Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2014-01-09
Sal (1) Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och <u>ringa in</u> vilken sal som avses	KÅRA
Tid	8-12
Kurskod	TDDD14
Provkod	TEN3
Kursnamn/benämning **Provnamn/benämning**	Formella språk och automatateori En skriftlig tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	12
Jour/Kursansvarig Ange vem som besöker salen	Jonas Wallgren
Telefon under skrivtiden	— *)
Besöker salen ca kl.	09.00, 11.00
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Liselotte Lundberg 281278 liselotte.lundberg@liu.se
Tillåtna hjälpmedel	se tentans försättsida
Övrigt	
Vilken typ av papper ska användas, rutigt eller linjerat	rutat
Antal exemplar i påsen	16

*) I tentalokalens närhet under tentaperioden

Examination
Formal Languages and Automata Theory
TDDD14

(Formella Språk och Automatateori)

2014–01–09, 08.00 – 12.00

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

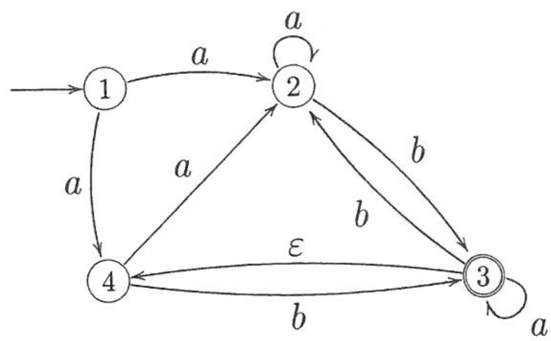
   (Tillåtna hjälpmedel:
   - Ett papper med valfria anteckningar - 2 sidor A5 eller 1 sida A4.
     Anteckningarna ska signeras på samma sätt som tentamensarken och bifogas tentamen vid inlämnandet.
   - Engelsk ordbok)

2. You may answer in Swedish or English.
3. Total number of credits is 31: 3: 15 p, 4: 20 p, 5: 25 p.
4. *Jour* (person on duty): Jonas Wallgren, tel. (013 28) 26 82

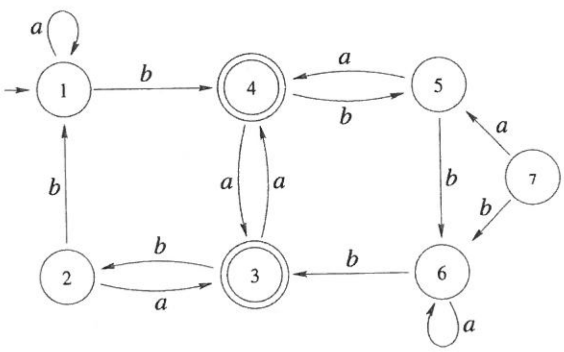
GOOD LUCK !

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)* Using a standard method, construct a DFA equivalent to the NFAϵ given by the diagram.



2. *(2p)* Construct a minimal DFA defining the same language as the DFA given by the diagram



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

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

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

4. *(1p)* Which of these pairs of regular expressions are equivalent (i.e. denote the same language)?

(a)	$((ab)^*c)^*$	$(ab + c)^*$
(b)	$(ab + a)^*a$	$a(ba + a)^*$

To show that two languages are distinct, give a string that is in one of them but not in the other.
In a case of equivalent regular expressions, an informal justification is sufficient.

5. *(2p)* Using a standard method, construct a regular expression defining the same language as the given NFA. (Its set of states is $Q = \{A, B, C, D, E, F\}$, the input alphabet $\Sigma = \{0, 1, 2\}$, the start state is A and the final states are A, D.)

	0	1	2
$\rightarrow A$ **F**	$\{B\}$	$\{D\}$	$\emptyset$
B	$\{C\}$	$\emptyset$	$\emptyset$
C	$\emptyset$	$\emptyset$	$\{B\}$
D **F**	$\{E\}$	$\emptyset$	$\emptyset$
E	$\emptyset$	$\{F\}$	$\emptyset$
F	$\{D\}$	$\emptyset$	$\{E\}$

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.)

Remember that $\#a(x)$ denotes the number of occurrences of symbol a in string x, and x^R denotes the string x reversed.

(a,b) The languages generated by the context-free grammars: L_1 by

$$S \rightarrow ASB \mid d \qquad A \rightarrow \epsilon \mid aA \qquad B \rightarrow a \mid b \mid c$$

L_2 by

$$S \rightarrow ASB \mid a \qquad A \rightarrow a \mid aA \qquad B \rightarrow b \mid c \mid d$$

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

(c) $L_3 = \left\{ w \in \{a, b, c\}^* \;\middle|\; \begin{array}{l} \text{each } b \text{ in } w \text{ is immediately preceded by } a, \\ w \text{ does not have a substring } acc, \\ \#a(w) \text{ is odd}, \;\; \#b(w) > \#c(w) \end{array} \right\}$.

(d) $L_4 = \{\, xcy^R cycx \mid x \in \{a, b\}^*,\; y \in \{c, d\}^* \,\}$.

(e) L_5 is the image of L_4 under the homomorphism $h\colon \{a, b, c, d\}^* \rightarrow \{d, e\}^*$ such that $h(a) = h(b) = 11,\; h(c) = 0,\; h(d) = 1$.

