_1,y_2,...y_n\rangle$, encontrar la subsecuencia más larga (LCS) entre a X y Y

- Si $x_m = y_n$, la solución del problema LCS(X_m, Y_n) será LCS(X_{m-1}, Y_{n-1}) y pegar al final x_m
- Si $x_m \neq y_n$, la solución del problema LCS(X_m, Y_n) se escoge entre LCS(X_{m-1}, Y_n) y LCS(X, Y_{n-1}), esto es, de estas dos se elije aquella de mayor longitud

Si $x_m = y_n$, la solución del problema LCS(X_m, Y_n) será LCS(X_{m-1}, Y_{n-1}) y pegar al final x_m

• Si $x_m \neq y_n$, la solución del problema LCS(X_m, Y_n) se escoge entre LCS(X_{m-1}, Y_n) y LCS(X, Y_{n-1}), esto es, de estas dos se elije aquella de mayor longitud

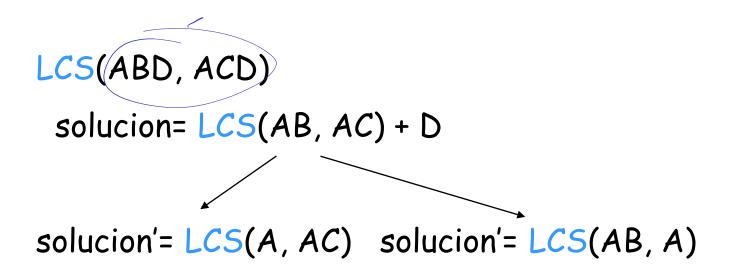
LCS(ABD, ACD)

Si $x_m = y_n$, la solución del problema LCS(X_m, Y_n) será LCS(X_{m-1}, Y_{n-1}) y pegar al final x_m

• Si $x_m \neq y_n$, la solución del problema LCS(X_m, Y_n) se escoge entre LCS(X_{m-1}, Y_n) y LCS(X, Y_{n-1}), esto es, de estas dos se elije aquella de mayor longitud

Si $x_m = y_n$, la solución del problema LCS(X_m, Y_n) será LCS(X_{m-1}, Y_{n-1}) y pegar al final x_m

• Si $x_m \neq y_n$, la solución del problema LCS(X_m, Y_n) se escoge entre LCS(X_{m-1}, Y_n) y LCS(X, Y_{n-1}), esto es, de estas dos se elije aquella de mayor longitud



```
•LCS(ABD, ACD)

solucion= LCS(AB, AC) + D

solucion= LCS(A, AC) solucion= LCS(AB, A)

solucion= LCS(, AC) solucion= LCS(A, A)
```

Se obtiene AD como una solución

•LCS(ABD, ACD)

solucion= LCS(AB, AC) + D

solucion'= LCS(A, AC) solucion'= LCS(AB, A)

solucion''= LCS(A, A) solucion''= LCS(AB,)

Ya fue calculado, no se calcula nuevamente.

Se retorna
$$A$$

Se obtiene AD como una solución

LCS (AFOR), AFKED) +D max (LCSCAFCK) AFKED)) ~

LCS(AF, AFKEJ) LCS(AFE, AFKEJ)

Note que hay subproblemas que se sobrelapan, para encontrar $LCS(X_m, Y_n)$ se podría necesitar encontrar $LCS(X_{m-1}, Y_n)$ y $LCS(X_m, Y_{n-1})$, pero cada uno de los subproblemas tendrá que encontrar $LCS(X_{m-1}, Y_{n-1})$

Cuando se comparten subproblemas, no se calculan nuevamente, ya se conoce el valor, simplemente se utiliza

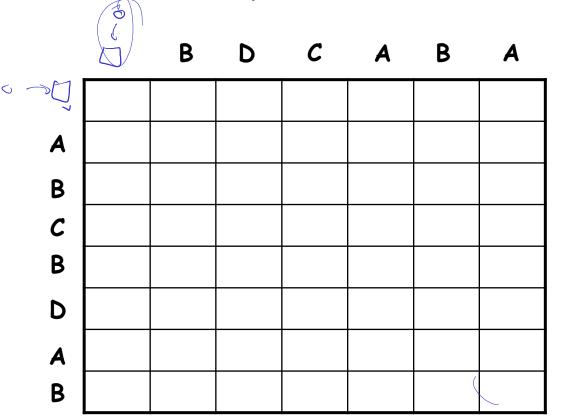
Estrategia: Divide y vencerás donde los subproblemas no son independientes sino que algunos son compartidos. Esto se aprovecha para no tener que hacer cálculos repetidos

Se tiene una función de costo que permite conocer la longitud de la solución de óptima:

