**Information Retrieval**

1. **June 2014**

**Ex 1 [ranks 3+5]** In the optimal solution to the RMQ problem on an array of integers, it is introduced a so called “(+1/-1) array”:

* Define it, and sketch how it is derived
* Describe the 4-russian trick detailing how it is used in the RMQ problem and the why.

**Ex 2 [ranks 4+3+4]** Given symbols and probabilities: p(a) = 0.5, p(b)=p(c) = 0.25.

* Decompress the first 4 symbols of the Arithmetic coded bit sequence: 10011.
* Indicate the formula for determining how many bits have to be emitted by Arithmetic coding and prove why this number ensures the correctness of the algorithm.
* Prove that the arithmetic coding with infinite precision arithmetic is close to Entropy

**Ex 3 [points 3+4+4]** Given the set of strings S={aba, abc, baac, babc}.

* Construct a trie T for S
* Construct the LOUDS representation of the tree structure of T
* Describe the procedure for computing the first-child of node x given LOUDS

**[Ex \*]** Define what is the “margin” in SVM and how can you solve classification problems which are not linearly separable?