Examination
Formal Languages and Automata Theory
TDDD14 & TDDD85

(Formella Språk och Automatateori)

2016–08–27, 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

   (Tillåtna hjälpmedel:
   - Ett papper med valfria anteckningar - 2 sidor A5 eller 1 sida A4. Anteckningarna ska signeras på samma sätt som tentamens-arken och bifogas tentamen vid inlämnandet.
   - Engelsk ordbok)
3. You may answer in Swedish or English.
4. Total number of credits is 32. Limits:

   3: 16 p, 4: 22 p, 5: 27 p.
5. *Jour* (contact person): Jonas Wallgren, tel. 013–178594

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 also explain that the grammar indeed generates the language. If you apply some known method then you should explain each step. And so on.)

1. *(6p)* The NFAϵ $N = (Q, \Sigma, \Delta, S, F)$ is defined as follows:

   $$Q = \{1, 2, 3, 4, 5, 6\} \qquad \Sigma = \{a, b\} \qquad S = \{1\} \qquad F = \{5, 6\}$$

   with the transition function Δ given by

q	ϵ	a	b
$\rightarrow$ 1	$\emptyset$	$\{1\}$	$\{2\}$
2	$\{3\}$	$\emptyset$	$\emptyset$
3	$\emptyset$	$\{3, 4\}$	$\{6\}$
4	$\emptyset$	$\emptyset$	$\{5\}$
5 F	$\emptyset$	$\emptyset$	$\{3\}$
6 F	$\emptyset$	$\{1\}$	$\{2, 3\}$

   (a) Draw the transition diagram for the NFAϵ N.

   (b) Using the standard method, construct an equivalent, minimal DFA M.

   (c) Let L be the language defined by M (and thus by N), and consider the relation $R_L \subseteq \Sigma^* \times \Sigma^*$ defined by

   $$xR_Ly \Leftrightarrow (\forall z \in \Sigma^*(xz \in L \Leftrightarrow yz \in L))$$

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

2. *(2p)* Consider an NFA N with the transition function Δ given by

q	a	b
$\rightarrow$ 1	$\{1, 2\}$	$\{2\}$
2 F	$\emptyset$	$\{3\}$
3	$\{4\}$	$\emptyset$
4	$\{2\}$	$\{1\}$

   where $\Sigma = \{a, b\}$ is the input alphabet, $1, 2, 3, 4$ are the states, 1 is the start state and the only final state is 2.

   Using a standard method, construct a regular expression defining the same language as N.

3. *(2p)* What does it mean that a context-free grammar is ambiguous? Give an example of an ambiguous grammar (and show that it is indeed ambiguous).

4. *(2p)* Provide a context-free grammar for the set of the strings over $\{0, 1\}$ containing at least two 0's, and such that the number of 1's before the first 0 is greater than the number of 1's after the last 0.

   While justifying your solution, explain the role of each production and explain which strings are generated by each nonterminal symbol.

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)

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

6. *(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) The language L_1 described as follows. Given two alphabets (of letters and digits) $\Sigma_1 = \{ a, b, \dots \}$, $\Sigma_2 = \{ 0, 1, \dots \}$, and a language $L_0 \subseteq \Sigma_1^*$ of 20 keywords, $L_0 = \{ \mathit{begin}, \mathit{end}, \mathit{if}, \dots \}$, L_1 contains any string over $\Sigma_1 \cup \Sigma_2$ beginning with a letter, provided the string is not in L_0.

   (b) $L_2 = \{ ww \mid w \in \{a, b\}^* \}$

(c) $L_3 = \{\, c^i w w^R d^j \mid w \in \{a,b\}^*,\ i > j \geq 0 \,\}$

(d) $L_4 = \{\, w w^R \mid w \in \{a,b\}^*,\ \#a(ww^R)$ is divisible by $3 \,\}$

(e) $L_5 = \{\, www \mid w \in \{a,b\}^*,\ |w| < 7 \,\}$

(x^R denotes the string x reversed, $\#a(x)$ denotes the number of a's in the string x.)

7. *(3p)* Use the appropriate pumping lemma to prove that ...

   - **only for TDDD14:** The language from problem 4 is not regular.
   - **only for TDDD85:** The language $L = \{\, a^i b^j c^{ij} \mid i, j > 0 \,\}$ is not context-free.

8. *(2p)* Explain briefly the notion of undecidable problem. Give an example of a decidable problem or of an undecidable problem.

9. *(4p)* Which of the following statements are true, which are false? Why?

   (a) There exists a language which is accepted by some PDA by final state but not accepted by any PDA by empty stack.

   (b) There exists a recursive language whose complement is not recursive.

   (c) If the language $L(r)$ defined by a regular expression r is infinite then r contains the $*$ symbol.

   (d) There exists an algorithm which finds out whether the languages defined by two DFA's are equal.

10. *(4p)* 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 labeling 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, D are nonterminal symbols of the grammars; S is the start symbol.

(a)

$S \to A{=}\bullet A$	$\{\$\}$
$A \to \bullet B$	$\{\$, +\}$
$A \to \bullet A{+}B$	$\{\$, +\}$
$B \to \bullet a$	$\{\quad\}$
$B \to \bullet B{*}a$	$\{\quad\}$

$\xrightarrow{B}$ [ ]

(b)

$S \to ba \bullet Da$	$\{\$\}$
$C \to a \bullet Ac$	$\{a, b\}$
$D \to \bullet Ac$	$\{\quad\}$
$A \to \bullet d$	$\{\quad\}$
$A \to \bullet aAc$	$\{\quad\}$

$\xrightarrow{A}$ [ ] $\longrightarrow$ [ ]

(c)

$A \to B \bullet a$	$\{\$\}$
$A \to B \bullet bA$	$\{\$\}$
$B \to B \bullet cd$	$\{a, b\}$

$\longleftarrow$ / $\longrightarrow$ [ ] $\xrightarrow{A}$ [ ]

In this case the productions of the grammar are:
$S \to A, \quad A \to Ba \mid BbA, \quad B \to cd \mid Bcd$