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:

\[ N \to L; \quad N \to NL; \quad N \to ND; \quad L \to a; \quad D \to 0; \quad D \to 1 \]

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

\[ N \to L; \quad N \to NL; \quad N \to ND; \quad L \to a; \quad D \to 0; \quad D \to 1; \quad S_0 \to N \]

**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 \( N \to L \). To eliminate this rule, for each rule \( L \to w \) that has the variable \( L \) is the left-hand side (for any right-hand side \( w \)), we add a rule \( N \to w \). In the current grammar, we have only one such rule: \( L \to a \), so we add the rule \( N \to a \). As a result, we get the following grammar:

\[ N \to NL; \quad N \to ND; \quad L \to a; \quad D \to 0; \quad D \to 1; \quad S_0 \to N; \quad N \to a \]

The next rule that need to be eliminated on this stage is \( S_0 \to N \). To eliminate this rule, for each rule \( N \to w \) that has the variable \( N \) 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 three such rules: \( N \to NL, N \to ND, \) and \( N \to a \), so we add rules \( S_0 \to NL, S_0 \to ND, \) and \( S_0 \to a \). As a result, we get the following grammar:

\[ N \to NL; \quad N \to ND; \quad L \to a; \quad D \to 0; \quad D \to 1; \quad S_0 \to N; \quad N \to a; \quad S_0 \to NL; \quad S_0 \to ND; \quad S_0 \to a \]
Step 2. On this step:

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

- 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 three terminal symbols 0, 1 and $a$. So, we introduce three new variables $V_0$, $V_1$, and $V_a$ and three new rules $V_0 \rightarrow 0$, $V_1 \rightarrow 1$, and $V_a \rightarrow a$. In this case, there is no need to replace, so we end up with the following grammar:

\[
N \rightarrow NL; \quad N \rightarrow ND; \quad L \rightarrow a; \quad D \rightarrow 0; \quad D \rightarrow 1; \quad S_0 \rightarrow N; \quad N \rightarrow a; \\
S_0 \rightarrow NL; \quad S_0 \rightarrow ND; \quad S_0 \rightarrow a; \quad V_0 \rightarrow 0; \quad V_1 \rightarrow 1; \quad V_a \rightarrow a
\]

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 $S_0 \rightarrow \varepsilon$, 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.