Examination
Formal Languages and Automata Theory
TDDD14 & TDDD85

(Formella Språk och Automatateori)

2014–06–03, 14.00 – 18.00

1. **NOT ALL PROBLEMS ARE FOR BOTH COURSES.**
   Pay attention to "only" comments.

2. Allowed help materials

   - A sheet of notes - 2 sided A5 or 1 sided A4.
     The contents is up to you.
     The notes should be signed in the same way as the exam sheets and
     returned together with the exam.

   - English dictionary

   $\Big($ Tillåtna hjälpmedel:

   - Ett papper med valfria anteckningar - 2 sidor A5 eller 1 sida A4.
     Anteckningarna ska signeras på samma sätt som tentamensarken
     och bifogas tentamen vid inlämnandet.

   - Engelsk ordbok $\Big)$

3. You may answer in Swedish or English.

4. Total number of credits is 32. Limits:

   3: 16 p,   4: 22 p,   5: 28 p.

5. *Jour* (person on duty): Johannes Schmidt, tel. 07 25 72 18 00

GOOD LUCK !

Make sure that you justify your answers! Unexplained answers will be granted 0 points. (For instance, if you are writing a grammar for a given language then you should explain that the grammar indeed generates the language. If you apply some known method then you should explain each step. And so on.)

1. *(2p)* The NFAϵ $N = (Q, \Sigma, \Delta, S, F)$ is defined as follows:

$$Q = \{1, 2, 3, 4\} \qquad \Sigma = \{a, b\} \qquad S = \{1\} \qquad F = \{3\}$$

with the transition function Δ given by

	ϵ	a	b
$\to 1$	$\{2\}$	$\{3\}$	$\{3\}$
2	$\emptyset$	$\{1,4\}$	$\{3\}$
$3\,\text{F}$	$\{4\}$	$\emptyset$	$\emptyset$
4	$\emptyset$	$\emptyset$	$\{2\}$

Using the standard method, construct an equivalent DFA M.

2. *(4p)* The DFA $M = (Q, \Sigma, \delta, s, F)$ is defined as follows:

$$Q = \{1, 2, 3, 4, 5\} \qquad \Sigma = \{a, b\} \qquad s = 1 \qquad F = \{1, 3, 4\}$$

with the transition function δ given by

	a	b
$\to 1\,\text{F}$	1	2
2	3	4
$3\,\text{F}$	1	5
$4\,\text{F}$	3	4
5	3	4

   (a) Using the standard method, construct an equivalent, minimal DFA M_{min}.

   (b) Let L be the language defined by M_{min} (and thus by M), and consider the relation $R_L \subseteq \Sigma^* \times \Sigma^*$ defined by

$$x R_L y \Leftrightarrow (\forall z \in \Sigma^* (xz \in L \Leftrightarrow yz \in L))$$

   How many equivalence classes does R_L have? Why? Choose two of the equivalence classes, and give two DFA's defining them.

3. *(2p)* Using the standard method, construct a regular expression defining the same language as the DFA whose transition function δ is given by

	a	b
$\to A$	A	C
$B\,\text{F}$	A	B
$C\,\text{F}$	B	A

4. *(2p)* Construct a context-free grammar for the language

$$L = \{\, a^i b c^j \mid 0 < i \leq j < 2i \,\}.$$

Hint: You may find it convenient to begin with $L_0 = \{\, a^i b c^j \mid 0 < i \leq j \leq 2i \,\}$.

5. *(2p)* Order the following formalisms according to their expressive power: placing A before B means that any language definable by A is definable by B. Also state which, if any, of them are equivalent.
(In this problem you are not required to provide a justification).

   - Context-free Grammars (CFG)
   - Deterministic Finite Automata (DFA)
   - Deterministic Pushdown Automata (DPDA)
   - LR(1) grammars
   - Nondeterministic Finite Automata (NFA)
   - Nondeterministic Finite Automata with ϵ-transitions (NFAϵ)
   - Nondeterministic Turing Machines (NTM)
   - Pushdown Automata (PDA)
   - Regular expressions (reg.exp.)
   - Turing Machines (TM)
   - Turing Machines with multiple tapes (TM+)

   Compare the expressive power of LR(0) grammars with the expressive power of the formalisms listed above.

6. *(2p)* Let Σ, Δ be alphabets and $h : \Sigma^* \to \Delta^*$ be a function. (Thus h maps strings over Σ into strings over Δ.) Show that if $h(xy) = h(x)h(y)$, for any $x, y \in \Sigma^*$, then $h(\epsilon) = \epsilon$.

7. *(5p)* For each of the following languages answer whether it is regular, context-free but not regular, or not context-free. (Here a brief explanation is sufficient.)

   (a) L_1 is the intersection of the language L from problem 4 and the language $\{\, a^i b^k c^i \mid i > 3k \,\}$.

   (b) L_2 is the language of those strings over alphabet $\{a, b, c, d\}$ which do not contain a substring cab and in which every substring ab is immediately followed by cd.

