# 4 Лабораторная работа № 4. Эквивалентные преобразования контекстно-свободных грамматик

Цель: - закрепить понятия «эквивалентные грамматики», «приведенная КС-грамматика»;

- сформировать умения и навыки эквивалентных преобразований контекстно-свободных грамматик.

## Основы теории

**Определение 4.1.** КС-грамматика называется приведенной, если она не имеет циклов, ε-правил и бесполезных символов.

Рассмотрим основные алгоритмы приведения КС-грамматик.

Перед всеми другими исследованиями и преобразованиями КС-грамматик выполняется проверка существования языка грамматики.

### Алгоритм 4.1. Проверка существования языка грамматики

Вход: КС-грамматика $G = (V_T, V_N, P, S)$.

Выход: заключение о существовании или отсутствии языка грамматики.

Определим множество нетерминалов, порождающих терминальные строки $N = \{Z \mid Z \in V_N, Z \Rightarrow {}^*x, x \in V_T^*\}$.

Шаг 1. Положить $N_0$=Ø.

Шаг 2. Вычислить $N_i = N_{i-1} \cup \{A \mid (A \to \alpha) \in P$ и $\alpha \in (N_{i-1} \cup V_T)^*\}$.

Шаг 3. Если $N_i \neq N_{i-1}$, то положить $i$=$i$+1 и перейти к пункту 2, иначе считать $N = N_i$.

Если $S \in N$, то выдать сообщение о том, что язык грамматики существует, иначе сообщить об отсутствии языка.

**Пример 4.1.** Дана грамматика $G = (\{0,1\}, \{S, A, B\}, P, S)$, где множество правил $P$: 1) $S \to AB$; 2) $A \to 0A$; 3) $A \to 0$; 4) $B \to 1$. Построим последовательность приближений множества $N$:

$N_0$ = Ø;
$N_1$ = $\{A, B\}$;
$N_2$ = $\{S, A, B\}$;
$N_3$ = $\{S, A, B\}$.

Т.к. $N_2$=$N_3$, то $N$ = $\{S, A, B\}$, следовательно, язык грамматики существует, потому что начальный символ $S \in N$.

**Определение 4.2.** Бесполезными символами грамматики называют:

а) нетерминалы, не порождающие терминальных строк, т.е. множество символов

$$\{X \mid X \in V_N, \neg\exists(X \Rightarrow {}^*x), x \in V_T^*\};$$

б) недостижимые нетерминалы, порождающие терминальные строки, т.е. множество символов

$$\{X \mid X \in V_N, \neg\exists(S \Rightarrow {}^*\alpha X\beta), \exists(X \Rightarrow {}^*x); \alpha, \beta \in V^*; x \in V_T^*\};$$

в) недостижимые терминалы, т.е. множество символов

$$\{X \mid X \in V_T, \neg\exists(S \Rightarrow {}^*\alpha X\beta); \alpha, \beta \in V^*\}.$$

## Алгоритм 4.2. Устранение нетерминалов, не порождающих терминальных строк

Вход: КС-грамматика $G = (V_T, V_N, P, S)$.

Выход: КС-грамматика $G' = (V_T, V_N', P', S)$, такая, что $L(G') = L(G)$ и для всех $Z \in V_N'$ существуют выводы $Z \Rightarrow {}^*x$, где $x \in V_T^*$.

Шаг 1. Определить множество нетерминалов, порождающих терминальные строки, с помощью алгоритма 4.1.

Шаг 2. Вычислить $V_N' = V_N \cap N, N_Б = V_N - V_N', P' = P - P_Б$, где $P_Б \subseteq P$ - это множество правил, содержащих бесполезные нетерминалы $X \in N_Б$.

**Пример 4.2.** Дана грамматика $G = (\{a, b, c\}, \{S, A, B, C\}, P, S)$ с правилами $P$: 1) $S \to ab$; 2) $S \to AC$; 3) $A \to AB$; 4) $B \to b$; 5) $C \to cb$.

