**Information Retrieval**

**01 February 2016**

**Ex 1 [rank 5]** Show how to synchronize via rsync the new file “bacaddabbb” (on the server) using the old file “acabbbdab” (on the client) and blocks of size 3 chars.

**Ex 2 [points 4+3]** Given the array B=0100 0111 1011 0010,

* Construct the Rank data structure over B by setting Z=4 and z=2.
* Describe how it is executed Rank1(10) by deploying that data structure.

**Ex 3 [ranks 5+3]** Given the dictionary D={bingo, bino, bon, bull}

* Construct an efficient data structure for the 1-error match problem (i.e. three admitted kinds of errors: INS-DEL-SUBST).
* Given the previous item, describe the algorithm that searches for the strings in D which are at 1-error from the pattern P=bin.

**Ex 4 [ranks 3+3]** Given two sets A and B,

* show a LSH-based solution to compute a representation (sketch) of the two sets having length L and such that the Jaccard distance between A and B can be estimated via the two sketches.
* let A={1, 4, 6, 7} and B={2, 3, 4, 6}, instantiate your solution above on this example by assuming that A and B are drawn from the Universe = {1, …, 7} and L=3.

**Ex 5 [ranks 4]** Given an annotator, such as TagMe, show how you could empower the TF-IDF vector representation in order to perform “semantic” comparisons between documents by using still the cosine similarity.

**Ex 6 [*lode*]** State and prove the Theorem for the 2-optimality of MTF.