БУ ВО Ханты-Мансийского автономного округа – Югры
«Сургутский государственный университет»

Теория языков программирования и методы трансляции

Часть 6

- Нормальная форма Хомского
- Нормальная форма Грейбах

Гришмановский Павел Валерьевич
доцент кафедры автоматики и компьютерных систем, к.т.н., доцент

Сургут, 2018

Нормальная форма Хомского

Грамматика G = ($\mathbf{V}^T$, $\mathbf{V}^N$, **P**, S) называется **грамматикой в нормальной форме Хомского** (в бинарной нормальной форме), если множество ее правил содержит только правила вида

- $A \to BC$, $A,B,C \in \mathbf{V}^N$
- $A \to a$, $A \in \mathbf{V}^N$, $a \in \mathbf{V}^T$
- $S \to \lambda$, если $\lambda \in L(G)$ и S не содержится в правых частях правил

Дерево вывода всегда является бинарным

Удобны для некоторых видов формального анализа

Для любой контекстно-свободной грамматики можно построить эквивалентную грамматику в нормальной форме Хомского (предварительно получив приведенную грамматику)

Вход: контекстно-свободная приведенная грамматика

$G = (\mathbf{V}^T, \mathbf{V}^N, \mathbf{P}, S)$

Выход: контекстно-свободная грамматика в нормальной форме Хомского

$G' = (\mathbf{V}^{T'}, \mathbf{V}^{N'}, \mathbf{P}', S')$

Нормальная форма Хомского

Алгоритм:

1. $\mathbf{V}^{T'} = \mathbf{V}^{T}$, $\mathbf{V}^{N'}_0 = \mathbf{V}^{N}$, $S' = S$

перенос терминальных и нетерминальных символов

2. $\mathbf{P}'_0 = \{ \forall(A \to a,\ A \to BC,\ S \to \lambda) \in \mathbf{P} : A,B,C \in \mathbf{V}^{N'},\ a \in \mathbf{V}^{T'} \}$

перенос «правильных» правил

3. $\mathbf{P}'_i = \mathbf{P}'_{i\text{-}1} \cup \{ (A \to XB,\ X \to a) : (A \to aB) \in \mathbf{P},\ \forall A,B \in \mathbf{V}^{N'},\ \forall a \in \mathbf{V}^{T'} \} \cup$
$\cup \{ (A \to BX,\ X \to a) : (A \to Ba) \in \mathbf{P},\ \forall A,B \in \mathbf{V}^{N'},\ \forall a \in \mathbf{V}^{T'} \}$,
$\mathbf{V}^{N'}_i = \mathbf{V}^{N'}_{i\text{-}1} \cup \{ X \}$
4. $\mathbf{P}'_i = \mathbf{P}'_{i\text{-}1} \cup \{ (A \to XY,\ X \to a\ ,\ Y \to b) : (A \to ab) \in \mathbf{P},\ \forall A \in \mathbf{V}^{N'},\ \forall a,b \in \mathbf{V}^{T'} \}$,
$\mathbf{V}^{N'}_i = \mathbf{V}^{N'}_{i\text{-}1} \cup \{ X, Y \}$

преобразование правил вида $A \to \gamma$, $|\gamma| = 2$

5. $\mathbf{P}'_i = \mathbf{P}'_{i\text{-}1} \cup \{ A \to X_1Y_1,\ Y_1 \to X_2Y_2,\ \ldots\ Y_{k\text{-}2} \to X_{k\text{-}1}X_k :$
$(A \to B_1B_2\ldots B_k) \in \mathbf{P},\ A \in \mathbf{V}^{N'} \} \cup \{ X_j \to B_j : B_j \in \mathbf{V}^{T'},\ j \in [1; k] \}$,
$\mathbf{V}^{N'}_i = \mathbf{V}^{N'}_{i\text{-}1} \cup \{ Y_1, Y_2, \ldots\ Y_{k\text{-}2} \}$,
$X_j = B_j : \forall B_j \in \mathbf{V}^{N'},\ j \in [1; k]$