Преобразуем ее в эквивалентную грамматику $G'$ по алгоритму 4.2:

$N_0 = \varnothing$;

$N_1 = \{S, B, C\}$;

$N_2 = \{S, B, C\}$.

Т.к. $N_1 = N_2$, то $N = \{S, B, C\}$. После удаления бесполезных нетерминалов и правил вывода, получим грамматику $G' = (\{a, b, c\}, \{S, B, C\}, P', S)$ с правилами $P'$: 1) $S \to ab$; 2) $B \to b$; 5) $C \to cb$.

## Алгоритм 4.3. Устранение недостижимых символов

Вход: КС-грамматика $G = (V_T, V_N, P, S)$.

Выход: КС-грамматика $G' = (V_T', V_N', P', S)$, такая, что $L(G') = L(G)$ и для всех $Z \in V'$ существует вывод $S \Rightarrow {}^*\alpha Z\beta$, где $\alpha, \beta \in (V')^*$.

Определим множество достижимых символов $Z$ грамматики $G$, т.е. множество

$$W = \{Z \mid Z \in V, \exists\,(S \Rightarrow {}^*\alpha Z\beta); \alpha, \beta \in V^*\}.$$

Шаг 1. Положить $W_0 = S$.

Шаг 2. Вычислить очередное приближение следующим образом:

$W_i = W_{i-1} \cup \{X \mid X \in V, (A \to \alpha X\beta) \in P, A \in W_{i-1}; \alpha, \beta \in V^*\}$.

Шаг 3. Если $W_i \neq W_{i-1}$, то положить $i$:=$i$+1 и перейти к шагу 2, иначе считать $W = W_i$.

Шаг 4. Вычислить $V'_N = V_N \cap W, V'_T = V_T \cap W, V_Б = V - W, P' = P - P_Б$, где $P_Б \subseteq P$ - это множество правил, содержащих недостижимые символы $X \in V_Б$.

**Пример 4.3.** Дана грамматика $G = (\{a, b, c\}, \{S, B, C\}, P, S)$ с правилами $P'$: 1) $S \to ab$; 2) $B \to b$; 5) $C \to cb$.

Преобразуем ее в эквивалентную грамматику $G'$ по алгоритму 4.3:

$W_0 = \{S\}$;

$W_1 = \{S, a, b\}$;

$W_2 = \{S, a, b\}$.

Т.к. $W_1$=$W_2$, то $W$=$\{S, a, b\}$. Множество недостижимых символов $V_Б = \{B, C, c\}$. Тогда после удаления недостижимых символов, получим грамматику $G' = (\{a, b\}, \{S\}, P, S)$ с правилом $P'$: $S \to ab$.

## Алгоритм 4.4. Устранение ε-правил

Вход: КС-грамматика $G = (V_T, V_N, P, S)$.

Выход: Эквивалентная КС-грамматика $G' = (V_T, V'_N, P', S')$ без ε-правил для всех нетерминальных символов, кроме начального, который не должен встречаться в правых частях правил грамматики.

Шаг 1. В исходной грамматике $G$ найти ε-порождающие нетерминальные символы $A \in V_N$, такие, что $A \Rightarrow {}^*\varepsilon$.

1.1 Положить $N_0 = \{A \mid (A \to \varepsilon) \in P\}$.

1.2 Вычислить $N_i = N_{i-1} \cup \{B \mid (B \to \alpha) \in P, \alpha \in N_{i-1}^*\}$.

1.3 Если $N_i \neq N_{i-1}$, то положить $i$:=$i$+1 и перейти к пункту 1.2, иначе считать $N = N_i$.

Шаг 2. Из множества $P$ правил исходной грамматики $G$ перенести во множество $P'$ все правила, за исключением ε-правил, т.е. $P' = P - \{(A \to \varepsilon) \in P$ для всех $A \in V_N\}$.

