Theoretical Computer Science 1
Exercise Sheet 2

Thomas Haas
Prof. Dr. Roland Meyer
TU Braunschweig
Winter semester 2020/21

Release: 17.11.2020 Due: 27.11.2020, 17:00

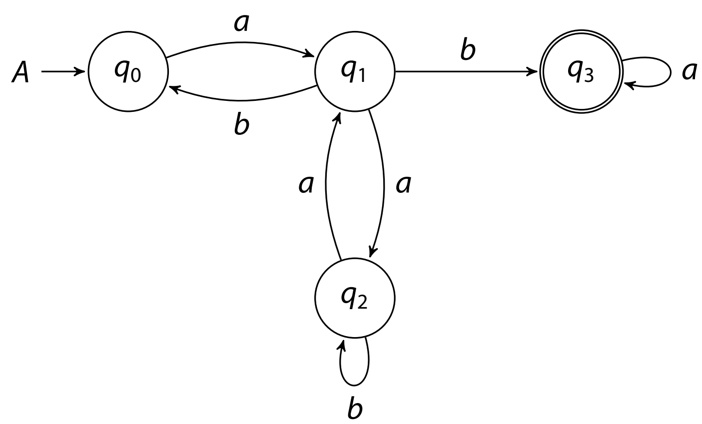
Hand in your solutions per E-Mail to your tutor until Friday, 13.11.2020 17:00 pm. You should provide your solutions either directly as .pdf file or as a readable scan/photo of your handwritten notes.

The references (e.g. Theorem 3.18) refer to the current version of the script.

Note: The power set construction will be content of next week's lecture.

Exercise 1: NFA to REG using Ardens Rule [7 Points]

Let A be the following NFA over the alphabet $\Sigma = \{a, b\}$:



a) [1 Point] Formulate the equation system associated with A.

b) [3 Points] Find a regular expression for $\mathcal{L}(A) = X_0$ by solving the equation system using Arden's Rule. Note: X_0 refers to the equation associated with the initial state of A.

c) [3 Points] From the regular expression, construct an NFA B with $\mathcal{L}(B) = X_0$. Compare the size (number of states) of A and B.

Exercise 2: Input sanitization [8 Punkte]

Check whether the following problems can be considered as problems over regular languages. Explain your answer by e.g. giving a regular expression, a finite automaton or a construction of such automaton, if possible, or by arguing that the language is indeed not regular. Correctness proofs are not needed.

The following problems are about sanitization of input. Consider the alphabet $\Sigma = L \cup U \cup D \cup S \cup W$, where L are lower case letters, U are upper case letters, D are digits, S are special characters and W are white spaces.

a) [1 Point] *Username:* Does the input text have **at least 4 symbols** and no **special characters**?

b) [2 Points] Does the input text satisfy the following property: **Every special character** is *either* **escaped** with a preceding **/** (e.g. // or /' or /!), *or* it is contained **between two (non-escaped) single quotes '** ('10/05/1998' or 'Th!s !s quoted and €scapes / do nothing here, but are also not needed /')

   Assume that escape symbols have no effect inside a quoted section.

c) [2 Points] *Parenthesization:* Is the input text **correctly parenthesized**, i.e. does every opening parenthesis have a matching closing parenthesis and vice versa? "(ri)(gh)t", "R(i(g)h)t" are correct, but "w(r)on)g" and "W)r)o(n(g" are not.

d) [3 Points] *Password*: Does the input text have **between 8 and 20 symbols**, where **every type of symbol** (L, U, D and S) is contained **at least once**?

Exercise 3: Theorem 3.18 [8 Points]

Let $A = (Q_A, q_0, \rightarrow, Q_F)$ be an NFA and let $A^{\text{det}} = (Q^{\text{det}}, q_0^{\text{det}}, \rightarrow_{\text{det}}, Q_F^{\text{det}})$ be the automaton constructed via the Rabin-Scott powerset construction, where $Q^{\text{det}} = \mathcal{P}(Q_A) = \{Q \mid Q \subseteq Q_A\}$, $q_0^{\text{det}} = \{q_0\}$ and $Q_F^{\text{det}} = \{Q \subseteq Q_A \mid Q \cap Q_F \neq \emptyset\}$. It holds $Q \xrightarrow{a}_{\text{det}} Q'$ if and only if $Q' = \{q' \in Q_A \mid \exists q \in Q{:}\, q \xrightarrow{a} q'\}$. Note that the automaton A^{det} is completely deterministic since for every pair of states Q and every input symbol a, there is a unique successor state Q'.

