

CCOD

Alexander Bryuzgin(AS-36)

Lab 10(10.29.2016)

Clusterisation of English alphabet

Here is a table of similarity percentage of letters

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
A	1	0.2	0.1	0.1	0.2	0.2	0.1	0.3	0.1	0.1	0.2	0.1	0.7	0.8	0.1	0.1	0.1	0.4	0.2	0.3	0.1	0.4	0.3	0.65	0.65	0.3	A
B	0.2	1	0.1	0.85	0.7	0.6	0.3	0.3	0.1	0.1	0.55	0.1	0.1	0.1	0.4	0.7	0.1	0.85	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.55	B
C	0.1	0.1	1	0.4	0.1	0.1	0.6	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.9	0.55	0.9	0.1	0.6	0.1	0.1	0.1	0.1	0.1	0.1	0.1	C
D	0.1	0.85	0.4	1	0.4	0.3	0.6	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.8	0.65	0.75	0.3	0.55	0.1	0.35	0.1	0.1	0.1	0.1	0.1	D
E	0.2	0.7	0.1	0.4	1	0.9	0.1	0.1	0.2	0.2	0.52	0.48	0.1	0.1	0.1	0.1	0.1	0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.6	E
F	0.2	0.6	0.1	0.3	0.9	1	0.1	0.1	0.1	0.1	0.55	0.4	0.1	0.1	0.1	0.1	0.1	0.95	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.55	F
G	0.1	0.3	0.6	0.6	0.1	0.1	1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.75	0.65	0.7	0.1	0.55	0.1	0.1	0.1	0.1	0.1	0.1	0.1	G
H	0.3	0.3	0.1	0.1	0.1	0.1	0.1	1	0.1	0.1	0.1	0.1	0.4	0.4	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	H
I	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	1	0.9	0.1	0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.8	0.1	0.1	0.1	0.1	0.1	0.1	I
J	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.9	1	0.1	0.75	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.75	0.1	0.1	0.1	0.1	0.1	0.1	J
K	0.2	0.55	0.1	0.1	0.52	0.55	0.1	0.1	0.1	0.1	1	0.1	0.3	0.2	0.1	0.1	0.1	0.8	0.1	0.1	0.1	0.1	0.1	0.6	0.55	0.56	K
L	0.1	0.1	0.1	0.1	0.48	0.4	0.1	0.1	0.9	0.75	0.1	1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.7	0.1	0.1	0.1	0.1	0.1	0.1	L
M	0.7	0.1	0.1	0.1	0.1	0.1	0.1	0.4	0.1	0.1	0.3	0.1	1	0.8	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.45	0.49	0.6	0.55	0.1	M
N	0.8	0.1	0.1	0.1	0.1	0.1	0.1	0.4	0.1	0.1	0.2	0.1	0.8	1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.3	0.45	0.55	0.55	0.1	N
O	0.1	0.4	0.9	0.8	0.1	0.1	0.75	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1	0.55	0.95	0.1	0.55	0.1	0.1	0.1	0.1	0.1	0.1	0.1	O
P	0.1	0.7	0.55	0.65	0.1	0.1	0.65	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.55	1	0.55	0.9	0.55	0.1	0.1	0.1	0.1	0.1	0.1	0.1	P
Q	0.1	0.1	0.9	0.75	0.1	0.1	0.7	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.95	0.55	1	0.1	0.51	0.1	0.1	0.2	0.1	0.1	0.1	0.1	Q
R	0.4	0.85	0.1	0.3	0.9	0.95	0.1	0.1	0.1	0.1	0.8	0.1	0.1	0.1	0.1	0.9	0.1	1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.55	R
S	0.2	0.1	0.6	0.55	0.1	0.1	0.55	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.55	0.55	0.51	0.1	1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	S
T	0.3	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.8	0.75	0.1	0.7	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1	0.1	0.1	0.1	0.1	0.1	0.1	T
U	0.1	0.1	0.1	0.35	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	0.1	1	0.9	0.7	0.1	0.45	0.1	U
V	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.45	0.3	0.1	0.1	0.2	0.1	0.1	0.1	0.9	1	0.85	0.45	0.49	0.1	V
W	0.3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.49	0.45	0.1	0.1	0.1	0.1	0.1	0.1	0.7	0.85	1	0.4	0.4	0.1	W
X	0.65	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.6	0.1	0.6	0.55	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.45	0.4	1	0.6	0.1	X
Y	0.65	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.55	0.1	0.55	0.55	0.1	0.1	0.1	0.1	0.1	0.1	0.45	0.49	0.4	0.6	1	0.1	Y
Z	0.3	0.55	0.1	0.1	0.6	0.55	0.1	0.1	0.1	0.1	0.56	0.1	0.1	0.1	0.1	0.1	0.1	0.55	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1	Z
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	