преобразование правил вида $A \to \gamma$, $|\gamma| > 2$

Нормальная форма Хомского
(пример)

Дано: контекстно-свободная приведенная грамматика G' = (**V**^{T}', **V**^{N}', **P**', S'):

V^{T} = { a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, _,
0, 1, 2, 3, 4, 5, 6, 7, 8, 9, +, -, *, /, =, ";" }

V^{N} = { S, I, E, E', T, T', M, C, A, K, D }

P = { S → I=E;
E → TE' | MT' | (E) | -M | AK | DC | a | b | c |…| z | _ | 0 | 1 |…| 9
E' → +T | -T | +TE' | -TE'
T → MT' | (E) | -M | AK | DC | a | b | c |…| z | _ | 0 | 1 |…| 9
T' → *M | /M | *MT' | /MT'
M → (E) | -M | AK | DC | a | b | c |…| z | _ | 0 | 1 |…| 9
I → AK | a | b | c |…| z | _
K → AK | DK | a | b | c |…| z | _ | 0 | 1 |…| 9
C → DC | 0 | 1 |…| 9
A → a | b | ... | z | _
D → 0 | 1 | ... | 9 }

Получить: контекстно-свободную грамматику G' = (**V**^{T}', **V**^{N}', **P**', S') в нормальной форме Хомского (в бинарной нормальной форме)

Нормальная форма Хомского
(пример – 2. Перенос правил)

P (все правила)	P' (перенесенные правила)
S → I=E;	-
E → TE'|MT'|(E)|-M|AK|DC
a|b|c|...|z|_|0|1|...|9	E → TE'|MT'|AK|DC
a|b|c|...|z|_|0|1|...|9
E' → +T|-T|+TE'|-TE'	-
T → MT'|(E)|-M|AK|DC
a|b|c|...|z|_|0|1|...|9	T → MT'|AK|DC
a|b|c|...|z|_|0|1|...|9
T' → *M|/M|*MT'|/MT'	-
M → (E)|-M|AK|DC
a|b|c|...|z|_|0|1|...|9	M → AK|DC
a|b|c|...|z|_|0|1|...|9
I → AK|a|b|c|...|z|_	I → AK|a|b|c|...|z|_
K → AK	DK
a|b|c|...|z|_|0|1|...|9	K → AK|DK
a|b|c|...|z|_|0|1|...|9
C → DC|0|1|...|9	C → DC|0|1|...|9
A → a|b|c|...|z|_	A → a|b|...|z|_
D → 0|1|...|9	D → 0|1|...|9

Обозначения: Удаляемое правило
Добавляемое правило

-15 правил

Нормальная форма Хомского

(пример – 3-4. Добавление правил вида A→XY)

P (оставшиеся правила)	P' (добавляемые правила)
S → I=E;	-
E → (E) | -M	E → X_1M X_1 → -
E' → +T | -T | +TE' | -TE'	E' → X_2T | X_3T X_2 → + X_3 → -
T → (E) | -M	T → X_4M X_4 → -
T' → *M | /M | *MT' | /MT'	T' → X_5M | X_6M X_5 → * X_6 → /
M → (E) | -M	M → X_7M X_7 → -

Обозначения:

Удаляемое правило
Заменяемое правило
Добавляемый нетерминальный символ

+14 правил
+7 символов

Нормальная форма Хомского

(пример – 5. Добавление правил вида $A \to X_1X_2\ldots X_n$)

