Theoretical Computer Science 1
Exercise Sheet 4

René Maseli
Prof. Dr. Roland Meyer
TU Braunschweig
Winter semester 2022/23

Release: 13.12.2022 Due: 23.12.2022, 23:59

Hand in your solutions to the Vips directory of the StudIP course until Friday, 23.12.2022 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.

Definition: Endlicher Transduktor
A finite-state transducer over a finite input alphabet Σ and a finite output alphabet Γ is formally a quadruple $T = \langle Q, q_0, \rightarrow, Q_F \rangle$ consisting of

1. a finite set of states Q,
2. an initial state $q_0 \in Q$
3. a transition relation $\rightarrow \subseteq Q \times (\Sigma \cup \{\tau\}) \times (\Gamma \cup \{\tau\}) \times Q$,
4. and a set of accepting states $Q_F \subseteq Q$

A transducer can be thought of as an NFA with spontaneous transitions, which not only accepts input words but also outputs new words. It translates input words from Σ^* to output words in Γ^*. Transducers are used in linguistics and the processing of natural languages.

In the following we fix notation and important definitions:

1. $\langle p, a, x, q \rangle \in \rightarrow$ is denoted by $p \xrightarrow{a/x} q$. When reading an a in state p, the transducer transitions to state q and outputs x. Intuitively, $a = \tau$ denotes a spontaneous transition, while $x = \tau$ denotes a transition without output.
2. $\rightarrow^* \subseteq Q \times (\Sigma \cup \{\tau\})^* \times (\Gamma \cup \{\tau\})^* \times Q$ denotes the reflexive, transitive closure of $\rightarrow$. It satisfies $q \xrightarrow{\varepsilon/\varepsilon}{}^* q$ and $q \xrightarrow{w/o}{}^* q_n \iff \exists q_1, \ldots, q_{n-1} : q \xrightarrow{w_1/o_1} q_1 \xrightarrow{w_2/o_2} \cdots \xrightarrow{w_{n-1}/o_{n-1}} q_{n-1} \xrightarrow{w_n/o_n} q_n$.
3. T induces a relation $[\![T]\!] \subseteq \Sigma^* \times \Gamma^*$ as follows:

   $$w \, [\![T]\!] \, o \iff \exists w' \in (\Sigma \cup \{\tau\})^*, o' \in (\Gamma \cup \{\tau\})^* : q_0 \xrightarrow{w'/o'}{}^* q_f \in Q_F \quad \text{and} \quad \pi_\Sigma(w') = w, \pi_\Gamma(o') = o$$

   Hereby, $\pi_\Sigma \colon (\Sigma \cup \{\tau\}) \rightarrow \Sigma^*$ with $\pi_\Sigma(\tau) = \varepsilon$ and $\forall a \in \Sigma \colon \pi_\Sigma(a) = a$ induces a homomorphism, which deletes τ from a word. τ denotes either spontaneous transitions or empty output and hence should not be visible.

   We say that $o \in \Gamma^*$ is an output of T on $w \in \Sigma^*$.
4. T does not only transduce single words, but whole languages. We define for any language $L \subseteq \Sigma^*$ the translation under T as $T(L) = \{o \in \Gamma^* \mid \exists w \in L : w \, [\![T]\!] \, o\} \subseteq \Gamma^*$.

Exercise 1: Finite transducers [20 points]

a) [3 points] Construct a transducer T that for any given word $w \in \{a, b, c\}^*$ works a follows: it prepends a b to every occurence of a and removes every second occurrence of c. Give a regular expression for $T((ac)^*)$.
A proof of correctness is not needed.

b) [3 points] We call a transducer deterministic if in any state and for any input, the transducer has **at most one** possible, and hence **unique**, transition; this transition may be spontaneous. For example, a state with an a-labeled transition may not have another a-labeled transition nor another spontaneous transition, because in either case there would be two possible transitions on a.
Show that it is **not** possible to determinize transducers in general. That means, there are transducers T which do not have any equivalent deterministic transducer T^{det} such that $T(L) = T^{\text{det}}(L)$ for all languages $L \subseteq \Sigma^*$.

