Theoretical Computer Science 1
Exercise 3

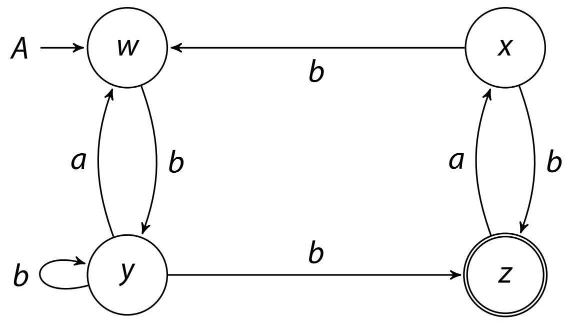
René Maseli
Prof. Dr. Roland Meyer
TU Braunschweig
Winter semester 2020/21

Release: 7.12.2021 Due: 16.12.2021, 23:59

Hand in your solutions per e-mail or **StudIP** to your tutor until Thursday, 16.12.2021 23:59 pm. You should provide your solutions either directly as .pdf file or as a readable scan/photo of your handwritten notes. Submit your results as a group of four.

Exercise 1: Homomorphisms [10 points]

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



Note: This automaton coincides with that of the last sheet.

a) [5 points] Consider the homomorphism $f\colon \Sigma \to \{0, 1\}$.

$$f(a) = 01$$
$$f(b) = 10$$

Construct the image-automaton $f(A)$ with $\mathcal{L}(f(A)) = f(\mathcal{L}(A))$.

Show $101001100110100110 \in \mathcal{L}(f(A))$ by giving a corresponding run through A.

b) [5 points] Consider the homomorphism $g\colon \{c, d, e\} \to \Sigma$.

$$g(c) = \varepsilon$$
$$g(d) = bbb$$
$$g(e) = ba$$

Construct the co-image-automaton $g^{-1}(A)$ with $\mathcal{L}(g^{-1}(A)) = g^{-1}(\mathcal{L}(A))$.

Show $ceddc \in \mathcal{L}(g^{-1}(A))$ by giving a corresponding run through A.

Exercise 2: Input sanitization [10 points]

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

Assume the alphabet $\Sigma = L \cup U \cup D \cup S \cup W$, partitioned into lower-case letters L, upper-case letters U, digits D, special characters S and white spaces W.

a) [1 point] Does the input have at least 4 symbols and at most 20?

b) [1 point] Does each class (L, U, D und S) occur at least once?

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] *Escaping:* We may enclose sequences with `'` $\in S$ and *escape* each directly following symbol with \ $\in S$.

Does every non-escaped opening `'` have a matching closing `'` and are all white space and special character either enclosed or escaped? (i.e. `\\`, `\'`, `\!`, `'10/05/1998'` or `'Th!s !s quo\ed and €scapes \\ do nothing \ere\'`)

e) [3 points] *Tables:* Do all rows (separated by newlines `\n`) have the same number of cells (separated by commata `,`)?

Exercise 3: Theorem 3.18 [8 points]

Let $A = (Q, q_0, \rightarrow, Q_F)$ be an NFA over Σ, and $A' = (\mathcal{P}(Q), Q_0, \rightarrow_B, Q_F')$ be the automaton constructed via the Rabin-Scott powerset construction, where $Q_0 = \{q_0\}$ and $X \xrightarrow{a}{}' \{q \in Q \mid \exists p \in X\colon p \xrightarrow{a} q\}$ for all $X \subseteq Q$ and $a \in \Sigma$ and $Q_F' = \{X \subseteq Q \mid X \cap Q_F \neq \emptyset\}$.

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

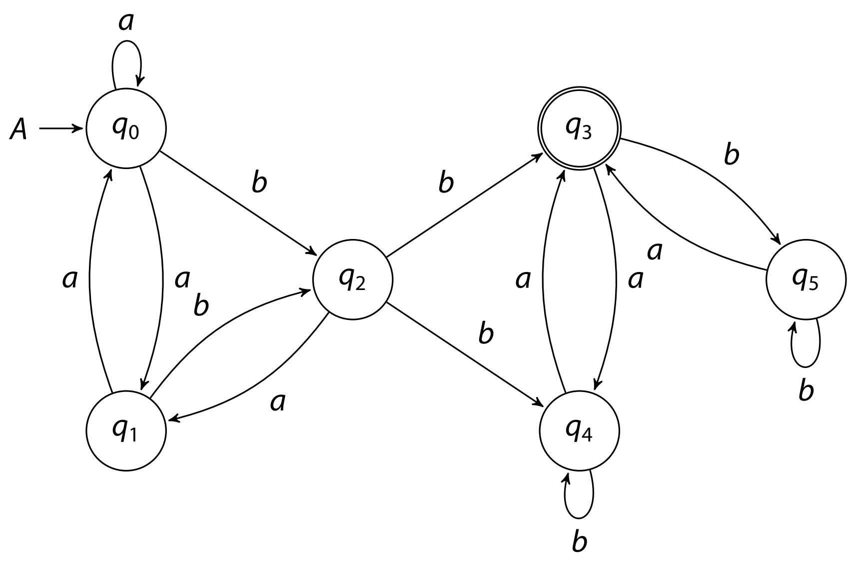
a) [3 points] Show by induction on i: For every run $q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} \dots \xrightarrow{a_i} q_i$ on A the (unique) run $Q_0 \xrightarrow{a_1}{}' Q_1 \xrightarrow{a_2}{}' \dots \xrightarrow{a_i}{}' Q_i$ on A', which reads the same word, satisfies $q_i \in Q_i$.

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

c) [2 points] Using the partial results of a) and b), prove that $\mathcal{L}(A) = \mathcal{L}(A')$ 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 B with $\mathcal{L}(A) = \mathcal{L}(B)$ 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 B with the worst-case-value of $|\mathcal{P}(\{q_0, \ldots, q_5\})|$.

c) [1 point] Construct an automaton $\overline{B}$ with $\mathcal{L}(\overline{B}) = \overline{\mathcal{L}(A)}$.

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