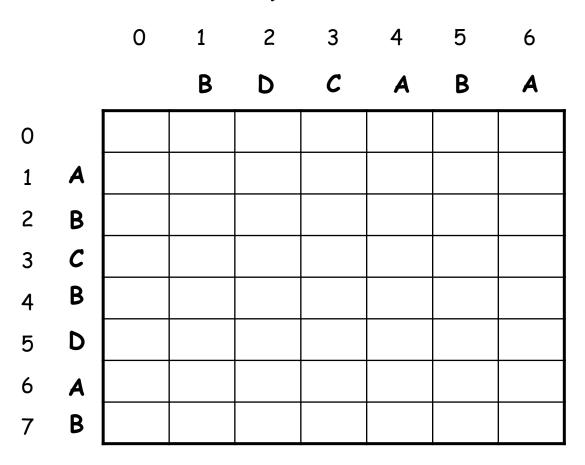
$$c[i,j] = \begin{cases} 0 & \text{, si } i=0 \text{ o } j=0 \\ c[i-1,j-1] + 1 & \text{, si } \underbrace{x_i = x_j} \text{ e } i,j>0 \\ max(c[i,j-1], c[i-1,j]), \text{ si } x_i \neq x_j \text{ e } i,j>0 \end{cases}$$

c es una matriz que indica la longitud de la secuencia más larga entre X_i y Y_i

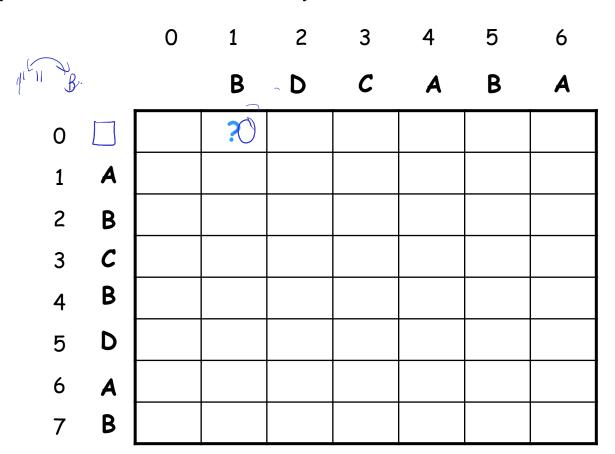
LCS(ABCBDAB, BDCABA)



LCS(ABCBDAB, BDCABA)

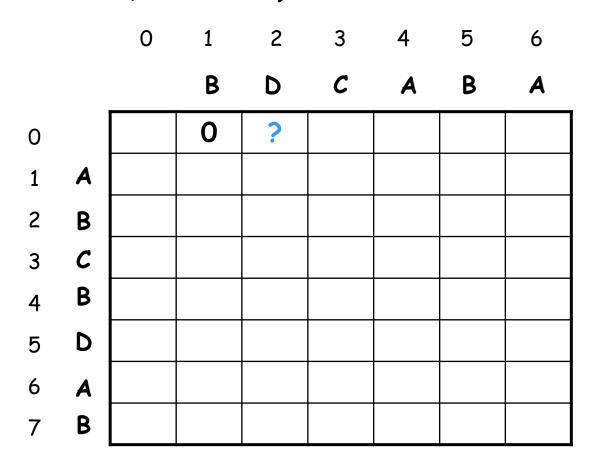


LCS(ABCBDAB, BDCABA)



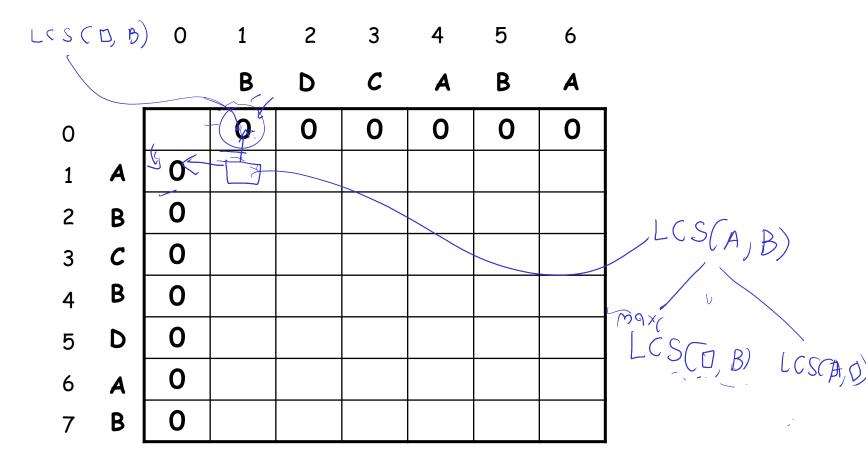
C[0,1] es la longitud de la subsecuencia más larga entre las secuencias X=" " y Y="B"

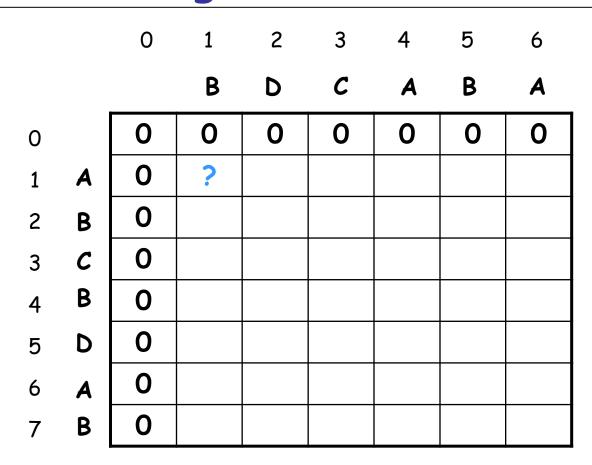
LCS(ABCBDAB, BDCABA)

