**Information Retrieval**

**5 September 2017**

**Name: Surname: Matricola:**

**Ex 1 [points 4+3+3]** Let us given a set of strings S = { atom, bad, baddy, zoo }.

* Build a 2-gram index over S
* Given pattern P = ady, show how the index executes the search for 1-edit error
* Given pattern P = ady, show how the index executes the search for 2-edit errors

**Ex 2 [points 3+3]** You are given the two files: F\_old = “cane gatto orso”, F\_new = “cane matto dorso”, and assume a block size B=3 chars.

* Show the execution of the algorithm rsync. *(comment the various steps)*
* Show the execution of the algorithm zsync. *(comment the various steps)*

**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)}:

* Compute one step of the PageRank of G’s nodes by assuming that the teleportation step occurs with probability 0.5 and the starting probability distribution is uniform.
* Comment how the similarity between node B and all the other G’s nodes can be estimated by using Personalized PageRank. Apply your algorithm over G for one step only and with the teleportation step occurring with probability 0.5.

**Ex 4 [points 2+4]** Describe a dynamic index based on a small-big solution or a cascade of indexes, specifying the pros/cons of each of these solutions to the problem of adding and searching documents into the index.