

# Contemporary Data Processing Technology (CCOD)

## Lab 10 (November 4, 2016)

Savchuk Artem, AS-36

Clustering Russian alphabet:  $I = \{A, Б, В, Г, Д, Е, Ё, Ж, З, И, Й, К, Л, М, Н, О, П, Р, С, Т, У, Ф, Х, Ц, Ч, Ш, Щ, Ъ, Ы, Ь, Э, Ю, Я\}$ ;

Table of similarity percentage of letters:

		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	Ъ	Ы	Ь	Э	Ю	Я		
1	А	1	0,1	0,1	0,1	0,8	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,95	0,5	0,7	0,1	0,7	0,2	0,1	0,1	0,1	0,1	0,2	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,7	1	А
2	Б	0,1	1	0,95	0,8	0,2	0,95	0,9	0,1	0,2	0,1	0,1	0,3	0,1	0,1	0,2	0,4	0,2	0,1	0,2	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,8	0,6	0,9	0,3	0,1	0,1	2	Б
3	В	0,1	0,95	1	0,2	0,1	0,95	0,9	0,1	0,9	0,1	0,1	0,4	0,1	0,1	0,1	0,1	0,5	0,1	0,3	0,1	0,2	0,2	0,1	0,3	0,1	0,1	0,7	0,6	0,8	0,4	0,1	0,2	3	В	
4	Г	0,1	0,8	0,2	1	0,1	0,5	0,4	0,1	0,1	0,2	0,1	0,3	0,2	0,1	0,2	0,1	0,3	0,4	0,1	0,85	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	4	Г
5	Д	0,8	0,2	0,1	0,1	1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,9	0,2	0,2	0,1	0,6	0,1	0,1	0,2	0,1	0,2	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,2	5	Д	
6	Е	0,1	0,95	0,95	0,5	0,1	1	0,9	0,2	0,2	0,1	0,1	0,4	0,2	0,2	0,3	0,1	0,3	0,1	0,1	0,2	0,1	0,2	0,1	0,1	0,1	0,2	0,1	0,2	0,1	0,3	0,2	0,1	0,2	6	Е
7	Ё	0,1	0,9	0,9	0,4	0,1	0,9	1	0,1	0,2	0,1	0,2	0,3	0,2	0,2	0,3	0,1	0,2	0,1	0,1	0,3	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,3	0,1	0,2	7	Ё	
8	Ж	0,1	0,1	0,1	0,1	0,1	0,2	0,1	1	0,1	0,2	0,1	0,6	0,2	0,2	0,2	0,1	0,2	0,2	0,1	0,1	0,1	0,2	0,9	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	8	Ж	
9	З	0,1	0,2	0,9	0,1	0,1	0,2	0,2	0,1	1	0,2	0,2	0,2	0,1	0,1	0,1	0,2	0,1	0,3	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,3	0,2	0,4	0,9	0,2	0,4	9	З
10	И	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,2	0,2	1	0,95	0,2	0,1	0,4	0,1	0,1	0,2	0,1	0,2	0,1	0,1	0,2	0,2	0,4	0,2	0,2	0,2	0,1	0,1	0,2	0,1	0,1	0,1	10	И
11	Й	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,2	0,95	1	0,2	0,1	0,3	0,1	0,1	0,3	0,3	0,2	0,2	0,2	0,2	0,2	0,4	0,3	0,3	0,2	0,1	0,2	0,1	0,1	0,1	0,1	11	Й
12	К	0,1	0,3	0,4	0,3	0,1	0,4	0,3	0,6	0,2	0,2	0,2	1	0,2	0,2	0,3	0,1	0,1	0,2	0,1	0,2	0,1	0,1	0,6	0,1	0,2	0,1	0,1	0,2	0,2	0,1	0,2	0,1	0,1	12	К
13	Л	0,95	0,1	0,1	0,2	0,9	0,2	0,2	0,2	0,1	0,1	0,1	0,2	1	0,7	0,1	0,2	0,9	0,3	0,2	0,2	0,1	0,1	0,2	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	13	Л
14	М	0,5	0,1	0,1	0,1	0,2	0,2	0,2	0,2	0,1	0,4	0,3	0,2	0,7	1	0,2	0,1	0,4	0,1	0,1	0,2	0,1	0,1	0,2	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,1	14	М
15	Н	0,7	0,2	0,1	0,2	0,2	0,3	0,3	0,2	0,1	0,1	0,3	0,1	0,2	1	0,2	0,8	0,2	0,1	0,4	0,2	0,1	0,2	0,1	0,7	0,2	0,1	0,3	0,2	0,4	0,5	0,5	0,1	15	Н	
16	О	0,1	0,4	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,2	0,1	0,2	1	0,7	0,2	0,6	0,2	0,1	0,9	0,1	0,3	0,2	0,1	0,1	0,2	0,1	0,1	0,1	0,8	0,1	16	О	
