

Edit Distance

Kwankamol Nongpong, Ph.D.

CSX3009 Algorithm Design

Department of Computer Science

Assumption University

Minimum Edit Distance

- Given two strings A and B (i.e., two sequences of characters).
- Goal: Find the minimum number of edits to transforms s_1 to s_2 .
- Edit operations include:
 - Insertion
 - Deletion
 - Substitution

Example: Different Alignments

F	O	O	D		
M	O	N	E	Y	
S		S	S	I	

4 edits

	F	O	O	D	
M	O	N	E	Y	
I	S	S	S	S	

5 edits

F	O	O	D		
M	O	N	E	Y	
D	S		S	I	I

5 edits

F		O	O	D	
M	O	N	E	Y	
S	I	S	S	S	

5 edits

F	O		O	D	
M	O	N	E	Y	
S		I	S	S	

4 edits

F	O	O		D	
M	O	N	E	Y	
S		S	I	S	

4 edits

Exercise:

Determine the number of edits in each alignment

I		N	T	E		N	T	I	O	N
E	X			E	C	U	T	I	O	N

? edits

I	N	T	E	N	T	I	O	N
E	X	E	C	U	T	I	O	N

? edits

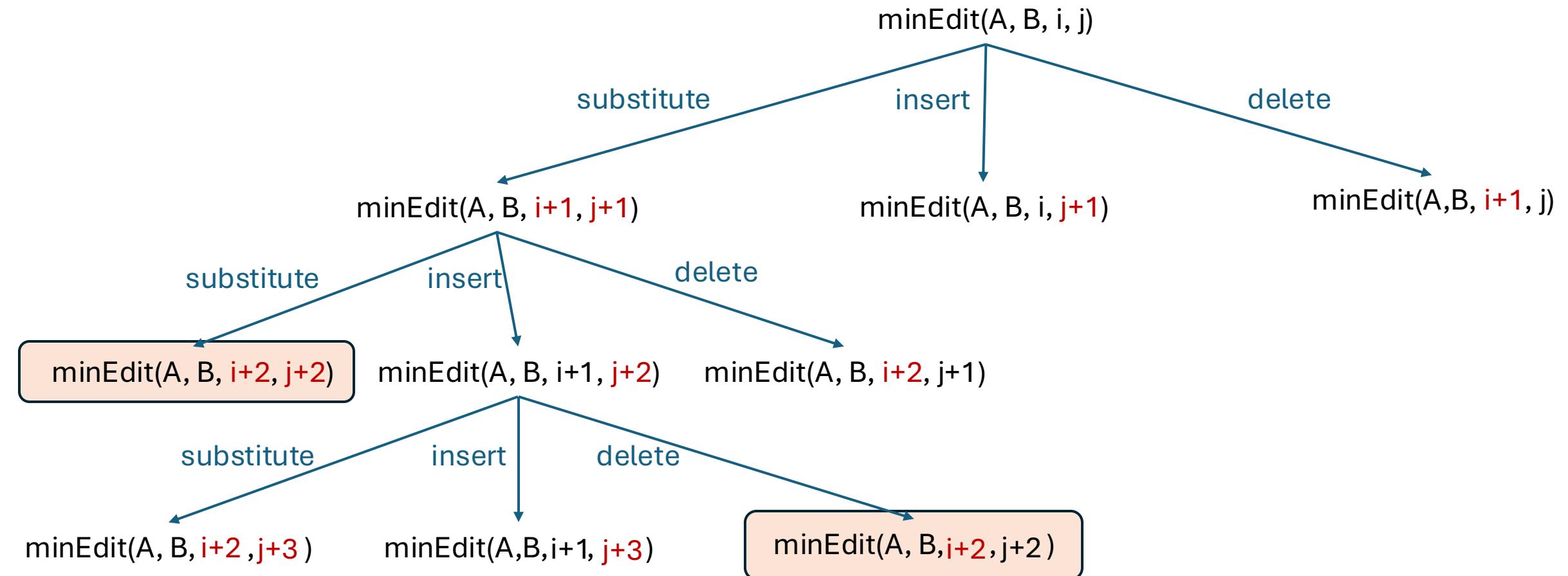
I	N	T	E	N	T	I	O	N
E	X	E	C	U	T	I	O	N

? edits

I	N	T	E	N	T	I	O	N
E	X		E	C	U	T	I	O

? edits

Recursion Tree



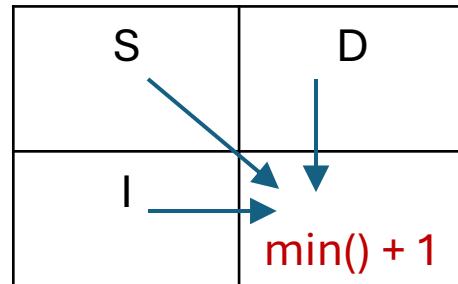
Edit Distance with Memoization

- Your task.

