T-61.5060 Algorithmic methods in data mining
Exercises November 8, 2007

1. Which properties of Kleinberg's does the k-means algorithm satisfy?

2. Look at the abstracts dataset available from the web page of the course and investigate empirically the relationship of Jaccard distance and L_2 distance.

3. Investigate the relationship of Jaccard distance and L_2 distance on randomly generated 0-1 vectors. Give an analytical result on the relationship between $J(x, y)$ and $d_2(x, y)$.

4. Prove that the Jaccard distance is a metric.

5. Suppose we define the distance between two strings as the smallest number of operations from a class of allowed changes that takes one string to another. Under what conditions is such a distance a metric?