Exercise session 7

1. Does the following grammar satisfy the LL(1) condition?

   $S \rightarrow (L)|p|q$
   $L \rightarrow LandS|LorS|S$

   What kind of language does it decribe?

2. Give (as pseudocode) a recursive parser for the language described by the previous grammar. (N.B. Transform it first to LL(1) if needed!)

3. Let L be a language, the words of which are constructed from any text and legal parantheses structures. I.e. as in the problem 8, but now there can be text in the middle of parantheses. E.g. sentence "a (big) cat animal (either lion or tiger {which are rare } or saber-toothed tiger {which extincts [read further Kurten] – shame –} or leopard) is an attractive (but dangerous!) friend" belongs to the language. You can suppose for simplisity that the text parts consist of only lowercase letters a..z and spaces.

   Give a LL(1) grammar, which describes the langauge, and design for it a recursive parser!

4. Show that the following languages are deterministic:
   a) $\{a^m b^n | m \neq n\}$
   b) $\{wcw^R | w \in \{a,b\}^*\}$
   c) $\{a^m cb^m\} \bigcup \{a^m db^{2m}\}$

5. Transfor the grammar

   $S \rightarrow (S)|A$
   $A \rightarrow SS|\epsilon$

   into Chomsky normal form. Give also the middle steps (removing ϵ-productions and unit productions)!

6. Transfor the grammar

   $S \rightarrow ABC|a$
   $A \rightarrow aAaa|\epsilon$
   $B \rightarrow bBbb|\epsilon$
   $C \rightarrow cCa|c$

   into Chomsky normal form. Give also the middle steps!

7. Simulate the CYK-algorithm, when it decides, if the strings bbaab, ababab and aabba belong to the language described by the grammar

   $S \to AS|b$
   $A \to SA|a$

   If the answer is yes, give also the corresponding parse tree.

8. How would you parse the following language? It is enough to give the basic strategy.
   a) HTML
   b) Clauses of propositional logic, which consist of atomic clauses $A, B, ..., Z$, operations $\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$, and parantheses.
   c) C- or Pascal-source code
   d) SQL-query

9. Let $L_1 = \{a^n b^{2n} c^m | n, m \geq o\}$ and $L_2 = \{a^n b^m c^{2m} | n, m \geq o\}$. Is $L_1 \bigcap L_2$ context-free language? Justify your answer!

10. Design an efficient algorithm, which recognizes, if the given nonterminal symbol is nullable (i.e. can it produce ϵ in some derivation)!

11. Prove that context-free langauges are closed under union, concatenation and closure. I.e. if L_1 and L_2 are context-free, then

    a) $L_1 \bigcup L_2$
    b) $L_1 L_2$
    c) L_1^*

    are context-free. (Hint: Suppose that there exists pushdown automata $M(L_1)$ and $M(L_2)$ and construct pushdown automata for combined languages.)

12. Language $\{a^n b^n c^n | n \geq 0\}$ cannot be recognized by a common pushdown automaton. Could it be recognized, if you had two stacks available? If it could, draw the transition diagram of the automaton and simulate its behaviour! If not, then justify, why not!

13. Create a grammar, which describes legal - and -liststructures in html. The listed items can be any text (no special characters) or new sublists. E.g. Cat Dogs <ul> Wolf Fox Hyena Rabbit Rat belongs to the language. What kind of parsing algorithm would you use to decide, if the given text belongs to the language?

More challenging:

14. Prove by the Pumpinh Lemma for context-free languages that $\{ww \ w \in \{a,b\}^*\}$ is not context-free!

15. Prove by the Pumpinh Lemma for context-free languages that $\{a^p | p \text{ is prime}\}$ is not context-free!