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:

$$D \to 0; \quad D \to 1; \quad E \to D; \quad E \to E + E; \quad E \to E \cdot E$$

Preliminary step. First, we introduce a new starting variable S_0 and a rule $S_0 \to S$, 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:

$$D \to 0; \quad D \to 1; \quad E \to D; \quad E \to E + E; \quad E \to E \cdot E; \quad \underline{S_0 \to E}$$

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

$$D \to 0; \quad D \to 1; \quad E \to E + E; \quad E \to E \cdot E; \quad S_0 \to E; \quad \underline{E \to 0}; \quad \underline{E \to 1}$$

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

$$D \to 0; \quad D \to 1; \quad E \to E + E; \quad E \to E \cdot E; \quad E \to 0; \quad E \to 1;$$

$$\underline{S_0 \to E + E}; \quad \underline{S_0 \to E \cdot E}; \quad \underline{S_0 \to 0}; \quad \underline{S_0 \to 1}.$$

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 $\cdot$. So, we introduce four new variables V_0, V_1, V_+, and $V_\cdot$ and four new rules $V_0 \to 0$, $V_1 \to 1$, $V_+ \to +$, and $V_\cdot \to \cdot$. So we end up with the following grammar:

$$D \to 0; \quad D \to 1; \quad \underline{E \to EV_+E}; \quad \underline{E \to EV_\cdot E}; \quad E \to 0; \quad E \to 1;$$

$$\underline{S_0 \to EV_+E}; \quad \underline{S_0 \to EV_\cdot E}; \quad \underline{S_0 \to 0}; \quad \underline{S_0 \to 1};$$

$$\underline{V_0 \to 0}; \quad \underline{V_1 \to 1}; \quad \underline{V_+ \to +}, \quad \underline{V_\cdot \to \cdot}$$

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:

$$D \to 0; \quad D \to 1; \quad \underline{E \to V_{E+}E}; \quad \underline{V_{E+} \to EV_+}; \quad \underline{E \to V_{E\cdot}E}; \quad \underline{V_{E\cdot} \to EV_\cdot};$$

$$E \to 0; \quad E \to 1; \quad \underline{S_0 \to V_{E+}E}; \quad \underline{S_0 \to V_{E\cdot}E}; \quad \underline{S_0 \to 0}; \quad \underline{S_0 \to 1};$$

$$V_0 \to 0; \quad V_1 \to 1; \quad V_+ \to +, \quad V_\cdot \to \cdot$$

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.