Levenshtein distance to calculate string similarity



Number of edits that will transform one string to another





- Number of edits that will transform one string to another
- These edits are:
 - Insertion
 - Deletion

















t -> c S c i e n t i s t i -> e S c i e n c e



t -> c
S c i e n t i s t
i -> e
s -> delete
t -> delete



t -> c
S c i e n t i s t
i -> e
s -> delete
S c i e n c e
t -> delete







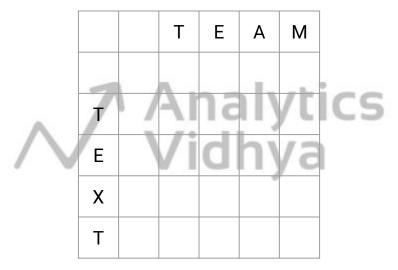








Team





Team

			Т	Е	Α	М
		0	1	2	3	4
		1/	41	٦ē		/[
/	E	2	/i	d	h)	/2
,	Χ	3				
	Т	4				



T E A M

0 1 2 3 4

T 1

E 2

X 3

T 4

Team



T E A M

0 1 2 3 4

T 1

E 2

X 3

T 4

Team



 T
 E
 A
 M

 0
 1
 2
 3
 4

 T
 1
 0
 4

 E
 2
 4
 4

 X
 3
 3
 4

4

Team



T E A M

0 1 2 3 4

T 1 0

E 2

X 3

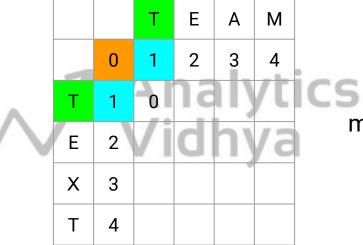
T 4

Team



Team

Text

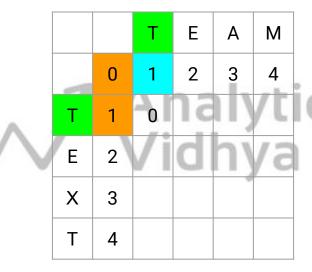


min(0,



Team

Text



min(0, 1+1,



Team



Text

min(0, 1+1, 1+1)



T E A M

0 1 2 3 4

T 1 0

E 2

X 3

T 4

Team



T E A M

0 1 2 3 4

T 1 0 1

E 2

X 3

T 4

Team



Team

		Т	Е	Α	М
	0	1	2	3	4
Т	1/	0	ηē		/[
E	2	۷i	d	h)	/a
X	3				
Т	4				

Text

min(0,1,2) + 1



T E A M
0 1 2 3 4
T 1 0 1
E 2
X 3
T 4

Team



Team

			Т	Ε	Α	М
		0	1	2	3	4
ð	Т	1/	0	1		/[
V	E	2	۷i	d	h)	/a
	X	3				
	Т	4				

Text

min(1,2,3) + 1



Team

		Т	Е	Α	М
	0	1	2	3	4
Т	1/	0	1	2	/[
Ε	2	/i	d	h)	/a
Χ	3				
Т	4				

Text

min(1,2,3) + 1



T E A M

0 1 2 3 4

T 1 0 1 2

E 2

X 3

T 4

Team



Team



T E A M

0 1 2 3 4

T 1 0 1 2 3

E 2 7 7 7

X 3 7 4

Team



Team

			Т	E	Α	М
		0	1	2	3	4
	Т	1	0	٦ē	2	3
d	Е	2	۷i	d	h)	/8
	Χ	3				
	Т	4				

Text

min(2,1,0) + 1



Team

			Т	Е	Α	М
		0	1	2	3	4
	Т	1	0	٦ē	2	3
d	Ε	2	Λi	d	h)	/a
	Χ	3				
	Т	4				

Text

min(2,1,0) + 1



T E A M

0 1 2 3 4

T 1 0 1 2 3

E 2 1 7

X 3

T 4

Team



Team

			Т	Ε	Α	М
		0	1	2	3	4
	Т	1/	0	٦ē	2	3
V	Е	2	Λi	d	h)	/ a
	X	3				
	Т	4				

Text

min(0,1+1,1+1)



Team

			Т	Ε	Α	М
		0	1	2	3	4
	Т	1/	0	٦ē	2	3
1	Е	2	/ 1 i	0	h)	/a
	Χ	3				
	Т	4				

Text

min(0,1+1,1+1)