Шаг 3. Пополнить множество $P'$ правилами, которые получаются из каждого правила этого множества путем исключения всевозможных комбинаций ε-порождающих нетерминалов в правой части. Полученные при этом ε-правила во множество $P'$ не включать.

Шаг 4. Если $S \in N$, то $P' = P \cup \{S' \to \varepsilon, S' \to S\}, V'_N = V_N \cup S'$, где $V \cap \{S'\} = \varnothing$; иначе $V'_N$ =$V_N$, $S' = S$.

**Пример 4.4.** Дана грамматика $G = (\{0,1\}, \{S, A, B\}, P, S)$ с правилами $P$: 1) $S \to AB$; 2) $A \to 0A \mid \varepsilon$; 3) $B \to 1B \mid \varepsilon$. Преобразуем ее в эквивалентную грамматику по алгоритму 4.4.

Шаг 1. $N_0 = \{A, B\}$;

$N_1 = \{S, A, B\}$;

$N_2 = \{S, A, B\}$.

Т.к. $N_1 = N_2$, то искомое множество построено и $N = \{S, A, B\}$.

Шаг 2, 3. Множество $P'$: 1) $S \to AB \mid A \mid B$; 2) $A \to 0A \mid 0$; 3) $B \to 1B \mid 1$.

Шаг 4. Т.к. $S \in N$, то введем новый нетерминал $C$ и пополним множество $P'$ правилом вида $C \to S \mid \varepsilon$. Результирующая грамматика будет иметь вид: $G' = (\{0,1\}, \{S, A, B, C\}, P, C)$ с правилами $P'$: 1) $C \to S \mid \varepsilon$; 2) $S \to AB \mid A \mid B$; 3) $A \to 0A \mid 0$; 4) $B \to 1B \mid 1$.

### Алгоритм 4.5. Устранение цепных правил

Вход: КС-грамматика $G = (V_T, V_N, P, S)$.

Выход: Эквивалентная КС-грамматика $G' = (V_T, V'_N, P', S')$ без цепных правил, т.е. правил вида $A \to B$, где $A, B \in V_N$.

Шаг 1. Для каждого нетерминала $A$ вычислить множество выводимых из него нетерминалов, т.е. множество $N^A = \{B \mid A \Rightarrow {*B},\ \text{где } B \in V_N\}$.

1.1 Положить $N_0^A = \{A\}$.

1.2 Вычислить $N_i^A = N_{i-1}^A \cup \{C \mid (B \to C) \in P, B \in N_{i-1}^A, C \in V_N\}$.

1.3 Если $N_i^A \neq N_{i-1}^A$, то положить $i{:=}i{+}1$ и перейти к пункту 1.2, иначе считать $N^A = N_i^A$.

Шаг 2. Построить множество $P'$ так: если $(B \to \alpha) \in P$ не является цепным правилом $(\alpha \notin V_N)$, то включить в $P'$ правило $A \to \alpha$ для каждого $A$, такого, что $B \in N^A$.

**Пример 4.5.** Грамматика $G = (\{+, n\}, \{L, M, N\}, P, L)$ с правилами $P$: 1) $L \to M$; 2) $M \to N$; 3) $N \to N + \mid n$. Преобразуем ее в эквивалентную грамматику $G'$ по алгоритму 4.5.

Шаг 1. $N_0^L = \{L\}$;

$N_1^L = \{L, M\}$;

$N_2^L = \{L, M, N\}$;

$N_3^L = \{L, M, N\}$.

Т.к. $N_2^L = N_3^L$, то $N^L = \{L, M, N\}$.

$N_0^M = \{M\}$;

$N_1^M = \{M, N\};$

$N_2^M = \{M, N\}.$

Т.к. $N_1^M = N_2^M$, то $N^M = \{M, N\}$.

