Universität Zürich
Institut für Informatik

Autumn 2020

Sven Seuken
Dennis Komm

Student Name:
Student Number:

Foundations of Computing II
Assignment 3

Context-Free Grammars and Languages

Distributed: 26.10.2020 – Due Date: 01.11.2020

Upload your solutions to the OLAT system.

3.1 Context-Free Grammars and Languages

As defined in the lecture, for a word w, w^{R} denotes the "reversal" of w; for instance, $(aabbba)^{\mathsf{R}}$ is $abbbaa$. Furthermore, as already defined in Exercise 2.1, $|w|_a$ denotes the number of occurrences of the letter a in w. Construct context-free grammars for the following languages.

a) $L_1 = \{w \in \{0,1\}^* \mid w \text{ ends with } 01 \text{ and } |w|_0 \text{ is even}\}$

b) $L_2 = \{w \in \{a,b\}^* \mid (2|w|_a + |w|_b) \bmod 5 = 0\}$

c) $L_3 = \{w \in \{a,b\}^* \mid w \text{ ends with } aa \text{ or } w = w^{\mathsf{R}}\}$

d) As defined in the lecture, a grammar is called a "regular grammar" if it has only productions of the form $X \to aY$, $X \to a$, $X \to \varepsilon$, where X and Y are non-terminals and a is a terminal. As the name suggests, exactly the regular languages allow for regular grammars. For which of the above languages L_1, L_2, and L_3 can you give regular grammars? How surprising is your result?

3.2 Parsing Strings

The CFG $G_4 = (\{S, A, B\}, \{0,1\}, P, S)$ with

$$\begin{aligned} P_4 = \{&S \to A1B, \\ &A \to 0A \mid \varepsilon, \\ &B \to 0B \mid 1B \mid \varepsilon\} \end{aligned}$$

generates the language L_4 which corresponds to the regular expression $0^*1(0+1)^*$.

a) Give both the leftmost and rightmost derivations of 1101.

b) Write down the parse tree of 001101.

3.3 Normal Forms

a) Use the method presented in the lecture to eliminate all ε-productions of the CFG $G_5 = (\{S, A, B, C, D\}, \{a, b, c\}, P_5, S)$ with

$$P_5 = \{S \to ABCD, A \to \varepsilon, A \to BB, B \to AA, A \to a, B \to b, C \to bc, D \to \varepsilon\}.$$

b) Use the method presented in the lecture to eliminate all unit productions in the CFG $G_6 = (\{S, A, B, C, D\}, \{a, b, c, d\}, P_6, S)$ with

$$P_6 = \{S \to ABC, A \to B, B \to C, B \to b, B \to bB, C \to D, D \to d\}.$$

c) Use the method presented in the lecture to eliminate all useless symbols in the CFG $G_7 = (\{S, A, B, C, D, E\}, \{a, b, c, d\}, P_7, S)$ with

$$\begin{aligned} P_7 = \{&S \to A, S \to AaB, S \to BbA, B \to bB, A \to aa, A \to Ab, \\ &C \to cD, D \to c, D \to Ad, D \to EE, D \to dd\}. \end{aligned}$$

d) Convert the CFG $G_8 = (\{S, A, B, C\}, \{a, b\}, P_8, S)$ with

$$\begin{aligned} P_8 = \{&S \to aS, S \to Sb, S \to Aa, S \to bbB, A \to aBb, A \to ab, \\ &B \to bCa, B \to ba, C \to b\} \end{aligned}$$

into Chomsky normal form.

3.4 CYK Algorithm

Consider the CFG $G_9 = (\{S, A, B, C\}, \{a, b\}, P_9, S)$ in Chomsky normal form with

$$\begin{aligned} P_9 = \{&S \to AB \mid BC, \\ &A \to BA \mid a, \\ &B \to CC \mid b, \\ &C \to AB \mid a\}. \end{aligned}$$

Use the CYK algorithm to determine whether each of the following strings is in $L(G_9)$.

a) $ababa$

b) $baaab$

c) $aabab$