**Information Retrieval**

1. **June 2019**

**Ex 1 [points 3+3+3]**

* Construct a 2-gram index for the set of strings S={dat, daz, dit, dox, zox}.
* Design an algorithm that finds all **candidate** strings of S to possibly have 1-edit distance from P[1,p] by using the **overlap distance**.
* Simulate the proposed algorithm over the string P=dix.

**Ex 3 [ranks 2+2+3+3]** Given the list of integers S=(2,4,10,13,14,21).

* Show how to compress S via gap-coding and then Gamma-code
* Show how to compress S via gap-coding and then Delta-code
* Show how to compress S via Elias-Fano coding.
* Describe how to select the 5th integer from S, by assuming to have a Select\_1 data structure over the array H.

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

* Simulate the execution of **one step** of the PageRank algorithm, starting with nodes 1 and 2 having score ¼ and node 3 having score ½, 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 anchor *a* and a Wikipedia page *p*.