Edit distance

The minimum edit distance is also used to implement both the Q&A problem and the Topic modelling. Given two strings, str1 and str2, and the operations delete, insert and substitute, you can find the minimum number of edits required to convert one string into the other. The delete, insert and substitute are equal cost.

Eg one

*input: str1=”run”, str 2= “ruin”*

output=1

we can convert str1 to str2 by inserting an “*I*”

Eg two

*input: str1=”cat”, str 2= “cut”*

output=1

we can convert str1 to str2 by replacing “*a*” with “*u*”

Eg three

*input: str1=”Execution”, str 2= “intention”*

output=5

Last four characters are the same

we can convert str1 to str2 by inserting an “*I*”, substituiting “*E”* with “*N*”, substituiting “*x*” with “*t*”, deleting c and substituiting “*u*” with “*n*”

Implementation

The idea it to process all characters starting either rfrom left or right side of both strings

Eg

M: length of str1

N: length of str2

If the last two strings are similar, we do nothing, ignore the last character and get count of remaining strings

If the last characters are however not same, consider all the operations on last character of first string and recursively compute cost of all the three operations. Take the minimum of the three values.

1. Insert: Recur for m and n-1
2. Delete: Recur for m-1 and n
3. Substitute: Recur for m-1 and n-1