René Maseli
Thomas Haas

Theoretical Computer Science 1
Exercise Sheet 3

TU Braunschweig
Winter Semester 2023/24

Release: 2023-11-24 Due: 2023-12-07 23:59

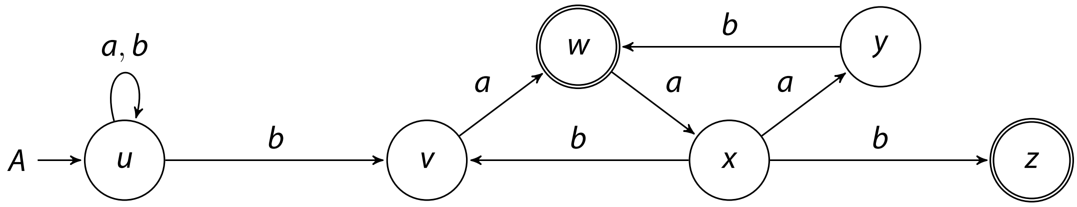
Hand in your solutions to the Vips directory of the StudIP course until Thursday, December 07th 2023 23:59. You should provide your solutions either directly as .pdf file or as a readable scan/photo of your handwritten notes. Submit your results as a group of four and state **all** members of your group with **student id, name and course**.

Homework Exercise 1: Regular languages and finite automata [11 points]

Show that the following statements are valid:

a) [5 points] For each pair of NFAs A and B, there are (other) NFAs $A \cup B$ und $A.B$, with $\mathcal{L}(A \cup B) = \mathcal{L}(A) \cup \mathcal{L}(B)$ and $\mathcal{L}(A.B) = \mathcal{L}(A).\mathcal{L}(B)$.
Hint: Give construction procedures, that work for any A's and B's, and show, that the respective languages are equal.

Now consider the following NFA A over the alphabet $\{a, b\}$:



b) [1 point] Formulate the equation system associated with A.

c) [3 points] Find a regular expression for $\mathcal{L}(A)$ by solving the system using Arden's Rule.

d) [2 points] Give expressions for all other variables of the equation system.

Homework Exercise 2: Input sanitization [8 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 3 symbols and at most 18?

b) [1 point] Does each class of non-space symbols (L, U, D and 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) [2 points] *String literals:* 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? (e.g. `'no'issue'with'\'\\'` or `'Robert\');DROP TABLE Students;--'`)

e) [2 points] *Tables:* Do all rows (separated by newlines `\n`) have the same number of cells (separated by commata `,`)?

Homework Exercise 3: Determinisation [8 points]

Let $A = \langle Q, q_0, \rightarrow, Q_F \rangle$ be an NFA over Σ, and $A' = \langle \mathcal{P}(Q), Q_0, \rightarrow_B, Q'_F \rangle$ be the automaton constructed via the Rabin-Scott powerset construction, with $Q_0 = \{q_0\}$, $X \xrightarrow{a}{}' \{ q \in Q \mid \exists p \in X\colon p \xrightarrow{a} q \}$ 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:

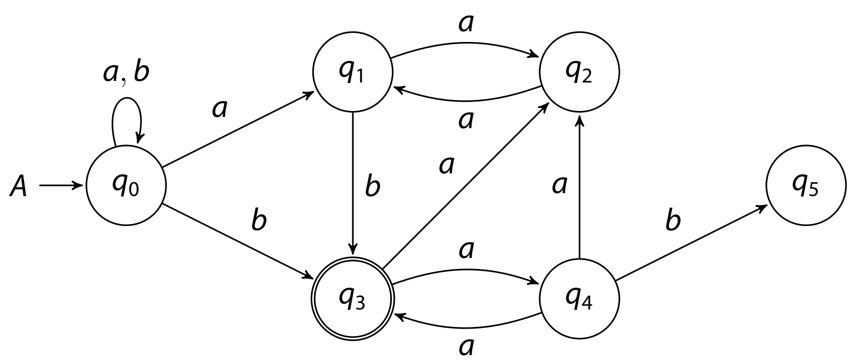
a) [3 points] Show by induction on i: For every run $q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} \ldots \xrightarrow{a_i} q_i$ on A the (unique) run $Q_0 \xrightarrow{a_1}{}' Q_1 \xrightarrow{a_2}{}' \ldots \xrightarrow{a_i}{}' Q_i$ on A', which reads the same word, satisfies $q_i \in Q_i$.

b) [3 points] Show by induction on i: For every run $Q_0 \xrightarrow{a_1}{}' Q_1 \xrightarrow{a_2}{}' \cdots \xrightarrow{a_i}{}' Q_i$ of A' and every state $q_i \in Q_i$ there exists a run $q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} \cdots \xrightarrow{a_i} q_i$ of A, reading 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.

