**Information Retrieval**

**13 January 2020**

**Name: Surname: Matricola:**

**Exercise [points 3+4+4]** Given a sequence of integers S = (1, 6, 11, 13, 15, 20, 21), encode them using (by motivating **which sequence** is compressed):

* Gamma-code (of Elias)
* PForDelta with base=1 and b=2
* Elias-Fano

**Exercise [points 5]** Consider the WAND algorithm by assuming that it is examining the head of the following three posting lists:

t1 🡪 (…, 5, 6, 7, 8, 11)

t2 🡪 (…, 2, 3, 5, 7, 8, 11)

t3 🡪 (…, 1, 4, 7, 13, 15)

At that time the current threshold equals 2, and the upper bounds of the scores in each posting list are: ub\_1 = 0.4, ub\_2 = 1, ub\_3 = 0.7.

Which is the next docID whose full score is computed? (Motivate your answer)

**Exercise [points 3+4+4+3]**

Answer to the following questions:

1. Define Precision, Recall and F1.
2. Draw the graph G with edges (1,2), (2,3), (3,1) and (2,1), and show the formulas of the HITS algorithm for the hubness and authority scores of every node of your graph.
3. Show how to compute via Bloom Filter the intersection between two sets A and B, each one stored on a different machine, and **estimate** its communication cost (in bits) and its expected error (as the average number of extra items returned).
4. Show and prove the probability of a collision via Karp-Rabin’s rolling hash for two binary strings A and B of m bits each