**CME4408 INTRODUCTION TO NATURAL LANGUAGE PROCESSING**

**Assignment 2**

**Name:** Furkan Kayar

**Number:** 2016510110

## Definition of Algorithm

### Methods

ReadAllLines method in FileManager class is implemented to read text file. Since the text file contains Turkish characters, file is opened with ISO-8859-9 encoding. This method takes file name as a parameter and returns list of words.

CalculateMED method in MEDCalculator class creates distance matrix according to rule below:

AnalyzeDistanceMatrix method finds operation to transform given words each other by using algorithm below:

*x <- length of word 1*

*y <- length of word 2*

***while*** *x is not 0 and y is not 0*

***if*** *min(matrix(x-1, y-1), matrix(x, x-1), matrix(x-1, y)) = matrix(x-1, y-1)*

***and*** *(matrix(x-1, y-1) = matrix(x, y)* ***or*** *matrix(x-1, y-1) = matrix(x, y) -1)*

***if*** *matrix(x, y) = matrix(x, y) – 1*

*Substitute word1[x] with word2[y]*

***else***

*No operation*

*x <- x-1*

*y <- y-1*

***else if*** *matrix(x, y-1) <= matrix(x-1, y)* ***and*** *(matrix(x-1, y) = matrix(x, y)*

***or*** *matrix(x-1, y) == matrix(x, y) -1)*

*Delete word1[x]*

*x <- x-1*

***else***

*Insert word2[y]*

*y = y-1*

MainLoop method in UserIntarface class manages program flow and user input, output. It also measures the runtimes of algorithms.

PrintMEDModel method in UserInterface class plots distance matrix with colorized cells and shows the operations to transform word to target word.

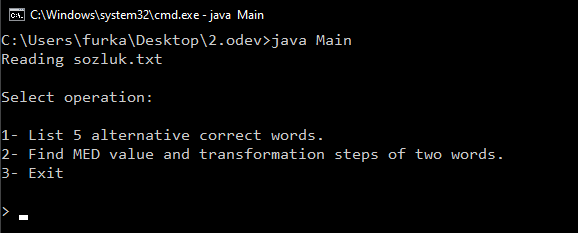
#### Inputs

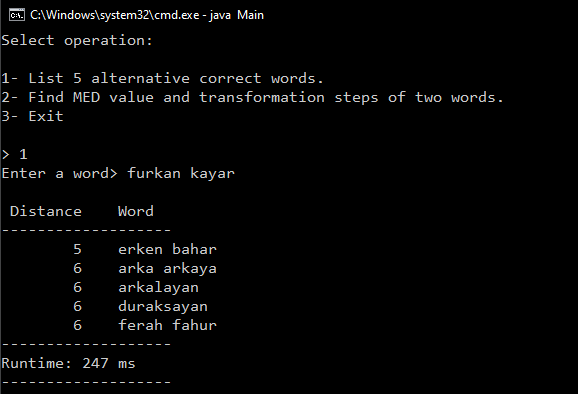
Program requires “sozluk.txt” file that is located at same location with the program. In the main loop, user should input to choose which operation to be applied. After operation is chosen, user should enter words.

#### Outputs

Program prints the minimum edit distance matrix with colorized cells that shows the path from beginning to end. Program also prints the 5 nearest word to the given word and runtimes of algorithms.

## Screenshots







## Test Results

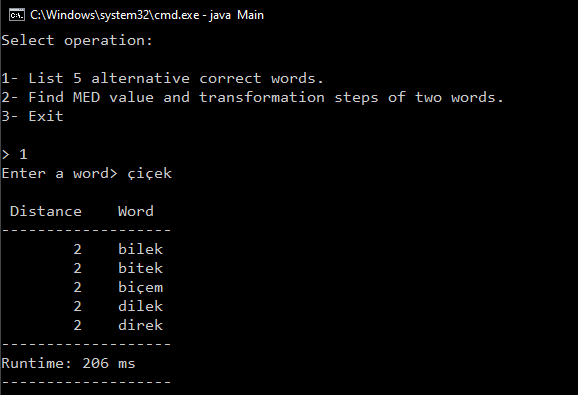
### Running Times

|  |  |
| --- | --- |
| **Part 1** | 247ms |
| **Part 2** | 25ms |

### Part 1 Results

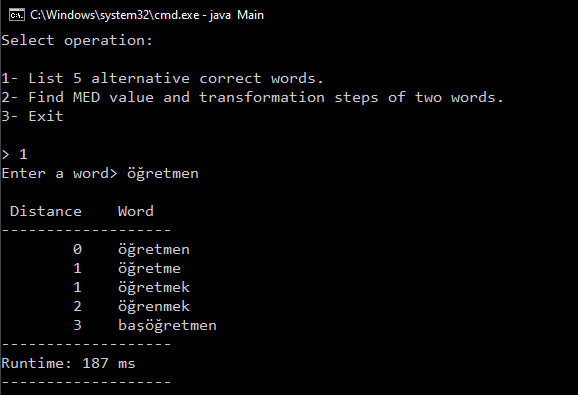
#### Word 1: Çiçek

|  |  |
| --- | --- |
| **Distance** | **Word** |
| 2 | bilek |
| 2 | bitek |
| 2 | biçem |
| 2 | dilek |
| 2 | direk |