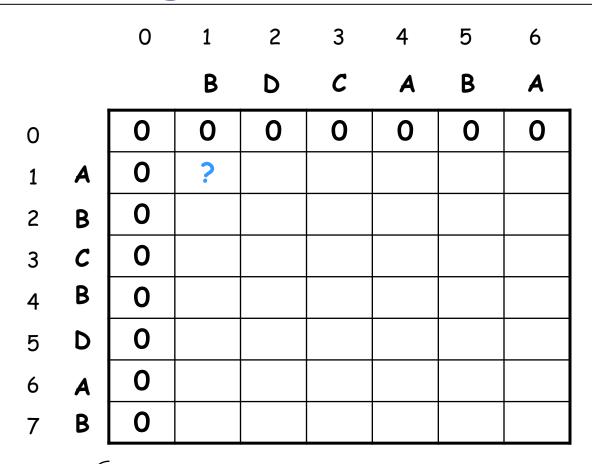


C[0,1] es la longitud de la subsecuencia más larga entre las secuencias X=" " y Y="B"

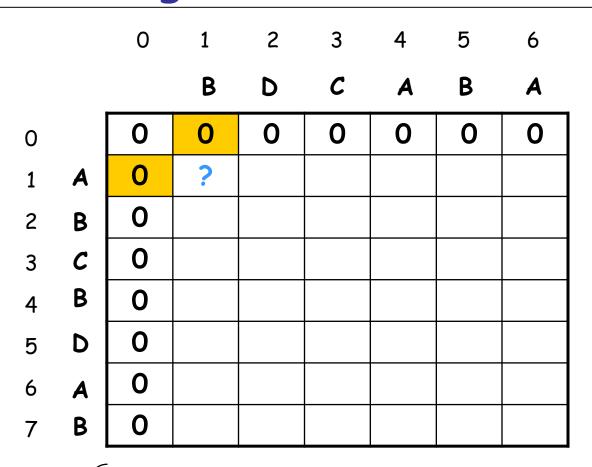
LCS(ABCBDAB, BDCABA)



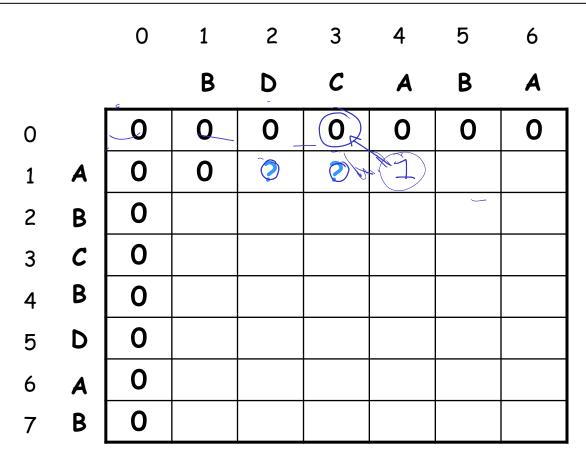




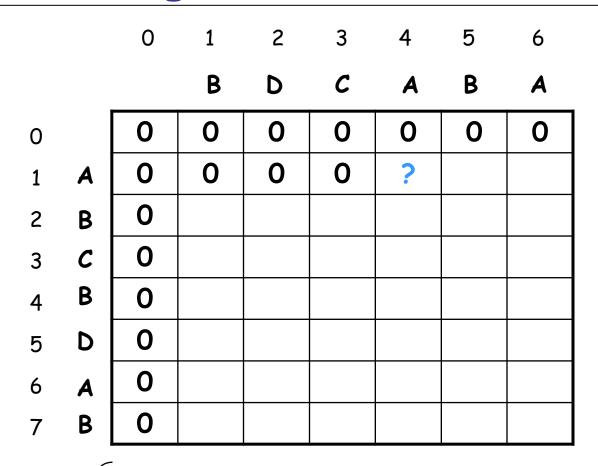
$$c[i,j] = \begin{cases} 0 & , \text{ si } i=0 \text{ o } j=0 \\ c[i-1,j-1]+1 & , \text{ si } x_i=x_j \text{ e } i,j>0 \\ max(c[i,j-1], c[i-1,j]), \text{ si } x_i\neq x_j \text{ e } i,j>0 \end{cases}$$



