Assignment 2
for Formal Languages and Automata Theory
TDDD14 and TDDD85

Deadline: Monday week 21 (May 23rd 2016) at 23:59:59.

For all the problems below it is not sufficient just to give a solution. **Justify your answers**. In the final exam 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). It is allowed to discuss the exercises with others, but you are supposed to solve each exercise individually. It is absolutely not allowed to copy solutions from others.

The solutions should be handed in to Jonas Wallgren for TDDD14 and to Johannes Schmidt for TDDD85 – on paper using the postbox in front of "Java Cafe" in building B (or at a lecture or a tutorial), or in a file in an email attachment (please only plain text, or pdf).

Your solutions can be written in Swedish or English.

1. Construct context-free grammars for the languages below. Explain which strings are generated by each nonterminal symbol of your grammars and what the role of each production is.

   (a)
   $$L_{\mathrm{np}} = \{\, x \in \{a,b\}^* \mid x \neq x^R \,\},$$
   where x^R is the reverse of the string x. (So L_{np} is the complement of the language of palindromes over $\{a,b\}^*$).

   (b) The language $L_2 \subseteq \{a,b,c\}^*$ of those strings of odd length whose first symbol, the middle one, and the last one are the same.

2. Consider a grammar
   $$S \to s \mid \mathbf{if}\, b\, \mathbf{then}\, S \mid \mathbf{if}\, b\, \mathbf{then}\, S\, \mathbf{else}\, S$$
   (with the terminal alphabet $\{b, s, \mathbf{if}, \mathbf{then}, \mathbf{else}\}$). Show that it is ambiguous. Construct an unambiguous grammar defining the same language.

   Here we accept informal justification of unambiguity. But it is better to show that the grammar is LL(1) or LR(1), in this way proving that it is unambiguous.

3. Which of the following languages are regular, context-free but not regular, or non-context-free? (Here a brief explanation is sufficient.)

   (a) $L_4 = \left\{ abw \;\middle|\; \begin{array}{l} w \in \{0,1\}^*,\ a,b \in \{0,1\}, \\ a \equiv \#1(w) \bmod 2,\ b \equiv \#0(w) \bmod 2 \end{array} \right\}$

   (b) $L_5 \in \{a,b,c\}^*$ is the language generated by the grammar

   $$S \to ASA \mid \epsilon \qquad\qquad A \to a \mid b \mid c$$

   (In this and in the next grammar the start symbol is S.)

   (c) $L_6 \in \{a,b,c\}^*$ is the language generated by the grammar

   $$S \to ASA \mid c \qquad\qquad A \to a \mid b \mid c$$

   (d) $L_7 = \{\, w \in \{a,b,c,d\}^* \mid \#a(w) > \#b(w) = \#c(w) \,\}$.

   (e) $L_8 = \{\, w \in \{a,b,c,d\}^* \mid \#a(w) > 7,\ \#b(w) = \#c(w) \,\}$.

   (f)
   $$L_9 = \left\{ x \in \{a,b,c,d\}^* \;\middle|\; \begin{array}{l} \#b(x) = \#c(x),\ \#d(x) \text{ is even}, \\ \#c(x) \equiv 3 \bmod 5,\ \text{each } b \text{ in } x \text{ is not} \\ \text{immediately followed by } a \end{array} \right\}.$$

   (g) $L_{10} \subseteq (\Sigma \cup \{\boldsymbol{\varepsilon}, \phi, (,), +, ^*\})^*$ is the set of regular expressions over an alphabet Σ (where $\boldsymbol{\varepsilon}, \phi, (,), +, ^* \notin \Sigma$).

   Hint: Note that we study here regular expressions themselves, not the regular languages they define. Consider regular expressions without abbreviations (skipped parentheses).

   (h) $L_{11} \subseteq (\Sigma \cup N \cup \{\leftarrow\})^*$ is the set of productions of all the context-free grammars with a terminal alphabet Σ and a nonterminal alphabet N.

4. (a) **TDDD85 only**

   Prove that the language

   $$L_5 = \{\, wc^j x \mid w, x \in \{a,b\}^*,\ |w| = |x| > j \,\}$$

   is not context-free. Use the appropriate pumping lemma.

(b) **TDDD14 only**

Prove that the language

$$L_6 = \left\{ x \in \{a,b,c,d\}^* \;\middle|\; \begin{array}{c} \#a(x) < \#b(x),\ \#c(x) < \#d(x), \\ \text{no } c \text{ occurs anywhere before a } b, \\ \text{no } d \text{ occurs anywhere before a } b \end{array} \right\}.$$

is not context-free. Use the appropriate pumping lemma.

5. Let G be the following context-free grammar (S is the start symbol):

$$\begin{aligned} S &\to aB \mid aDc \\ B &\to bBc \mid c \\ D &\to bc \mid c \end{aligned}$$

(a) Describe the language $L(G)$.

(b) Show that G is not an LR(1) grammar.

(c) Modify G so that an LR(1) grammar G', equivalent to G, is obtained. Explain why G' is an LR(1) grammar and why G and G' are equivalent.

6. Order the following formalisms (but one) 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. Point out the formalism that does not fit into the ordering.
(In this problem a brief justification – by referring to known properties – is sufficient.)

- Context-free Grammars (CFG)
- Deterministic Finite Automata (DFA)
- Deterministic Pushdown Automata (DPDA)
- LR(0) grammars
- 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 two heads (TM2h)