

Levenshtein distance to calculate
string similarity

Levenshtein Distance

- Number of edits that will transform one string to another



Levenshtein Distance

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

 Analytics
Vidhya

Levenshtein Distance

Scientist

Science

Levenshtein Distance

t -> c

Scientist

Science

Levenshtein Distance

Scientist

Science

t -> c

i -> e

Levenshtein Distance

Scientist

Science

t -> c

i -> e

s -> delete

Levenshtein Distance

Scientist

Science

t -> c

i -> e

s -> delete

t -> delete

Levenshtein Distance

Scientist

Science

t -> c

i -> e

s -> delete

t -> delete

Levenshtein distance = 4

Levenshtein Distance



T e a m

T e x t

Levenshtein Distance

T e a m

a -> x

T e x t

m -> t

Levenshtein Distance

Team a -> x
Text m -> t

Levenshtein distance = 2

Levenshtein Distance

Team

Text

		T	E	A	M
T					
E					
X					
T					

Levenshtein Distance

Team

Text

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

Levenshtein Distance

Team

Text

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

Levenshtein Distance

Team

Text

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

Levenshtein Distance

Team

Text

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

Levenshtein Distance

Team

Text

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

Levenshtein Distance

Team

Text

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

min(0,

Levenshtein Distance

Team

Text

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

$\min(0, 1+1,$

Levenshtein Distance

Team

Text

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

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

Levenshtein Distance

Team

Text

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

Levenshtein Distance

Team

Text

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

Levenshtein Distance

Team

Text

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

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

Levenshtein Distance

Team

Text

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

Levenshtein Distance

Team

Text

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

$$\min(1, 2, 3) + 1$$

Levenshtein Distance

Team

Text

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

$$\min(1, 2, 3) + 1$$

Levenshtein Distance

Team

Text

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

Levenshtein Distance

Team

Text

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

Levenshtein Distance

Team

Text

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

Levenshtein Distance

Team

Text

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

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

Levenshtein Distance

Team

Text

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

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

Levenshtein Distance

Team

Text

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

Levenshtein Distance

Team

Text

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

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

Levenshtein Distance

Team

Text

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

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

Levenshtein Distance

Team

Text

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

Levenshtein Distance

Team

Text

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

Levenshtein Distance

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				
T	4				

Levenshtein Distance

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

Levenshtein Distance

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

Levenshtein distance = 2

Levenshtein Distance

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

Levenshtein distance = 2

Levenshtein Distance

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

Levenshtein distance = 2

Levenshtein Distance

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

Levenshtein distance = 2

Levenshtein Distance

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

m -> t

Levenshtein distance = 2

Levenshtein Distance

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

m -> t

Levenshtein distance = 2

Levenshtein Distance

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

m -> t

a -> x

Levenshtein distance = 2

Levenshtein Distance

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

m -> t

a -> x

Levenshtein distance = 2

Levenshtein Distance

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

m -> t

a -> x

Levenshtein distance = 2

Levenshtein Distance

To o

To



Levenshtein Distance

To o

To

		T	O	O
	0	1	2	3
T	1			
O	2			

Levenshtein Distance

To o

To

		T	0	0
	0	1	2	3
T	1			
0	2			

Levenshtein Distance

To o

To

		T	0	0
	0	1	2	3
T	1	0		
0	2			

Levenshtein Distance

To o

To

		T	O	O
	0	1	2	3
T	1	0		
O	2			

Levenshtein Distance

To o

To

		T	O	O
	0	1	2	3
T	1	0	1	
O	2			

Levenshtein Distance

T o o

T o

		T	O	O
	0	1	2	3
T	1	0	1	2
O	2			

Levenshtein Distance

To o

To

		T	0	0
	0	1	2	3
T	1	0	1	2
O	2			

Levenshtein Distance

To o

To

		T	0	0
	0	1	2	3
T	1	0	1	2
O	2	1		

Levenshtein Distance

To o

To

		T	O	O
	0	1	2	3
T	1	0	1	2
O	2	1		

Levenshtein Distance

To o

To

		T	O	O
	0	1	2	3
T	1	0	1	2
O	2	1	0	

Levenshtein Distance

To o

To

		T	O	O
	0	1	2	3
T	1	0	1	2
O	2	1	0	

Levenshtein Distance

T o o

T o

		T	O	O
	0	1	2	3
T	1	0	1	2
O	2	1	0	1

Levenshtein Distance

To o

T o

		T	O	O
	0	1	2	3
T	1	0	1	2
O	2	1	0	1

Levenshtein distance = 1

Levenshtein Distance

To o

To

		T	O	O
	0	1	2	3
T	1	0	1	2
O	2	1	0 ← 1	

Levenshtein distance = 1

Levenshtein Distance

T o o

T o

		T	O	O
	0	1	2	3
T	1	0	1	2
O	2	1	0 ← 1	

o -> delete

Levenshtein distance = 1

Levenshtein Distance

T o o

T o

		T	O	O
	0	1	2	3
T	1	0	1	2
O	2	1	0	1

o -> delete

Levenshtein distance = 1

Levenshtein Distance

T o o

T o

		T	O	O
	0	1	2	3
T	1	0	1	2
O	2	1	0	1

o -> delete

Levenshtein distance = 1



Thank You