17	П	0,7	0,2	0,1	0,3	0,6	0,3	0,2	0,2	0,1	0,2	0,3	0,1	0,9	0,4	0,8	0,7	1	0,1	0,2	0,2	0,1	0,2	0,1	0,3	0,2	0,1	0,1	0,2	0,1	0,2	0,1	0,1	17	П	
18	Р	0,2	0,1	0,5	0,4	0,1	0,1	0,1	0,2	0,3	0,1	0,3	0,2	0,3	0,1	0,2	0,2	0,1	1	0,2	0,1	0,2	0,5	0,2	0,1	0,2	0,1	0,1	0,1	0,2	0,1	0,1	0,1	18	Р	
19	С	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,2	0,1	0,2	0,1	0,1	0,6	0,2	0,2	1	0,2	0,1	0,4	0,1	0,2	0,1	0,1	0,2	0,1	0,1	0,2	0,1	0,2	0,1	19	С
20	Т	0,1	0,1	0,3	0,85	0,2	0,2	0,3	0,1	0,1	0,1	0,2	0,2	0,2	0,2	0,4	0,2	0,2	0,1	0,2	1	0,2	0,2	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,1	20	Т
21	У	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,2	0,1	0,1	0,2	0,1	0,2	1	0,1	0,8	0,1	0,8	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	21	У
22	Ф	0,1	0,2	0,2	0,1	0,2	0,2	0,2	0,2	0,1	0,2	0,2	0,1	0,1	0,1	0,9	0,2	0,5	0,4	0,2	0,1	1	0,2	0,2	0,1	0,1	0,1	0,1	0,3	0,1	0,2	0,2	0,1	22	Ф	
23	Х	0,2	0,1	0,2	0,1	0,1	0,1	0,1	0,9	0,1	0,2	0,2	0,6	0,2	0,2	0,2	0,1	0,1	0,2	0,1	0,1	0,8	0,2	1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	23	Х
24	Ц	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,4	0,4	0,1	0,1	0,1	0,3	0,3	0,1	0,2	0,1	0,1	0,2	0,1	1	0,5	0,8	0,8	0,1	0,1	0,1	0,1	0,1	0,1	24	Ц	
25	Ч	0,2	0,1	0,3	0,1	0,2	0,1	0,1	0,1	0,1	0,2	0,3	0,2	0,2	0,1	0,7	0,2	0,2	0,2	0,1	0,1	0,8	0,1	0,1	0,5	1	0,6	0,5	0,1	0,1	0,1	0,1	0,3	25	Ч	
26	Ш	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,2	0,3	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,8	0,6	1	0,9	0,2	0,2	0,2	0,1	0,2	0,1	26	Ш
27	Щ	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,2	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,8	0,5	0,9	1	0,2	0,1	0,1	0,1	0,1	0,1	27	Щ	
28	Ъ	0,1	0,8	0,7	0,1	0,1	0,2	0,1	0,1	0,3	0,1	0,1	0,2	0,1	0,1	0,3	0,2	0,1	0,1	0,2	0,2	0,1	0,1	0,1	0,1	0,1	0,2	0,2	1	0,8	0,9	0,3	0,1	0,1	28	Ъ
29	Ы	0,1	0,6	0,6	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,2	0,1	0,1	0,2	0,1	0,2	0,1	0,1	0,1	0,1	0,3	0,1	0,1	0,1	0,2	0,1	0,8	1	0,8	0,2	0,1	0,1	29	Ы
30	Ь	0,1	0,9	0,8	0,1	0,1	0,3	0,2	0,1	0,4	0,2	0,2	0,1	0,1	0,1	0,4	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,9	0,8	1	0,1	0,1	0,1	0,1	30	Ь
31	Э	0,1	0,3	0,4	0,1	0,1	0,2	0,3	0,1	0,9	0,1	0,1	0,2	0,1	0,1	0,5	0,1	0,2	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,3	0,2	0,1	1	0,3	0,1	0,1	31	Э
32	Ю	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,5	0,8	0,1	0,1	0,2	0,1	0,1	0,2	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,3	1	0,3	0,1	32	Ю
33	Я	0,7	0,1	0,2	0,1	0,2	0,2	0,2	0,1	0,4	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,3	0,1	0,1	0,1	0,1	0,3	1	0,3	0,1	33	Я
		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		

Then I repeated  $R^{(n+1)} = R^{(n)} * R^{(n)}$  till  $R^{(n)} = R^{(n+1)}$ ;

And assuming  $\alpha = 0.6$ , I got table that shown below:

