REPORT

The smith-water algorithm is used for searching in files and also searching in given strings.

Substitution, deletion, and insertion cost are:

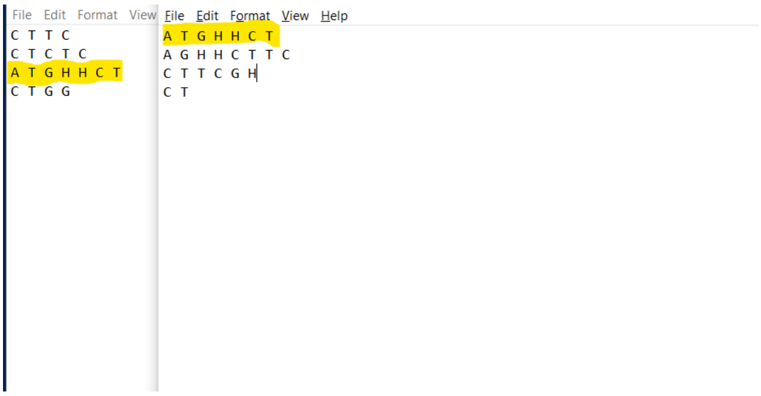
substitutionMatchReward = 3 → when the words to be replaced are the same

substitutionNotMatchPenalty = -3 → when the words to be replaced are not the same

horizontalMovementPenalty = -2 →Insertion

verticalMovementPenalty = -2 → Deletion

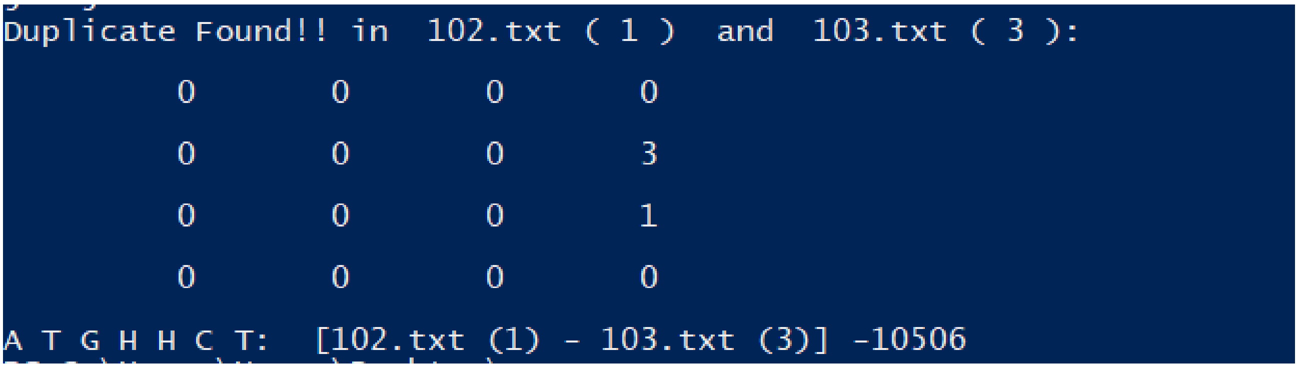
When these files given the function:



```
C T T C
C T C T C
A T G H H C T
C T G G
```

```
A T G H H C T
A G H H C T T C
C T T C G H
C T
```

It will produce this output:



```
Duplicate Found!! in  102.txt ( 1 )  and  103.txt ( 3 ):
         0       0       0       0
         0       0       0       3
         0       0       0       1
         0       0       0       0
A T G H H C T:  [102.txt (1) - 103.txt (3)] -10506
```

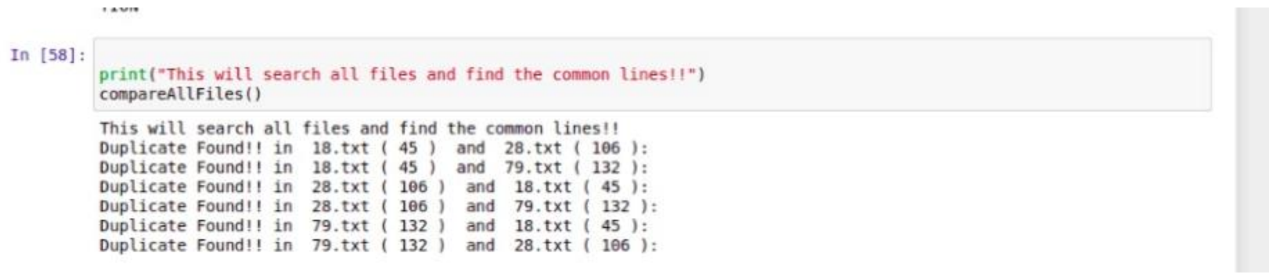
We can also look commons in all files:

