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 $\{+, \cdot, ^*, (,)\}$, and the concatenation operator $\cdot$ can be omitted.

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)

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

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

1 Formal languages (6 pts)

Formally describe the following languages:

a) all words $w \in \Sigma^*_{\text{ASCII}}$ containing one million instances of the word `million`,

b) all binary numbers representing powers of 2,

c) all words adhering to the Base64 encoding scheme (i.e., all the words containing an arbitrary combination of alphanumeric characters or the special characters `+` or `/`; furthermore, words need to be terminated by a single occurrence of the special character `=`).

2 Regular expressions (9 pts)

a) Give a regular expression for all the words $w \in \Sigma^*_{\text{abc}}$ containing no more than one instance of the character `a`, no less than one instance of `b`, and no more than one instance of `c`.

b) Give a regular expression for all the words $w \in \Sigma^*_{\text{bin}}$ not containing the substring `0000`.

c) Specify a regular expression r representing the language $L(r) := L_1 \cap L_2$ with

$$\begin{aligned} L_1 &:= L((\varepsilon + \mathtt{1} + \mathtt{10})^*), \\ L_2 &:= L(\mathtt{10}(\mathtt{1}^*\mathtt{00}^*)^*). \end{aligned}$$

Hint: The *intersection* of the sets A and B is defined as

$$A \cap B := \{x | x \in A \wedge x \in B\}.$$

3 Algebraic operations (6 pts)

Prove that

$$r^*r^* \doteq r^*.$$

For this task, use the algebraic operations listed in the General Notes section above. Indicate which rules you are using by referring to their ID (1-17) at every step you take.
Hint: Provide a proof for an arbitrary r including the case that $\varepsilon \in L(r)$.

4 Transforming NFAs into DFAs (6 pts)

Convert the following NFA into a minimized DFA:

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

with

1. $Q = \{0, 1, 2, 3, 4\}$
2. $\Sigma = \Sigma_{\text{bin}}$
3. $\delta = \{\langle 2, \texttt{1}, 3\rangle\ \langle 4, \texttt{1}, 3\rangle, \langle 2, \varepsilon, 4\rangle, \langle 3, \texttt{1}, 0\rangle, \langle 4, \texttt{1}, 1\rangle,$
   $\langle 4, \varepsilon, 3\rangle, \langle 0, \texttt{1}, 2\rangle, \langle 3, \texttt{0}, 3\rangle, \langle 4, \varepsilon, 2\rangle, \langle 4, \varepsilon, 4\rangle\}$
4. $q_0 = 2$
5. $F = \{1, 2, 3, 4\}$

5 Complements of regular languages (10 pts)

Prove that the regular expression

$$r = (\texttt{a}^*\texttt{c} + \texttt{b})^*\texttt{a}^*$$

matches all strings $w \in \Sigma^*_{\text{abc}}$ *not* containing the substring `ab`. Do so by

i) defining a regular expression r' matching all strings that *do* contain the substring `ab`,

ii) determining the complete minimized DFA A representing r,

iii) determining the complete minimized DFA A' representing r', and

iv) showing that $\text{neg}(A) = A'$.

6 Formal grammars (12 pts)

a) Define a formal grammar representing the language L specified in Task 7.

b) Consider the grammar

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

with

1. $V_N = \{X, Y, Z\}$,
2. $V_T = \{\texttt{a}, \texttt{b}, \texttt{c}\}$,

3. $P:$

$$
\begin{array}{lll}
X \to \mathtt{b} & & 1 \\
X \to \mathtt{a}Y & & 2 \\
X \to Z\mathtt{c} & & 3 \\
Y \to \mathtt{a}Y & & 4 \\
Y \to \mathtt{b} & & 5 \\
Z \to Z\mathtt{c} & & 6 \\
Z \to \mathtt{b} & & 7
\end{array}
$$

4. $S = X$.

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

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

7 The pumping lemma (4 pts)

Prove whether or not the following language is regular:

$$L = \{vwv | v, w \in \Sigma_{\text{bin}}^*\}.$$