Minimum Edit Distance.

In natural language processing, minimum edit distance is a way to measure how similar two strings are.

It is also known as Levenshtein distance.

How is it calculated?

- The minimum edit distance is the minimum number of operations needed to change one string into another.

- These operations include inserting, deleting, or substituting a character.

- Each operation can be assigned a cost or weight.

- The Levenshtein distance is calculated using dynamic programming.

Ex:-

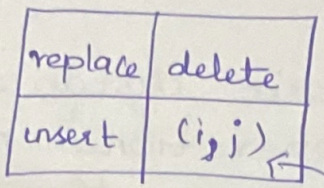
The minimum edit distance between "ab" and "abc" is 1.

→ The Russian mathematician Vladimir Levenshtein first introduced the concept in 1965.

Application :-

↳ Spelling Correction
↳ plagarism detection
↳ DNA Analysis i.e. Computational Biology.

Operations



- Initialization

$D(i, 0) = i$
$D(0, j) = j$

- Recurrence Relation.

for each $i = 1 \ldots M$
for each $j = 1 \ldots N$

$$D(i,j) = \min \begin{cases} D(i-1, j) + 1 \\ D(i, j-1) + 1 \\ D(i-1, j-1) + \begin{cases} 2 ; (\text{if } x(i) \neq y(j)) \\ 0 ; (\text{if } (x(i) = y(j)) \end{cases} \end{cases}$$

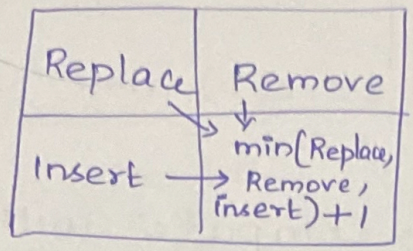
- Termination

$D(N, M)$ is distance

(simple) :-

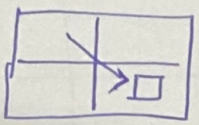
1. If $r \neq c$

r = row
c = column



2. If $r == c$

Just copy the diagram element



convert adceg to abcfg.

	Null	a	b	c	f	g
Null	0	1	2	3	4	5
a	1	0	min(0,1,2)+1 = 1	1+1 = 2	min(3,4,2)+1 = 2+1 = 3	min(3,4,5)+1 = 3+1 = 4
b	2	1	1	2	3	4
c	3	2	1+1 = 2	1	2	3
e	4	3	2+1 = 3	2	1+1 = 2	3
g	5	4	4	3	3	2

row col
Null & Null are same so 0

	Null	a	b	c	f	g
Null	0					
a						
b						
⋮						

Null row → compare with 'a' col ~~so~~ (it is different) so add 1 with 0.
a value gets 1.

	Null	a	b	c	d
Null	0 →	0+1=1 →	1+1=2 →	2+1=3 →	3+1=4
a	0+1=1 ↓				
b	1+1=2				
⋮					

row 'a' compared with col 'a' ⇒ both are same so take diagonal value

	Null	a
Null	0	1
a	1	→0

row 'a' compared with col a & b

	null	a	b	c	f	g
null	0	1	2	3		9
a	1	0 →	1			

$\min(0, 1, 2) + 1$

$= 0 + 1 = 1$

~~=0~~

Here ⇒ a & a are same, b is inserted

row 'a' compared with col a, b & c

	null	a	b	c
null	0	1	2	3
a	1	0	1	2

$$\min(1,2,3)+1$$
$$= 1+1$$
$$= \underline{2}$$

Row a = $\underset{a}{\text{col}}$ are same b & c are inserted.

and so on.