Universität Bielefeld
Technische Fakultät

AG Genominformatik
Prof. Dr. Jens Stoye
Dr. Lucas Robidou

Sequence Analysis 3

Summer 2024

Exercises

Number 2, Discussion: 2024-April-25

1. Develop a polynomial-time algorithm that finds the longest increasing subsequence in an array $X[1, 2, \ldots, n]$ of n numbers.
2. Draw a deterministic finite automaton that recognizes the set of strings over the alphabet $\{\mathtt{a}, \mathtt{b}\}$ that start with an `a` and do not contain `ba`.
3. Provide an infinite family of strings for which each strings has a trie of its conjugates (rotations) that is of quadratic size.
4. Make deterministic and minimize the following two nondeterministic finite automata:
   (a) $Q = \{1, 2\}$; $\delta(1, \mathtt{a}) = \emptyset$, $\delta(1, \mathtt{b}) = \{1, 2\}$, $\delta(1, \varepsilon) = \{1\}$, $\delta(2, \mathtt{a}) = \{2\}$, $\delta(2, \mathtt{b}) = \emptyset$, $\delta(2, \varepsilon) = \{1, 2\}$; $q_0 = 1$; $F = \{2\}$.
   (b) $Q = \{1, 2, 3\}$, $\delta(1, \mathtt{a}) = \{2\}$, $\delta(1, \mathtt{b}) = \emptyset$, $\delta(2, \mathtt{a}) = \{2\}$, $\delta(2, \mathtt{b}) = \{2, 3\}$, $\delta(3, \mathtt{a}) = \emptyset$, $\delta(3, \mathtt{b}) = \emptyset$; $q_0 = 1$, $F = \{3\}$.
5. Analyse the efficiency of the Aho-Corasick automaton:
   (a) the time to construct the automaton for a set of k patterns of maximal length l;
   (b) the time to search in a text of length n.
6. Give the matching statistics of the pattern $p = \mathtt{abaaba}$ for the text $t = \mathtt{aabcbabaaabaabaab}$.