7. *(3p)* Prove that the language L_6 generated by the grammar

$$S \rightarrow \epsilon \mid ASbb \qquad A \rightarrow aa \mid aaa$$

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

$$L_7 = \{\, xcydx^R \mid x, y \in \{a, b\}^*,\; |y| > |x| \,\}$$

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

In the grammar, S is the start symbol, A is the other nonterminal, and a, b are terminal symbols. The alphabet of L_7 is $\{a, b, c, d\}$.

8. *(2p)* Explain briefly the notion of a recursive language and of a recursively enumerable language.

   Is the complement of a context-free language recursive?

9. *(3p)* Let A be a symbol from the tape alphabet of a Turing machine M. We say that M writes A on input x, when M writes A on the tape at some step of computation started with x on the tape, (Assume that this includes writing A into a cell already containing A.)

   Show that the problem "a Turing machine M on input x writes symbol A on the tape" is undecidable. In other words, show that the language of the problem,

$$NS = \left\{ \langle M, x, A \rangle \;\middle|\; \begin{array}{l} \text{Turing machine } M \\ \text{on input } x \text{ writes } A \end{array} \right\},$$

   is not recursive. Use the fact that the membership problem is undecidable. The language of the membership problem is

$$MP = \{ \langle M, x \rangle \mid x \in L(M) \}.$$

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

10. *(4p)* Which of the following statements are true, which are false? Justify your answers.

    (a) If L_1, L_2 are languages and L_1L_2 is empty then L_1 is empty or L_2 is empty.

    (b) There exist context-free languages L_1, L_2 such that $(L_1 \cup L_2)^*L_1$ is not context-free.

    (c) If a context-free grammar G generates strings which can be represented as uwy, $uvwxy$, uv^3wx^3y, uv^4wx^4y,... (i.e. uv^iwx^iy for any natural number i except $i = 2$) then G generates uv^2wx^2y.

    (d) The grammar

$$S \to \epsilon \mid AS \qquad\qquad A \to Aa \mid abA$$

    generates the empty language $\emptyset$. (S, A are the nonterminal symbols, a, b the terminal symbols, the start symbol is S)

11. *(1p)* Consider a grammar

$$S \to abAc \mid A \qquad\qquad A \to baAc \mid a$$

    Explain why it is not LL(1). (S is the start symbol, A is a nonterminal, the remaining symbols are terminal ones.)

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.

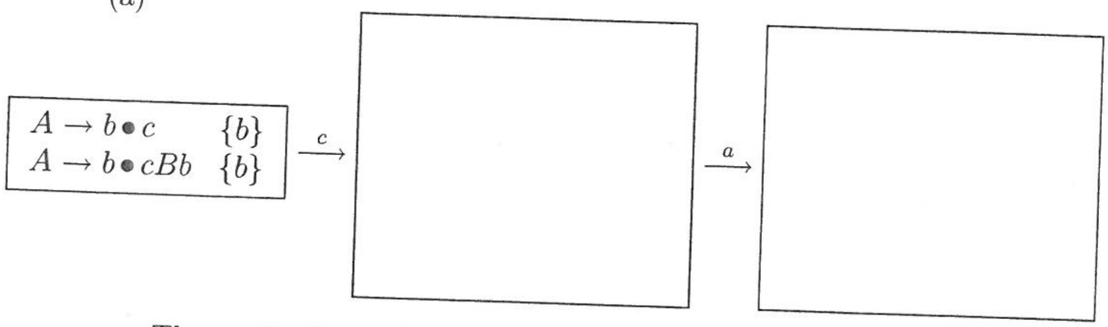
Complete the missing items in the given states, the missing lookahead sets and the missing symbols labelling the arrows. In each case answer the following questions. Justify your answers.

- 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.

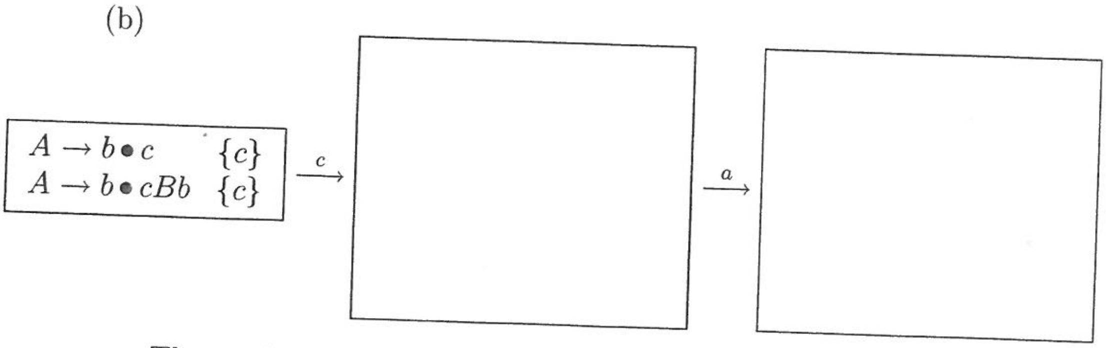
a, b, c are terminal symbols and S, A, B are nonterminal symbols of the grammars; S is the start symbol.

(a)



The productions of the grammar beginning with B are $B \to a \mid aA$.

(b)



The productions of the grammar beginning with B are $B \to a \mid Ac$.