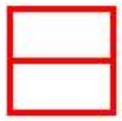
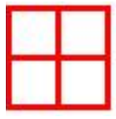
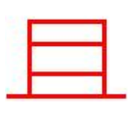
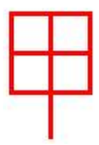
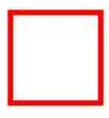
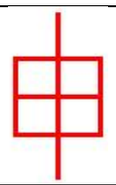
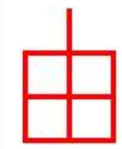
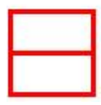
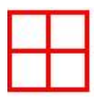
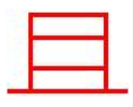
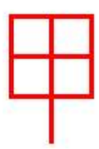
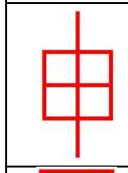
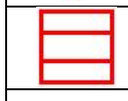
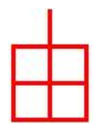
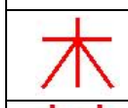
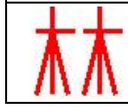
CCOD

Professor: Akira Imada

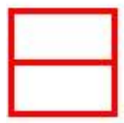
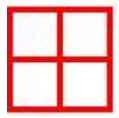
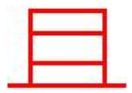
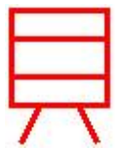
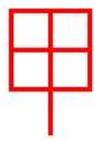
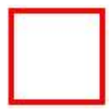
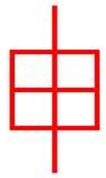
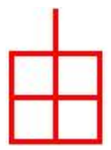
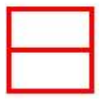
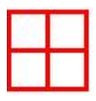
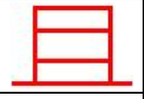
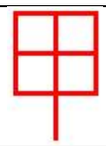
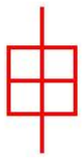
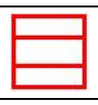
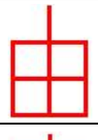
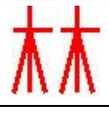
Student: Katerina Gritsuk (AS-37) 04.11.2017

Clustering of 13 Japanese characters

The table below shows the similarity of characters:

													
	1	0,9	0,6	0,55	0,3	0,8	0,9	0,6	0,8	0,1	0,8	0,1	0,1
	0,9	1	0,65	0,4	0,45	0,9	0,7	0,8	0,6	0,1	0,9	0,1	0,1
	0,6	0,65	1	0,95	0,5	0,5	0,6	0,6	0,9	0,1	0,5	0,1	0,1
	0,55	0,4	0,95	1	0,4	0,4	0,5	0,5	0,9	0,1	0,6	0,1	0,1
	0,3	0,45	0,5	0,4	1	0,5	0,9	0,6	0,25	0,1	0,45	0,1	0,1
	0,8	0,9	0,5	0,4	0,5	1	0,7	0,85	0,7	0,1	0,85	0,1	0,1
	0,9	0,7	0,6	0,5	0,9	0,7	1	0,65	0,75	0,1	0,45	0,1	0,1
	0,6	0,8	0,6	0,5	0,6	0,85	0,65	1	0,55	0,1	0,9	0,1	0,1
	0,8	0,6	0,9	0,9	0,25	0,7	0,6	0,4	1	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	1	0,1	0,2	0,15
	0,8	0,9	0,5	0,6	0,45	0,85	0,55	0,9	0,7	0,1	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	1	0,8
	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,15	0,1	0,8	1

Final table: (main diagonal and values < 0.55 become zeros)

