Contemporary Data Processing Technology (CCOD)

Lab 10 (October 28, 2016)

Zayats Igor, AS-36

In this work I took the ten Russian letters: **Б, В, Е, О, С, Э, И, Н, У, Ж.**

$\alpha = 0.5$

Table of similarity percentage of letters:

R0

	Б	В	Е	О	С	Э	И	Н	У	Ж
Б	1	0.8	0.7	0.4	0.1	0.1	0.2	0.3	0.1	0.1
В	0.8	1	0.6	0.2	0.1	0.1	0.2	0.2	0.1	0.1
Е	0.7	0.6	1	0.1	0.1	0.1	0.2	0.4	0.1	0.1
О	0.4	0.2	0.1	1	0.9	0.8	0.1	0.1	0.1	0.1
С	0.1	0.1	0.1	0.9	1	0.6	0.1	0.1	0.1	0.1
Э	0.1	0.1	0.1	0.8	0.6	1	0.1	0.2	0.1	0.1
И	0.2	0.2	0.2	0.1	0.1	0.1	1	0.8	0.4	0.1
Н	0.3	0.2	0.4	0.1	0.1	0.2	0.8	1	0.4	0.2
У	0.1	0.1	0.1	0.1	0.1	0.1	0.4	0.4	1	0.1
Ж	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	1

There will be 4 steps of modifying this table:

R1

	Б	В	Е	О	С	Э	И	Н	У	Ж
Б	1	0.8	0.7	0.4	0.4	0.4	0.3	0.4	0.3	0.2
В	0.8	1	0.7	0.4	0.2	0.2	0.2	0.4	0.2	0.2
Е	0.7	0.7	1	0.4	0.1	0.2	0.4	0.4	0.4	0.2
О	0.4	0.4	0.4	1	0.9	0.8	0.2	0.3	0.1	0.1
С	0.4	0.2	0.1	0.9	1	0.8	0.1	0.2	0.1	0.1
Э	0.4	0.2	0.2	0.8	0.8	1	0.2	0.2	0.2	0.2
И	0.3	0.2	0.4	0.2	0.1	0.2	1	0.8	0.4	0.2
Н	0.4	0.4	0.4	0.3	0.2	0.2	0.8	1	0.4	0.2
У	0.3	0.2	0.4	0.1	0.1	0.2	0.4	0.4	1	0.2
Ж	0.2	0.2	0.2	0.1	0.1	0.2	0.2	0.2	0.2	1

R2

	Б	В	Е	О	С	Э	И	Н	У	Ж
Б	1	0.8	0.7	0.4	0.4	0.4	0.4	0.4	0.4	0.2
В	0.8	1	0.7	0.4	0.4	0.4	0.4	0.4	0.4	0.2
Е	0.7	0.7	1	0.4	0.4	0.4	0.4	0.4	0.4	0.2
О	0.4	0.4	0.4	1	0.9	0.8	0.4	0.4	0.4	0.2
С	0.4	0.4	0.4	0.9	1	0.8	0.3	0.4	0.3	0.2
Э	0.4	0.4	0.4	0.8	0.8	1	0.3	0.4	0.3	0.2
И	0.4	0.4	0.4	0.4	0.3	0.3	1	0.8	0.4	0.2
Н	0.4	0.4	0.4	0.4	0.4	0.4	0.8	1	0.4	0.2
У	0.4	0.4	0.4	0.4	0.3	0.3	0.4	0.4	1	0.2
Ж	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	1

R3										
	Б	В	Е	О	С	Э	И	Н	У	Ж
Б	1	0.8	0.7	0.4	0.4	0.4	0.4	0.4	0.4	0.2
В	0.8	1	0.7	0.4	0.4	0.4	0.4	0.4	0.4	0.2
Е	0.7	0.7	1	0.4	0.4	0.4	0.4	0.4	0.4	0.2
О	0.4	0.4	0.4	1	0.9	0.8	0.4	0.4	0.4	0.2
С	0.4	0.4	0.4	0.9	1	0.8	0.4	0.4	0.4	0.2
Э	0.4	0.4	0.4	0.8	0.8	1	0.4	0.4	0.4	0.2
И	0.4	0.4	0.4	0.4	0.4	0.4	1	0.8	0.4	0.2
Н	0.4	0.4	0.4	0.4	0.4	0.4	0.8	1	0.4	0.2
У	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	1	0.2
Ж	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	1

