Text

Description automatically generated

SEMESTER 1, 2021/2022

CSCI 4342 NLP Section 01

NATURAL LANGUAGE PROCESSING

**Exercise 4**

**PREPARED BY:**

|  |  |
| --- | --- |
| **NAME** | **MATRIC NO.** |
| MOHAMED MOUBARAK MOHAMED MISBAHOU MKOUBOI | 1820705 |
| ISLAM MD SHARIFUL | 1720601 |

LECTURER:

Dr. SURIANI BT. SULAIMAN

1. Using the file *distance.py*, compute the distance between the following words:

translating vs traversing

precaution vs prevention

reliability vs responsibility

**Code:**

def iterative\_levenshtein(s, t):

"""

iterative\_levenshtein(s, t) -> ldist

ldist is the Levenshtein distance between the strings

s and t.

For all i and j, dist[i,j] will contain the Levenshtein

distance between the first i characters of s and the

first j characters of t

"""

rows = len(s)+1

cols = len(t)+1

dist = [[0 for x in range(cols)] for x in range(rows)]

# source prefixes can be transformed into empty strings

# by deletions:

for i in range(1, rows):

dist[i][0] = i

# target prefixes can be created from an empty source string

# by inserting the characters

for i in range(1, cols):

dist[0][i] = i

for col in range(1, cols):

for row in range(1, rows):

if s[row-1] == t[col-1]:

cost = 0

else:

cost = 1

dist[row][col] = min(dist[row-1][col] + 1, # deletion

dist[row][col-1] + 1, # insertion

dist[row-1][col-1] + cost) # substitution

for r in range(rows):

print(dist[r])

return dist[row][col]

print("\ntranslating vs traversing")

print(iterative\_levenshtein("translating", "traversing"))

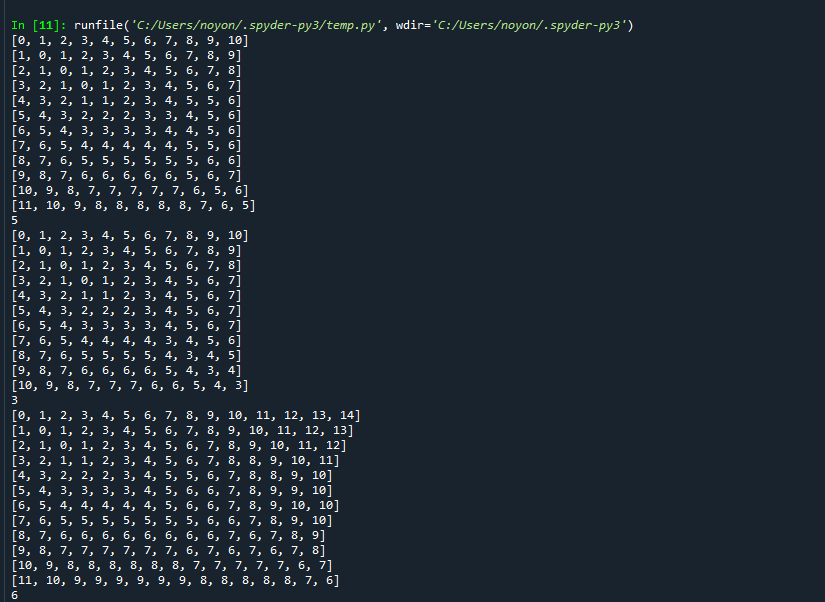
print("\nprecaution vs prevention")

print(iterative\_levenshtein("precaution", "prevention"))

print("\nreliability vs responsibility")

print(iterative\_levenshtein("reliability", "responsibility"))

Output:



1. For each of the word pairs above, show your manual calculation of the operations costs for changing from one word to another using the Levenshtein algorithm

