

Word Similarity Distance

u^b — Mohammad Ghafari

Edit distance (D)

The minimum number of operations required to convert one string to another string.

studnt

→

student

+

studnt

→

studnty

+

+

gold

→

sold

+

+

From “studnt” to “study”

	n	s	t	u	d	y
nul	0	1	2	3	4	5
s	1	0				
t	2					
u	3					
d	4					
n	5					
t	6					

1

From “studnt” to “study”

		n	s	t	u	d	y
n	0	1	2	3	4	5	
s	1	0	1	2	3	4	
t	2	1	0	1	2	3	
u	3	2	1	0	1	2	
d	4	3	2	1	0	1	
n	5	4	3	2	1	1	
t	6	5	4	3	2	2	



5

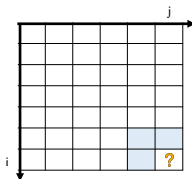
From “studnt” to “study”

		n	s	t	u	d	y
n	0	1	2	3	4	5	
s	1	0	1	2	3	4	
t	2	1	0	1	2	3	
u	3	2	1	0	1	2	
d	4	3	2	1	0	1	
n	5	4	3	2	1	1	
t	6	5	4	3	2	2	



6

Algorithm



$$D(i, j) = \min \begin{cases} D(i, j-1) + 1 & \text{+} \\ D(i-1, j) + 1 & \text{-} \\ D(i-1, j-1) + \delta & \begin{cases} 0 & s_i = s_j \\ 1 & s_i \neq s_j \end{cases} \end{cases}$$

7

Exercise

Implement the similarity distance algorithm, and then compute the similarity between two given sentences.

8