Solution to Problem 10

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

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

- $D \to 1$
- $L \to a$
- $S \to D$
- $S \to SD$
- $S \to SL$

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 S, so we end up with the following rules:

- $D \to 1$
- $L \to a$
- $S \to D$
- $S \to SD$
- $S \to SL$
- $\underline{S_0 \to S}$

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 first such rule is $S_0 \to S$. To eliminate this rule, for each rule $S \to w$ that has the variable S 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:

- $S \to D$
- $S \to SD$
- $S \to SL$

So, we add three rules:

- $S_0 \to D$
- $S_0 \to SD$
- $S_0 \to SL$

As a result, we get the following grammar:

- $D \to 1$
- $L \to a$
- $S \to D$
- $S \to SD$
- $S \to SL$
- $\underline{S_0 \to D}$
- $\underline{S_0 \to SD}$
- $\underline{S_0 \to SL}$

Another rule that needs to be eliminated on this stage is $S \to D$. 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 $S \to w$. In the current grammar, we have one such rule:

- $D \to 1$

So, we add one new rule:

- $S \to 1$

As a result, we get the following grammar:

- $D \to 1$
- $L \to a$
- $S \to SD$
- $S \to SL$
- $S_0 \to D$

- $S_0 \to SD$
- $S_0 \to SL$
- $\underline{S \to 1}$

One more rule needs to be eliminated on this stage: $S_0 \to D$. 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 $S_0 \to w$. In the current grammar, we have one such rule:

- $D \to 1$

So, we add one new rule:

- $S_0 \to 1$

As a result, we get the following grammar:

- $D \to 1$
- $L \to a$
- $S \to SD$
- $S \to SL$
- $S_0 \to SD$
- $S_0 \to SL$
- $S \to 1$
- $\underline{S_0 \to 1}$

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 two terminal symbols 1 and a. So, we introduce two new variables V_1 and V_a, and two new rules

- $V_1 \to 1$
- $V_a \to a$

So we end up with the following grammar:

- $D \to 1$

- $L \to a$
- $S \to SD$
- $S \to SL$
- $S_0 \to SD$
- $S_0 \to SL$
- $S \to 1$
- $S_0 \to 1$
- $\underline{V_1 \to 1}$
- $\underline{V_a \to a}$

There are no rules that need to be eliminated on this stage.

Step 3. At this step, we deal with the rules in which the right-hand side has length 3 or larger. We do not have any such rules, so we do not need to do anything.

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

- $D \to 1$
- $L \to a$
- $S \to SD$
- $S \to SL$
- $S_0 \to SD$
- $S_0 \to SL$
- $S \to 1$
- $S_0 \to 1$
- $V_1 \to 1$
- $V_a \to a$

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.

Comment. Since we do not have any rules with right-hand side of length 2 which are not in Chomsky normal form, we could skip Step 2 and not all the rules $V_1 \to 1$ and $V_a \to a$.