Homework 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 $w \in \mathcal{L}(A)$?

Exercise 5:

Show that the Kleene-Star is indeed a closure operator: $(L^*)^* = L^*$.

Exercise 6:

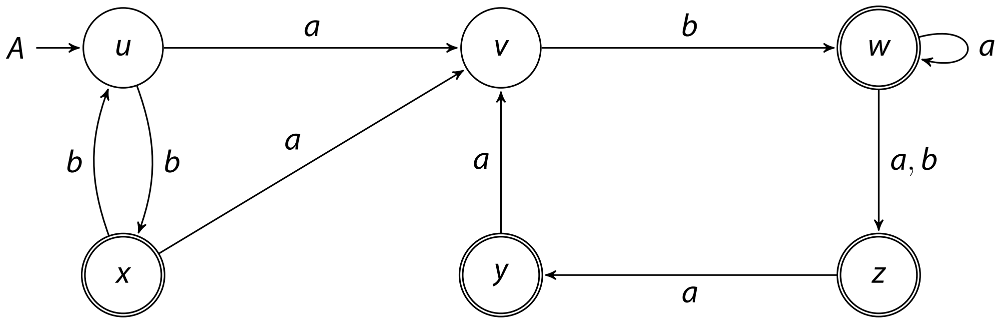
Step-by-step, construct a finite automaton for the regular expression over $\{a, b, c\}$

$$\big(c + a(b + ca)^*(\varepsilon + cc)\big)\big((a + b)c^* + a^*b(cb)^*\big)^* .$$

Give an automaton A for $(a + b)c^*$.
Give an automaton B for $a^*b(cb)^*$.
Give an automaton C for $\mathcal{L}(A) \cup \mathcal{L}(B)$.
Give an automaton D for $\mathcal{L}(C)^*$.
Give an automaton E for $c + a(b + ca)^*(\varepsilon + cc)$.
Give an automaton F for $\mathcal{L}(E).\mathcal{L}(D)$.

Exercise 7:

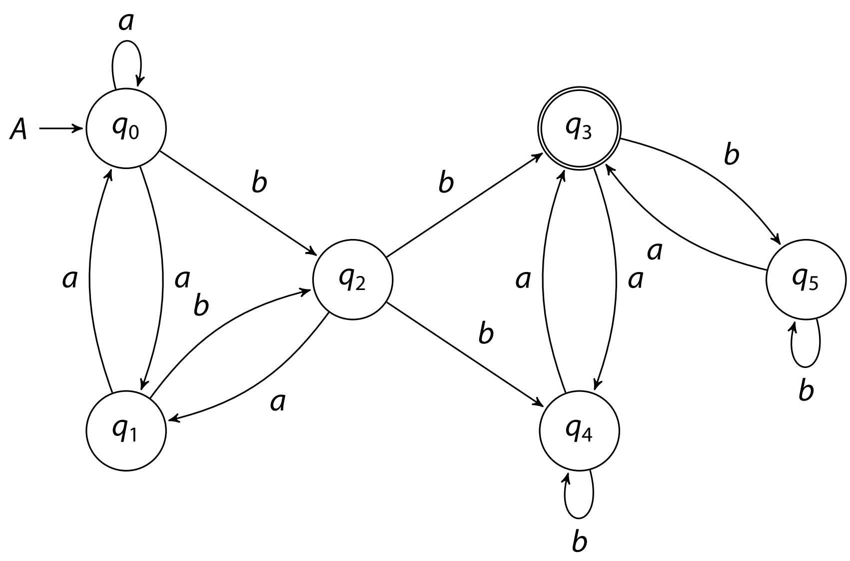
Consider the following NFA A over the alphabet $\{a, b\}$:



Formulate the equation system associated with A.
Find a regular expression for $\mathcal{L}(A)$ by solving the equation system using Arden's Rule.
Give expressions for all other variables of the equation system.
Find a DFA B with $\mathcal{L}(B) = \mathcal{L}(A)$ using the construction by Rabin & Scott.
Give a regular expression for $\overline{\mathcal{L}(A)}$.
Describe what happens in this procedure, if no accepting state is reachable from the initial state.
How does this affect the solution space of the equation system?

Exercise 8:

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



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. Compare the size of the state space of B with the worst-case-value of $|\mathcal{P}(\{q_0, \ldots, q_5\})|$.

Construct an automaton $\overline{B}$ with $\mathcal{L}(\overline{B}) = \overline{\mathcal{L}(A)}$.

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 $w \in \mathcal{L}(A)$?