* **translating vs traversing**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **#** | **t** | **r** | **a** | **v** | **e** | **r** | **s** | **i** | **n** | **g** |
| **#** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **t** | **1** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **r** | **2** | **1** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **a** | **3** | **2** | **1** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **n** | **4** | **3** | **2** | **1** | **2** | **3** | **4** | **5** | **6** | **5** | **6** |
| **s** | **5** | **4** | **3** | **2** | **3** | **4** | **5** | **4** | **5** | **6** | **7** |
| **l** | **6** | **5** | **4** | **3** | **4** | **5** | **6** | **5** | **6** | **7** | **8** |
| **a** | **7** | **6** | **5** | **4** | **5** | **6** | **7** | **6** | **7** | **8** | **9** |
| **t** | **8** | **7** | **6** | **5** | **6** | **7** | **8** | **7** | **8** | **9** | **10** |
| **i** | **9** | **8** | **7** | **6** | **7** | **8** | **9** | **8** | **7** | **8** | **9** |
| **n** | **10** | **9** | **8** | **7** | **8** | **9** | **10** | **9** | **8** | **7** | **8** |
| **g** | **11** | **10** | **9** | **8** | **9** | **10** | **11** | **10** | **9** | **8** | **7** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| t | r | a |  |  | n | s | l | a | t | i | n | g |
| t | r | a | v | e | r | s |  |  |  | i | n | g |

* **precaution vs prevention**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **#** | **p** | **r** | **e** | **v** | **e** | **n** | **t** | **i** | **o** | **n** |
| **#** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **p** | **1** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **r** | **2** | **1** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **e** | **3** | **2** | **1** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **c** | **4** | **3** | **2** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **a** | **5** | **4** | **3** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **u** | **6** | **5** | **4** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **t** | **7** | **6** | **5** | **4** | **5** | **6** | **7** | **6** | **7** | **8** | **9** |
| **i** | **8** | **7** | **6** | **5** | **6** | **7** | **8** | **7** | **6** | **7** | **8** |
| **o** | **9** | **8** | **7** | **6** | **7** | **8** | **9** | **8** | **7** | **6** | **7** |
| **n** | **10** | **9** | **8** | **7** | **8** | **9** | **8** | **9** | **8** | **7** | **6** |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| p | r | e |  |  | c | a | u | t | i | o | n |
| p | r | e | v | e | n |  |  | t | i | o | n |

* **reliability vs responsibility**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **#** | **r** | **e** | **s** | **p** | **o** | **n** | **s** | **i** | **b** | **i** | **l** | **i** | **t** | **y** |
| **#** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** |
| **r** | **1** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** |
| **e** | **2** | **1** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| **l** | **3** | **2** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **8** | **9** | **10** | **11** |
| **i** | **4** | **3** | **2** | **3** | **4** | **5** | **6** | **7** | **6** | **7** | **8** | **9** | **8** | **9** | **10** |
| **a** | **5** | **4** | **3** | **4** | **5** | **6** | **7** | **8** | **7** | **8** | **9** | **10** | **9** | **10** | **11** |
| **b** | **6** | **5** | **4** | **5** | **6** | **7** | **8** | **9** | **8** | **7** | **8** | **9** | **10** | **11** | **12** |
| **i** | **7** | **6** | **5** | **6** | **7** | **8** | **9** | **10** | **9** | **8** | **7** | **8** | **9** | **10** | **11** |
| **l** | **8** | **7** | **6** | **7** | **8** | **9** | **10** | **11** | **10** | **9** | **8** | **7** | **8** | **9** | **10** |
| **i** | **9** | **8** | **7** | **8** | **9** | **10** | **11** | **12** | **11** | **10** | **9** | **8** | **7** | **8** | **9** |
| **t** | **10** | **9** | **8** | **9** | **10** | **11** | **12** | **13** | **12** | **11** | **10** | **9** | **8** | **7** | **8** |
| **y** | **11** | **10** | **9** | **10** | **11** | **12** | **13** | **14** | **13** | **12** | **11** | **10** | **9** | **8** | **7** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| r | e |  |  |  |  | l | i | a | b | i | l | i | t | y |
| r | e | s | p | o | n | s | i |  | b | i | l | i | t | y |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |