Examination
Formal Languages and Automata Theory
TDD14

(Formella Språk och Automatateori)

2013–08–31, 14.00 – 18.00

1. 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 tentamensarken och bifogas tentamen vid inlämnandet.
   - Engelsk ordbok)

2. You may answer in Swedish or English.
3. Total number of credits is 30: 3: 15 p, 4: 20 p, 5: 25 p.
4. *Jour* (person on duty): Włodek Drabent, tel. (013 28) 89 29.

GOOD LUCK !

Make sure that you justify your answers! Unexplained answers will be granted 0 points. (For example, assume that you are writing a grammar for a given language. Then you should also explain that the grammar indeed generates the language).

1. *(2p)* Consider the NFAϵ whose transition function is given by the table. (Its set of states is $Q = \{1, 2, 3, 4, 5\}$, the input alphabet $\Sigma = \{a, b, c\}$, the start state and the final state is 1.) Using a standard method construct an equivalent DFA.

	ϵ	a	b	c
$\rightarrow$ 1 **F**	$\emptyset$	$\{2\}$	$\{4\}$	$\emptyset$
2	$\emptyset$	$\{3\}$	$\{5\}$	$\{4\}$
3	$\emptyset$	$\emptyset$	$\emptyset$	$\{5\}$
4	$\{3\}$	$\{5\}$	$\emptyset$	$\emptyset$
5	$\{1\}$	$\emptyset$	$\emptyset$	$\emptyset$

2. *(2p)*

	a	b
$\rightarrow$ 1 **F**	2	6
2	1	5
3 **F**	2	4
4	4	3
5 **F**	6	2
6	5	5

Using a standard method, construct the minimal DFA equivalent to the DFA given by the table. (Its set of states is $Q = \{1, 2, 3, 4, 5, 6\}$, the input alphabet $\Sigma = \{a, b\}$, the start state is 1, and $\{1, 3, 5\}$ is the set of final states.)

3. *(2p)* Let L be the language accepted by the DFA from the previous problem. Consider the relation $\equiv_L$ on strings over $\Sigma = \{a, b\}$, defined by

$$x \equiv_L y \;\Leftrightarrow\; \forall z \in \Sigma^* \,(xz \in L \Leftrightarrow yz \in L).$$

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

4. *(2p)* Using a standard method, construct a regular expression defining the same language as the given DFA. (Its set of states is $Q = \{A, B, C, D\}$, the input alphabet $\Sigma = \{0, 1\}$, the start state and the final state is A.)

	0	1
$\rightarrow A$ **F**	B	C
B	D	C
C	A	C
D	D	D

5. *(1p)* We know from the lectures that $X = A^*B$ is a solution of an equation $X = AX \cup B$, where A, B are given languages and $\epsilon \notin A$. Moreover, there are no other solutions.

   Why is the requirement $\epsilon \notin A$ needed? What goes wrong when $\epsilon \in A$?

Hint: Find an example with $\epsilon \in A$ so that one of the two properties is violated (there are more than one solutions, or $X \neq AX \cup B$ for $X = A^*B$). Begin with A, B which are as simple as possible.

6. *(6p)* 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) $L_1 = \{\, a^i b^{i+1} c^j d^{2i} \mid i, j > 0 \,\}$,

   (b) $L_2 = \{\, a^i b^{j+1} c^j d^{2i} \mid i, j > 0 \,\}$,

   (c) $L_3 = \{\, a^i b^{i+1} c^j d^{2j} \mid i, j > 0 \,\}$,

   (d) $L_4 = \{\, a^i b^j c^{i+1} d^{2j} \mid i, j > 0 \,\}$,

   (e) L_5 is the image of L_4 under the homomorphism $h\colon \{a, b, c, d\}^* \to \{0, 1, 2\}^*$ such that $h(a) = 0101,\ h(b) = \epsilon,\ h(c) = 01,\ h(d) = 2$.

7. *(3p)* Prove that the language

$$L_6 = \{\, x \in \{a, b, c\}^* \mid \#a(x) = \#b(x) + 2 \,\}$$

is not regular, <u>or</u> that

$$L_7 = \{\, a^{2j} b^m c^j \mid 0 < m < j \,\}$$

is not context-free. Use the appropriate pumping lemma or employ reasoning similar to the proof of the lemma.

$\#a(w)$ denotes the number of occurrences of symbol a in string w.

8. *(1p)* Explain briefly Church's thesis (also called the Church-Turing thesis).

9. *(3p)* Show that the problem "a Turing machine M on input x does not reach state q" is undecidable. In other words, show that the language of the problem,

$$NS = \left\{ \langle M, x, q \rangle \;\middle|\; \begin{array}{l} \text{Turing machine } M \text{ on input } x \\ \text{does not reach state } q \end{array} \right\},$$

is not recursive. Use the fact that the halting problem is undecidable.

The language of the halting problem is

$$MP = \{\, \langle M, x \rangle \mid \text{Turing machine } M \text{ halts on input } x \,\}.$$

Brackets $\langle\ \rangle$ denote the encoding of a string, a TM, etc. as a string over alphabet $\{0, 1\}$.

10. *(3p)* Which of the following statements are true, which are false? Justify your answers.

    (a) If L_1, L_2 are languages and L_1 is infinite then $L_1 L_2$ is infinite.

    (b) There exists a context-free language L such that $LLL \cup L$ is not context-free.

    (c) If a context-free grammar G generates strings uwy, $uvwxy$ and uv^2wx^2y (where u, v, w, x, y are strings of terminal symbols of G, and $vx \neq \epsilon$) then it generates uv^3wx^3y.

11. *(2p)* Construct a grammar which is LL(1), has two nonterminal and three terminal symbols, has no useless symbols, and generates an infinite language.

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.

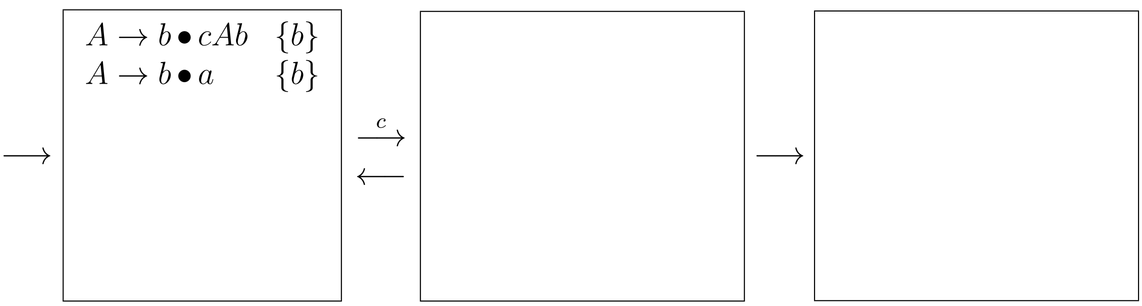
    Complete the missing items in the given states, the missing lookahead sets and the missing symbols labelling the arrows. In each case answer the following questions. Justify your answers.

    - 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.

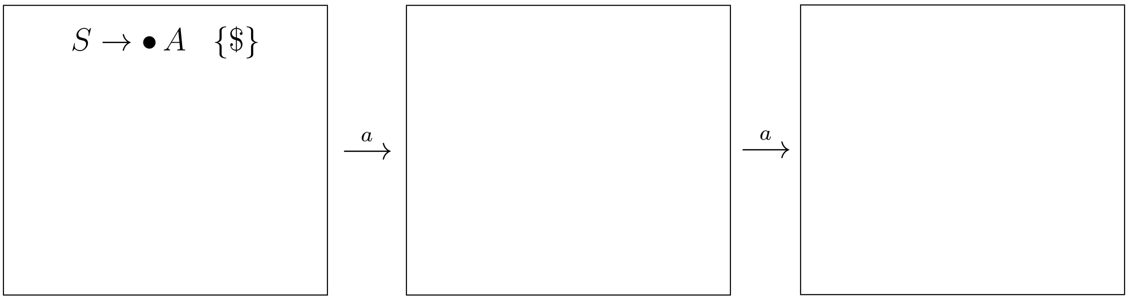
    a, b, c are terminal symbols and S, A, B are nonterminal symbols of the grammars; S is the start symbol.

    (a)



The productions of the grammar are $S \to A$, $A \to bcAb \mid ba \mid B$, $B \to ca$.

(b)



The productions of the grammar are $S \to A$, $A \to aAb \mid aAc \mid a$.