Final table, where all values less than α and main diagonal will become zeros.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	0	0.6	0.6	0.6	0.6	0.6	0.6	0	0	0	0.6	0	0.8	0.8	0.6	0.6	0.6	0.6	0.6	0	0	0	0	0.65	0.65	0.6
B	0.6	0	0.8	0.85	0.85	0.85	0.75	0	0	0	0.8	0	0.6	0.6	0.8	0.85	0.8	0.85	0.6	0	0	0	0	0.6	0.6	0.6
C	0.6	0.8	0	0.8	0.8	0.8	0.75	0	0	0	0.8	0	0.6	0.6	0.9	0.8	0.9	0.8	0.6	0	0	0	0	0.6	0.6	0.6
D	0.6	0.85	0.8	0	0.85	0.85	0.75	0	0	0	0.8	0	0.6	0.6	0.8	0.85	0.8	0.85	0.6	0	0	0	0	0.6	0.6	0.6
E	0.6	0.85	0.8	0.85	0	0.9	0.75	0	0	0	0.8	0	0.6	0.6	0.8	0.9	0.8	0.9	0.6	0	0	0	0	0.6	0.6	0.6
F	0.6	0.85	0.8	0.85	0.9	0	0.75	0	0	0	0.8	0	0.6	0.6	0.8	0.9	0.8	0.95	0.6	0	0	0	0	0.6	0.6	0.6
G	0.6	0.75	0.75	0.75	0.75	0.75	0	0	0	0	0.75	0	0.6	0.6	0.75	0.75	0.75	0.75	0.6	0	0	0	0	0.6	0.6	0.6
H	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
I	0	0	0	0	0	0	0	0	0	0.9	0	0.9	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0.9	0	0	0.9	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0
K	0.6	0.8	0.8	0.8	0.8	0.8	0.75	0	0	0	0	0	0.6	0.6	0.8	0.8	0.8	0.8	0.6	0	0	0	0	0.6	0.6	0.6
L	0	0	0	0	0	0	0	0	0.9	0.9	0	0	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0
M	0.8	0.6	0.6	0.6	0.6	0.6	0.6	0	0	0	0.6	0	0	0.8	0.6	0.6	0.6	0.6	0.6	0	0	0	0	0.65	0.65	0.6
N	0.8	0.6	0.6	0.6	0.6	0.6	0.6	0	0	0	0.6	0	0.8	0	0.6	0.6	0.6	0.6	0.6	0	0	0	0	0.65	0.65	0.6
O	0.6	0.8	0.9	0.8	0.8	0.8	0.75	0	0	0	0.8	0	0.6	0.6	0	0.8	0.95	0.8	0.6	0	0	0	0	0.6	0.6	0.6
P	0.6	0.85	0.8	0.85	0.9	0.9	0.75	0	0	0	0.8	0	0.6	0.6	0.8	0	0.8	0.9	0.6	0	0	0	0	0.6	0.6	0.6
Q	0.6	0.8	0.9	0.8	0.8	0.8	0.75	0	0	0	0.8	0	0.6	0.6	0.95	0.8	0	0.8	0.6	0	0	0	0	0.6	0.6	0.6
R	0.6	0.85	0.8	0.85	0.9	0.95	0.75	0	0	0	0.8	0	0.6	0.6	0.8	0.9	0.8	0	0.6	0	0	0	0	0.6	0.6	0.6
S	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0	0	0	0.6	0	0.6	0.6	0.6	0.6	0.6	0	0	0	0	0	0	0.6	0.6	0.6
T	0	0	0	0	0	0	0	0	0.8	0.8	0	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9	0.85	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9	0	0.85	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.85	0.85	0	0	0	0
X	0.65	0.6	0.6	0.6	0.6	0.6	0.6	0	0	0	0.6	0	0.65	0.65	0.6	0.6	0.6	0.6	0.6	0	0	0	0	0	0.65	0.6
Y	0.65	0.6	0.6	0.6	0.6	0.6	0.6	0	0	0	0.6	0	0.65	0.65	0.6	0.6	0.6	0.6	0.6	0	0	0	0	0.65	0	0.6
Z	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0	0	0	0.6	0	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0	0	0	0	0.6	0.6	0