P (оставшиеся правила)	P' (добавляемые правила)	
$S \to I{=}E;$	$S \to IY_1$	
	$Y_1 \to X_8Y_2$	$X_8 \to =$
	$Y_2 \to EX_9$	$X_9 \to ;$
$E \to (E)$	$E \to X_{10}Y_3$	$X_{10} \to ($
	$Y_3 \to EX_{11}$	$X_{11} \to)$
$E' \to +TE' \mid -TE'$	$E' \to X_{12}Y_4 \mid X_{13}Y_5$	
	$Y_4 \to TE'$	$X_{12} \to +$
	$Y_5 \to TE'$	$X_{13} \to -$
$T \to (E)$	$T \to X_{14}Y_6$	$X_{14} \to ($
	$Y_6 \to EX_{15}$	$X_{15} \to)$
$T' \to *MT' \mid /MT'$	$T' \to X_{16}Y_7 \mid X_{17}Y_8$	
	$Y_7 \to MT'$	$X_{16} \to *$
	$Y_8 \to MT'$	$X_{17} \to /$
$M \to (E)$	$M \to X_{18}Y_9$	$X_{18} \to ($
	$Y_9 \to EX_{19}$	$X_{19} \to)$

Обозначения:
Удаляемое правило
Заменяемое правило
Добавляемый нетерминальный символ

+29 правил
+21 символ

Нормальная форма Хомского
(пример – Результат)

Результат: грамматика G' = ($\mathbf{V}^{T}$', $\mathbf{V}^{N}$', **P**', S') в нормальной форме Хомского:

$\mathbf{V}^T$ = { a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, _,
0, 1, 2, 3, 4, 5, 6, 7, 8, 9, +, -, *, /, =, ";" }

$\mathbf{V}^N$ = { S, I, E, E', T, T', M, C, A, K, D, Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7, Y_8, Y_9,
X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}, X_{18}, X_{19} }

P = { S → IY_1
E → TE' | MT' | AK | DC | X_1M | $X_{10}Y_3$ |
| a | b | c |…| z | _ | 0 | 1 |…| 9
E' → X_2T | X_3T | $X_{12}Y_4$ | $X_{13}Y_5$
T → MT' | AK | DC | X_4M | $X_{14}Y_6$ |
| a | b | c |…| z | _ | 0 | 1 |…| 9
T' → X_5M | X_6M | $X_{16}Y_7$ | $X_{17}Y_8$
M → AK | DC | X_7M | $X_{18}Y_9$ |
| a | b | c |…| z | _ | 0 | 1 |…| 9
I → AK | a | b | c |…| z | _
K → AK | DK |
| a | b | c |…| z | _ | 0 | 1 |…| 9
C → DC | 0 | 1 |…| 9
A → a | b | ... | z | _
D → 0 | 1 | ... | 9

$Y_1 \to X_8Y_2$
$Y_2 \to EX_9$
$Y_3 \to EX_{11}$
$Y_4 \to TE'$
$Y_5 \to TE'$
$Y_6 \to EX_{15}$
$Y_7 \to MT'$
$Y_8 \to MT'$
$Y_9 \to EX_{19}$

$X_1 \to -$
$X_2 \to +$
$X_3 \to -$
$X_4 \to -$
$X_5 \to *$
$X_6 \to /$
$X_7 \to -$
$X_8 \to =$
$X_9 \to ;$

$X_{10} \to ($
$X_{11} \to)$
$X_{12} \to +$
$X_{13} \to -$
$X_{14} \to ($
$X_{15} \to)$
$X_{16} \to *$
$X_{17} \to /$
$X_{18} \to ($
$X_{19} \to)$
}

$X_1 = X_3 = X_4 = X_7 = X_{13}$
$X_2 = X_{12}$, $X_5 = X_{16}$, $X_6 = X_{17}$
$X_{10} = X_{14} = X_{18}$, $X_{11} = X_{15} = X_{19}$
$Y_3 = Y_6 = Y_9$, $Y_4 = Y_5$, $Y_7 = Y_8$

+28 правил
+28 символов

Нормальная форма Хомского
(пример – Результат)

Результат: грамматика G' = ($\mathbf{V}^{T}$', $\mathbf{V}^{N}$', **P**', S') в нормальной форме Хомского:

$\mathbf{V}^T$ = { a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, _,
0, 1, 2, 3, 4, 5, 6, 7, 8, 9, +, -, *, /, =, ";" }

$\mathbf{V}^N$ = { S, I, E, E', T, T', M, C, A, K, D, Y_1, Y_2, Y_3, Y_4, Y_7, X_1, X_2, X_5, X_6, X_8, X_9, X_{10}, X_{11} }

