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

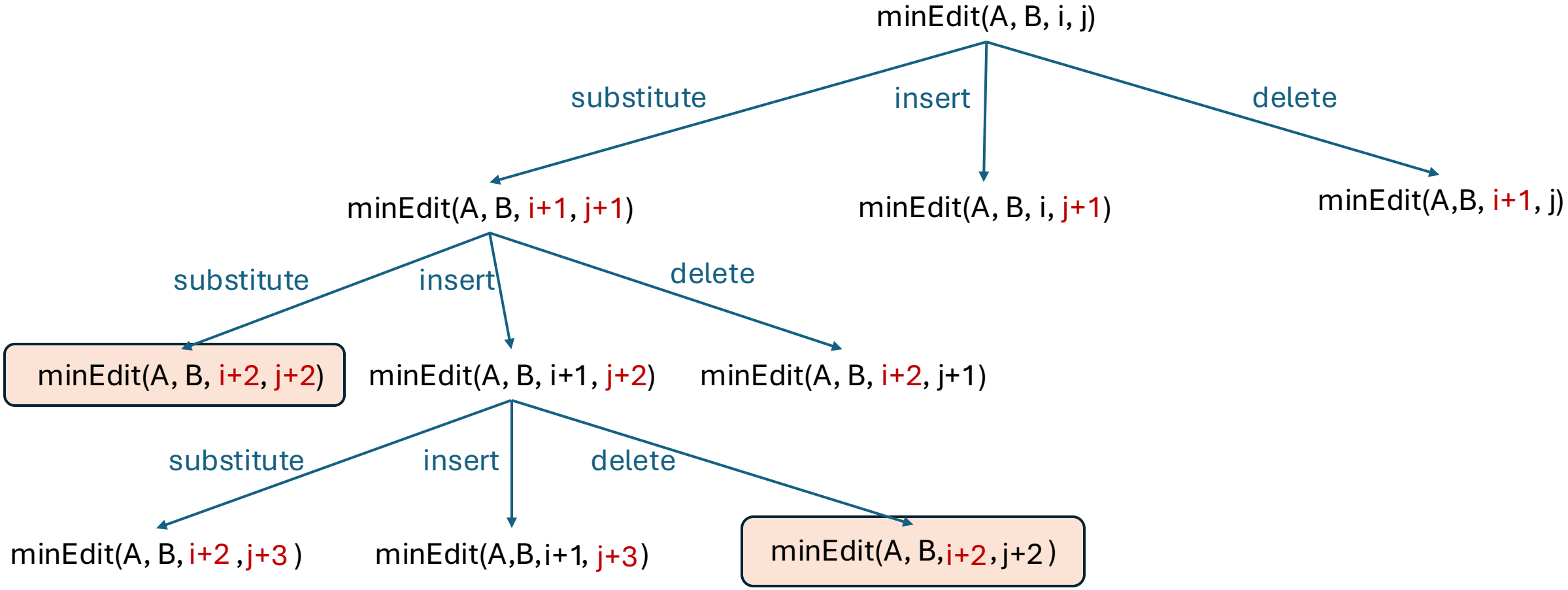
I	N	T	E	N		T	I	O	N
	E	X	E	C	U	T	I	O	N

? edits

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

? edits

Recursion Tree



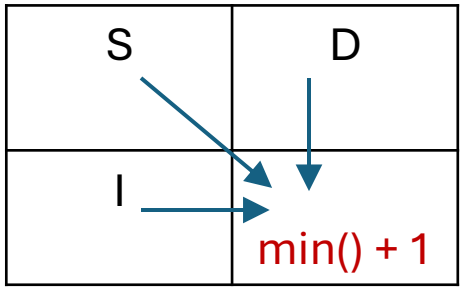
Edit Distance with Memoization

- Your task.

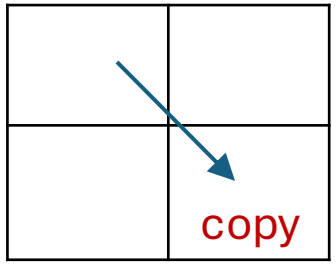
Edit Distance with Dynamic Programming

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

if A[i] ≠ B[j]

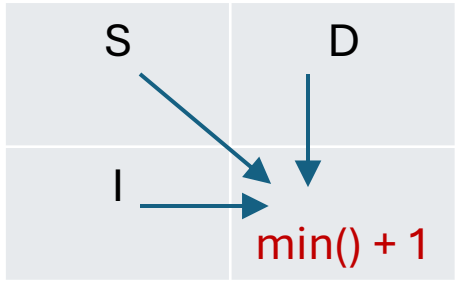


if A[i] = B[j]

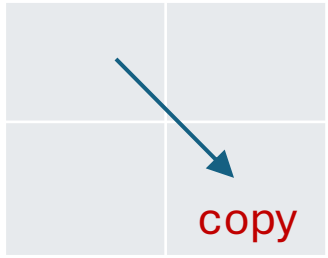


Example

if A[i] ≠ B[j]



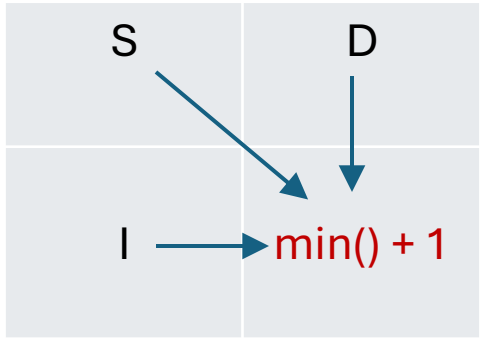
if A[i] = B[j]



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

Example

if A[i] ≠ B[j]



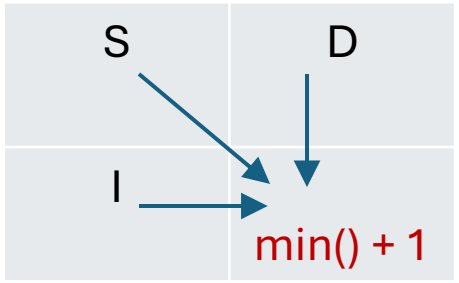
if A[i] = B[j]

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

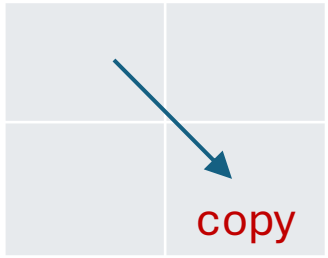
A

Exercise

if A[i] ≠ B[j]



if A[i] = B[j]



		E	X	E	C	U	T	I	O	N
	0	1	2	3	4	5	6	7	8	9
I	1									
N	2									
T	3									
E	4									
N	5									
T	6									
I	7									
O	8									
N	9									