Longest Common Subsequence

subsequence

- An ordered combination of each member of the sequence
- Sequence = (w,a,l,k,i,n,g)
  - Subsequence Ex1 = (w,a,l,k) >> (**w,a,l,k**,i,n,g)
  - Subsequence Ex2 = (k,i,n,g) >> (w,a,l,**k,i,n,g**)
  - Subsequence Ex3 = (w,g) >> (**w**,a,l,k,i,n,**g**)
  - Subsequence Ex4 = (w,l,n,g) >> (**w**,a,**l**,k,i,**n,g**)

The problem

- Given two sequences A,B
  - Find a subsequence s of both A and B such that the length of s is longest
- Example
  - A = (w,a,l,k,i,n,g)
  - B = (a,l,i,e,n)
  - Longest Common Subsequence = (a,l,i,n)
    - (a,l,i,n) is a subsequence of A (w,**a,l**,k,**i,n**,g)
    - (a,l,i,n) is a subsequence of B (**a,l,i**,e,**n**)

Notation

- Let the first index of A and B be 1
  - E.g., A[1] = 'w', A[2] = 'a', A[3] = 'l' , …
- Let |A| = n
- Let |B| = m
- Let A_i be the substring from position 1 to i of A
  - E.g. A_1 = 'w'
  - E.g. A_2 = 'wa'
  - E.g. A_5 = 'walki'
  - A_0 = ''

The sub-problem

- If we wish to know LCS(A,B)
  - Does LCS of (A_x,B_y) helps us?
  - What sub problem shall we use?

Think Backward (or forward?)

- If we know LCS(A,B)
- How does it help?
  - i.e., where LCS(A,B) contribute to?
  - Try the very obvious case…
  - Does it help us solve
    - LCS(A + 'c',B + 'c') ?
  - Sure!
    - LCS(A + 'c',B + 'c') = LCS(A,B) + 'c'
    - Because they both ends with 'c'
    - E.g. A = 'walking', B = 'alien'
    - What is LCS('walkingC' , 'alienC')?
      - alinC

Think Backward (or forward?)

- Any more case to consider?
- If we know LCS(A,B)
  - Does it help us solve
    - LCS(A,B + ‘c’) ?
  - Yes
    - Adding ‘c’ would have only two outcomes
      - it *does not* change the LCS
        - » So LCS(A,B + ‘c’) = LCS(A,B)
      - It *does* change the LCS
        - » So LCS(A,B + ‘c’) = something ending with ‘c’
        - » To be continue…

Using LCS(A,B + ‘c’)

- The case that LCS is changed
  - Is that possible?
  - Yes, when there are ‘c’ in A that comes after LCS(A,B)
    - Assume that that point is at A[k] (hence, A[k] = c)
    - LCS(A,B + ‘c’) would be LCS(A_{k-1},B) + ‘c’
  - Check that LCS(A_{k-1},B) + ‘c’ is actually LCS(A,B + ‘c’)
    - LCS(A_{k-1},B +’c’) will be the same as
      - LCS(A_k,B+’c’)
      - LCS(A_{k+1},B+’c’)
      - LCS(A_{k+2},B+’c’)
      - …
      - LCS(A,B+’c’)

Notice that, in this case, what comes after both B and A_{k-1} is ‘c’

This means that LCS(A,B + ‘c’) = LCS(A_{n-1},B+’c’).. So, LCS(A,B) ***does not contribute*** to LCS(A,B + ‘c’)

Think Backward (or forward?)

- The remaining case
- If we know LCS(A,B)
  - Does it help us solve
    - LCS(A + ‘c’,B) ?
  - Yes
    - Similar to the case of A,B + ‘c’
- In conclusion, this means that
  - LCS(A + ‘c’,B) = LCS(A+’c’,B_{n-1})

conclusion

- LCS(A,B) will be LCS(A,B + ‘c’) when ‘c’ does not give a longer common subsequence
  - If ‘c’ is in the longer common subsequence
  - LCS(A,B + ‘c’) will be LCS(A_{n-1},B + ‘c’) instead!!!!
    - Not our case
- So, backwardly,
  - LCS(A,B) is either
    - LCS(A_{n-1},B)
      - Or
    - LCS(A,B_{m-1})
  - Just select the longer one!!!!

