CSE 654 HW1 Muhammed Alperen Karaçete 171044052

Firstly I downloaded 20 documents and deleted all empty lines in this documents. Then with split file algortihm I split 20 documents to 100 documents that they have all maximum 400 line.Then with insert new line algorithm, I inserted all 100 documents a string between 20-30 times.With random insert code, that string placed all of text randomly.Then I take 2 text file to 2 array and then I sended this array to Smith-Waterman Algorithm.In this algorithm, I finded lines which are same and I hide this same line coordinates in icoor array and jcoor array.Then I printed the lines that have same input and that inputs coordinates.Then I print with icoor and jcoor submatrix that is created by algorithm.

I defined:

Substitution as -2,

Deletion as -2,

Insertion as -2.

Some matching examples:

This is line match in physics1.txt and physics3.txt

```
Matched lines are
1:  Dünyamız da dâhil olmak üzere ,j:Coordinate:  366
2:  Dünyamız da dâhil olmak üzere ,i:Coordinate:  414
[[0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 3. 1. 0. 0.]
 [0. 0. 0. 0. 0. 0. 1. 1. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]]
*******************************************************
```

This is match in same text match as previous text files,I inserted this line all of the documents.

```
 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Matched lines are
1:  ThisIsTestLineForAlgorithm
 ,j:Coordinate:  81
2:  ThisIsTestLineForAlgorithm
 ,i:Coordinate:  238
[[0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 3. 3. 1. 0.]
 [0. 0. 0. 0. 0. 0. 3. 6. 4. 2.]
 [0. 0. 0. 0. 0. 0. 1. 4. 4. 2.]
 [0. 0. 0. 0. 0. 0. 0. 2. 2. 2.]]
 ****************************************************
```

Match in basic_religious_knowledge1.txt and basic_religious_knowledge3.txt

```
 ****************************************************************
Matched lines are
1:  İSLAM VE SOSYAL HAYAT
 ,j:Coordinate:  105
2:  İSLAM VE SOSYAL HAYAT
 ,i:Coordinate:  421
[[0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 3. 1. 0. 0. 0.]
 [0. 0. 0. 0. 0. 1. 6. 4. 2. 0.]
 [0. 0. 0. 0. 0. 0. 4. 4. 2. 0.]
 [0. 0. 0. 0. 0. 0. 2. 2. 2. 0.]
 [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]]
 ****************************************************
```

This is match in basic_religious_knowledge1.txt and physics1.txt.All first text files has 'İstiklal Marşı' and 'Atatürk'ün Gençliğe Hitabesi' so matrix look like this.

```
Matched lines are
1:  Mehmet Âkif ERSOY
 ,j:Coordinate:  64
2:  Mehmet Âkif ERSOY
 ,i:Coordinate:  58
[[125. 123. 121. 119. 117. 115. 113. 111. 109. 107.]
 [130. 128. 126. 124. 122. 120. 118. 116. 114. 112.]
 [128. 133. 131. 129. 127. 125. 123. 121. 119. 117.]
 [126. 131. 136. 134. 132. 130. 128. 126. 124. 122.]
 [124. 129. 134. 139. 137. 135. 133. 131. 129. 127.]
 [122. 127. 132. 137. 142. 140. 138. 136. 134. 132.]
 [120. 125. 130. 135. 140. 140. 143. 141. 139. 137.]
 [118. 123. 128. 133. 138. 138. 141. 146. 144. 142.]
 [116. 121. 126. 131. 136. 136. 139. 144. 144. 142.]
 [114. 119. 124. 129. 134. 139. 137. 142. 142. 142.]]
***************************************************************
```