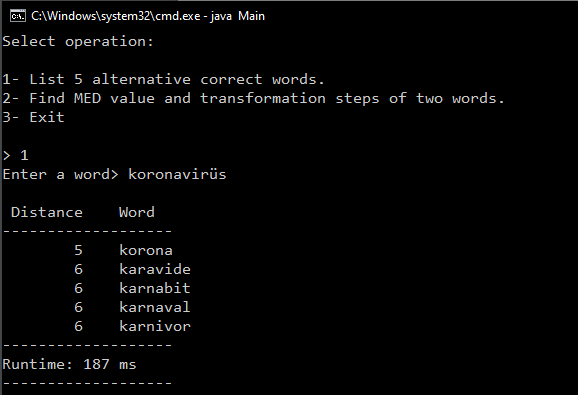
#### Word 2: Öğretmen

|  |  |
| --- | --- |
| **Distance** | **Word** |
| 0 | öğretmen |
| 1 | öğretme |
| 1 | öğretmek |
| 2 | öğrenmek |
| 3 | başöğretmen |



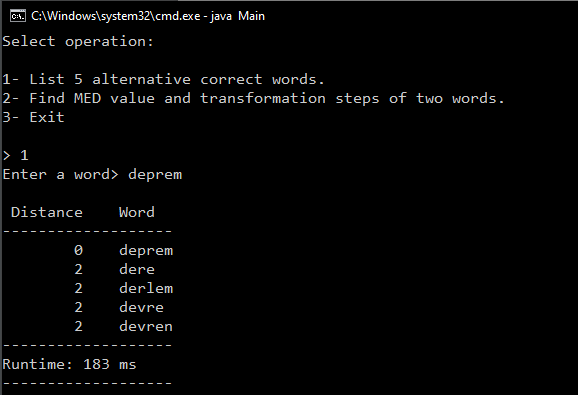
#### Word 3: Koronavirüs

|  |  |
| --- | --- |
| **Distance** | **Word** |
| 5 | korona |
| 6 | karavide |
| 6 | karnabit |
| 6 | karnaval |
| 6 | karnivor |



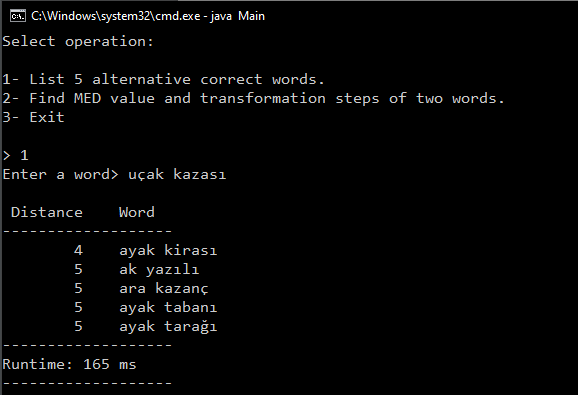
#### Word 4: Deprem

|  |  |
| --- | --- |
| **Distance** | **Word** |
| 0 | deprem |
| 2 | dere |
| 2 | derlem |
| 2 | devre |
| 2 | devren |



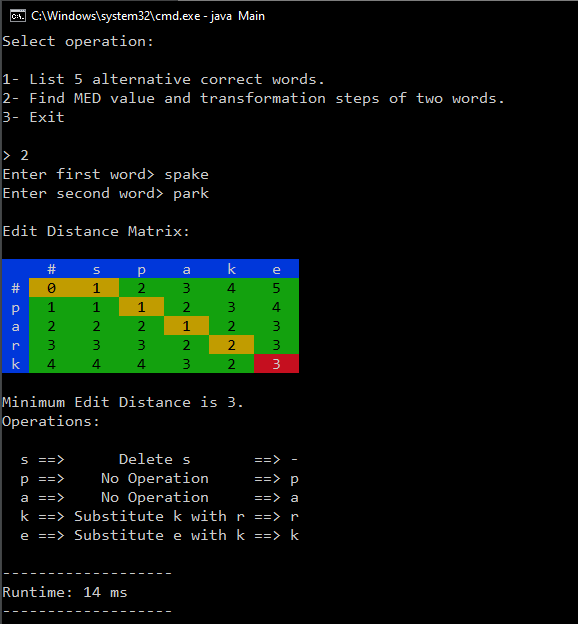
|  |  |
| --- | --- |
| **Distance** | **Word** |
| 4 | ayak kirası |
| 5 | ak yazılı |
| 5 | ara kazanç |
| 5 | ayak tabanı |
| 5 | ayak tarağı |

#### Word 5: Uçak Kazası



### Part 2 Results

#### Example 1: Spake and Park



#### Example 2: Firm and Factory