P = { S → IY_1
E → TE' | MT' | AK | DC | X_1M | $X_{10}Y_3$ |
| a | b | c |…| z | _ | 0 | 1 |…| 9
E' → X_2T | X_1T | X_2Y_4 | X_1Y_4
T → MT' | AK | DC | X_1M | $X_{10}Y_3$ |
| a | b | c |…| z | _ | 0 | 1 |…| 9
T' → X_5M | X_6M | X_5Y_7 | X_6Y_7
M → AK | DC | X_1M | $X_{10}Y_3$ |
| a | b | c |…| z | _ | 0 | 1 |…| 9
I → AK | a | b | c |…| z | _
K → AK | DK |
| a | b | c |…| z | _ | 0 | 1 |…| 9
C → DC | 0 | 1 |…| 9
A → a | b | ... | z | _
D → 0 | 1 | ... | 9

$Y_1 \to X_8Y_2$
$Y_2 \to EX_9$
$Y_3 \to EX_{11}$
$Y_4 \to$ TE'
$Y_7 \to$ MT'
$X_1 \to$ -
$X_2 \to$ +
$X_5 \to$ *
$X_6 \to$ /
$X_8 \to$ =
$X_9 \to$;
$X_{10} \to$ (
$X_{11} \to$)
}

+13 правил
+13 символов

Нормальная форма Грейбах

Грамматика G = ($\mathbf{V}^T$, $\mathbf{V}^N$, **P**, S) называется **грамматикой в нормальной форме Грейбах**, если множество ее правил содержит только правила вида

- $A \to a\alpha$, $A \in \mathbf{V}^N$, $a \in \mathbf{V}^T$, $\alpha \in \mathbf{V}^{N*}$
- $S \to \lambda$, если $\lambda \in L(G)$ и S не содержится в правых частях правил

Для любой контекстно-свободной грамматики можно построить эквивалентную грамматику в нормальной форме Грейбах (предварительно получив приведенную грамматику)

Вход: контекстно-свободная приведенная грамматика

G = ($\mathbf{V}^T$, $\mathbf{V}^N$, **P**, S)

Выход: контекстно-свободная грамматика в нормальной форме Грейбах

G' = ($\mathbf{V}^{T}$', $\mathbf{V}^{N}$', **P**', S')

Суть алгоритма преобразования: все правила вида $A \to B\gamma$ заменяются на $A \to \beta_1\gamma \mid \beta_2\gamma \mid \ldots \mid \beta_m\gamma$: $(B \to \beta_1 \mid \beta_2 \mid \ldots \mid \beta_m) \in \mathbf{P}$ до тех пор, пока все правила грамматики не примут вид $A \to a\alpha$, $A \in \mathbf{V}^N$, $a \in \mathbf{V}^T$, $\alpha \in \mathbf{V}^*$

(необходимо, чтобы $\forall i : A_i \to a\alpha$ или $A_i \to A_j\alpha$, $j > i$)

Затем вместо правил $A \to a\alpha b\beta$, $a,b \in \mathbf{V}^T$, $\alpha \in \mathbf{V}^{N*}$, $\beta \in \mathbf{V}^*$ записываются $A \to a\alpha A'$, $A' \to b\beta$, где A' – новый нетерминальный символ, правила которого также преобразуются и т.д.