The task of this exercise is to proof Theorem 3.18. Towards this, proceed as follows:

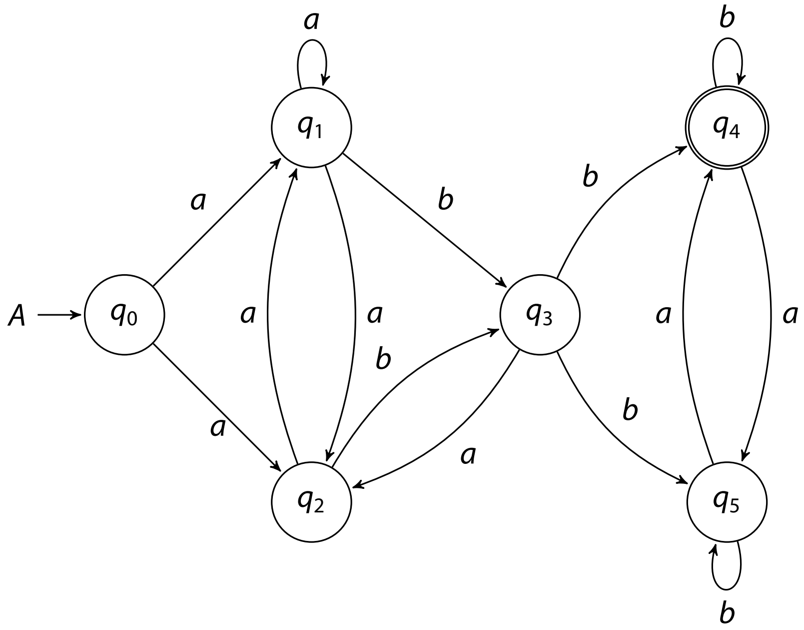
a) [3 Points] Show by induction on i: For every run $q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} \dots \xrightarrow{a_i} q_i$ of A, the (unique) run $Q_0 = q_0 \xrightarrow{a_1}_{\text{det}} Q_1 \xrightarrow{a_2}_{\text{det}} \dots \xrightarrow{a_i}_{\text{det}} Q_i$ of A^{det}, which reads the same word, satisfies $q_i \in Q_i$.

b) [3 Points] Show by induction on i: For every run $Q_0 = q_0 \xrightarrow{a_1}_{\text{det}} Q_1 \xrightarrow{a_2}_{\text{det}} \dots \xrightarrow{a_i}_{\text{det}} Q_i$ of A^{det} and every state $q_i^{\text{det}} \in Q_i$ there exists a run $q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} \dots \xrightarrow{a_i} q_i = q_i^{\text{det}}$ of A, which reads the same word and stops in q_i^{det}.

c) [2 Punkte] Using the partial results of a) and b), prove that $\mathcal{L}(A) = \mathcal{L}(A^{\text{det}})$ holds.

Exercise 4: Powerset construction and complementation [7 Points]

Let A be the following NFA over the alphabet $\Sigma = \{a, b\}$.



a) [2 Points] Determinize A, that is, find a DFA A^{det} with $\mathcal{L}(A) = \mathcal{L}(A^{\text{det}})$ by using the Rabin-Scott powerset construction.

   Note: You can restrict to the states reachable from the initial state $\{q_0\}$. For this, start with $\{q_0\}$ as the only state and then iteratively construct for the current set of states all possible direct successors until no more states are added.

b) [1 Point] Compare the size of the state space of A^{det} with the worst-case-value of $2^{|\{q_0,\ldots,q_4\}|}$.

c) [1 Point] Construct an automaton $\overline{A^{\text{det}}}$ with $\mathcal{L}(\overline{A^{\text{det}}}) = \overline{\mathcal{L}(A)}$.

d) [3 Points] For the word $w = aababba$, give all possible runs of A on w and the unique run of A^{det} on w. How many different runs on $w = aababba$ are there in A? Is $w \in \mathcal{L}(A)$?