T E A M

0 1 2 3 4

T 1 0 1 2 3

E 2 1 0

X 3

T 4

Team



T E A M

0 1 2 3 4

T 1 0 1 2 3

E 2 1 0 1

X 3

T 4

Team

Text



 T
 E
 A
 M

 0
 1
 2
 3
 4

 T
 1
 0
 1
 2
 3

 E
 2
 1
 0
 1
 2

 X
 3
 0
 0
 0
 0
 0

4

Team

Text



Team

Text

		Т	Е	Α	М
	0	1	2	3	4
Т	1/	0	٦ē	2	3
Е	2	Λi	0	h)	2
X	3	2	1	1	2
Т	4	3	2	2	2



T E A M

0 1 2 3 4

T 1 0 1 2 3

E 2 1 0 1 2

X 3 2 1 1 2

T 4 3 2 2 2

Team

Text



T E A M

0 1 2 3 4

T 1 0 1 2 3

E 2 1 0 1 2

X 3 2 1 1 2

T 4 3 2 2 2

Team

Text



 T
 E
 A
 M

 0
 1
 2
 3
 4

 T
 1
 0
 1
 2
 3

 E
 2
 1
 0
 1
 2

 X
 3
 2
 1
 1
 2

 T
 4
 3
 2
 2
 2

Team

Text



T E A M

0 1 2 3 4

T 1 0 1 2 3

E 2 1 0 1 2

X 3 2 1 1 2

T 4 3 2 2 2

Team

Text



Text

Text

Total and the second of the seco

Levenshtein distance = 2

3



Text

Text

Text

Total

Text

Total

Total

Text

Total

Levenshtein distance = 2



Text

Text

Text

Total

Text

Total

Text

Text

Total

Text

Levenshtein distance = 2



Text

Text

Total

Tota

Levenshtein distance = 2



Text

0 1 2 3 4

T 1 0 1 2 3

E 2 1 0 1 2

X 3 2 1 1 2

T 4 3 2 2 2

Levenshtein distance = 2

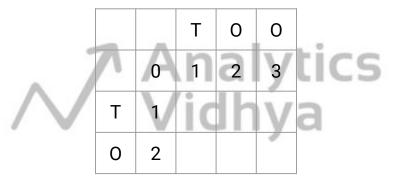


Too



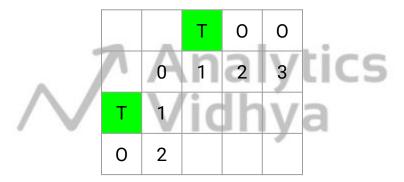


Тоо



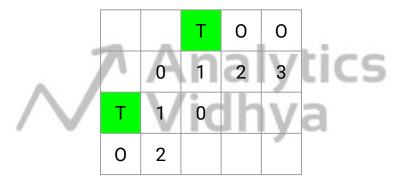


Тоо



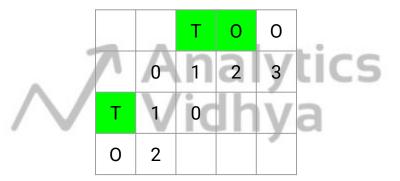


Тоо



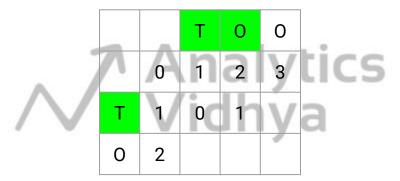


Тоо



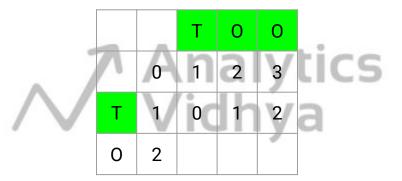


Тоо



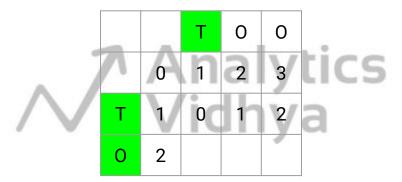


Тоо



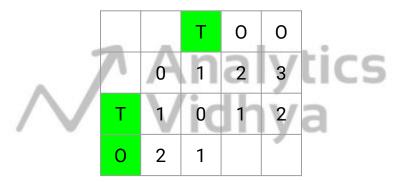


Тоо



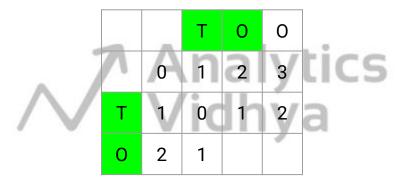


Тоо



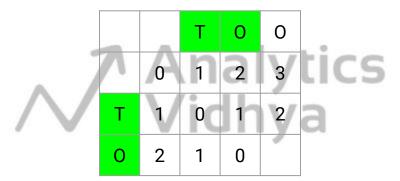


Тоо



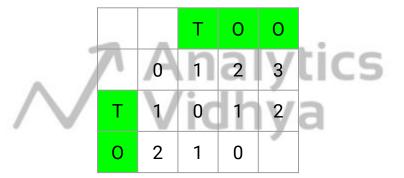


Тоо



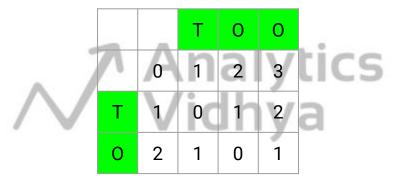


Тоо





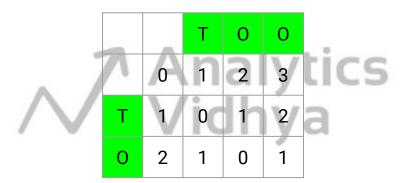
Too





Too

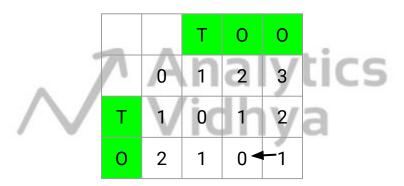
То





Too

То





Too

To

0 1 2 3 T 1 0 1 2 O 2 1 0 ←1

o -> delete



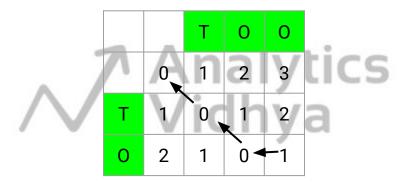
Levenshtein distance = 1

2



Too

То



o -> delete