		1	2	3	4	5	6	7	8	9	10	11	12	13
														
1		0	0.9	0.8	0.8	0.9	0.9	0.9	0.9	0.8	0	0.9	0	0
2		0.9	0	0.8	0.8	0.9	0.9	0.9	0.9	0.8	0	0.9	0	0
3		0.85	0.85	0	0.95	0.85	0.85	0.85	0.85	0.9	0	0.85	0	0
4		0.85	0.85	0.95	0	0.85	0.85	0.85	0.85	0.9	0	0.85	0	0
5		0.9	0.9	0.8	0.8	0	0.9	0.9	0.9	0.8	0	0.9	0	0
6		0.9	0.9	0.8	0.8	0.9	0	0.9	0.9	0.8	0	0.9	0	0
7		0.9	0.9	0.8	0.8	0.9	0.9	0	0.9	0.8	0	0.9	0	0
8		0.9	0.9	0.8	0.8	0.9	0.9	0.9	0	0.8	0	0.9	0	0
9		0.85	0.85	0.9	0.9	0.85	0.85	0.85	0.85	0	0	0.85	0	0
10		0	0	0	0	0	0	0	0	0	0	0	0	0
11		0.9	0.9	0.8	0.8	0.9	0.9	0.9	0.9	0.8	0	0	0	0
12		0	0	0	0	0	0	0	0	0	0	0	0	0.8
13		0	0	0	0	0	0	0	0	0	0	0	0.8	0

First Iteration:

First, set I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13} and C1 = {},

C1 = {1, 2} (selected at random from the maximum = 0,9),

1	2	3	4	5	6	7	8	9	10	11	12	13
-	-	1.6	1.6	1.8	1.8	1.8	1.8	1.6	-	1.8	-	-

S_{max} = 1.8, select at random 5. Now C1 = {1, 2, 5}.

1	2	3	4	5	6	7	8	9	10	11	12	13
-	-	2.4	2.4	-	2.7	2.7	2.7	2.4	-	2.7	-	-

S_{max} = 2.7, select at random 6. Now C1 = {1, 2, 5, 6}.

1	2	3	4	5	6	7	8	9	10	11	12	13
-	-	3.2	3.2	-	-	3.6	3.6	3.2	-	3.6	-	-

S_{max} = 3.6, select at random 7. Now C1 = {1, 2, 5, 6, 7}.

1	2	3	4	5	6	7	8	9	10	11	12	13
-	-	4	4	-	-	-	4.5	4	-	4.5	-	-

S_{max} = 4.5, select at random 8. Now C1 = {1, 2, 5, 6, 7, 8}.

1	2	3	4	5	6	7	8	9	10	11	12	13
-	-	4.8	4.8	-	-	-	-	4.8	-	5.4	-	-

S_{max} = 5.4, select 11. Now C1 = {1, 2, 5, 6, 7, 11}.

1	2	3	4	5	6	7	8	9	10	11	12	13
-	-	5.6	5.6	-	-	-	-	5.6	-	-	-	-

S_{max} = 5.6, select at random 3. Now C1 = {1, 2, 3, 5, 6, 7, 11}.

1	2	3	4	5	6	7	8	9	10	11	12	13
-	-	-	6.4	-	-	-	-	6.4	-	-	-	-

S_{max} = 6.4, select at random 4. Now C1 = {1, 2, 3, 4, 5, 6, 7, 11}.

1	2	3	4	5	6	7	8	9	10	11	12	13
-	-	-	-	-	-	-	-	7.2	-	-	-	-

S_{max} = 7.2, select 9. Now C1 = {1, 2, 3, 4, 5, 6, 7, 9, 11}.

The final C1 = {1, 2, 3, 4, 5, 6, 7, 9, 11}.
After we delete 1, 2, 3, 4, 5, 6, 7, 9, 11 rows and columns from the final table and go to the next iteration.

Second Iteration:

New table after deleting rows and columns:

		10	12	13
		今	木	林
10	今	0	0	0
12	木	0	0	0.8
13	林	0	0.8	0

C2 = {12, 13} (selected at random from the maximum = 0,8),

10	12	13
-	-	-

Now C2 = {12, 13}.

C3 = {10}.

As a result, we have 3 clusters: C1 = {1, 2, 3, 4, 5, 6, 7, 9, 11}, C2 = {12, 13}, C3 = {10}.

{{日, 田, 且, 貝, 回, 甲, 口, 申, 目, 由}, {木, 林}, {今}}.