Solution to Problem 10

Task. Transform the grammar from Homework 7 into Chomsky normal form. Assume that we are only using digits 0 and 1.

Solution. The grammar from Homework 7 has the following rules:

\[
D \rightarrow 0; \quad D \rightarrow 1; \quad E \rightarrow D; \quad E \rightarrow E + E; \quad E \rightarrow E \cdot E
\]

Preliminary step. First, we introduce a new starting variable \( S_0 \) and a rule \( S_0 \rightarrow S \), where \( S \) is the starting variable of the original grammar. In our grammar, the starting variable is \( E \), so we end up with the following rules:

\[
D \rightarrow 0; \quad D \rightarrow 1; \quad E \rightarrow D; \quad E \rightarrow E + E; \quad E \rightarrow E \cdot E; \quad S_0 \rightarrow E
\]

Step 0. On this step, we eliminate non-Chomsky rules with right-hand side of length 0, i.e., with right-hand side an empty string and the left-hand side is not a starting variable.

In the above grammar, there are no such rules, so we do not do anything on this step.

Step 1. On this step, we eliminate non-Chomsky rules in which the right-hand side has length 1, i.e., in which the right-hand side is a variable. In the above grammar, there are several such rules, we will eliminate them one by one.

The first such rule is \( E \rightarrow D \). To eliminate this rule, for each rule \( D \rightarrow w \) that has the variable \( D \) is the left-hand side (for any right-hand side \( w \)), we add a rule \( E \rightarrow w \). In the current grammar, we have two such rule: \( D \rightarrow 0 \) and \( D \rightarrow 1 \), so we add two rules \( E \rightarrow 0 \) and \( E \rightarrow 1 \). As a result, we get the following grammar:

\[
D \rightarrow 0; \quad D \rightarrow 1; \quad E \rightarrow E + E; \quad E \rightarrow E \cdot E; \quad S_0 \rightarrow E; \quad E \rightarrow 0; \quad E \rightarrow 1
\]

The next rule that need to be eliminated on this stage is \( S_0 \rightarrow E \). To eliminate this rule, for each rule \( E \rightarrow w \) that has the variable \( E \) is the left-hand side (for any right-hand side \( w \)), we add a rule \( S_0 \rightarrow w \). In the current grammar, we have four such rules: \( E \rightarrow E + E, E \rightarrow E \cdot E, E \rightarrow 0, \) and \( E \rightarrow 1 \), so we add rules \( S_0 \rightarrow E + E \) and \( S_0 \rightarrow E \cdot E \). As a result, we get the following grammar:

\[
D \rightarrow 0; \quad D \rightarrow 1; \quad E \rightarrow E + E; \quad E \rightarrow E \cdot E; \quad E \rightarrow 0; \quad E \rightarrow 1
\]
\[ S_0 \to E + E; \quad S_0 \to E \cdot E; \quad S_0 \to 0; \quad S_0 \to 1. \]

**Step 2.** On this step:

- For each terminal symbol \( a \), we introduce an auxiliary variable \( V_a \) and a rule \( V_a \to a \).
- Then, in each rule in which the right-hand side has 2 or more symbols and at least one of them is a terminal symbol, we replace each terminal symbol with the corresponding variable.

In our grammar, we have four terminal symbols 0, 1, +, and ·. So, we introduce four new variables \( V_0, V_1, V_+, \) and \( V \) and four new rules \( V_0 \to 0, V_1 \to 1, V_+ \to +, \) and \( V \to \cdot \). So we end up with the following grammar:

\[
\begin{align*}
D & \to 0; \quad D \to 1; \\
E & \to EV_+E; \quad E \to EV_E; \quad E \to 0; \quad E \to 1; \\
S_0 & \to EV_+E; \quad S_0 \to EV_E; \quad S_0 \to 0; \quad S_0 \to 1; \\
V_0 & \to 0; \quad V_1 \to 1; \quad V_+ \to +; \quad V \to \cdot.
\end{align*}
\]

**Step 3.** At this step, we deal with the rules in which the right-hand side has length 3 or larger. In line with the general algorithm, e.g., the rule \( E \to EV_+E \) is replaced by two rules: \( E \to V_E+E \) and \( V_{E+} \to EV_+. \) So, we get the following set of rules in Chomsky normal form:

\[
\begin{align*}
D & \to 0; \quad D \to 1; \\
E & \to V_{E+}E; \quad V_{E+} \to EV_+; \quad E \to V_E E; \quad V_E \to EV; \\
E & \to 0; \quad E \to 1; \\
S_0 & \to V_{E+}E; \quad S_0 \to V_E E; \quad S_0 \to 0; \quad S_0 \to 1; \\
V_0 & \to 0; \quad V_1 \to 1; \quad V_+ \to +; \quad V \to \cdot.
\end{align*}
\]

**Reminder.** In Chomsky normal form, only the following three types of rules are allowed:

- rules of the type \( S_0 \to \varepsilon \), where \( S_0 \) is the starting variable;
- rules of the type \( V \to a \), where \( V \) is a variable and \( a \) is a terminal symbol; and
- rules of the type \( V \to AB \), where \( V, A, \) and \( B \) are variables.