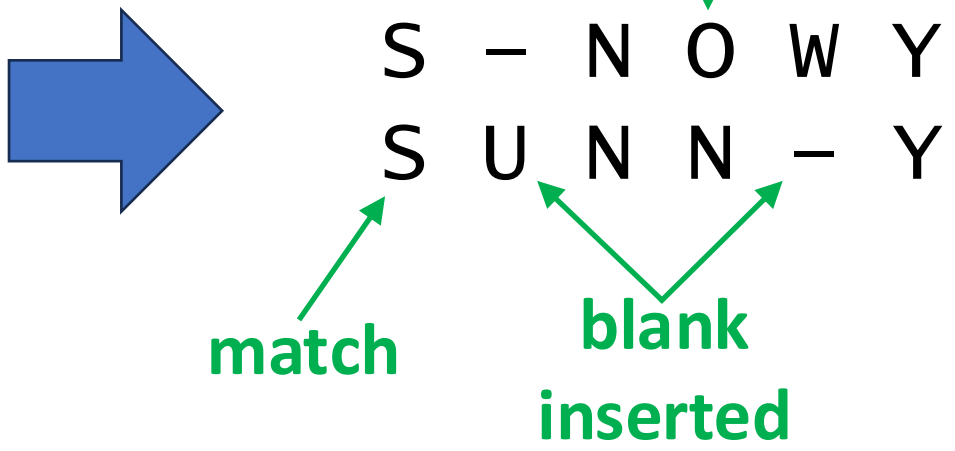
Workshop #6
CSCI 432

Edit Distance

Given two strings, how many edits are needed to turn one string into another?

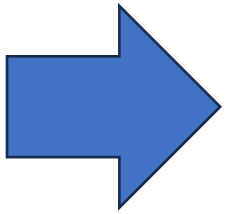
SNOWY vs SUNNY



Longest Common Subsequence

Given two strings, what is the longest (not necessarily contiguous) sequence of characters that is in both?

SNOWY vs SUNNY



S N Y

S N O W Y
S U N N Y

Edit Distance

We want to align two strings, $x = [x_1, \ldots, x_n]$ and $y = [y_1, \ldots, y_m]$.

$E(i, j)$ = optimal cost of aligning $[x_1, \ldots, x_i]$ and $[y_1, \ldots, y_j]$.

Can we say anything about optimal alignment of $[x_1, \ldots, x_i]$ and $[y_1, \ldots, y_j]$?

Alignment	Cost
x_i y_j	$\{0,1\} + E(i-1, j-1)$
x_i $-$	$1 + E(i-1, j)$
$-$ y_j	$1 + E(i, j-1)$

Longest Common Subsequence

$L(i, j)$ = Length of LCS between $[x_1, \ldots, x_i]$ and $[y_1, \ldots, y_j]$.

$\ldots \; x_{i-3} \; x_{i-2} \; x_{i-1} \; x_i$

$\ldots \; y_{j-3} \; y_{j-2} \; y_{j-1} \; y_j$

Longest Common Subsequence

$L(i, j)$ = Length of LCS between $[x_1, \ldots, x_i]$ and $[y_1, \ldots, y_j]$.

$$\begin{array}{llllll} \ldots & x_{i-3} & x_{i-2} & x_{i-1} & x_i \\ \ldots & y_{j-3} & y_{j-2} & y_{j-1} & y_j \end{array}$$

Possibility	LCS Length

How could x_i and y_j be used in the optimal LCS?

Longest Common Subsequence

$L(i, j)$ = Length of LCS between $[x_1, \ldots, x_i]$ and $[y_1, \ldots, y_j]$.

$$\begin{array}{llllll} \ldots & x_{i-3} & x_{i-2} & x_{i-1} & x_i \\ \ldots & y_{j-3} & y_{j-2} & y_{j-1} & y_j \end{array}$$

Possibility	LCS Length
LCS includes both	
LCS excludes y_j	
LCS excludes x_i	
LCS excludes both	

How could x_i and y_j be used in the optimal LCS?

Longest Common Subsequence

$L(i,j)$ = Length of LCS between $[x_1, \ldots, x_i]$ and $[y_1, \ldots, y_j]$.

$$\begin{array}{llllll} \ldots & x_{i-3} & x_{i-2} & x_{i-1} & x_i \\ \ldots & y_{j-3} & y_{j-2} & y_{j-1} & y_j \end{array}$$

$L(i,j) =$?