$N_0^N = \{N\};$

$N_1^N = \{N\}.$

Т.к. $N_1^N = N_0^N$, то $N^N = \{N\}$.

Шаг 2. Преобразовав правила вывода грамматики, получим грамматику $G' = (\{+, n\}, \{L, M, N\}, P', L)$ с правилами $P'$:

$$1)\ L \to N+ \mid n;\quad 2)\ M \to N+ \mid n;\quad 3)\ N \to N+ \mid n.$$

## Алгоритм 4.6. Устранение левой факторизации правил

Вход: КС-грамматика $G = (V_T, V_N, P, S)$.

Выход: Эквивалентная КС-грамматика $G' = (V_T, V_N', P', S')$ без одинаковых префиксов в правых частях правил, определяющих нетерминалы.

Шаг 1. Записать все правила для нетерминала $X$, имеющие одинаковые префиксы $\alpha \in V^*$, в виде одного правила с альтернативами: $X \to \alpha\beta_1 \mid \alpha\beta_2 \mid \ldots \mid \alpha\beta_n;\ \beta_1, \beta_2, \ldots, \beta_n \in V^*$.

Шаг 2. Вынести за скобки влево префикс $\alpha$ в каждой строке-альтернативе: $X \to \alpha(\beta_1 \mid \beta_2 \mid \ldots \mid \beta_n)$.

Шаг 3. Обозначить новым нетерминалом $Y$ выражение, оставшееся в скобках: $X \to \alpha Y, Y \to \beta_1 \mid \beta_2 \mid \ldots \mid \beta_n$.

Шаг 4. Пополнить множество нетерминалов новым нетерминалом $Y$ и заменить правила, подвергшиеся факторизации, новыми правилами для $X$ и $Y$.

Шаг 5. Повторить шаги 1-4 для всех нетерминалов грамматики, для которых это возможно и необходимо.

**Пример 4.6.** Дана грамматика $G = (\{k,l,m,n\}, \{S\}, P, S)$ с правилами $P$: $1)\ S \to kSl;\quad 2)\ S \to kSm;\quad 3)\ S \to n$. Преобразуем ее в эквивалентную грамматику $G'$ по алгоритму 4.6:

Шаг 1. $S \to kSl \mid kSm \mid n$.

Шаг 2. $S \to kS(l \mid m) \mid n$.

Шаг 3,4. Пополнив множество нетерминалов новым нетерминалом $C$ и заменив правила, подвергшиеся факторизации, получим грамматику $G' = (\{k,l,m,n\}, \{S, C\}, P', S)$ с правилами $P'$: $1)\ S \to kSC;\quad 2)\ S \to n;\quad 3)\ C \to l;$ $4)\ C \to m.$

## Алгоритм 4.7. Устранение прямой левой рекурсии

Вход: КС-грамматика $G = (V_T, V_N, P, S)$.

Выход: Эквивалентная КС-грамматика $G' = (V_T, V_N', P', S')$ без прямой левой рекурсии, т.е. без правил вида $A \to A\alpha,\ A \in V_N, \alpha \in V^*$.

Шаг 1. Вывести из грамматики все правила для рекурсивного нетерминала $X$:

$X \to X\alpha_1 \mid X\alpha_2 \mid \ldots \mid X\alpha_m \quad (X \in V_N; \alpha_1, \alpha_2, \ldots, \alpha_m \in V^*)$

$X \to \beta_1 \mid \beta_2 \mid \ldots \mid \beta_n \quad (\beta_1, \beta_2, \ldots, \beta_n \in V^*)$.

Шаг 2. Внести новый нетерминал $Y$ так, чтобы он описывал любой «хвост» строки, порождаемой рекурсивным нетерминалом $X$:

$Y \to \alpha_1 Y \mid \alpha_2 Y \mid \ldots \mid \alpha_m Y$

$Y \to \alpha_1 \mid \alpha_2 \mid \ldots \mid \alpha_m$.