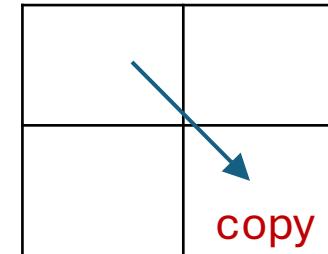
Edit Distance with Dynamic Programming

- Create an $(n+1) \times (m+1)$ table
- where
 - n is the length of A , and
 - m is the length of B

if $A[i] \neq B[j]$



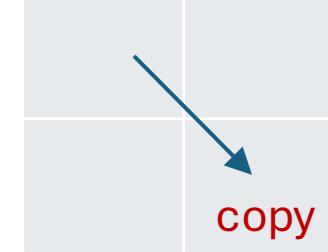
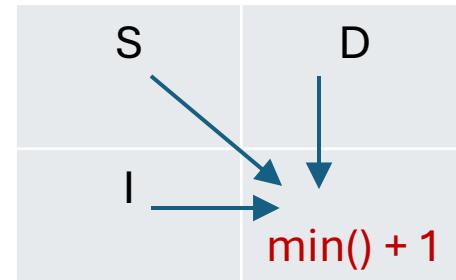
if $A[i] = B[j]$



Example

if $A[i] \neq B[j]$

if $A[i] = B[j]$

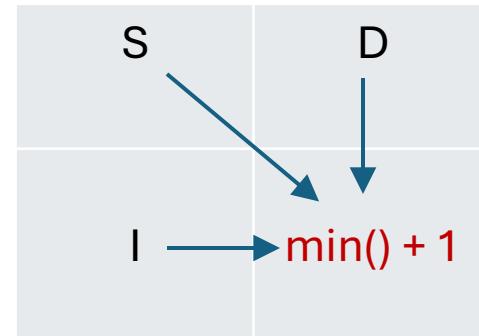


The diagram shows two arrays, A and B, being compared character by character. Array A (the source) has rows labeled F, O, O, D. Array B (the destination) has columns labeled M, O, N, E, Y. The first four rows of A are mapped to the first four columns of B. A curly brace on the left groups the first four rows of A, and another curly brace above B groups the first four columns of B. The fifth row of A (D) is shown below the fourth row of B (E).

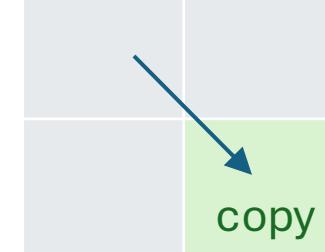
		M	O	N	E	Y
	0	1	2	3	4	5
F	1					
O	2					
O	3					
D	4					

Example

if $A[i] \neq B[j]$



if $A[i] = B[j]$

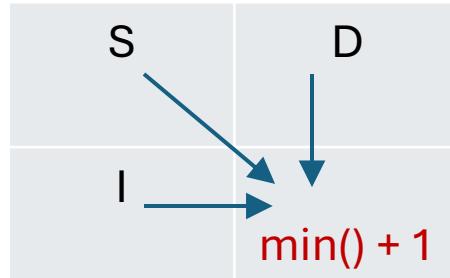


		M	O	N	E	Y
		0	1	2	3	4
A		1	1	2	3	4
F	0	1	2	3	4	5
O	1	2	1	2	3	4
O	2	2	1	2	3	4
D	3	3	2	2	3	4
	4	4	3	3	3	4

The table shows two arrays, A and B, being compared. Array A (left) has elements F, O, O, D. Array B (right) has elements M, O, N, E, Y. The comparison starts at index 0. The first four comparisons result in matches (green cells), while the fifth results in a mismatch (red cell). Arrows indicate the movement of pointers S, D, and I during the comparison process.

Exercise

if $A[i] \neq B[j]$



if $A[i] = B[j]$