Recurrence

- LCS(A_i,B_j) =

LCS(A_{i-1},B_{j-1}) + A[i]	A[i] = B[j]
max(LCS(A_{i-1},B_j) , LCS(A_i,B_{j-1}))	A[i]!= B[j]

Choose to neglect A[i]

Choose to neglect B[j]

Solution to the LCS

- Simplify problem
  - To find the length of LCS
- Let c(i,j) be the length of LCS(A_i,B_j)

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\ c(i-1,j)\ ,\ c(i,j-1)\ ) & \text{if } i>0,\ j>0 \text{ and } A[i]!= B[j] \end{cases}$$

Example

C[0,0]

C[7,5]

		W	A	L	K	I	N	G
A								
L								
I								
E								
N								

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i\text{-}1,j\text{-}1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i\text{-}1,j)\,,\ c(i,j\text{-}1)\,) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

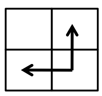
Fill the trivial case

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0							
L	0							
I	0							
E	0							
N	0							

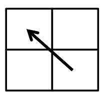
$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i\text{-}1,j\text{-}1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i\text{-}1,j)\,,\ c(i,j\text{-}1)\,) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

$$c(i,j) = \begin{cases} 0 & \text{if } I>0,\ j>0 \\ c(i\text{-}1,j\text{-}1) + 1 & \text{if } I>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i\text{-}1,j)\,,\ c(i,j\text{-}1)\,) & \text{if } I>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$



A[i] != b[j]



A[i] = b[j]

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0						
L	0							
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i\text{-}1,j\text{-}1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i\text{-}1,j)\,,\ c(i,j\text{-}1)\,) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0						
L	0							
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i\text{-}1,j\text{-}1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i\text{-}1,j)\,,\ c(i,j\text{-}1)\,) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1					
L	0							
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i\text{-}1,j\text{-}1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i\text{-}1,j)\,,\ c(i,j\text{-}1)\,) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1				
L	0							
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0, j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i-1,j)\, ,\, c(i,j-1)\,) & \text{if } i>0, j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1			
L	0							
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0, j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i-1,j)\, ,\, c(i,j-1)\,) & \text{if } i>0, j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1		
L	0							
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0, j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i-1,j)\, ,\, c(i,j-1)\,) & \text{if } i>0, j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	
L	0							
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0, j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i-1,j)\, ,\, c(i,j-1)\,) & \text{if } i>0, j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0							
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0, j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i-1,j)\, ,\, c(i,j-1)\,) & \text{if } i>0, j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0						
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0, j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i-1,j)\, ,\, c(i,j-1)\,) & \text{if } i>0, j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1					
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\ c(i-1,j)\ ,\ c(i,j-1)\ ) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1				
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\ c(i-1,j)\ ,\ c(i,j-1)\ ) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1	2			
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\ c(i-1,j)\ ,\ c(i,j-1)\ ) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1	2	2		
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\ c(i-1,j)\ ,\ c(i,j-1)\ ) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1	2	2	2	
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\ c(i-1,j)\ ,\ c(i,j-1)\ ) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1	2	2	2	2
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\ c(i-1,j)\ ,\ c(i,j-1)\ ) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1	2	2	2	2
I	0							
E	0							
N	0							

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\ c(i-1,j)\ ,\ c(i,j-1)\ ) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

Example

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1	2	2	2	2
I	0	0	1	2	2	2+1	3	3
E	0	0	1	2	2	3	3	3
N	0	0	1	2	2	3	3+1	4

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\ c(i-1,j)\ ,\ c(i,j-1)\ ) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

Recovering the Actual Solution

- We know particulaliry which case c(i, j) is from

$$c(i,j) = \begin{cases} 0 & \text{if } I>0,\ j>0 \\ c(i-1,j-1) + 1 & \text{if } I>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\ c(i-1,j)\ ,\ c(i,j-1)\ ) & \text{if } I>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

- If it is the second case, it simply means that A[i] is the last member in LCS

What is the LCS?