Шаг 3. Заменить в рекурсивном правиле для $X$ правую часть, используя новый нетерминал и все нерекурсивные правила для $X$ так, чтобы генерируемый язык не изменился:

$X \to \beta_1 Y \mid \beta_2 Y \mid \ldots \mid \beta_n Y$

$X \to \beta_1 \mid \beta_2 \mid \ldots \mid \beta_n$

$Y \to \alpha_1 Y \mid \alpha_2 Y \mid \ldots \mid \alpha_m Y$

$Y \to \alpha_1 \mid \alpha_2 \mid \ldots \mid \alpha_m$.

Шаг 4. Пополнить множество нетерминалов грамматики новым нетерминалом $Y$. Пополнить множество правил грамматики правилами, полученными на шаге 3.

Шаг 5. Повторить действия шагов 1-4 для всех рекурсивных нетерминалов грамматики, после чего полученные множества нетерминалов и правил принять в качестве $V_N'$ и $P'$.

**Пример 4.7.** Дана грамматика $G = (\{a, b, c, d, z\}, \{S, A, B, C\}, P, S)$ с правилами $P$: 1) $S \to Aa$; 2) $A \to Bb$; 3) $B \to Cc \mid d$; 4) $C \to Ccbz \mid dbz$. После устранения прямой левой рекурсии получим эквивалентную грамматику $G' = (\{a, b, c, d, z\}, \{S, A, B, C, Z\}, P', S)$ с правилами $P'$:

1) $S \to Aa$; 2) $A \to Bb$; 3) $B \to Cc \mid d$; 4) $C \to dbzZ \mid dbz$; 5) $Z \to cbzZ \mid cbz$.

## Постановка задачи к лабораторной работе № 4

Разработать программное средство, автоматизирующее процесс эквивалентного преобразования КС-грамматик. Программное средство должно выполнять следующие функции:

1) организация ввода грамматики и проверка ее на принадлежность к классу КС-грамматик;

2) проверка существования языка КС-грамматики;

3) реализация эквивалентных преобразований грамматики, направленных на удаление:

а) бесполезных символов;

б) недостижимых символов;

в) ε-правил;

г) цепных правил;

д) левой факторизации правил;

е) прямой левой рекурсии.

Варианты индивидуальных заданий представлены в таблице 4.1.

