**Information Retrieval**

**29 January 2014**

**Ex 1 [ranks 3+4]** Given a binary array B of size n,

* define Rank and Select primitives over B,
* describe a possible use of these primitives, commenting on the time/space complexity advantage and possible disadvantages.

**Ex 2 [ranks 3]** Given the string S=aaba, compute the length of its Arithmetic encoding **without running** the entire algorithm, and comment the answer.

**Ex 3 [points 3+3+2]**

* Given the array of scores A = [2, 5, 4, 3, 1, 10 , 8 , 9, 7 , 6] show the steps performed by the algorithm that identifies the top-5 elements in A (i.e., the five elements with the largest values in A).
* Indicate and discuss the time complexity of this algorithm.
* Explain how to modify the solution above to identify the k elements with the smallest values.

**Ex 4 [points 4]** Given a term-document matrix (m *x* n), describe how to project via LSI-technique a query vector of size m onto a reduced space of size k.

**Ex 5 [points 4+2]** Giventhe array of integers (1, 2, 5, 11, 12, 15, 16, 20, 23, 26, 31),

* Compress with Elias-Fano
* Comment on the space occupancy of the proposed solution.

**Ex 6 [points 2]** Prove that the algorithm K-means converges to a local minimum.