**Information Retrieval**

**11 January 2016**

**Ex 1 [rank 3]** Given the two strings S1=baaa and S2=baacac use Z-Delta to compress S2 with respect to S1.

**Ex 2 [points 2+3+4]** Given a sequence of integers S=(1, 6, 15, 18, 21, 24, 30), encode them using:

* Gamma encoding
* Delta encoding
* Elias-Fano encoding

**Ex 3 [ranks 2+4]** Given the dictionary D={bingo, bins, box, bull, cat}

* construct a 3-gram index for D (by guaranteeing L 3-grams per L-long strings);
* describe the algorithm that searches for the strings in D which are at 1-edit distance from P=binno.

**Ex 4 [ranks 5+3]** Let us given a graph G of directed edges {(1,3), (3,1), (3,2), (1,4), (2,1)}.

* Simulate the execution of **two steps** of the PageRank algorithm, starting with all nodes having score 1, and assuming a teleportation step which favors node 3 (hence, it jumps only to it).
* Simulate the HITS computation of a(3) and h(3) by assuming that all a- and h-scores are initially set to 1.

**Ex 5 [ranks 2+2]** In the context of a text annotator like TagMe, define

* the *link probability* of an anchor text *a* with respect to the Wikipedia knowledge graph.
* the *commonness* of the link between an achor *a* and a Wikipedia page *p*.

**Ex 6 [*lode*]** Show how it is possible to efficiently identify whether two bit-vectors of size d and hamming distance *k* are *highly similar* using the LSH approach, and compute the probability of detection as a function of *d* and *k*.