c) [3 points] Let $h : \Sigma^* \to \Gamma^*$ be an arbitrary homomorphism between words. Construct a transducer T_h such that $T_h(L) = h(L)$ holds for all languages $L \subseteq \Sigma^*$. Prove the correctness of your construction.

d) [3 points] Now prove that there is also a transducer $T_{h^{-1}}$ such that $T_{h^{-1}}(L) = h^{-1}(L)$ holds for all $L \subseteq \Gamma^*$. Prove the correctness of your construction.

e) [2 points] Show that for any regular language M, there is a transducer T_M with $T_M(L) = L \cap M$.
Remark: In this exercise you have shown that transducers are capable of representing many typical operations on languages. If a class of languages is closed under translations of transducers, then it follows directly that it is also closed under the above mentioned operations.

f) [6 points] Now show that the converse also holds true, i.e. a class of languages that is closed under those three mentioned operations is also closed under translations of transducers.
Remark: You have to show that for any transducer T, you can express the translation of a language L under T in terms of those three operations.

Exercise 2: Unique minimal DFAs [9 points]

Let $L \subseteq \Sigma^*$ be a regular language and $A = \langle Q_A, i_A, \to_A, F_A \rangle$ its Equivalence-class-DFA with $Q_A = \Sigma^*|_{\equiv_L}$, $i_A = [\varepsilon]_{\equiv_L}$, $[v]_{\equiv_L} \xrightarrow{s}_A [w]_{\equiv_L} \iff v.s = w$ and $F_A = \{[w]_{\equiv_L} \mid w \in L\}$. At this point, it was proven that A satisfies $\mathcal{L}(A) = L$ and that A is minimal for L. Let $B = \langle Q_B, i_B, \to_B, F_B \rangle$ another DFA with $\mathcal{L}(B) = L$ and $|Q_B| = [\equiv_L]$.

a) [2 points] Show that all states in B are reachable. This means that for each state $q \in Q_B$, there is at least one word $w \in \Sigma^*$, such that there is a run $i_B \xrightarrow{w}_B^* q$.

Let $f : Q_B \to Q_A$ be defined inductively with $f(i_B) = [\varepsilon]_{\equiv_L}$ and $\forall p \xrightarrow{s}_B q$: $f(q) = [w]_{\equiv_L}.s$.

You will have to show that B is isomorphic to A ($B \sim A$).

b) [3 points] Show that for all $w \in \Sigma^*$, the image is $f(q_w) = [w]_{\equiv_L}$, where $q_w \in Q_B$ denotes the unique state with $i_B \xrightarrow{w}_B^* q_w$.

c) [2 points] Show that f is bijective.

d) [2 points] Let $q \in Q_B$ be a state of B. Show $q \in F_B \iff f(q) \in F_A$.

Exercise 3: Costs of Determinization [6 points]

In this exercise we want to show that some languages that admit a description by small NFAs do not admit a description by small DFAs; every DFA for that language is necessarily large.

For a number $k \in \mathbb{N}, k > 0$ let $L_{a@k} = \Sigma^*.a.\Sigma^{k-1}$ be the language of words over $\Sigma = \{a, b\}$ that have an a at the k-th last position.

a) [1 point] Show how to construct for any $k \in \mathbb{N}, k > 0$ an NFA $A_k = \langle Q_k, q_0, \rightarrow_k, F_k \rangle$ with $\mathcal{L}(A_k) = L_{a@k}$ and $|Q_k| = k + 1$. Give the automaton formally as a tuple.
You do not have to show correctness of your construction.

b) [2 points] Now draw A_3 and its determinization A_3^{det} via Rabin-Scott-power set construction. Compare the number of states of A_3 and A_3^{det}.

c) [3 points] Let $k \in \mathbb{N}, k > 0$ be arbitrary. Prove that for $L_{a@k}$ there is no DFA B with less than 2^k many states such that $\mathcal{L}(B) = L_{a@k}$ holds.
Hint: Proceed as follows:

1. Assume there is a DFA $B = (Q', q_0', \rightarrow', Q_F')$ with $\mathcal{L}(B) = L_{a@k}$ and $|Q'| < 2^k$.
2. Consider the set Σ^k of words of length k. How many such words are there?
3. Now consider to each word $w \in \Sigma^k$ the (unique) state q_w in the DFA B after it read the word w.
4. Now derive a contradiction.