Possibility	LCS Length
LCS includes both	
LCS excludes y_j	
LCS excludes x_i	
LCS excludes both	

How could x_i and y_j be used in the optimal LCS?

Longest Common Subsequence

$L(i,j)$ = Length of LCS between $[x_1, \ldots, x_i]$ and $[y_1, \ldots, y_j]$.

$$\begin{array}{llllll} \ldots & x_{i-3} & x_{i-2} & x_{i-1} & x_i \\ \ldots & y_{j-3} & y_{j-2} & y_{j-1} & y_j \end{array} \qquad L(i,j) = ?$$

Possibility	LCS Length
LCS includes both	$1 + L(i-1, j-1)$
LCS excludes y_j	
LCS excludes x_i	
LCS excludes both	

How could x_i and y_j be used in the optimal LCS?

Longest Common Subsequence

$L(i,j)$ = Length of LCS between $[x_1, \ldots, x_i]$ and $[y_1, \ldots, y_j]$.

$$\begin{array}{llllll} \ldots & x_{i-3} & x_{i-2} & x_{i-1} & x_i \\ \ldots & y_{j-3} & y_{j-2} & y_{j-1} & y_j \end{array} \qquad L(i,j) = ?$$

Possibility	LCS Length
LCS includes both	$1 + L(i-1, j-1)$
LCS excludes y_j	$L(i, j-1)$
LCS excludes x_i	
LCS excludes both	

How could x_i and y_j be used in the optimal LCS?

Longest Common Subsequence

$L(i,j)$ = Length of LCS between $[x_1, \ldots, x_i]$ and $[y_1, \ldots, y_j]$.

$$\begin{array}{llllll} \ldots & x_{i-3} & x_{i-2} & x_{i-1} & x_i \\ \ldots & y_{j-3} & y_{j-2} & y_{j-1} & y_j \end{array} \qquad L(i,j) = ?$$

Possibility	LCS Length
LCS includes both	$1 + L(i-1, j-1)$
LCS excludes y_j	$L(i, j-1)$
LCS excludes x_i	$L(i-1, j)$
LCS excludes both	

How could x_i and y_j be used in the optimal LCS?

Longest Common Subsequence

$L(i, j)$ = Length of LCS between $[x_1, \ldots, x_i]$ and $[y_1, \ldots, y_j]$.

$$\begin{array}{llllll} \ldots & x_{i-3} & x_{i-2} & x_{i-1} & x_i \\ \ldots & y_{j-3} & y_{j-2} & y_{j-1} & y_j \end{array} \qquad L(i,j) = ?$$

Possibility	LCS Length
LCS includes both	$1 + L(i-1, j-1)$
LCS excludes y_j	$L(i, j-1)$
LCS excludes x_i	$L(i-1, j)$
LCS excludes both	$L(i-1, j-1)$

How could x_i and y_j be used in the optimal LCS?

Longest Common Subsequence

$L(i,j)$ = Length of LCS between $[x_1, \dots, x_i]$ and $[y_1, \dots, y_j]$.

$$\dots \; x_{i-3} \; x_{i-2} \; x_{i-1} \; x_i$$
$$\dots \; y_{j-3} \; y_{j-2} \; y_{j-1} \; y_j$$

$$L(i,j) = \max \begin{cases} 1 + L(i-1, j-1), \text{ if } x_i = y_j \\ L(i, j-1) \\ L(i-1, j) \\ L(i-1, j-1) \end{cases}$$

Possibility	LCS Length
LCS includes both	$1 + L(i-1, j-1)$
LCS excludes y_j	$L(i, j-1)$
LCS excludes x_i	$L(i-1, j)$
LCS excludes both	$L(i-1, j-1)$

How could x_i and y_j be used in the optimal LCS?