Таблица 4.1 – Варианты индивидуальных заданий к лабораторной работе № 4

| Вариант | Контекстно-свободная грамматика |
|---|---|
| 1 | $G$=({$S, A, B, D, E$}, {$a, b, c, e$}, $P, S$), где $P$:<br>1) $S \to AB \mid \varepsilon$; 2) $A \to Aa \mid S \mid a$; 3) $B \to bD \mid bS \mid b$; 4) $D \to ccD$; 5) $E \to eE \mid e$. |
| 2 | $G$=({$E, T, F, G, H$}, {+, -, *, /, $n, m, h$}, $P, E$), где $P$:<br>1) $E \to T \mid E+T \mid E-T \mid \varepsilon$; 2) $T \to F \mid F*T \mid F/T \mid \varepsilon$; 3) $F \to G \mid Fn \mid n$;<br>4) $G \to Gm$; 5) $H \to Hh \mid h$. |
| 3 | $G$=({$S, R, T, X, Y$}, {$a, b, p, g, y$}, $P, S$), где $P$:<br>1) $S \to R \mid T$; 2) $R \to pX \mid paR \mid paT \mid \varepsilon$ 3) $T \to Tg \mid g$; 4) $X \to aXb$;<br>5) $Y \to aYa \mid y$. |
| 4 | $G$=({$Q, A, B, C, D$}, {$a, b, c, d$}, $P, Q$), где $P$:<br>1) $Q \to acA \mid acB \mid \varepsilon$; 2) $B \to A \mid Cb \mid \varepsilon$; 3) $A \to Aa \mid Ab \mid a$; 4) $C \to dCc$<br>5) $D \to dc$ |
| 5 | $G$=({$R, T, F, G, K$}, {$m, i, j, k$, ^, ~, ⊥}, $P, R$), где $P$:<br>1) $R \to R\sim T\perp \mid R\hat{}T\perp \mid \varepsilon$; 2) $T \to F \mid Fi \mid Fj \mid Gk \mid \varepsilon$; 3) $G \to GkG$;<br>4) $K \to Ki \mid Km \mid m$. |
| 6 | $G$=({$S, X, Y, Z, K$}, {$x, y, z, k$, #, $}, $P, S$), где $P$:<br>1) $S \to X \mid Y \mid Z$; 2) $X \to x\#X \mid x\#Y \mid \varepsilon$; 3) $Y \to Yy\$ \mid Yz\$ \mid \$ \mid \varepsilon$; 4) $Z \to Zz\$$;<br>5) $K \to Kk\$ \mid k\$$. |
| 7 | $G$=({$S, L, M, P, N$}, {$n, m, l, p$, @, ⊥}, $V, S$), где $V$:<br>1) $S \to @nL \mid @mM \mid P$; 2) $L \to M \mid Ll\perp \mid Lm\perp \mid \varepsilon$; 3) $M \to L \mid Mm \mid mm$;<br>4) $N \to pN@ \mid @$; 5) $P \to nmP$. |
| 8 | $G$=({$X, Y, Z, K, L$}, {$a, b, l$, =, <, >, ∧, ∨, ¬}, $V, X$), где $V$:<br>1) $X \to Y \mid Y=Y \mid Y<Y \mid Y>Y \mid K$; 2) $Y \to Y\wedge Z \mid Y\vee Z \mid \varepsilon$; 3) $Z \to \neg a \mid \neg b \mid \varepsilon$;<br>4) $K \to \neg K$; 5) $L \to l \mid a \mid b$. |
| 9 | $G$=({$Q, A, B, C, D$}, {0, 1, -}, $P, Q$), где $P$:<br>1) $Q \to 01A \mid 01B \mid A$; 2) $A \to 0B1 \mid B \mid 1 \mid \varepsilon$; 3) $B \to BA0 \mid B1 \mid C \mid \varepsilon$; |

Продолжение таблицы 4.1 – Варианты индивидуальных заданий к лабораторной работе № 4

| Вариант | Контекстно-свободная грамматика |
|---|---|
| | 4) $C \to 0C11$;   5) $D \to$ - $D1$ \| -0 \| -1. |
| 10 | $G$=({$R$, $T$, $U$, $W$, $V$}, {0, 1, +, -, *, /}, $P$, $R$), где $P$:<br>1) $R \to T1T$ \| $T1U$ \| $W$ \| ε;   2) $T \to U$ \| $T01$ \| $T10$ \| ε;   3) $U \to +U$ \| +0 \| +1<br>4) $W \to W$-$W$ \| $W$+$W$;   5) $V \to$ *0 \| /1. |
| 11 | $G$=({$S$, $R$, $T$, $F$, $E$}, {$a$, $b$, $k$, {, [, }, ], ⊥}, $P$, $S$), где $P$:<br>1) $S \to${$R$ \| [ $R$;   2) $R \to Ra$} \| $Ra$] \| $a$ \| $T$ \| $F$ \| ε;   3) $F \to${$F$} \| $bb$;   4) $T \to$[$T$];<br>5) $E \to k$⊥. |
| 12 | $G$=({$Y$, $K$, $M$, $L$, $S$}, {$a$, $b$, *, /, ^}, $P$, $Y$), где $P$:<br>1) $Y \to KS$ \| $KM$;   2) $K \to K$* \| $K$/ \| $S$;   3) $S \to Sa$/ \| $Sb$/ \| ε;   4) $M \to$ *$M$*;<br>5) $L \to L$^ \| ^$a$. |