CMPE493 - Information Retrieval - Assignment 1

Berk Atıl - 2016400102

Introduction

In this assignment, we are asked to write a program that calculates the Levenshtein and Damerau-Levenshtein Distances between given two strings. In addition to distance values, it is also asked to show edit tables and needed operations to transform $string_1$ into $string_2$. In the next section, 3 sample inputs and outputs of the program is provided.

Sample Inputs - Outputs

First example

Input: berk, kreb
Output :

```
berk@berk:~/Desktop/Course Books and Slides/CMPE493/InformationRetrieval/Assignment 1$ python3 assignment1.py berk kreb
List of operations needed to trasform berk into kreb: replace b to k, delete e, copy r, replace k to e, insert b
Levenshtein edit distance is 4
    k r e b
  0 1 2 3 4
b 1 1 2 3 3
e 2 2 2 2 3
r 3 3 2 3 3
k 4 3 3 3 4

List of operations needed to trasform berk into kreb: replace b to k, replace e to r, transpose r and k
Damerau-Levenshtein edit distance is 3
    k r e b
  0 1 2 3 4
b 1 1 2 3 3
e 2 2 2 2 3
r 3 3 2 2 3
k 4 3 3 3 3
```

Figure 1: Sample Input - Output 1

Second example

Input: oslo,snow
Output :

```
berk@berk:~/Desktop/Course Books and Slides/CMPE493/InformationRetrieval/Assignment 1$ python3 assignment1.py oslo snow
List of operations needed to trasform oslo into snow: delete o, copy s, replace l to n, copy o, insert w
Levenshtein edit distance is 3
    s n o w
  0 1 2 3 4
o 1 1 2 2 3
s 2 1 2 3 3
l 3 2 2 3 4
o 4 3 3 2 3

List of operations needed to trasform oslo into snow: delete o, copy s, transpose l and o, insert w
Damerau-Levenshtein edit distance is 3
    s n o w
  0 1 2 3 4
o 1 1 2 2 3
s 2 1 2 3 3
l 3 2 2 3 4
o 4 3 3 2 3
```

Figure 2: Sample Input - Output 2

Last example

Input: as gbkl, sa bgsk
Output :

```
berk@berk:~/Desktop/Course Books and Slides/CMPE493/InformationRetrieval/Assignment 1$ python3 assignment1.py "as gbkl" "sa bgsk"
List of operations needed to trasform as gbkl into sa bgsk: delete a, copy s, insert a, copy  , insert b, copy g, replace b to s, copy k, delete l
Levenshtein edit distance is 5
    s a   b g s k
  0 1 2 3 4 5 6 7
a 1 1 1 2 3 4 5 6
s 2 1 2 2 3 4 4 5
  3 2 2 2 3 4 5 5
g 4 3 3 3 3 3 4 5
b 5 4 4 4 3 4 4 5
k 6 5 5 5 4 4 5 4
l 7 6 6 6 5 5 5 5

List of operations needed to trasform as gbkl into sa bgsk: transpose a and s, copy  , insert b, copy g, transpose b and k, delete l
Damerau-Levenshtein edit distance is 4
    s a   b g s k
  0 1 2 3 4 5 6 7
a 1 1 1 2 3 4 5 6
s 2 1 1 2 3 4 4 5
  3 2 2 1 2 3 4 5
g 4 3 3 2 2 2 3 4
b 5 4 4 3 2 2 3 4
k 6 5 5 4 3 3 3 3
l 7 6 6 5 4 4 4 4
```

Figure 3: Sample Input - Output 3