Longest Common Subsequence

$L(i,j)$ = Length of LCS between $[x_1, \dots, x_i]$ and $[y_1, \dots, y_j]$.

$$\dots \; x_{i-3} \; x_{i-2} \; x_{i-1} \; x_i$$
$$\dots \; y_{j-3} \; y_{j-2} \; y_{j-1} \; y_j$$

$$L(i,j) = \max \begin{cases} 1 + L(i-1, j-1), \text{ if } x_i = y_j \\ L(i, j-1) \\ L(i-1, j) \\ L(i-1, j-1) \end{cases}$$

Possibility	LCS Length
LCS includes both	$1 + L(i-1, j-1)$
LCS excludes y_j	$L(i, j-1)$
LCS excludes x_i	$L(i-1, j)$
LCS excludes both	$L(i-1, j-1)$

Could we ever have:

$$L(i, j-1) < L(i-1, j-1)$$

Longest Common Subsequence

$L(i, j)$ = Length of LCS between $[x_1, \dots, x_i]$ and $[y_1, \dots, y_j]$.

$$\dots \; x_{i-3} \; x_{i-2} \; x_{i-1} \; x_i$$
$$\dots \; y_{j-3} \; y_{j-2} \; y_{j-1} \; y_j$$

$$L(i,j) = \max \begin{cases} 1 + L(i-1, j-1), \text{ if } x_i = y_j \\ L(i, j-1) \\ L(i-1, j) \\ L(i-1, j-1) \end{cases}$$

Possibility	LCS Length
LCS includes both	$1 + L(i-1, j-1)$
LCS excludes y_j	$L(i, j-1)$
LCS excludes x_i	$L(i-1, j)$
LCS excludes both	$L(i-1, j-1)$

Could we ever have:

$$L(i, j-1) < L(i-1, j-1)$$

No. The LCS cannot decrease by considering more characters.

Longest Common Subsequence

$L(i,j)$ = Length of LCS between $[x_1, \dots, x_i]$ and $[y_1, \dots, y_j]$.

$$\dots \; x_{i-3} \; x_{i-2} \; x_{i-1} \; x_i$$
$$\dots \; y_{j-3} \; y_{j-2} \; y_{j-1} \; y_j$$

$$L(i,j) = \max \begin{cases} 1 + L(i-1, j-1), \text{ if } x_i = y_j \\ L(i, j-1) \\ L(i-1, j) \\ \sout{L(i-1, j-1)} \end{cases}$$

Possibility	LCS Length
LCS includes both	$1 + L(i-1, j-1)$
LCS excludes y_j	$L(i, j-1)$
LCS excludes x_i	$L(i-1, j)$
~~LCS excludes both~~	~~$L(i-1, j-1)$~~

Could we ever have:

$$L(i, j-1) < L(i-1, j-1)$$

No. The LCS cannot decrease by considering more characters.

Longest Common Subsequence

$L(i,j)$ = Length of LCS between $[x_1, \dots, x_i]$ and $[y_1, \dots, y_j]$.

$$\dots \; x_{i-3} \; x_{i-2} \; x_{i-1} \; x_i$$
$$\dots \; y_{j-3} \; y_{j-2} \; y_{j-1} \; y_j$$

$$L(i,j) = \max \begin{cases} 1 + L(i-1, j-1), \text{ if } x_i = y_j \\ L(i, j-1) \\ L(i-1, j) \\ \sout{L(i-1, j-1)} \end{cases}$$

Possibility	LCS Length
LCS includes both	$1 + L(i-1, j-1)$
LCS excludes y_j	$L(i, j-1)$
LCS excludes x_i	$L(i-1, j)$
~~LCS excludes both~~	~~$L(i-1, j-1)$~~

If $x_i = y_j$, could we ever have:

$$1 + L(i-1, j-1) < L(i-1, j-1)$$

Longest Common Subsequence

$L(i,j)$ = Length of LCS between $[x_1, \dots, x_i]$ and $[y_1, \dots, y_j]$.

$$\begin{array}{llllll} \dots & x_{i-3} & x_{i-2} & x_{i-1} & x_i \\ \dots & y_{j-3} & y_{j-2} & y_{j-1} & y_j \end{array}$$

$$L(i,j) = \max \begin{cases} 1 + L(i-1, j-1), \text{ if } x_i = y_j \\ L(i, j-1) \\ L(i-1, j) \\ \sout{L(i-1, j-1)} \end{cases}$$

Possibility	LCS Length
LCS includes both	$1 + L(i-1, j-1)$
LCS excludes y_j	$L(i, j-1)$
LCS excludes x_i	$L(i-1, j)$
~~LCS excludes both~~	~~$L(i-1, j-1)$~~

If $x_i = y_j$, could we ever have:

$1 + L(i-1, j-1) < L(i-1, j-1)$

No. Adding 1 makes things bigger.

