

## Contemporary Data Processing Technology (CCOD)

## Lab 10 (November 04, 2016)

**Shurpo Dmitry, AS-36**

## Clustering

We have 33 russian letters l: = {А,Б,В,Г,Д,Е,Ё,Ж,З,И,Й,К,Л,М,Н,О,П,Р,С,Т,У,Ф,Х,Ц,Ч,Ш,Щ,Ь,Э,Ю,Я}

After modifying this matrix and replacing all values less than 0.3 and main diagonal by 0 we get

such matrix:

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33			
		А	Б	В	Г	Д	Е	Ё	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ь	Ы	Б	Э	Ю	Я			
1	А	0	0	0	0	0,9	0	0	0,7	0	0	0	0	0,9	0,7	0,8	0,7	0,9	0	0,7	0	0,7	0,7	0,7	0	0	0,7	0	0	0	0,7	0	0,7				
2	Б	0	0	0,9	0,7	0	0,9	0,9	0	0,9	0	0	0	0	0	0	0	0	0,7	0	0	0	0	0	0	0	0,9	0,8	0,9	0,9	0	0,7					
3	В	0	0,9	0	0,7	0	0,9	0,9	0	0,9	0	0	0	0	0	0	0	0	0,7	0	0	0	0	0	0	0	0,9	0,8	0,9	0,9	0	0,7					
4	Г	0	0,7	0,7	0	0	0,7	0,7	0	0,7	0	0	0	0	0	0	0	0	0,7	0	0	0	0	0	0	0	0,7	0,7	0,7	0,7	0	0,7					
5	Д	0,9	0	0	0	0	0	0	0,7	0	0	0	0	0,9	0,7	0,8	0,7	0,9	0	0,7	0	0,7	0,7	0,7	0	0	0	0	0	0	0	0,7	0	0,7			
6	Е	0	0,9	0,9	0,7	0	0	0,9	0	0,9	0	0	0	0	0	0	0	0	0,7	0	0	0	0	0	0	0	0,9	0,8	0,9	0,9	0	0,7					
7	Ё	0	0,9	0,9	0,7	0	0,9	0,9	0	0,9	0	0	0	0	0	0	0	0	0,7	0	0	0	0	0	0	0	0,9	0,8	0,9	0,9	0	0,7					
8	Ж	0,7	0	0	0	0,7	0	0	0	0	0	0	0	0,7	0,7	0,7	0,7	0,7	0	0,7	0	0,8	0,7	0,9	0	0,8	0	0	0	0	0	0,7	0	0,7			
9	З	0	0,9	0,9	0,7	0	0,9	0,9	0	0	0	0	0	0	0	0	0	0	0	0,7	0	0	0	0	0	0	0,9	0,8	0,9	0,9	0	0,7					
10	И	0	0	0	0	0	0	0	0	0	0	0	0	0,9	0,3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
11	Й	0	0	0	0	0	0	0	0	0	0	0	0	0,9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	К	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	
13	Л	0,9	0	0	0	0	0,9	0	0	0,7	0	0	0	0	0	0	0	0,7	0,8	0,7	0,9	0	0,7	0	0,7	0,7	0,7	0	0,7	0	0	0	0	0,7	0	0,7	
14	М	0,7	0	0	0	0,7	0	0	0,7	0	0	0	0	0	0,7	0	0,7	0,7	0,7	0	0,7	0	0,7	0,7	0	0	0	0	0	0	0	0	0	0,7	0	0	
15	Н	0,8	0	0	0	0,8	0	0	0,7	0	0	0	0	0,8	0,7	0	0,7	0,8	0	0,7	0	0,7	0	0,7	0	0	0	0	0	0	0	0	0	0,7	0	0	
16	О	0,7	0	0	0	0,7	0	0	0,7	0	0	0	0	0,7	0,7	0,7	0,7	0,7	0	0,7	0,8	0	0,7	0,8	0,7	0	0,7	0	0	0	0	0	0,8	0	0,8		
17	П	0,9	0	0	0	0,9	0	0	0,7	0	0	0	0	0,9	0,7	0,8	0,7	0,7	0	0	0,7	0	0,7	0,7	0	0,7	0	0	0	0	0	0	0	0,7	0	0,7	
18	Р	0	0,7	0,7	0,7	0	0,7	0,7	0	0,7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7	0,7	0,7	0,7	0	0,7				
19	С	0,7	0	0	0	0,7	0	0	0,7	0	0	0	0	0,7	0,7	0,7	0,8	0,7	0	0	0	0,7	0	0	0	0	0	0	0	0	0	0	0,8	0	0		
20	Т	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	
21	У	0,7	0	0	0	0,7	0	0	0,8	0	0	0	0	0,7	0,7	0,7	0,7	0,7	0	0,7	0	0,7	0,8	0,8	0	0,8	0	0	0	0	0	0,7	0	0			
22	Ф	0,7	0	0	0	0	0,7	0	0	0	0	0	0	0,7	0,7	0,7	0,8	0,7	0	0,8	0	0,7	0	0,7	0	0	0	0	0	0	0	0	0	0,8	0	0,8	
23	Х	0,7	0	0	0	0,7	0	0	0,9	0	0	0	0	0,7	0,7	0,7	0,7	0,7	0	0,7	0	0,8	0,7	0	0,8	0	0	0	0	0	0	0	0	0,7	0	0	
24	Ц	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	0,8	0,8	0	0	0	0	0	0	0	0	0
25	Ч	0,7	0	0	0	0	0,7	0	0	0,8	0	0	0	0	0,7	0,7	0,7	0,7	0,7	0	0,7	0	0,8	0,7	0,8	0	0	0	0	0	0	0	0	0	0,7	0	0
26	Ш	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	0,8	0	0	0,9	0	0	0	0	0	0	
27	Щ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8	0	0,9	0	0	0	0	0	0	0	0	
28	Ь	0	0,9	0,9	0,7	0	0,9	0,9	0	0,9	0	0	0	0	0	0	0	0	0,7	0	0	0	0	0	0	0	0	0	0	0,8	0,9	0,9	0	0,7			
29	Ы	0	0,8	0,8	0,7	0	0,8	0,8	0	0,8	0	0	0	0	0	0	0	0	0,7	0	0	0	0	0	0	0	0,8	0,8	0,8	0,8	0	0,7					
30	Ь	0	0,9	0,9	0,7	0	0,9	0,9	0	0,9	0	0	0	0	0	0	0	0	0,7	0	0	0	0	0	0	0	0,9	0,8	0	0,9	0,9	0	0,7				
31	Э	0	0,9	0,9	0,7	0	0,9	0,9	0	0,9	0	0	0	0	0	0	0	0	0,7	0	0	0	0	0	0	0	0,9	0,8	0,9	0,9	0	0,7					
32	Ю	0,7	0	0	0	0,7	0	0	0,7	0	0	0	0	0,7	0,7	0,7	0,8	0,7	0,7	0	0,8	0	0,7	0,8	0,7	0	0,7	0	0	0	0	0	0	0			
33	Я	0	0,7	0,7	0,7	0	0,7	0	0,7	0	0	0	0	0	0	0	0	0	0,7	0	0	0	0	0	0	0	0,7	0,7	0,7	0,7	0	0					

