SAARLAND UNIVERSITY
COMPUTER SCIENCE

Prof. Dr. Sebastian Hack
Johannes Doerfert, B.Sc.

# Compiler Construction WS15/16

# Exercise Sheet 2

## Exercise 2.1 Regular Expressions and Languages

The lecture defined regular expressions using the metacharacters $\underline{\emptyset}$ and $\underline{\varepsilon}$. Show that they are the neutral elements with respect to the alternative and concatenation operations in regular expressions. This means show that:

- $\underline{(}r_1\underline{|}\underline{\emptyset}\underline{)}$ describes the same language as $r_1$
- $\underline{(}r_1\underline{\varepsilon}\underline{)}$ describes the same language as $r_1$

only by reasoning about the described languages as shown in the lecture. Assume the regular expression $r_1$ to denote the language $R_1$.

## Exercise 2.2 Finite Automata Reloaded

In this exercise we take a closer look at recognising common language structures like comments. Consider comments in XML which start with $<!--$ and end with the first occurrence of $-->$. However, XML comments are not nestable. So the first $-->$ ends the comment no matter how many $<!--$ it contained. We can define the construct $<!--\ \mathit{until}\ -->$ to describe such comments.

- Create a minimal deterministic finite automaton that accepts XML comments over an alphabet $\Sigma$, where $\{<, >, -, !\} \subseteq \Sigma$. You may label an automaton edge with $\Sigma \setminus \{x, y\}$ to express that there are in fact edges for all of the alphabet's symbols except $\{x, y\}$.

## Exercise 2.3 Grammar Flow Analysis

Let $G = (\{S, A, B, C, D, E, F, G, H, K, L, M\}, \{a, b, c, d, e\}, P, S)$ describe a context-free grammar with productions $P$ defined as follows:

$$
\begin{array}{rcl}
S & \rightarrow & KA \mid BK \\
A & \rightarrow & abA \mid BcH \\
B & \rightarrow & eBd \mid aGd \mid c \\
C & \rightarrow & dAb \mid aa \\
D & \rightarrow & S \mid \varepsilon \\
E & \rightarrow & FB \\
F & \rightarrow & FA \mid Ec \\
H & \rightarrow & CD \mid eEd \\
K & \rightarrow & cd \\
L & \rightarrow & aLa \mid b \\
M & \rightarrow & Lb \mid cd
\end{array}
$$

- Compute the set of reachable non-terminals of $G$.

- Compute the set of productive non-terminals of $G$.
- Formally describe the reduced grammar $G_r$ accepting the same language as $G$.
- Compute the set $\mathit{first}_1(T)$ for each non-terminal $T$ in $G_r$.
- Compute the set $\mathit{follow}_1(T)$ for each non-terminal $T$ in $G_r$.

You have to use the algorithms from the lecture. Provide the corresponding system of equations for each analysis subtask.

## Exercise 2.4 Push-Down Automata

Let $(\{S, A, B, C, D, H, K\}, \{a, b, c, d, e\}, P, S)$ be a context-free grammar with the following productions $P$:

$$\begin{array}{lcl} S & \to & KA \mid BK \\ A & \to & abA \mid BcH \mid \varepsilon \\ B & \to & eBd \mid c \\ C & \to & dAb \mid aa \\ D & \to & S \mid \varepsilon \\ H & \to & CD \\ K & \to & cd \end{array}$$

Write down a successful run of the push-down automaton constructed for this grammar (using the algorithms presented in the lecture) on the input word $cdeecddcaaccd$.