Нормальная форма Грейбах

Алгоритм:

1. Получить нелеворекурсивную грамматику G' = ($\mathbf{V}^{T'}$, $\mathbf{V}^{N'}{}_0$, $\mathbf{P}'_0$, S'), сохранив отношение порядка на множестве $\mathbf{V}^{N'}$ (используется алгоритм преобразования грамматики в нелеворекурсивную)
2. Для i = n-1..1: (последовательно, в обратном порядке от n-1 до 1)
   $\mathbf{P}'_i = \mathbf{P}'_{i\text{-}1} \setminus \{ \forall(A_i \to A_j\gamma) \in \mathbf{P}'\} \cup \{ A_i \to \beta_1\gamma \mid \beta_2\gamma \mid \ldots \mid \beta_m\gamma \}$:
   $\exists(A_i \to A_j\gamma) \in \mathbf{P}', \exists(A_j \to \beta_1 \mid \beta_2 \mid \ldots \mid \beta_m) \in \mathbf{P}'\ , \forall j \}$
   - если правая часть правила начинается с **нетерминального** символа A_j, то вместо него подставляются все правые части его правил
   - выполнение строго в обратном порядке (от n-1 до 1) гарантирует, что правые части всех правил будут начинаться с **терминальных** символов
   - ***j* > *i*** по определению, т.к. в п.1 установлено отношение порядка на $\mathbf{V}^{N'}$
3. $\mathbf{P}'_i = \mathbf{P}'_{i\text{-}1} \setminus \{ \forall(A \to a\alpha b\beta) \in \mathbf{P}' \} \cup \{ (A \to a\alpha A', A' \to b\beta) \}$:
   $\exists(A \to a\alpha b\beta) \in \mathbf{P}', a,b \in \mathbf{V}^T, \alpha \in \mathbf{V}^{N*}, \beta \in \mathbf{V}^*,$
   $\mathbf{V}^{N'}{}_i = \mathbf{V}^{N'}{}_{i\text{-}1} \cup \{ A' \}$ A' – новый нетерминальный символ
   - правая часть правила «разбивается» по второму терминальному символу, начиная с которого часть правила переносится в новое правило
4. Если $\mathbf{P}'_i \neq \mathbf{P}'_{i\text{-}1}$, то повторить п. 3

Нормальная форма Грейбах
(пример)

Дано: контекстно-свободная приведенная нелеворекурсивная грамматика $G = (\mathbf{V}^T, \mathbf{V}^N, \mathbf{P}, S)$:

$\mathbf{V}^T$ = { a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, _, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, +, -, *, /, =, ";" }

$\mathbf{V}^N$ = { S, I, E, E', T, T', M, C, A, K, D }

P = { S → I=E;
E → TE' | MT' | (E) | -M | AK | DC | a | b | c |…| z | _ | 0 | 1 |…| 9
E' → +T | -T | +TE' | -TE'
T → MT' | (E) | -M | AK | DC | a | b | c |…| z | _ | 0 | 1 |…| 9
T' → *M | /M | *MT' | /MT'
M → (E) | -M | AK | DC | a | b | c |…| z | _ | 0 | 1 |…| 9
I → AK | a | b | c |…| z | _
K → AK | DK | a | b | c |…| z | _ | 0 | 1 |…| 9
C → DC | 0 | 1 |…| 9
A → a | b | … | z | _
D → 0 | 1 | … | 9 }

Отношение порядка

$S = A_1$
$E = A_2$
$E' = A_3$
$T = A_4$
$T' = A_5$
$M = A_6$
$I = A_7$
$K = A_8$
$C = A_9$
$A = A_{10}$
$D = A_{11}$

Получить: грамматику $G' = (\mathbf{V}^{T}{}', \mathbf{V}^{N}{}', \mathbf{P}', S')$ в нормальной форме Грейбах

Исходная грамматика удовлетворяет требованиям к грамматике G', получаемой на первом шаге алгоритма

Нормальная форма Грейбах

(пример – 2. Замена правил вида $A_i \to A_j\beta$)