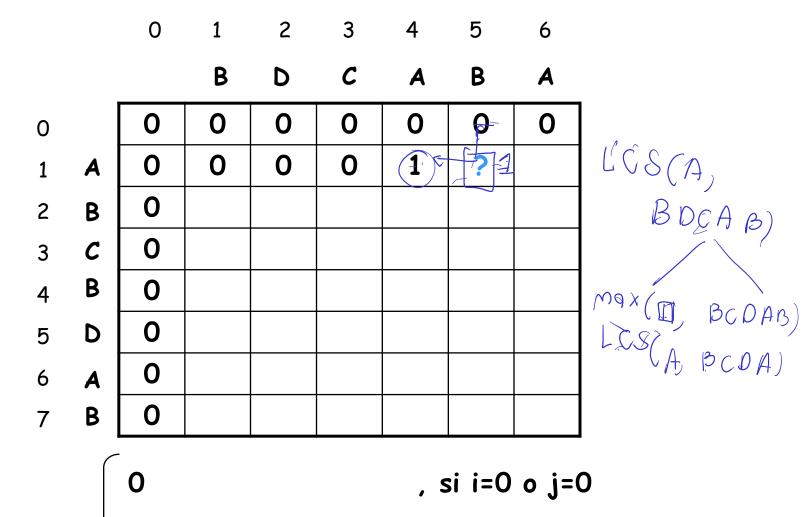
$$c[i,j] = \begin{cases} 0 & , \text{ si } i=0 \text{ o } j=0 \\ c[i-1,j-1]+1 & , \text{ si } x_i=x_j \text{ e } i,j>0 \\ max(c[i,j-1], c[i-1,j]), \text{ si } x_i\neq x_j \text{ e } i,j>0 \end{cases}$$



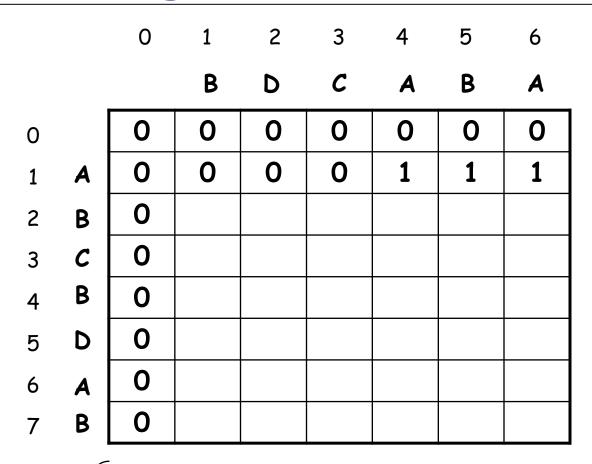
$$c[i,j] = \begin{cases} 0 & , \text{ si } i=0 \text{ o } j=0 \\ c[i-1,j-1] + 1 & , \text{ si } x_i=x_j \text{ e } i,j>0 \\ max(c[i,j-1], c[i-1,j]), \text{ si } x_i\neq x_j \text{ e } i,j>0 \end{cases}$$



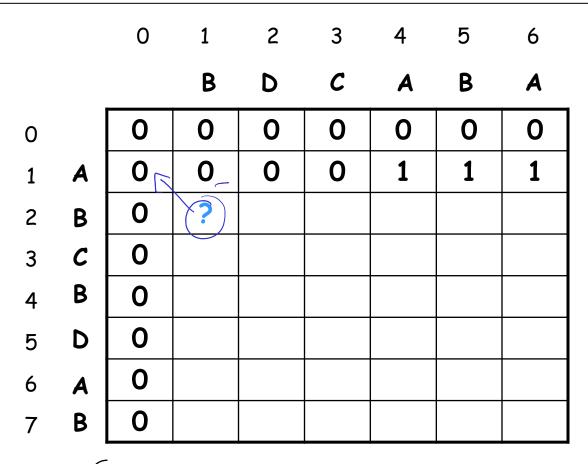
$$c[i,j] = \begin{cases} 0 & , \text{ si } i=0 \text{ o } j=0 \\ c[i-1,j-1]+1 & , \text{ si } x_i=x_j \text{ e } i,j>0 \\ max(c[i,j-1], c[i-1,j]), \text{ si } x_i\neq x_j \text{ e } i,j>0 \end{cases}$$



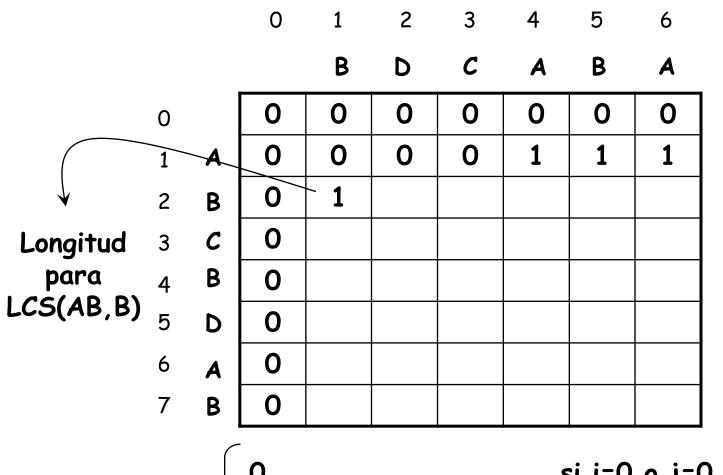
$$c[i,j] = \begin{cases} 0 & , \text{ si } i=0 \text{ o } j=0 \\ c[i-1,j-1] + 1 & , \text{ si } x_i=x_j \text{ e } i,j>0 \\ max(c[i,j-1], c[i-1,j]), \text{ si } x_i\neq x_j \text{ e } i,j>0 \end{cases}$$



