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:

- $D \to 0$
- $L \to a$
- $S \to L = D;$
- $S \to L = L;$
- $S \to SS$

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:

- $D \to 0$
- $L \to a$
- $S \to L = D;$
- $S \to L = L;$
- $S \to SS$
- $\underline{S_0 \to S}$

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 only 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:

- $S \to L = D;$
- $S \to L = L;$
- $S \to SS$

So, we add three rules:

- $S_0 \to L = D;$
- $S_0 \to L = L;$
- $S_0 \to SS$

As a result, we get the following grammar:

- $D \to 0$
- $L \to a$
- $S \to L = D;$
- $S \to L = L;$
- $S \to SS$
- $\underline{S_0 \to L = D;}$
- $\underline{S_0 \to L = L;}$
- $\underline{S_0 \to SS}$

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, a, =, and ;. So, we introduce four new variables V_0, V_a, $V_=$, and $V_;$ and four new rules

- $V_0 \to 0$
- $V_a \to a$
- $V_= \to =$
- $V_; \to ;$

So we end up with the following grammar:

- $D \to 0$
- $L \to a$
- $\underline{S \to LV_{=}DV_{;}}$
- $\underline{S \to LV_{=}LV_{;}}$
- $S \to SS$
- $\underline{S_0 \to LV_{=}DV_{;}}$
- $\underline{S_0 \to LV_{=}LV_{;}}$
- $S_0 \to SS$
- $\underline{V_0 \to 0}$
- $\underline{V_a \to a}$
- $\underline{V_{=} \to =}$
- $\underline{V_{;} \to ;}$

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 $S \to LV_{=}LV_{;}$ is replaced by three rules:

- $E \to V_{L=L}V_{;}$
- $V_{L=L} \to V_{L=}L$
- $V_{L=} \to LV_{=}$

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

- $D \to 0$
- $L \to a$
- $\underline{S \to V_{L=D}V_{;}}$
- $\underline{V_{L=D} \to V_{L=}D}$
- $\underline{V_{L=} \to LV_{=}}$
- $\underline{S \to V_{L=L}V_{;}}$
- $\underline{V_{L=L} \to V_{L=}L}$
- $S \to SS$
- $\underline{S_0 \to V_{L=D}V_{;}}$
- $\underline{S_0 \to V_{L=L}V_{;}}$

- $S_0 \to SS$
- $V_0 \to 0$
- $V_a \to a$
- $V_= \to =$
- $V_; \to ;$

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.