<table>
<tr><th>Исходная грамматика</th><th>Преобразованная грамматика</th></tr>
<tr><td>D → 0|1|...|9</td><td>D → 0|1|...|9</td></tr>
<tr><td>A → a|b|c|...|z|_</td><td>A → a|b|c|...|z|_</td></tr>
<tr><td>C → DC|0|1|...|9</td><td>C → 0C|1C|...|9C|0|1|...|9</td></tr>
<tr><td>K → AK|DK|
|a|b|c|...|z|_|0|1|...|9</td><td>K → aK|bK|cK|...|zK|_K|0K|1K|...|9K|
|a|b|c|...|z|_|0|1|...|9</td></tr>
<tr><td>I → AK|a|b|c|...|z|_</td><td>I → aK|bK|cK|...|zK|_K|a|b|c|...|z|_</td></tr>
<tr><td>M → (E)|-M|AK|DC|
|a|b|c|...|z|_|0|1|...|9</td><td>M → aK|bK|cK|...|zK|_K|0C|1C|...|9C|
|(E)|-M|a|b|c|...|z|_|0|1|...|9</td></tr>
<tr><td>T'→ *M|/M|*MT'|/MT'</td><td>T'→ *M|/M|*MT'|/MT'</td></tr>
<tr><td>T → MT'|(E)|-M|AK|DC|
a	b	c	...	z	_	0	1	...	9</td><td>T → aKT'	bKT'	cKT'	...	zKT'	_KT'
0CT'	1CT'	...	9CT'
(E)T'	-MT'
aT'	bT'	cT'	...	zT'	_T'
0T'	1T'	...	9T' 	aK	bK	cK	...	zK	_K	0C	1C	...	9C
(E)	-M	a	b	c	...	z	_	0	1	...	9</td></tr>
<tr><td>E'→ +T|-T|+TE'|-TE'</td><td>E'→ +T|-T|+TE'|-TE'</td></tr>
</table>

-9 правил

+224 правил

Нормальная форма Грейбах

(пример – 2. Замена правил вида $A_i \to A_j\beta$)

Исходная грамматика	Преобразованная грамматика
E → TE'	E → aKT'E'|bKT'E'|cKT'E'|...|zKT'E'| |_KT'E'|0CT'E'|1CT'E'|...|9CT'E'| |(E)T'E'|-MT'E'| |aT'E'|bT'E'|cT'E'|...| |zT'E'|_T'E'| |0T'E'|1T'E'|...|9T'E'| |aKE'|bKE'|cKE'|...|zKE'|_KE'| |0CE'|1CE'|...|9CE'| |(E)E'|-ME'| |aE'|bE'|cE'|...|zE'|_E'| |0E'|1E'|...|9E'
E → MT'	E → aKT'|bKT'|cKT'|...|zKT'|_KT'| |0CT'|1CT'|...|9CT'|(E)T'|-MT'| |aT'|bT'|cT'|...|zT'|_T'| |0T'|1T'|...|9T'
E → AK|DC E → (E)|-M| |a|b|c|...|z|_|0|1|...|9	E → aK|bK|cK|...|zK|_K|0C|1C|...|9C E → (E)|-M| |a|b|c|...|z|_|0|1|...|9
S → I=E;	S → aK=E;|bK=E;|cK=E;|...|zK=E;| |_K=E;|a=E;|b=E;|c=E;|...|z=E;|_=E;

-5 правил

+319 правил

Нормальная форма Грейбах

(пример – 3. Замена правил вида $A_i \to a\alpha b\beta$)

Удаляемые правила	Добавляемые правила и символы
S → aK=E;|bK=E;|cK=E;|...| |zK=E;|_K=E; | |a=E;|b=E;|c=E;|...| |z=E;|_=E;	S → aKS₁|bKS₁|cKS₁|...|zKS₁|_KS₁| |aS₁|bS₁|cS₁|...|zS₁|_S₁ S₁ → =ES₂ S₂ → ;
E → (E)T'E'|(E)E'|(E)T'|(E)	E → (EE₁|(EE₂|(EE₃|(EE₄ E₁ →)T'E' E₂ →)E' E₃ →)T' E₄ →)
T → (E)T'|(E)	T → (ET₁|(ET₂ T₁ →)T' T₂ →)
M → (E)	T → (EM₁ M₁ →)

-61 правил

+68 правил
+9 символов

- объединить нетерминальные символы с совпадающими правилами
- выполнить левую факторизацию (с последующим приведением грамматики к нормальной форме Грейбах)
- удалить несущественные символы и соответствующие правила