# Algorithm Engineering -- EXERCISES
# 15 January 2024 – 1 hour

## Name and Surname: #matricola:

**Question #1 [score 4+3+3]** Given the set of strings

S = {BAA, BAB, BACAA, BACAB, BACAD, BACB, CA, CB},

index S via a two-level scheme with block size of 2 strings each and a Patricia trie in internal memory.

Then show how to perform:

- A **lexicographic** search for the string BB
- A **prefix** search for the string BAC.

**Question #2 [score 5].** Given the sequence of integers (2, 5, 7), compress it via Interpolative Coding.

**Question #3 [rank 5].** Given a sequence of strings (BACAB, ABB, BBC, DD, DF), sort them via multikey quicksort by assuming that the pivot is taken as the first string of each (sub-)sequence to be sorted.

**Question #4 [score 6].** Given the text T = ABRABRA, apply the pipeline BWT+MTF+RLE0 (with Wheeler's code) and finally apply Arithmetic coding on the first 3 numbers of the output of this pipeline.

**Question #5 [score 4]** Assume you are given 5 strings (aa, ab, bb, bc, cc) and you wish to construct a minimal ordered perfect hash function (MOPHF).
Assume that rank(c) is the ordered position of '*c*' in the alphabet, counting letter *a* from *1*.
We let the two random functions required by the design of MOPHF as
h1(c' c'') = 2 * rank(c') + rank(c'') mod 11 and h2(c' c'') = 3 * rank(c') * rank(c'') mod 11. Construct the final h(t).

# Algorithm Engineering -- THEORY
# 15 January 2024 – 45 minutes

## Name and Surname: #matricola:

**Question #1 [score 5+3]**

- Prove that the expected length of an ordered sequence produced by the algorithm Snow Plow is 2M.
- What is that expected length if the probability for an item to go in the "unsorted bucket" is ¼ instead of ½ ?

**Question #2 [score 5+4+3].**

- Show how to COUNT in a text T[1,n] all occurrences of a pattern P[1,p], by assuming that T has been indexed via a Suffix Array data structure, built off-line and residing in memory.
- Show and prove the time complexity of the above COUNT operation.
- What is the I/O-cost of performing the RETRIEVAL of the positions of all pattern occurrences in the case that the Suffix Array is stored on disk?

**Question #3 [rank 5+5].** Given two sorted lists of integers, say L1 and L2 of lengths n and m respectively:

- Describe the "doubling algorithm" to compute their intersection and state its time complexity.
- Prove the time complexity of the previous point.