Solution to Problem 10

**Task.** Transform the grammar from Homework 7 into Chomsky normal form. Assume that we are only using digit 0 and letter a.

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

- $D \rightarrow 0$
- $L \rightarrow a$
- $S \rightarrow L = D$
- $S \rightarrow L = L$
- $S \rightarrow SS$

**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 $S$, so we end up with the following rules:

- $D \rightarrow 0$
- $L \rightarrow a$
- $S \rightarrow L = D$
- $S \rightarrow L = L$
- $S \rightarrow SS$
- $S_0 \rightarrow S$

**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 only such rule is $S_0 \rightarrow S$. To eliminate this rule, for each rule $S \rightarrow w$ that has the variable $S$ 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 three such rules:
• $S \rightarrow L = D$
• $S \rightarrow L = L$
• $S \rightarrow SS$

So, we add three rules:
• $S_0 \rightarrow L = D$
• $S_0 \rightarrow L = L$
• $S_0 \rightarrow SS$

As a result, we get the following grammar:

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 four terminal symbols 0, $a$, $=$, and $;$: So, we introduce four new variables $V_0$, $V_a$, $V_=$, and $V;$ and four new rules

• $V_0 \rightarrow 0$
• $V_a \rightarrow a$
• $V_\rightarrow \rightarrow$
• $V_\rightarrow ;$

So we end up with the following grammar:
- $D \rightarrow 0$
- $L \rightarrow a$
- $S \rightarrow LV \rightarrow DV$
- $S \rightarrow LV \rightarrow LV$
- $S \rightarrow SS$
- $S_0 \rightarrow LV \rightarrow DV$
- $S_0 \rightarrow LV \rightarrow LV$
- $S_0 \rightarrow SS$
- $V_0 \rightarrow 0$
- $V_a \rightarrow a$
- $V_e \rightarrow \epsilon$
- $V \rightarrow \epsilon$

**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 $S \rightarrow LV \rightarrow LV$ is replaced by three rules:

- $E \rightarrow V_L = LV$
- $V_{L=L} \rightarrow V_{L=L}$
- $V_{L=a} \rightarrow LV$

So, we get the following set of rules in Chomsky normal form:

- $D \rightarrow 0$
- $L \rightarrow a$
- $S \rightarrow V_L = DV$
- $V_L = D \rightarrow V_L = D$
- $V_L = \epsilon \rightarrow LV$
- $S \rightarrow V_L = LV$
- $V_L = L \rightarrow V_L = L$
- $S \rightarrow SS$
- $S_0 \rightarrow V_L = DV$
- $S_0 \rightarrow V_L = LV$
• $S_0 \rightarrow SS$
• $V_0 \rightarrow 0$
• $V_a \rightarrow a$
• $V_\rightarrow \rightarrow$
• $V_i \rightarrow$

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

• 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.