Solution to Problem 10

Task. Transform the grammar from Homework 7 into Chomsky normal form.

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

\[
\begin{align*}
I & \rightarrow +U; \quad I \rightarrow -U; \quad I \rightarrow U; \quad U \rightarrow DU; \quad U \rightarrow D; \quad D \rightarrow 0; \\
D & \rightarrow 1
\end{align*}
\]

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 \( I \), so we end up with the following rules:

\[
\begin{align*}
I & \rightarrow +U; \quad I \rightarrow -U; \quad I \rightarrow U; \quad U \rightarrow DU; \quad U \rightarrow D; \quad D \rightarrow 0; \\
D & \rightarrow 1; \quad S_0 \rightarrow I
\end{align*}
\]

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

\[
\begin{align*}
I & \rightarrow +U; \quad I \rightarrow -U; \quad U \rightarrow DU; \quad U \rightarrow D; \quad D \rightarrow 0; \quad D \rightarrow 1; \\
S_0 & \rightarrow I; \quad I \rightarrow DU; \quad I \rightarrow D
\end{align*}
\]

Next rule that need to be eliminated on this stage is \( U \rightarrow D \). 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 \( U \rightarrow w \). In the current grammar, we have
two such rules: $D \rightarrow 0$ and $D \rightarrow 1$, so we add rules $U \rightarrow 0$ and $U \rightarrow 1$. As a result, we get the following grammar:

\[
I \rightarrow +U; \quad I \rightarrow -U; \quad U \rightarrow DU; \quad D \rightarrow 0; \quad D \rightarrow 1; \quad S_0 \rightarrow I;
\]

\[
I \rightarrow DU; \quad I \rightarrow D; \quad U \rightarrow 0; \quad U \rightarrow 1
\]

Next rule that need to be eliminated on this stage is $S_0 \rightarrow I$. To eliminate this rule, for each rule $I \rightarrow w$ that has the variable $I$ 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: $I \rightarrow +U$, $I \rightarrow -U$, $I \rightarrow DU$, and $I \rightarrow D$, so we add the rules $S_0 \rightarrow +U$, $S_0 \rightarrow -U$, $S_0 \rightarrow DU$, and $S_0 \rightarrow D$. As a result, we get the following grammar:

\[
I \rightarrow +U; \quad I \rightarrow -U; \quad U \rightarrow DU; \quad D \rightarrow 0; \quad D \rightarrow 1; \quad I \rightarrow DU;
\]

\[
I \rightarrow D; \quad U \rightarrow 0; \quad U \rightarrow 1; \quad S_0 \rightarrow +U; \quad S_0 \rightarrow -U; \quad S_0 \rightarrow DU;
\]

\[
S_0 \rightarrow D
\]

Next rule that need to be eliminated on this stage is $I \rightarrow D$. 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 $I \rightarrow w$. In the current grammar, we have two such rules: $D \rightarrow 0$ and $D \rightarrow 1$, so we add the rules $I \rightarrow 0$ and $I \rightarrow 1$. As a result, we get the following grammar:

\[
I \rightarrow +U; \quad I \rightarrow -U; \quad U \rightarrow DU; \quad D \rightarrow 0; \quad D \rightarrow 1; \quad I \rightarrow DU;
\]

\[
U \rightarrow 0; \quad U \rightarrow 1; \quad S_0 \rightarrow +U; \quad S_0 \rightarrow -U; \quad S_0 \rightarrow DU; \quad S_0 \rightarrow D;
\]

\[
I \rightarrow 0; \quad I \rightarrow 1
\]

The last rule that need to be eliminated on this stage is $S_0 \rightarrow D$. 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 $S_0 \rightarrow w$. In the current grammar, we have two such rules: $D \rightarrow 0$ and $D \rightarrow 1$, so we add the rules $S_0 \rightarrow 0$ and $S_0 \rightarrow 1$. As a result, we get the following grammar:

\[
I \rightarrow +U; \quad I \rightarrow -U; \quad U \rightarrow DU; \quad D \rightarrow 0; \quad D \rightarrow 1; \quad I \rightarrow DU;
\]

\[
U \rightarrow 0; \quad U \rightarrow 1; \quad S_0 \rightarrow +U; \quad S_0 \rightarrow -U; \quad S_0 \rightarrow DU; \quad I \rightarrow 0;
\]

\[
I \rightarrow 1; \quad S_0 \rightarrow 0; \quad S_0 \rightarrow 1
\]

**Step 2.** On this step:

- For each terminal symbol $a$, we introduce an auxiliary variable $V_a$ and a rule $V_a \rightarrow a$. 

2
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. So, we introduce four new variables $V_+$, $V_0$, $V_1$, and $V_2$ and four new rules $V_+ \rightarrow +$, $V_- \rightarrow -$, $V_0 \rightarrow 0$, and $V_1 \rightarrow 1$. In the rule $I \rightarrow +U$, we replace $+$ with $V_+$ and get the new rule $I \rightarrow V_+U$. We do the same replacement with all other rules in which the right-hand side has 2 or more symbols and at least one of them is a terminal symbol. As a result, we get the following grammar:

$$
I \rightarrow V_+U; \quad I \rightarrow V_-U; \quad U \rightarrow DU; \quad D \rightarrow 0; \quad D \rightarrow 1; \quad I \rightarrow DU;
$$

$$
U \rightarrow 0; \quad U \rightarrow 1; \quad S_0 \rightarrow V_+U; \quad S_0 \rightarrow V_-U; \quad S_0 \rightarrow DU; \quad I \rightarrow 0;
$$

$$
I \rightarrow 1; \quad S_0 \rightarrow 0; \quad S_0 \rightarrow 1; \quad V_+ \rightarrow +; \quad V_- \rightarrow -; \quad V_0 \rightarrow 0;
$$

$$
V_1 \rightarrow 1.
$$

**Step 3.** At this step, we deal with the rules in which the right-hand side has length 3 or larger. In our grammar, there are not such rules, so the grammar that we obtained after Step 2 is already in Chomsky normal form, i.e., it only has three types of rules:

- rules of the type $S_0 \rightarrow e$, where $S_0$ is the starting variable;
- rules of the type $V \rightarrow a$, where $V$ is a variable and $a$ is a terminal symbol; and
- rules of the type $V \rightarrow AB$, where $V$, $A$, and $B$ are variables.