Trace from the back

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1	2	2	2	2
I	0	0	1	2	2	2+1	3	3
E	0	0	1	2	2	3	3	3
N	0	0	1	2	2	3	3+1	4

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\ c(i-1,j)\ ,\ c(i,j-1)\ ) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

What is the LCS?

Trace from the back

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1	2	2	2	2
I	0	0	1	2	2	2+1	3	3
E	0	0	1	2	2	3	3	3
N	0	0	1	2	2	3	3+1	4

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\ c(i-1,j)\ ,\ c(i,j-1)\ ) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

What is the LCS?

Trace from the back

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1	2	2	2	2
I	0	0	1	2	2	2+1	3	3
E	0	0	1	2	2	3	3	3
N	0	0	1	2	2	3	3+1	4

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0,\ j>0 \text{ and } A[i] = B[j] \\ \max(\ c(i-1,j)\ ,\ c(i,j-1)\ ) & \text{if } i>0,\ j>0 \text{ and } A[i] != B[j] \end{cases}$$

What is the LCS?

Trace from the back

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1	2	2	2	2
I	0	0	1	2	2	2+1	3	3
E	0	0	1	2	2	3	3	3
N	0	0	1	2	2	3	3+1	4

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0, j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i-1,j)\, , c(i,j-1)\,) & \text{if } i>0, j>0 \text{ and } A[i] != B[j] \end{cases}$$

What is the LCS?

Trace from the back

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1	2	2	2	2
I	0	0	1	2	2	2+1	3	3
E	0	0	1	2	2	3	3	3
N	0	0	1	2	2	3	3+1	4

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0, j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i-1,j)\, , c(i,j-1)\,) & \text{if } i>0, j>0 \text{ and } A[i] != B[j] \end{cases}$$

What is the LCS?

Trace from the back

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1	2	2	2	2
I	0	0	1	2	2	2+1	3	3
E	0	0	1	2	2	3	3	3
N	0	0	1	2	2	3	3+1	4

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0, j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i-1,j)\, , c(i,j-1)\,) & \text{if } i>0, j>0 \text{ and } A[i] != B[j] \end{cases}$$

What is the LCS?

Trace from the back

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1	2	2	2	2
I	0	0	1	2	2	2+1	3	3
E	0	0	1	2	2	3	3	3
N	0	0	1	2	2	3	3+1	4

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0, j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i-1,j)\, , c(i,j-1)\,) & \text{if } i>0, j>0 \text{ and } A[i] != B[j] \end{cases}$$

What is the LCS?

Trace from the back

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1	2	2	2	2
I	0	0	1	2	2	2+1	3	3
E	0	0	1	2	2	3	3	3
N	0	0	1	2	2	3	3+1	4

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0, j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i-1,j)\, , c(i,j-1)\,) & \text{if } i>0, j>0 \text{ and } A[i] != B[j] \end{cases}$$

What is the LCS?

Trace from the back

		W	A	L	K	I	N	G
	0	0	0	0	0	0	0	0
A	0	0	0+1	1	1	1	1	1
L	0	0	1	1+1	2	2	2	2
I	0	0	1	2	2	2+1	3	3
E	0	0	1	2	2	3	3	3
N	0	0	1	2	2	3	3+1	4

$$c(i,j) = \begin{cases} 0 & \text{if } i=0 \text{ or } j=0 \\ c(i-1,j-1) + 1 & \text{if } i>0, j>0 \text{ and } A[i] = B[j] \\ \max(\, c(i-1,j)\, , c(i,j-1)\,) & \text{if } i>0, j>0 \text{ and } A[i] != B[j] \end{cases}$$