Nooreldean Koteb
Professor Anatasopoulos
CS499 NLP
February 11, 2021

Part 4

Credit 1:

```
[[0, 1, 2, 3, 4, 5],
 [1, 0, 1, 2, 3, 4],
 [2, 1, 0, 1, 2, 3],
 [3, 2, 1, 0, 1, 2],
 [4, 3, 2, 1, 0, 1]]
```
Edit distance Matrix:

Credit 2:
Average edit distance: 2.5560802833530105

Min distance: 0 **Pairs with minimum:** 396

Max distance: 10 **Pairs with maximum:** 46

Credit 3:
Ankara edit distance: 0

Ankara Province edit distance: 10

How do you explain the second value?

 The word province and its grammatical affiliation with (Ankara) are different in Kinyarwanda. The name "Ankara" however remains the same.

Credit 4:
Country at distance 5: Australia | Ositaraliya

Country at distance 5 Matrix:

[[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11],

[1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11],

 [2, 2, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11],

[3, 3, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11],

[4, 4, 3, 3, 3, 4, 5, 6, 7, 8, 9, 10],

 [5, 5, 4, 4, 4, 4, 4, 5, 6, 7, 8, 9],

[6, 6, 5, 5, 5, 4, 5, 4, 5, 6, 7, 8],

[7, 7, 6, 6, 6, 5, 5, 5, 4, 5, 6, 7],

[8, 8, 7, 6, 7, 6, 6, 6, 5, 4, 5, 6],

[9, 9, 8, 7, 7, 7, 7, 6, 6, 5, 5, 5]]

Back-traced:

[(0, 0), (1, 1), (2, 2), (3, 2), (4, 3), (4, 4), (5, 5), (5, 6), (6, 7), (7, 8), (8, 9), (9, 10), (9, 11)]

Credit 5:

Most Common inserted letter: a Inserted: 166 times.

Second Common inserted letter: i Inserted: 122 times.

Credit 6:
Modified Cost Average edit distance: 2.47874852420307

References:
The NLTK code was used in the creation of edit distance and backtracing.
https://www.nltk.org/_modules/nltk/metrics/distance.html

I also found code from my grepper extension provided by an anonymous person. This was used in edit
distance as well. https://www.codegrepper.com/app/profile.php?id=114123