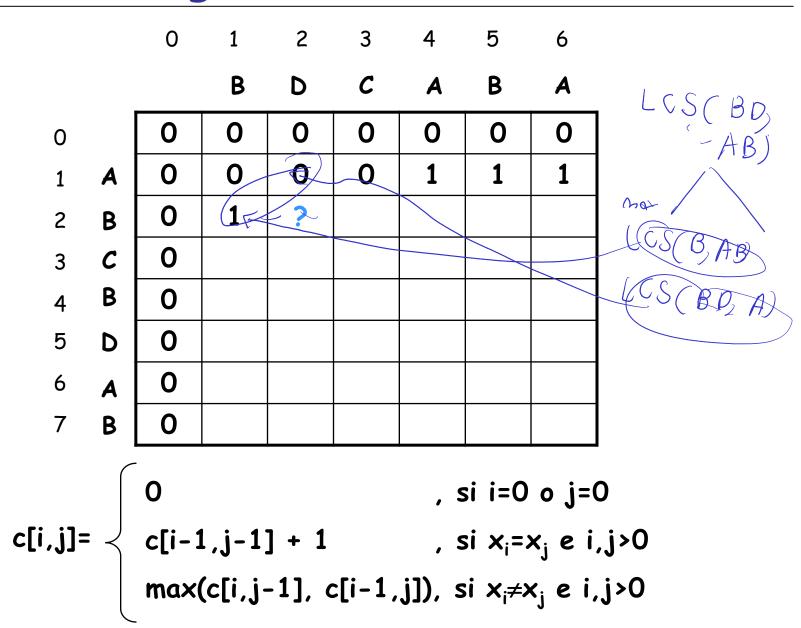
$$c[i,j] = \begin{cases} 0 & , \text{ si } i=0 \text{ o } j=0 \\ c[i-1,j-1] + 1 & , \text{ si } x_i=x_j \text{ e } i,j>0 \\ max(c[i,j-1], c[i-1,j]), \text{ si } x_i\neq x_j \text{ e } i,j>0 \end{cases}$$

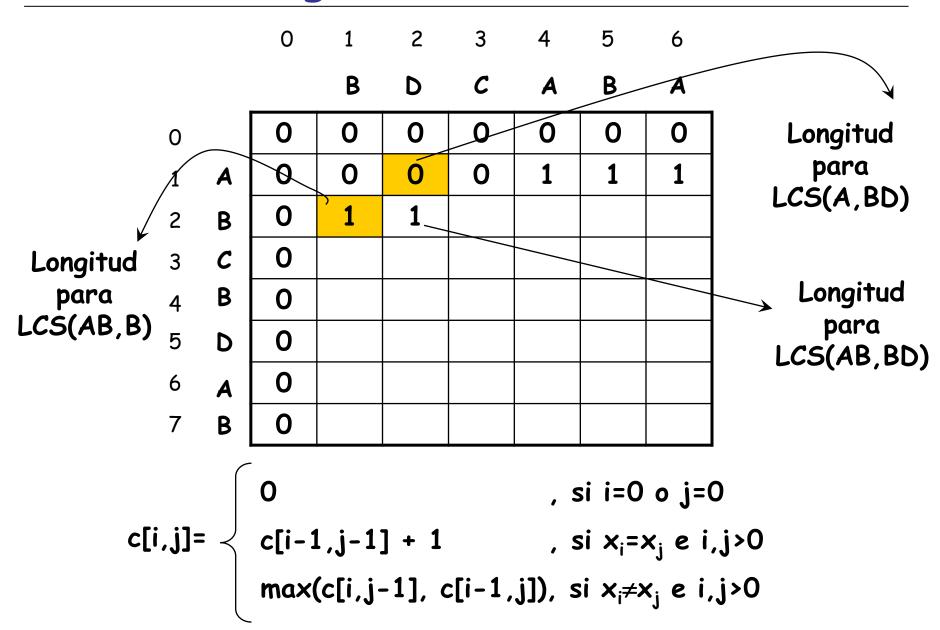


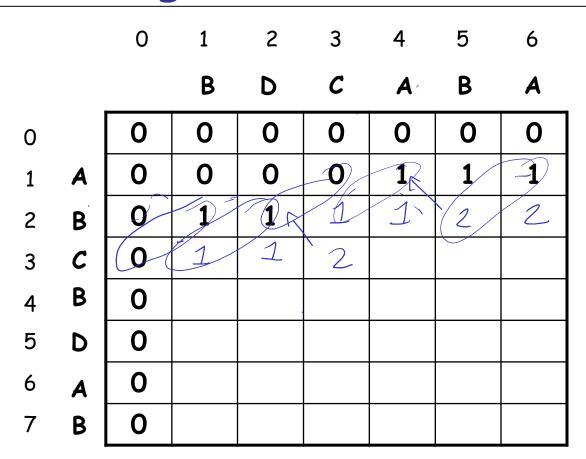
$$c[i,j] = \begin{cases} 0 & , \text{ si } i=0 \text{ o } j=0 \\ c[i-1,j-1] + 1 & , \text{ si } x_i=x_j \text{ e } i,j>0 \\ max(c[i,j-1], c[i-1,j]), \text{ si } x_i\neq x_j \text{ e } i,j>0 \end{cases}$$



$$c[i,j] = \begin{cases} 0 & , \text{ si } i=0 \text{ o } j=0 \\ c[i-1,j-1] + 1 & , \text{ si } x_i=x_j \text{ e } i,j>0 \\ max(c[i,j-1], c[i-1,j]), \text{ si } x_i\neq x_j \text{ e } i,j>0 \end{cases}$$





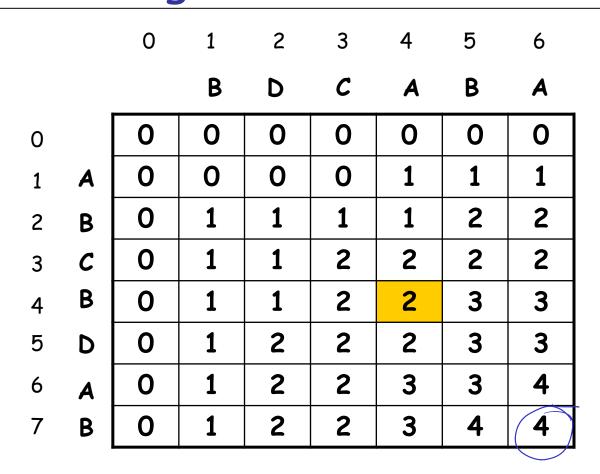


Completar la tabla

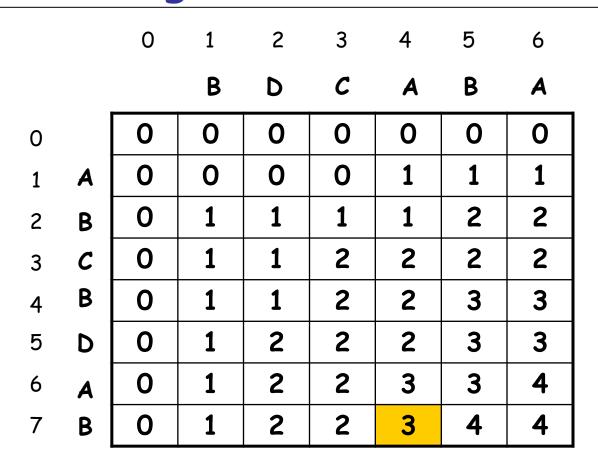
		0	1	2	3	4	5	6
			В	D	C	A	В	A
0		0	0	0	0	0	0	0
1	A	0	0	0	0	1	1	1
2	В	0	1	1	1	1	2	2
3	C	0	1	1	2	2	2	2
4	В	0	1	1	2	2	3	3
5	D	0	1	2	2	2	3	3
6	A	0	1	2	2	3	3	4
7	В	0	1	2	2	3	4	4

		0	1	2	3	4	5	6
			В	D	C	A	В	A
0		0	0	0	0	0	0	0
1	A	0	0	0	0	1	1	1
2	В	0	1	1	1	1	2	2
3	C	0	1	1	2	2	2	2
4	В	0	1	1	2	2	3	3
5	D	0	1	2	2	2	3	3
6	A	0	1	2	2	3	3	4
7	В	0	1	2	2	3	4	4

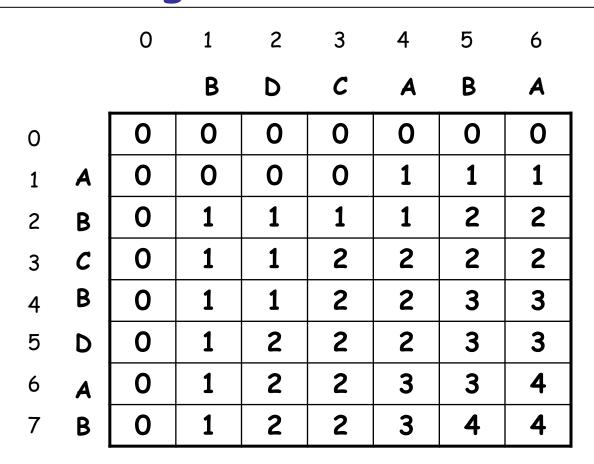
Qué significado tiene el hecho de que c[4,4]=2



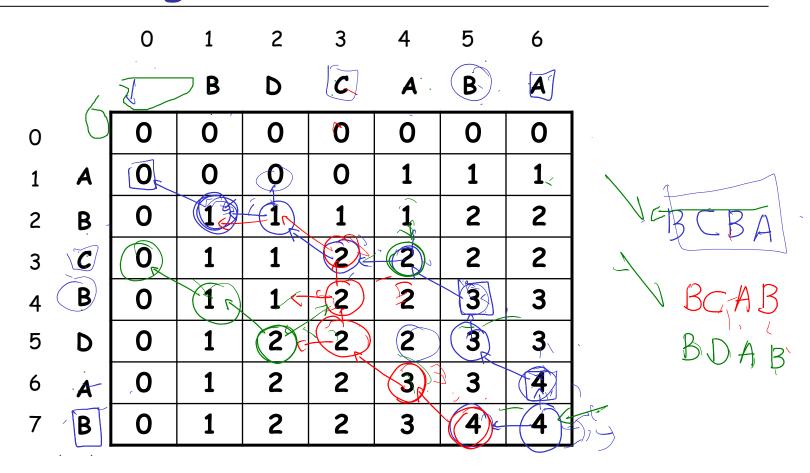
Qué significado tiene el hecho de que c[4,4]=2 Indica que la longitud de LCS(ABCB,BDCA) es 2



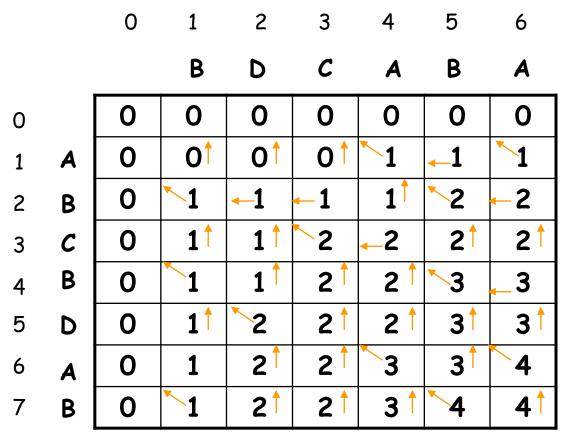
Qué significado tiene el hecho de que c[7,4]=3 Indica que la longitud de LCS(ABCBDAB, BDCA) es 3



Qué casilla de la matriz guarda la longitud de la solución al problema original

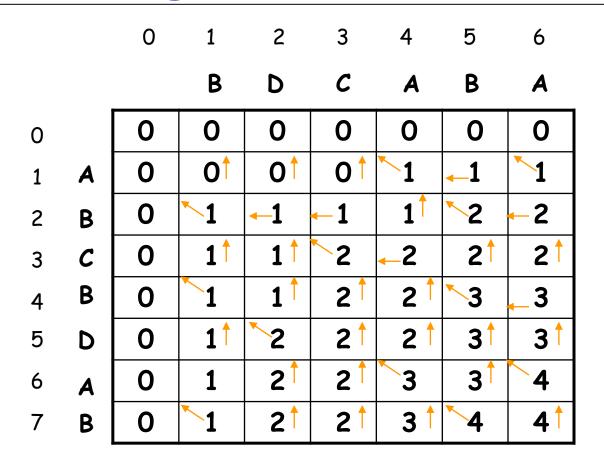


Cuál es la solución, es decir, cual es la subsecuencia común?

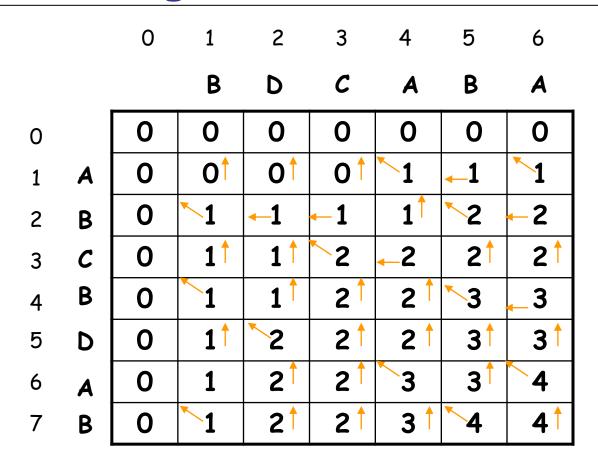


Las direcciones se guardan en otro arreglo llamado B

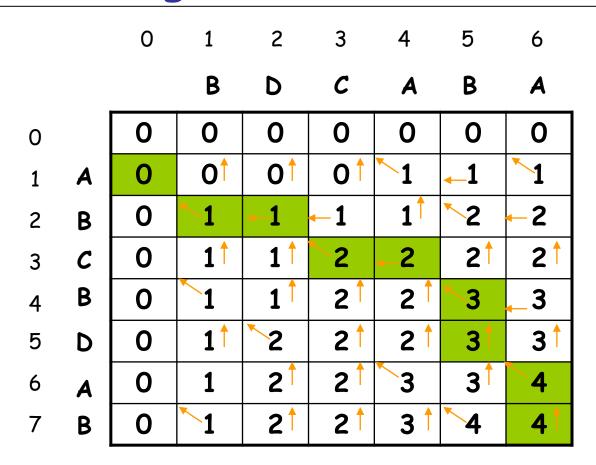
$$c[i,j] = \begin{cases} 0 & \text{, si } i=0 \text{ o } j=0 \\ c[i-1,j-1] + 1 & \text{, si } x_i=x_j \text{ e } i,j>0 \text{ (} \\ max(c[i,j-1], c[i-1,j]), \text{ si } x_i\neq x_j \text{ e } i,j>0 \text{ (} \\ \end{pmatrix}$$



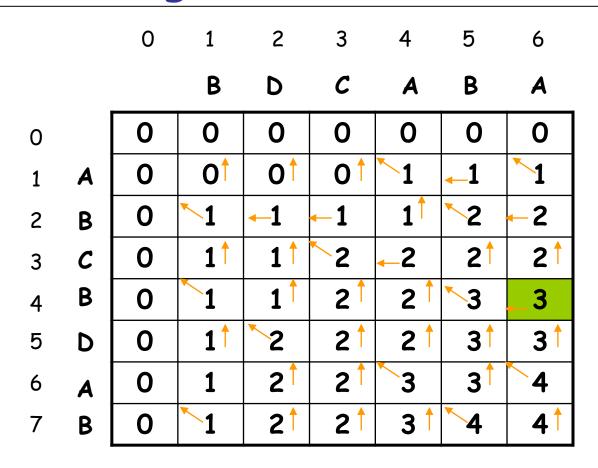
Solo > indica que es un símbolo común, éstos se imprimen. En los demás casos, se sigue la flecha, ↑ o ←



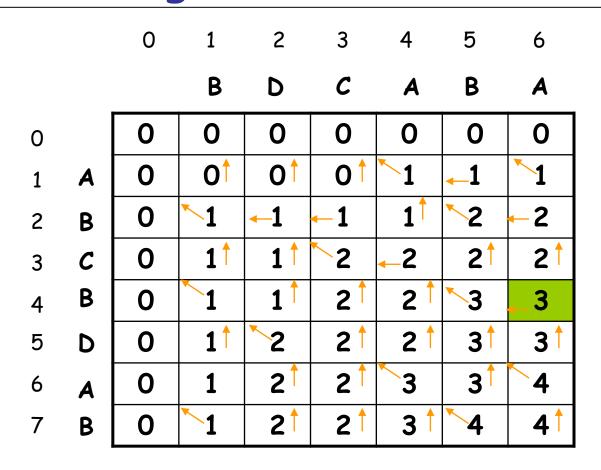
Muestre la solución al problema



ABCB, se invierte y se obtiene BCBA



Muestre la solución al problema LCS(ABCB, BDCABA)



Muestre la solución al problema LCS(ABCB, BDCABA)

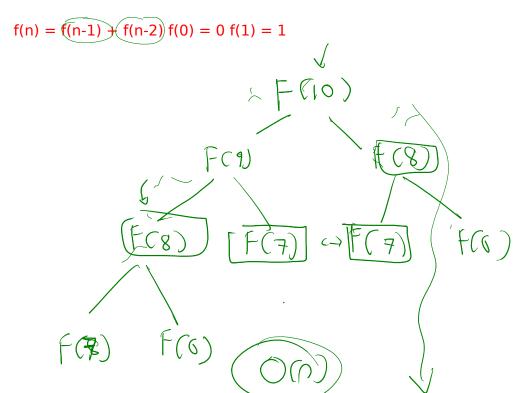
BCB, que en orden invertido es BCB

```
LCS-LENGTH(X,Y)
m \leftarrow length[X]
n \leftarrow length[Y]
for i←1 to m
 do c[i,0]←0
for j \leftarrow 0 to n
 do c[0,j] \leftarrow 0
for i←1 to m
 do for j \leftarrow 1 to n
        do if x_i = y_i
               then c[i,j] \leftarrow c[i-1,j-1]+1
                      b[i,j]←" \_"
             else if c[i-1,j] \ge c[i,j-1]
                      then c[i,j] \leftarrow c[i-1,j]
                             b[i,j]←"↑"
                      else c[i,j] \leftarrow c[i,j-1]
                             b[i,j]←"	"
 return c and b
```

Indique la longitud de LCS(AFCEA, CFEHA)

```
PRINT-LCS(b, X, i, j)
 if i=0 or j=0
    then return
 if b[i,j]=" \ "
    then PRINT-LCS(b, X, i-1, j-1)
    print xi
 else if b[i,j]=" | "
        then PRINT-LCS(b, X, i-1, j)
 else PRINT-LCS(b, X, i, j-1)
```

Resuelva LCS(AFCEA, CFEHA)



$$\frac{1}{2}(n) = \frac{1}{2}(n-1) + \frac{1}{2}(n-2)$$

$$O\left(\frac{s}{1+\sqrt{e}}\right)$$

Subestructura optima vector.

Regla de este vector,

$$f(n) = \begin{cases} n = 0 \\ n = 1 \end{cases}$$

$$f(n-1) + f(n-1)$$