18.txt-> 45

28.txt-> 106

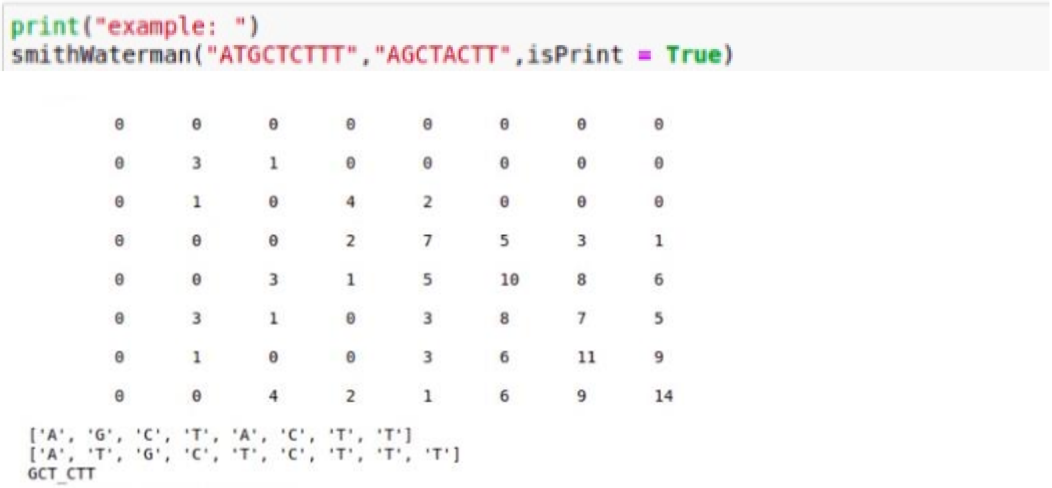
79.txt-> 132

The line is: “Dene! Denenmemiş her bilgi, yanlış ile doğru arasında asılı kalır. (Cezeri)”



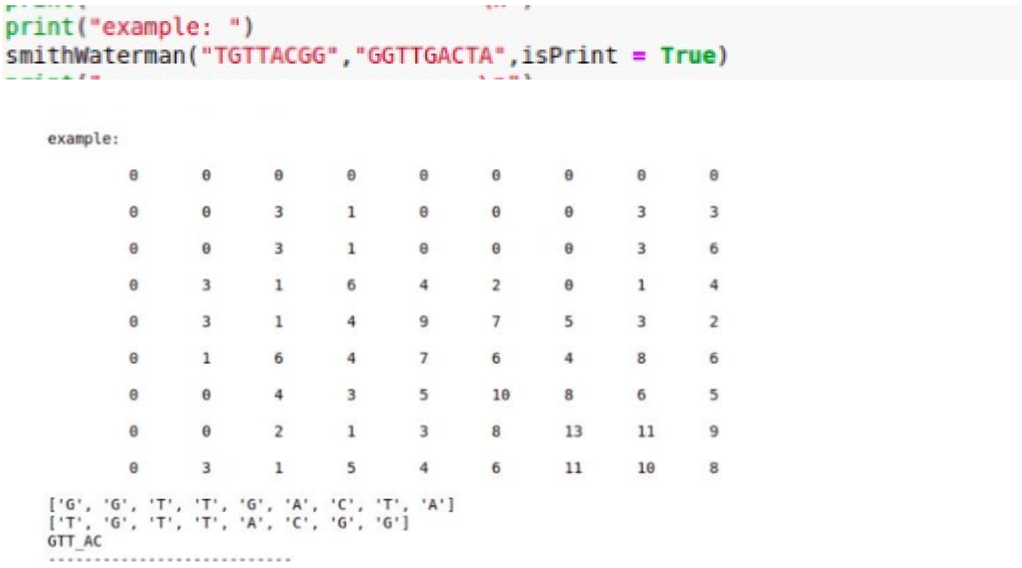
```
In [58]:
print("This will search all files and find the common lines!!")
compareAllFiles()

This will search all files and find the common lines!!
Duplicate Found!! in  18.txt ( 45 )  and  28.txt ( 106 ):
Duplicate Found!! in  18.txt ( 45 )  and  79.txt ( 132 ):
Duplicate Found!! in  28.txt ( 106 )  and  18.txt ( 45 ):
Duplicate Found!! in  28.txt ( 106 )  and  79.txt ( 132 ):
Duplicate Found!! in  79.txt ( 132 )  and  18.txt ( 45 ):
Duplicate Found!! in  79.txt ( 132 )  and  28.txt ( 106 ):
```

Example Executions:



```
print("example: ")
smithWaterman("ATGCTCTTT","AGCTACTT",isPrint = True)
```

```
        0       0       0       0       0       0       0       0

        0       3       1       0       0       0       0       0

        0       1       0       4       2       0       0       0

        0       0       0       2       7       5       3       1

        0       0       3       1       5       10      8       6

        0       3       1       0       3       8       7       5

        0       1       0       0       3       6       11      9

        0       0       4       2       1       6       9       14

['A', 'G', 'C', 'T', 'A', 'C', 'T', 'T']
['A', 'T', 'G', 'C', 'T', 'C', 'T', 'T', 'T']
GCT_CTT
```



```
print("example: ")
smithWaterman("TGTTACGG","GGTTGACTA",isPrint = True)
```

```
example:
        0       0       0       0       0       0       0       0       0

        0       0       3       1       0       0       0       3       3

        0       0       3       1       0       0       0       3       6

        0       3       1       6       4       2       0       1       4

        0       3       1       4       9       7       5       3       2

        0       1       6       4       7       6       4       8       6

        0       0       4       3       5       10      8       6       5

        0       0       2       1       3       8       13      11      9

        0       3       1       5       4       6       11      10      8

['G', 'G', 'T', 'T', 'G', 'A', 'C', 'T', 'A']
['T', 'G', 'T', 'T', 'A', 'C', 'G', 'G']
GTT_AC
---------------------------
```

Hacı Hasan Savan

1901042704