   (c) $L_4 = \{\, a^i b^j c^k a^k b^j c^i \mid i, j, k \geq 0 \,\}$.

   (d) $L_5 = \{\, a^i b^j c^k a^j b^k c^i \mid i, j, k \geq 0 \,\}$.

8. *(3p)* **only for TDDD14**

   Consider the language $L_1 \subseteq \{a, b, c\}^*$ generated by the grammar

   $$S \to aSb \mid aSc \mid \epsilon$$

   (where S is the start symbol) and the language

   $$L_2 = L_1 \cap \{\, w \in \{a, b, c\}^* \mid \#b(w) = \#c(w) \,\}.$$

   (So L_2 is the set of those strings from L_1 that have the same number of b's and c's).

   Prove that L_1 is not regular <u>or</u> that L_2 is not context-free. Use the appropriate pumping lemma or employ reasoning similar to the proof of the lemma.

9. *(3p)* **only for TDDD85**

   (a) Give arguments for the context-freeness of the language

   $$L_1 = \{(ab)^n(ab)^n(cd)^n | n \in N\}$$

   (b) Prove, using the pumping lemma, that

   $$L_2 = \{(ab)^n(cd)^n(ab)^n | n \in N\}$$

   is not a context-free language.

10. *(2p)* Explain briefly the notions of a recursive language and of a recursively enumerable language.

    What do we know about a recursively enumerable language L such that its complement $\sim L$ is recursively enumerable?

11. *(5p)* Which of the following statements are true, which are false? Why?

    (a) If a regular expression r does not contain the symbol * then the language $L(r)$ defined by r is finite.

    (b) There exists a regular language L such that $a^i b^i c^i \in L$ for any natural number i.

    (c) If a pushdown automaton accepts by empty stack then it does not accept the empty string.

    (d) If a finite automaton has n states and accepts a string of length n then it accepts an infinite language.

    (e) If L is a context-free language then its complement $\sim L$ is context-free.

12. *(3p)* In an attempt to construct LR parsers for certain grammars, we applied the standard method of constructing a DFA for the viable prefixes of a grammar. Some fragments of the obtained DFA's are given below.

Complete the missing items in the given states, the missing lookahead sets, and the missing symbols labelling the arrows. In each case, answer the following questions. Justify your answers.

- Does the fragment of a DFA satisfy the conditions for the grammar to be LR(0)?
- The same question about the conditions for LR(1).

You may skip adding missing items or lookahead sets if they are not needed to answer the questions. For instance if you find the items in some state to violate the LR(1) conditions then you do not need to complete the other states. In such a case just make an appropriate comment.
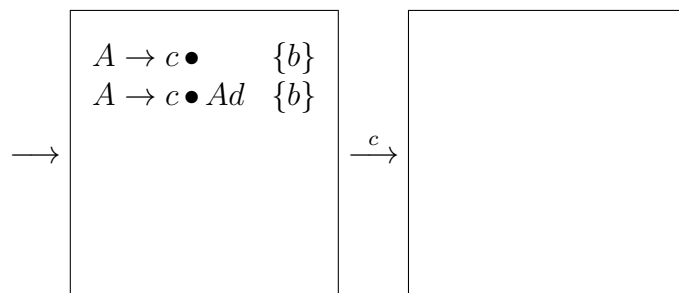
a, b, c, d are terminal symbols and S, A, B, C are nonterminal symbols of the grammars; S is the start symbol.

(a)

$$
\begin{array}{ll}
A \to \bullet ab & \{\$\} \\
A \to \bullet CB & \{\$\}
\end{array}
$$

$\xrightarrow{a}$ $\qquad$ $\longrightarrow$
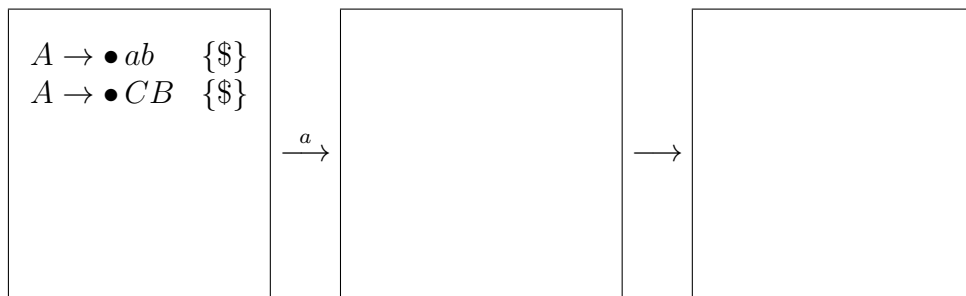
The productions of the grammar are $S \to A$, $A \to ab \mid CB$, $B \to d \mid dA$, $C \to ab$.

(b)

$$
\begin{array}{ll}
A \to c \bullet & \{b\} \\
A \to c \bullet Ad & \{b\}
\end{array}
$$

$\longrightarrow$ $\qquad$ $\xrightarrow{c}$

In this grammar the productions beginning with A are $A \to c \mid cAd$.

THE END. Remember the warning from the beginning of page 2. Check that your answers are justified.