		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	Ъ	Ы	Ь	Э	Ю	Я			
1	А	0	0	0	0	0,9	0	0	0,7	0	0	0	0	0,95	0,7	0,8	0,7	0,9	0	0	0	0,7	0,7	0,7	0	0,7	0	0	0	0	0	0	0,7	0,7	1	А	
2	Б	0	0	0,95	0,8	0	0,95	0,9	0	0,9	0	0	0	0	0	0	0	0	0	0	0	0,8	0	0	0	0	0	0	0,9	0,8	0,9	0,9	0	0	2	Б	
3	В	0	0,95	0	0,8	0	0,95	0,9	0	0,9	0	0	0	0	0	0	0	0	0	0	0	0,8	0	0	0	0	0	0	0,9	0,8	0,9	0,9	0	0	3	В	
4	Г	0	0	0,8	0,8	0	0	0,8	0,8	0	0,8	0	0	0	0	0	0	0	0	0	0	0,85	0	0	0	0	0	0	0,8	0,8	0,8	0,8	0	0	4	Г	
5	Д	0,9	0	0	0	0	0	0	0	0,7	0	0	0	0	0,9	0,7	0,8	0,7	0,9	0	0	0	0,7	0,7	0,7	0	0,7	0	0	0	0	0	0,7	0,7	5	Д	
6	Е	0	0,95	0,95	0,8	0	0	0,9	0	0,9	0	0	0	0	0	0	0	0	0	0	0	0	0,8	0	0	0	0	0	0,9	0,8	0,9	0,9	0	0	6	Е	
7	Ё	0	0,9	0,9	0,8	0	0,9	0	0	0,9	0	0	0	0	0	0	0	0	0	0	0	0,8	0	0	0	0	0	0	0,9	0,8	0,9	0,9	0	0	7	Ё	
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	0	0,8	0,7	0,9	0	0,8	0	0	0	0	0	0	0,7	0,7	8	Ж	
9	З	0	0,9	0,9	0,8	0	0,9	0,9	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8	0	0	0	0	0	0	0,9	0,8	0,9	0,9	0	0	9	З	
10	И	0	0	0	0	0	0	0	0	0	0	0	0,95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	И	
11	Й	0	0	0	0	0	0	0	0	0	0	0,95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	Й	
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	12	К	
13	Л	0,95	0	0	0	0,9	0	0	0,7	0	0	0	0	0	0,7	0,8	0,7	0,9	0	0	0	0	0,7	0,7	0,7	0	0,7	0	0	0	0	0	0,7	0,7	13	Л	
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	0,9	0	0	0	0,7	0,7	0,7	0,7	0,7	0	0	0	0	0	0,7	0,7	14	М	
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	0	0	0,7	0,7	0,7	0	0,7	0	0	0	0	0	0,7	0,7	15	Н	
16	О	0,7	0	0	0	0,7	0	0	0,7	0	0	0	0	0	0,7	0,7	0,7	0	0,7	0	0	0	0,7	0,9	0,7	0	0,7	0	0	0	0	0	0	0,8	0,7	16	О
17	П	0,9	0	0	0	0,9	0	0	0,7	0	0	0	0	0	0,9	0,7	0,8	0,7	0	0	0	0	0,7	0,7	0,7	0	0,7	0	0	0	0	0	0	0,7	0,7	17	П
18	Р	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	18	Р	
19	С	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	19	С	
20	Т	0	0,8	0,8	0,85	0	0,8	0,8	0	0,8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8	0,8	0,8	0,8	0	0	20	Т
21	У	0,7	0	0	0	0,7	0	0	0,8	0	0	0	0	0	0,7	0,7	0,7	0,7	0	0	0	0	0	0,7	0,8	0	0,8	0	0	0	0	0	0	0,7	0,7	21	У
22	Ф	0,7	0	0	0	0,7	0	0	0,7	0	0	0	0	0	0,7	0,7	0,7	0,9	0,7	0	0	0	0,7	0	0,7	0	0,7	0	0	0	0	0	0	0,8	0,7	22	Ф
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	0	0	0,8	0,7	0	0	0,8	0	0	0	0	0	0	0,7	0,7	23	Х
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	0,8	0,8	0	0	0	0	0	0	24	Ц
25	Ч	0,7	0	0	0	0,7	0	0	0,8	0	0	0	0	0	0,7	0,7	0,7	0,7	0,7	0	0	0	0,8	0,7	0,8	0	0	0	0	0	0	0	0	0,7	0,7	25	Ч
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	0,8	0	0	0,9	0	0	0	0	0	0	26	Ш
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	27	Щ
28	Ъ	0	0,9	0,9	0,8	0	0,9	0,9	0	0,9	0	0	0	0	0	0	0	0	0	0	0	0,8	0	0	0	0	0	0	0	0	0,8	0,9	0,9	0	0	28	Ъ
29	Ы	0	0,8	0,8	0,8	0	0,8	0,8	0	0,8	0	0	0	0	0	0	0	0	0	0	0	0,8	0	0	0	0	0	0	0,8	0	0,8	0,8	0	0	29	Ы	
30	Ь	0	0,9	0,9	0,8	0	0,9	0,9	0	0,9	0	0	0	0	0	0	0	0	0	0	0	0,8	0	0	0	0	0	0	0	0,9	0,8	0,9	0	0	0	30	Ь
31	Э	0	0,9	0,9	0,8	0	0,9	0,9	0	0,9	0	0	0	0	0	0	0	0	0	0	0	0,8	0	0	0	0	0	0	0,9	0,8	0,9	0	0	0	31	Э	
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	0	0	0	0,7	0,8	0,7	0	0,7	0	0	0	0	0	0	0,7	0,7	32	Ю
33	Я	0,7	0	0	0	0,7	0	0	0,7	0	0	0	0	0	0,7	0,7	0,7	0,7	0,7	0	0	0	0,7	0,7	0,7	0	0,7	0	0	0	0	0	0	0,7	0	33	Я
		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			

