Algorithm Engineering – exercises
10 Febuary 2023 – time 60 minutes

Name and Surname: #matricola:

Question #1 [scores 5+5] Given the sequence a,b,c,d,e,f,g,h,i,l simulate:

- The sampling algorithm for m=2 which knows the sequence length n=10, assuming probabilities for the parameter p = [0.5, 0.5, 0.5, 1, 1, 0.1, 0.5, 1, 0.1, 1]
- The sampling algorithm for m=2 which does not know the sequence length, assuming values for the parameter h = [1, 3, 4, 2, 1, 5, 4, 6]

Question #2 [scores 4+3+3] Given the ordered set of strings

S = { AB, ACA, ACB, CA, CB }

- Build the UNcompacted trie T for S by assuming an alphabet of 3 characters (Σ = {A,B,C}) and branching implemented via arrays.
- Show how to succinctly encode the structure of T in a binary array B, by assuming that pointers to strings are leaves of the tree T, and branching nodes are the internal nodes of T;
- Write a pseudo-code that, given a binary array succinctly encoding the structure of a binary tree (with its corresponding rank/select data structures), establishes the length of its left-only path (or, equivalently, the depth of its leftmost NULL pointer.

Question #3 [scores 5] Build a Treap by inserting the following sequence of pairs <key,priority> and assuming that the MIN priority is in the root (the order among the keys is the alphabetic one):

<E,1><C,14><M,5><A,12><B,8>

Question #4 [scores 5]. Decode the compressed sequence <4, 011110> produced by arithmetic code, by assuming probabilities P[a]=P[c]=1/4 and P[b]=1/2.

Algorithm Engineering – theory
10 Febuary 2023 – time 60 minutes

Name and Surname: #matricola:

Question #1 [scores 5+2]

- State and prove the main theorem used in the time complexity of cuckoo hashing, about the probability of having a shortest path of length L between two nodes i,j in the cuckoo graph for a table T of size m containing n keys.
- Comment on the impact of the formulas on the loading factor of the cuckoo hash table T

Question #2 [scores 5+3] Given a dataset S of n strings of total length N, drawn from an alphabet of size a:

- Write the pseudocode and prove the time complexity of Multi-key Quicksort when applied on the dataset S
- Prove the lower bound for sorting S

Question #3 [scores 3+3+3] State the upper bound for:

- Sorting n atomic items in a two-level memory model with internal memory of size M and disk-page size B, and 1 disk.
- Permuting those items in the two-level memory model, and 1 disk
- Sorting the atomic items in D disks by using the disk striping technique

Question #4 [scores 4+4]

- Define the Treap data structure when built over n pairs <key, priority>, and having the minimum priority in its root
- Describe how it is implemented the split operation over a key K not occurring in the Treap.