General notes

Conventions used throughout this exam

- $\Sigma_{\text{bin}} = \{\texttt{0}, \texttt{1}\}$.
- $\Sigma_{\text{abc}} = \{\texttt{a}, \texttt{b}, \texttt{c}\}$.
- Σ_{ASCII} is the ASCII alphabet.
- $\Sigma_{\text{cap}} \subset \Sigma_{\text{ASCII}}$ is the set of all capitals.
- $\Sigma_{\text{dig}} \subset \Sigma_{\text{ASCII}}$ is the set of all decimal digits.
- "Formally describe a language" means that a formal language is to be described in a set-theoretic way, e.g.

$$L = \{\, w \in \Sigma_{\text{bin}}^* \,|\, |w| < 5 \}$$

- *Binary numbers* are arbitrary words of Σ_{bin}^* containing at least one symbol, however, they must not begin with `0` unless the word is `0` itself.
- If not noted otherwise, regular expressions are based on the operator set $\{+, \cdot, ^*, (,)\}$.

Algebraic operations on regular expressions

1. $r_1 + r_2 \doteq r_2 + r_1$ (commutative law)
2. $(r_1 + r_2) + r_3 \doteq r_1 + (r_2 + r_3)$ (associative law)
3. $(r_1 r_2) r_3 \doteq r_1 (r_2 r_3)$ (associative law)
4. $\emptyset r \doteq \emptyset$
5. $\varepsilon r \doteq r$
6. $\emptyset + r \doteq r$
7. $(r_1 + r_2) r_3 \doteq r_1 r_3 + r_2 r_3$ (distributive law)
8. $r_1 (r_2 + r_3) \doteq r_1 r_2 + r_1 r_3$ (distributive law)
9. $r + r \doteq r$
10. $(r^*)^* \doteq r^*$
11. $\emptyset^* \doteq \varepsilon$
12. $\varepsilon^* \doteq \varepsilon$
13. $r^* \doteq \varepsilon + r^* r$

14. $r^* \doteq (\varepsilon + r)^*$

15. $\{r \doteq rs + t \quad \text{with} \quad \varepsilon \notin L(s)\} \longrightarrow r \doteq ts^*$ (proof by Arto Salomaa)

58. $r^*r \doteq rr^*$

65. $r^* \doteq \varepsilon + r^*$

1 Formal languages (6 pts)

Formally describe the following languages:

a) all binary numbers containing the string `101`,

b) all binary numbers whose multiplication by 24 is even,

c) a Voice Genie Session ID (consisting of eight hexadecimal digits followed by a dash followed by eight hexadecimal digits).

2 Regular expressions (8 pts)

a) Give a regular expression for all the words $w \in \Sigma_{\text{bin}}^*$ containing at least two of each character in Σ_{bin}.

b) Give a regular expression for all the words $w \in \Sigma_{\text{bin}}^*$ *not* containing the substring `001`.

c) Determine the shortest word $w_s \in \Sigma_{\text{abc}}^*$ containing equally many `a`s and `b`s matched by

$$r = (\texttt{ca}) + \texttt{bbb}((\texttt{aa})^*\texttt{c})\texttt{bb}^*\texttt{b}\varepsilon\texttt{b}\varepsilon\texttt{c}.$$

3 Algebraic operations (8 pts)

For this section, use the 15 algebraic operations on regular expressions introduced in the lecture. You can also use Equations (58) and (65) as we have proven them before. For the sake of simplicity, the rules are listed in the General Notes section above.

a) Simplify as much as possible:

$$r = ((\varepsilon\varepsilon^*)s + (\varepsilon + (\emptyset^* + (\varepsilon + t) + \varepsilon t)^*((\varepsilon + t) + t) + \varepsilon)^*(t\varepsilon^*s)) + (\varepsilon t\varepsilon^*s)$$

b) Prove that

$$\texttt{1}(\varepsilon^* + \texttt{01})^* \doteq (\varepsilon + (\texttt{10})^*\texttt{10})\texttt{1}.$$

In subtask b), indicate which rules you are using by referring to their ID (1-15, 58, 65) at every step you take.

4 JFlex (6 pts)

Give regular expressions using the JFlex operator set matching

a) the language described in Task 1 c),

b) a date of the format YYYYMMDD (no need to ensure that there are no more than 12 months or more than 31 days),

c) a table row whose second cell contains the word "second".
For the sake of simplicity, assume that

- each table row uses one line of the input text,
- there are at least three cells per row,
- cell delimiter is `\t`.

5 Transforming NFAs into DFAs (8 pts)

Convert the following NFA into a minimized DFA:

$$A = \langle Q, \Sigma, \delta, q_0, F \rangle$$

with

1. $Q = \{0, 1, 2\}$
2. $\Sigma = \Sigma_{\mathrm{abc}}$
3. $\delta = \{\langle 0, \mathtt{a}, 0\rangle\ \langle 2, \mathtt{c}, 2\rangle, \langle 1, \mathtt{a}, 0\rangle, \langle 0, \mathtt{a}, 2\rangle, \langle 1, \mathtt{a}, 1\rangle,$
$\langle 1, \varepsilon, 1\rangle, \langle 0, \varepsilon, 1\rangle, \langle 2, \mathtt{b}, 2\rangle, \langle 0, \mathtt{b}, 1\rangle, \langle 1, \mathtt{b}, 0\rangle\}$
4. $q_0 = 1$
5. $F = \{2\}$

6 Transforming regular expressions into FSMs (4 pts)

Determine a minimized DFA representing the same language as the regular expressions whose equivalence was to be proven in Task 3 b).

7 Formal grammars (8 pts)

a) Define a formal grammar representing the language specified in Task 1 c).
Note: Use quotes to represent terminal symbols (e.g. "A", "1", "-").

b) Consider the grammar

$$G = \langle V_N, V_T, P, S \rangle$$

with

1. $V_N = \{A, B\}$,
2. $V_T = \{\mathtt{1}\}$,
3. P :
$$
\begin{array}{ll}
A \to \varepsilon & 1 \\
A \to \mathtt{1}B & 2 \\
B \to \varepsilon & 3 \\
B \to \mathtt{11}A & 4
\end{array}
$$
4. $S = A$.

i. What is G's highest type? Justify your answer.

ii. Formally describe the language $L(G)$.

iii. If G is not itself regular, define a regular grammar G' with

$$L(G') = L(G).$$