Longest Common Subsequence

$L(i, j)$ = Length of LCS between $[x_1, \ldots, x_i]$ and $[y_1, \ldots, y_j]$.

$$\begin{array}{llllll} \ldots & x_{i-3} & x_{i-2} & x_{i-1} & x_i \\ \ldots & y_{j-3} & y_{j-2} & y_{j-1} & y_j \end{array}$$

$$L(i,j) = \begin{cases} 1 + L(i-1, j-1), \text{ if } x_i = y_j \\ \max\{L(i, j-1), L(i-1, j)\}, \text{ otherwise} \end{cases}$$

Possibility	LCS Length
LCS includes both	$1 + L(i-1, j-1)$
LCS excludes y_j	$L(i, j-1)$
LCS excludes x_i	$L(i-1, j)$

If $x_i = y_j$, could we ever have:

$1 + L(i-1, j-1) < L(i-1, j-1)$

No. Adding 1 makes things bigger.

Longest Common Subsequence

i	j	0	1	2	3	4	5
			S	U	N	N	Y
0							
1	S						
2	N						
3	O						
4	W						
5	Y						

$$L(i,j) = \begin{cases} 1 + L(i-1, j-1), \text{ if } x_i = y_j \\ \max\{L(i, j-1), L(i-1, j)\}, \text{ otherwise} \end{cases}$$

Longest Common Subsequence

i	j	0	1	2	3	4	5
			S	U	N	N	Y
0		0	0	0	0	0	0
1	S	0					
2	N	0					
3	O	0					
4	W	0					
5	Y	0					

$$L(i,j) = \begin{cases} 1 + L(i-1, j-1), \text{ if } x_i = y_j \\ \max\{L(i, j-1), L(i-1, j)\}, \text{ otherwise} \end{cases}$$

Longest Common Subsequence

i	j	0	1	2	3	4	5
			S	U	N	N	Y
0		0	0	0	0	0	0
1	S	0	1	1	1	1	1
2	N	0	1	1	?		
3	O	0	1	1			
4	W	0	1	1			
5	Y	0	1	1			

$$L(i,j) = \begin{cases} 1 + L(i-1, j-1), \text{ if } x_i = y_j \\ \max\{L(i, j-1), L(i-1, j)\}, \text{ otherwise} \end{cases}$$

Longest Common Subsequence

i	j	0	1	2	3	4	5
			S	U	N	N	Y
0		0	0	0	0	0	0
1	S	0	1	1	1	1	1
2	N	0	1	1	?		
3	O	0	1	1			
4	W	0	1	1			
5	Y	0	1	1			

$$L(i,j) = \begin{cases} 1 + L(i-1, j-1), \text{ if } x_i = y_j \\ \max\{L(i, j-1), L(i-1, j)\}, \text{ otherwise} \end{cases}$$

Longest Common Subsequence

i	j	0	1	2	3	4	5
			S	U	N	N	Y
0		0	0	0	0	0	0
1	S	0	1	1	1	1	1
2	N	0	1	1	?		
3	O	0	1	1			
4	W	0	1	1			
5	Y	0	1	1			

$$L(i,j) = \begin{cases} 1 + L(i-1, j-1), \text{ if } x_i = y_j \\ \max\{L(i, j-1), L(i-1, j)\}, \text{ otherwise} \end{cases}$$

$$\boldsymbol{L(2,3) = 1 + L(2-1, 3-1)}$$

Longest Common Subsequence

	j	0	1	2	3	4	5
i			S	U	N	N	Y
0		0	0	0	0	0	0
1	S	0	1	1	1	1	1
2	N	0	1	1	2	2	2
3	O	0	1	1	2	2	2
4	W	0	1	1	2	2	2
5	Y	0	1	1	2	2	3

$$L(i,j) = \begin{cases} 1 + L(i-1, j-1), \text{ if } x_i = y_j \\ \max\{L(i, j-1), L(i-1, j)\}, \text{ otherwise} \end{cases}$$

$$\boldsymbol{L(2,3)} = 1 + \boldsymbol{L(2-1, 3-1)}$$
$$= 1 + 1 = 2$$