**Information Retrieval**

**19 December 2018**

**Name: Surname: Matricola:**

**Ex 1 [points 5+5+3]**

* Given a binary array B[1,n] describe the data structure that is built to solve the Rank\_1 query taking o(n) bits in addition to the storage of B.
* Describe the solution based on champion lists to efficiently solve AND-queries over q terms and TF-IDF relevance score.
* Define Link Probability and Commonness as they are used by entity annotators, like TagMe.

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

Given the words {abaco, adagio, baco, alano}

* construct a 2-gram index,
* Show the algorithm that, based on the overlap distance, selects “good candidates” for the 1-edit distance problem over the two query strings P1=abco, P2=acabi.

**Ex 3 [points 4+4]**

Given the directed graph G consisting of nodes {A, B, C, D} and edges {(B,A), (A,C), (D,C), (A,B), (C, B), (D,A), (B,D)}.

* Compute two steps of PageRank by assuming uniform starting probability and teleportation step with **parameter ¼, hence alpha = ¾**  (*hint:* try to operate only on fractions).
* Compute one step of Personalized PageRank for the node A by assuming starting probability [1, 0, 0, 0] and teleportation step with **parameter ½** (*hint:* try to operate only on fractions).