

# Algorithm Engineering

15 July 2010

**Exercise 1 [2+3 points].** Given a set of  $N$  keys drawn from a Universe  $U$ , design an algorithm that prints the duplicate keys and evaluate its I/O complexity by distinguishing two cases:

- Case 1: assume that  $|U|$  is arbitrarily large
- Case 2: assume that  $|U|$  is smaller than  $M$

**Exercise 2 [5 points].** Write the pseudocode of an algorithm that builds the Suffix Array of a string  $S[1,n]$ , and comment on its time and I/O-complexity.

**Exercise 3 [3 points].** Assume that  $\text{BuildSA}(S,n)$  is an algorithm that returns the SuffixArray of the string  $S[1,n]$ . Write an algorithm that uses  $\text{BuildSA}$  as a black-box to build the BWT of the string  $S$ .

**Exercise 4 [3+3 points].** Given the string  $S="abcabdabcad"$ :

- Compute its LZ77-parsing
- Compute its LZ78-parsing

**Exercise 5 [3 points].** Given a dictionary  $D$  of  $2^{16}$  strings, compute the error rate of a Bloom Filter which uses an array of  $2^{20}$  bits and an optimal number of hash functions. *[Assume that logs are in base 2]*

**Exercise 6 [4 points].** Specify for which distributions of the positive integers the Gamma- and Delta-codes are optimal. *[Pls motivate your answer]*

**Exercise 7 [6 points].** Given the set of keys  $S=\{1, 5, 6, 7, 11, 12, 14, 17, 18, 19\}$  design a perfect hash table for  $S$ , by assuming that the table of the first level has size 5. *[Pls detail and comment all of your design choices.]*