## 1 cluster:

$I = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33\}$ ;

$C_1 = \{\}$

$a_{2,9} = 0.9$  is maximum;

$C_1 = \{2, 9\}$ ;

After our calculating we got:

$C_1 = \{2, 6, 3, 7, 9, 31, 4, 20\}$ ;

$C_1 = \{Б, Е, В, Ё, 3, Э, Г, Т\}$ ;

## 2 cluster:

$I = \{1, 5, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 32, 33, 28, 29, 30\}$ ;

$C_2 = \{\}$ ;

$a_{29,28} = 0.9$  – maximum;

$C_2 = \{29, 28\}$ ;

After our calculating we got:

$C_2 = \{29, 28, 30\}$ ;

$C_2 = \{б, ё, б\}$ ;

## 3 cluster:

$I = \{1, 5, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 32, 33\}$ ;

$C_3 = \{\}$ ;

$a_{10,11} = 0.9$  – maximum;

$C_3 = \{10, 11\}$ ;

Now we got:

$C_3 = \{10, 11\}$ ;

Letters:  $C_3 = \{И, Й\}$ ;

## 4 cluster:

$I = \{1, 5, 8, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 32, 33\}$ ;

$C_4 = \{\}$ ;

$a_{1,5} = 0.9$  – maximum;

$C_4 = \{1, 5\}$ ;

After our calculating we got:

$C_4 = \{1, 13, 5, 17, 8, 15, 25, 14\}$ ;

Letters:  $C_4 = \{А, Л, Д, П, Ж, Н, Ч, М\}$ ;

### **5 cluster:**

$I = \{12, 18, 19, 24, 26, 27, 16, 22, 32, 21, 33, 23\}$ ;

$C_5 = \{\}$ ;

$a_{16, 22} = 0.9$  – maximum;

$C_5 = \{16, 22\}$ ;

After our calculating we got:

$C_5 = \{16, 22, 32\}$ ;

Letters:  $C_5 = \{О, Ф, ІО\}$ ;

### **6 cluster:**

$I = \{12, 18, 19, 24, 26, 27, 21, 33, 23\}$ ;

$C_6 = \{\}$ ;

$a_{21, 23} = 0.9$  – maximum;

$C_6 = \{21, 23\}$ ;

After our calculating we got:

$C_6 = \{21, 23, 33\}$ ;

Letters:  $C_6 = \{У, Х, Я\}$ ;

### **7 cluster:**

$I = \{12, 18, 19, 24, 26, 27\}$ ;

$C_7 = \{\}$ ;

$a_{24, 27} = 0.9$  – maximum;

$C_7 = \{24, 27\}$ ;

After our calculating we got:

$C_7 = \{24, 27, 26\}$

Letters:  $C_7 = \{Ш, Щ, Ц\}$ ;

### **Remaining clusters:**

$I = \{12, 18, 19\}$ ;

$C_8 = \{12\}$ ,  $C_9 = \{18\}$  and  $C_{10} = \{19\}$ ;

Letters:  $C_8 = \{К\}$ ,  $C_9 = \{Р\}$  and  $C_{10} = \{С\}$ ;

**Result:**

C={

{Б, Е, В, Ё, З, Э, Г, Т}

{б, ё, з}

{и, й}

{А, Л, Д, П, Ж, Н, Ч, М}

{о, ф, ю}

{у, х, я}

{ш, щ, ц}

{к}

{р}

{с}

}.