### 1<sup>st</sup> cluster:

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

$C_1 = \{\}$

$a_{26} = a_{36} = 0.95 - \text{maximum};$

Take  $a_{26}, C_1 = \{2, 6\};$

$a_{23} + a_{63} = 1.9 - \text{maximum};$

$C_1 = \{2, 6, 3\};$

$a_{27} + a_{37} + a_{67} = 2.7 - \text{maximum};$

$C_1 = \{2, 6, 3, 7\};$

$a_{29} + a_{39} + a_{69} + a_{79} = 3.6 - \text{maximum};$

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

Similarly, we include: 28, 30, 31, 4, 20, 29;

Finally we got 1<sup>st</sup> cluster:  $C_1 = \{2, 6, 3, 7, 9, 28, 30, 31, 4, 20, 29\};$

Letters:  $C_1 = \{\text{Б, Е, В, Ё, З, Ъ, Ь, Э, Г, Т, Ы}\};$

### 2<sup>nd</sup> cluster:

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

$C_2 = \{\};$

$a_{1011} = 0.95 - \text{maximum};$

$C_2 = \{10, 11\};$

There aren't maximum values, because almost all values in 10 and 11 row are 0;

So  $C_2 = \{10, 11\};$

Letters:  $C_2 = \{\text{И, Й}\};$

### 3<sup>rd</sup> cluster:

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

$C_2 = \{\};$

$a_{1\ 13} = 0.95 - \text{maximum};$

$C_3 = \{1, 13\};$

$a_{1\ 5} + a_{13\ 5} = 1.8 - \text{maximum};$

$C_3 = \{1, 13, 5\}$

Similarly, we include: 17, 8, 15, 25, 14, 16, 22, 32, 21, 23, 33;

Finally we got 3<sup>rd</sup> cluster:  $C_3 = \{1, 13, 5, 17, 8, 15, 25, 14, 16, 22, 32, 21, 23, 33\};$

Letters:  $C_3 = \{\text{А, Л, Д, П, Ж, Н, Ч, М, О, Ф, Ю, У, Х, Я}\};$

### 4<sup>th</sup> cluster:

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

$C_4 = \{\};$

$a_{26\ 27} = 0.9 - \text{maximum};$

$C_4 = \{26, 27\};$

$a_{26\ 24} + a_{27\ 24} = 1.6 - \text{maximum};$

$C_4 = \{26, 27, 24\};$

There isn't any maximum, so we stop append new letters into 4<sup>th</sup> cluster.

$C_4 = \{26, 27, 24\}$

Letters:  $C_4 = \{\text{Ш, Щ, Ц}\};$

### 5<sup>th</sup>, 6<sup>th</sup> and 7<sup>th</sup> clusters:

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

Now all values in I is zero, so the remaining letters goes into different classes:

$C_5 = \{12\}, C_6 = \{18\}$  and  $C_7 = \{19\};$

Letters:  $C_5 = \{\text{К}\}, C_6 = \{\text{Р}\}$  and  $C_7 = \{\text{С}\};$

### Result:

$C = \{\{\text{Б, Е, В, Ё, З, Ъ, Ь, Э, Г, Т, Ы}\}, \{\text{И, Й}\}, \{\text{А, Л, Д, П, Ж, Н, Ч, М, О, Ф, Ю, У, Х, Я}\}, \{\text{Ш, Щ, Ц}\}, \{\text{К}\}, \{\text{Р}\} \text{ and } \{\text{С}\}\}$