**Information Retrieval**

**03 February 2017**

**Ex 1 [points 4+3]** Given the binary array B = 0110 1010 1111 1100

* Construct the Rank data structure over B with Z=4 and z=2
* Describe how it is executed Rank1(10)

**Ex 2 [points 2+3+3]** Show the compressed encoding of the following integers:

* 5 and 12 with DELTA-code
* the sequence (1, 3, 7, 4, 2, 3, 1, 1, 2, 9) with PForDelta: base=0 and b=2 bits
* Given the previous sequence, how you would turn it into a sequence that could be encoded via Elias-Fano? Show the overall result.

**Ex 3 [ranks 5]** Describe and comment the Shingling and Min-hashing techniques to compute the near-duplicate score between two textual documents.

**Ex 4 [points 4]** Given the directed graph G consisting of nodes {A, B, C, D} and edges {(A,B), (B,C), (D,C), (C,A)}: Describe how you compute the similarity between node B and node A via Personalized PageRank, and then show the result by iterating it once.

**Ex 5 [points 3+3]** You are given a binary tree T formed by n=6 nodes {a,b,c,d,e,f}, rooted in “a”, and having the following structure {(a,b), (a,c), (c,d), (c,e), (d,f)}.

* Show a binary encoding E(T) of T that takes 2n+1 bits
* Describe how it is possible to navigate E(T) for left/right children