Principles of Programming Languages [PLP]
Exercises on Lexical Analysis

1. List the typical components of an abstract machine $\mathbf{M_L}$ for a programming language **L**. Describe also the phases of the interpreter loop.

2. The typical way of implementing a programming language **L** is to *compile* its programs to an intermediate language **I**, and then to *interpret* the translated program on a target machine $\mathbf{M_H}$.
   In which languages can be written the compiler and the interpreter mentioned in the previous sentence?

3. Write a regular expression or a regular definition that describes all integers constants in Java. Java integer constants can be in decimal, octal or hexadecimal notation. Octal constants are made of a leading zero followed by digits from 0 to 7. Hexadecimal constants have a prefix "0x" or "0X" followed by characters which are either digits or letters from A to F, lower or upper case.

4. The extended regular expression **`r{n,m}`** matches any string that matches from **`n`** to **`m`** occurrences of pattern **`r`**. For example **`(ab){1,3}`** matches **`ab`**, **`abab`** and **`ababab`**. Show that each regular expression containing operator **`_{_,_}`** has an equivalent regular expression without such operation.

5. Consider the following grammar, whose terminals are {id, :}:

   $$
   \begin{aligned}
   S &\rightarrow P \mid P\ S \\
   P &\rightarrow id : R \\
   R &\rightarrow \varepsilon \mid id\ R
   \end{aligned}
   $$

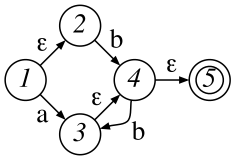
   For each of the following strings, say if they belong to the generated language or not. In the positive case, depict the corresponding parse tree and abstract syntax tree.
   a) id : id
   b) id id : id
   c) id : id id id :

6. In words, describe the languages denoted by the following regular expressions:
   (a) (0*10*)*
   (b) (0 | 1)* 1 (0 | 1)* 0 (0 | 1)
   (c) (0|01)*

7. Write regular expressions over the set of symbols {0,1} that describe:
   (a) the language of all strings having an even number of 0's
   (b) the language of all strings having an even number of 0's and of 1's

8. Use Thompson's algorithm to build an NFA for the regular expression ((a|ε)b)*

9. Given the NFA with S = {1,2,3,4,5}, Σ = {a,b}, s_0 = 1, F = {5} and the transition graph shown below, convert the NFA to a DFA using the subset construction algorithm (do not attempt to minimize the DFA). Express your answer as a transition graph and identify the start and final states.



10. Consider the following state transition table of a DFA with S = {0, 1, 2, 3, 4}, Σ = {a, b}, s_0 = 0, F = {3, 4}.

State	**a**	**b**
0	1	2
1	3	4
2	1	2
3	4	-
4	3	-

   (a) Draw the transition graph.
   (b) Minimize the DFA using the algorithm illustrated in class. Identify the start and final states of the minimized DFA.
   (c) Write an equivalent regular expression that represents the same language as defined by the (minimized) DFA.

11. Show that the regular expressions **ab*c** and **ac|abb*c** denote the same regular language. To this aim, for both regular expressions:
   (a) depict the syntax tree
   (b) use Thompson's algorithm to construct the corresponding NFAs
   (c) convert the NFAs to DFAs using the Subset Construction algorithm
   (d) minimize the DFAs, and check that they are isomorphic.

12. When is a grammar ambiguous?

13. Is the grammar of Exercise 5 ambiguous?

14. Show that the following grammar is ambiguous:

S → aS | A
A → aAS | b