Assignment 6

Remarks:

1. Please submit your work until **January, 13 23:59** on Ilias.
2. If you need any **help**, send an email to `max.kisselew@ims.uni-stuttgart.de` or drop in at my office: 1.013 (Pfaffenwaldring 5b).

Exercise 1 (IIR 18) [3 P.]

Given the singular value decomposition of the matrix C from the lecture.

C	d_1	d_2	d_3	d_4	d_5	d_6
ship	1	0	1	0	0	0
boat	0	1	0	0	0	0
ocean	1	1	0	0	0	0
wood	1	0	0	1	1	0
tree	0	0	0	1	0	1

Σ	1	2	3	4	5
1	2.16	0.00	0.00	0.00	0.00
2	0.00	1.59	0.00	0.00	0.00
3	0.00	0.00	1.28	0.00	0.00
4	0.00	0.00	0.00	1.00	0.00
5	0.00	0.00	0.00	0.00	0.39

U	1	2	3	4	5
1	0.44	-0.30	-0.57	0.58	-0.25
2	0.13	-0.33	0.59	0.00	-0.73
3	0.48	-0.51	0.37	0.00	0.61
4	0.70	0.35	-0.15	-0.58	-0.16
5	0.26	0.65	0.41	0.58	0.09

V	1	2	3	4	5
1	0.75	-0.29	-0.28	0.00	0.53
2	0.28	-0.53	0.75	0.00	-0.29
3	0.20	-0.19	-0.45	0.58	-0.63
4	0.45	0.63	0.20	0.00	-0.19
5	0.33	0.22	-0.12	0.58	-0.41
6	0.12	0.41	0.33	0.58	0.22

1. Calculate the reduced matrix C_3. That is the term-document matrix C reduced to 3 dimensions (see slides).
2. Compare the rankings of the query "ship ocean" for the matrices C and C_3: Rank the documents after relevance.

Exercise 2 (IIR 19) [3 P.]

The shingle representations of three documents are as follows: $d_3 = (0,0,1,0,0,0,1)^T$, $d_4 = (0,0,1,0,0,0,0)^T$, $d_5 = (1,1,1,0,1,1,1)^T$

We will use sketches of size 2. The two elements of a sketch are defined by the permutations $(2 \times n + 2) \mod 7$ and $(4 \times n + 1) \mod 7$. Based on this setup, what are the estimates of the three Jaccard coefficients $J(d_3, d_4)$, $J(d_3, d_5)$, and $J(d_4, d_5)$? Use the kind of table introduced in class to visualize the permutations and to calculate the final sketches.

Due date: Sunday, January 13, 2013, 23:59