$I = \{A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z\}$

$C_1 = \{\}$

$a_{OQ} = a_{QO} = 0.95$

$C_1 = \{O, Q\}$

$\text{Max sum} = C = 0.9 \cdot 2 = 1.8 \Rightarrow C_1 = \{C, O, Q\}$

$I = \{A, B, D, E, F, G, H, I, J, K, L, M, N, P, R, S, T, U, V, W, X, Y, Z\}$

$C_2 = \{\}$

$a_{UV} = a_{VU} = 0.9$

$C_2 = \{U, V\}$

$\text{Max sum} = W = 0.85 \cdot 2 = 1.7 \Rightarrow C_2 = \{U, V, W\}$

$I = \{A, B, D, E, F, G, H, I, J, K, L, M, N, P, R, S, T, X, Y, Z\}$

$C_3 = \{\}$

$a_{FR} = a_{RF} = 0.95$

$C_3 = \{F, R\}$

$\text{Max sum} = E, P = 0.9 \cdot 2 = 1.8 \Rightarrow C_3 = \{E, F, P, R\}$

$I = \{A, B, D, G, H, I, J, K, L, M, N, S, T, X, Y, Z\}$

$C_4 = \{\}$

$a_{BD} = a_{DB} = 0.85$

$C_4 = \{B, D\}$

$\text{Max sum} = K = 0.8 \cdot 2 = 1.6 \Rightarrow C_4 = \{B, D, K\}$

$I = \{A, G, H, I, J, L, M, N, S, T, X, Y, Z\}$

$C_5 = \{\}$

$a_{MN} = a_{NM} = 0.8$

$C_5 = \{M, N\}$

$\text{Max sum} = A = 0.8 \cdot 2 = 1.6 \Rightarrow C_5 = \{A, M, N\}$

$I = \{G, H, I, J, L, S, T, X, Y, Z\}$

$C_6 = \{\}$

$$a_{IJ} = a_{JI} = 0.9$$

$$C_6 = \{I, J\}$$

$$\text{Max sum} = L = 0.9 * 2 = 1.8 \Rightarrow C_6 = \{I, J, L\}$$

$$I = \{G, H, S, T, X, Y, Z\}$$

$$C_7 = \{\}$$

$$a_{XY} = a_{YX} = 0.65$$

$$C_7 = \{X, Y\}$$

$$\text{Max sum} = G, S, Z = 0.6 * 2 = 1.2 \Rightarrow C_7 = \{G, S, X, Y, Z\}$$

$$C_8 = \{H\}$$

$$C_9 = \{T\}$$

$$C = \{\{C, O, Q\}, \{U, V, W\}, \{E, F, P, R\}, \{B, D, K\}, \{A, M, N\}, \{I, J, L\}, \{G, S, X, Y, Z\}, \{H\}, \{T\}\}$$