R4										
	Б	В	Е	О	С	Э	И	Н	У	Ж
Б	1	0.8	0.7	0.4	0.4	0.4	0.4	0.4	0.4	0.2
В	0.8	1	0.7	0.4	0.4	0.4	0.4	0.4	0.4	0.2
Е	0.7	0.7	1	0.4	0.4	0.4	0.4	0.4	0.4	0.2
О	0.4	0.4	0.4	1	0.9	0.8	0.4	0.4	0.4	0.2
С	0.4	0.4	0.4	0.9	1	0.8	0.4	0.4	0.4	0.2
Э	0.4	0.4	0.4	0.8	0.8	1	0.4	0.4	0.4	0.2
И	0.4	0.4	0.4	0.4	0.4	0.4	1	0.8	0.4	0.2
Н	0.4	0.4	0.4	0.4	0.4	0.4	0.8	1	0.4	0.2
У	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	1	0.2
Ж	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	1

In the final table, where all values that are less than α and **main diagonal** will become zeros.

Final										
	Б	В	Е	О	С	Э	И	Н	У	Ж
Б	0	0.8	0.7	0	0	0	0	0	0	0
В	0.8	0	0.7	0	0	0	0	0	0	0
Е	0.7	0.7	0	0	0	0	0	0	0	0
О	0	0	0	0	0.9	0.8	0	0	0	0
С	0	0	0	0.9	0	0.8	0	0	0	0
Э	0	0	0	0.8	0.8	0	0	0	0	0
И	0	0	0	0	0	0	0	0.8	0	0
Н	0	0	0	0	0	0	0.8	0	0	0
У	0	0	0	0	0	0	0	0	0	0
Ж	0	0	0	0	0	0	0	0	0	0

The iterations begin from maximum meaning:

First step:

I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10} $C_1 = \{\}$

$a_{45} = a_{54} = 0.9$ $C_1 = \{4, 5\}$

$Max = Sum_6 \rightarrow$ $C_1 = \{4, 5, 6\}$

Final										
	Б	В	Е	О	С	Э	И	Н	У	Ж
Б	0	0.8	0.7	0	0	0	0	0	0	0
В	0.8	0	0.7	0	0	0	0	0	0	0
Е	0.7	0.7	0	0	0	0	0	0	0	0
О	0	0	0	0	0.9	0.8	0	0	0	0
С	0	0	0	0.9	0	0.8	0	0	0	0
Э	0	0	0	0.8	0.8	0	0	0	0	0
И	0	0	0	0	0	0	0	0.8	0	0
Н	0	0	0	0	0	0	0.8	0	0	0
У	0	0	0	0	0	0	0	0	0	0
Ж	0	0	0	0	0	0	0	0	0	0

Second step:

I = {1, 2, 3, 7, 8, 9, 10} $C_2 = \{\}$

$a_{12} = a_{21} = 0.8$ $C_2 = \{1, 2\}$

$Max = Sum_3 \rightarrow C_1 = \{1, 2, 3\}$

Final							
	Б	В	Е	И	Н	У	Ж
Б	0	0.8	0.7	0	0	0	0
В	0.8	0	0.7	0	0	0	0
Е	0.7	0.7	0	0	0	0	0
И	0	0	0	0	0.8	0	0
Н	0	0	0	0.8	0	0	0
У	0	0	0	0	0	0	0
Ж	0	0	0	0	0	0	0

Third step:

I = {7, 8, 9, 10} $C_3 = \{\}$

$a_{78} = a_{87} = 0.8$ $C_3 = \{7, 8\}$

Final				
	И	Н	У	Ж
И	0	0.8	0	0
Н	0.8	0	0	0
У	0	0	0	0
Ж	0	0	0	0

Last Step:

I = {9, 10}

Final		
	У	Ж
У	0	0
Ж	0	0

Because of matrix contain only 0, we can decide that these two letters belong to different classes.

C_4 = {9}

C_5 = {10}

Result:

***C* = {{1, 2, 3}, {4, 5, 6}, {7, 8}, {9}, {10}}**

OR

***C* = {{Б, В, Е}, {О, С, Э}, {И, Н}, {У}, {Ж}}**