58093 String Processing Algorithms (Autumn 2014)
Course Exam, 17 December 2014 at 17-20
Lecturer: Juha Kärkkäinen

Please write on each sheet: your name, student number or identity number, signature, course name, exam date and sheet number. You can answer in English, Finnish or Swedish.

1. \[ 4+4+4 \text{ points} \] Each of the following pairs of concepts are somehow connected. Describe the main connecting factors or commonalities as well as the main separating factors or differences.
   (a) String quicksort and MSD radix sort.
   (b) Horspool algorithm and BNDM algorithm.
   (c) Aho–Corasick automaton and suffix tree.

A few lines for each part is sufficient.

2. \[ 6+6 \text{ points} \]
   (a) Compute the edit distance between strings \textit{tukholma} and \textit{stockholm} using the dynamic programming algorithm described on the course.
   (b) Give all optimal alignments between \textit{tukholma} and \textit{stockholm}, i.e., alignments with the same cost as the edit distance.

3. \[ 6+7 \text{ points} \] Let \( A, B, B' \) and \( C \) be strings such that \( A \leq B \leq C \) and \( A \leq B' \leq C \).
   (a) Prove that \( \text{lcp}(B, B') \geq \text{lcp}(A, C) \). You may assume only basic definitions from the course to be known, i.e., do not use any lemmas or theorems from the course.
   (b) Describe in detail how the above result can be used for speeding up string binary searching.

4. \[ 13 \text{ points} \] The reverse of the string \( A = a_1a_2 \ldots a_m \) is the string \( A^R = a_m \ldots a_1 \). Describe an algorithm that, given two strings \( S \) and \( T \), finds the shortest string \( X \) such that \( X \) occurs in \( S \) but neither \( X \) nor \( X^R \) occurs in \( T \). The time complexity should be linear on a constant size alphabet.