

Глава 4. Восходящий синтаксический анализ

4.2. Грамматики простого предшествования

Одним из наиболее легких подходов к решению проблемы поиска и своевременной свертки основы является реализация восходящего синтаксического анализа для небольшого класса КС-грамматик, называемых *грамматиками простого предшествования*. Технология синтаксического анализа для таких грамматик предполагает введение специальных бинарных отношений между каждой парой символов грамматики (как терминалов, так и нетерминалов). Эти отношения управляют выбором основ сентенциальных форм для их последующей свертки.

Основным недостатком этого метода является применимость его лишь в узком классе грамматик простого предшествования.

4.2.1. Отношения предшествования

Пусть $G = (V_T, V_N, P, S)$ – КС-грамматика, а строка $\alpha XY\beta$ – правосторонняя сен-тенциальная форма, где $\alpha, \beta \in (V_T \cup V_N)^*$, $X, Y \in V_T \cup V_N$. В некоторый момент (на одном из этапов процесса последовательных сверток сентенциальной формы) возникает одна из следующих возможных ситуаций:

1. Y – самый левый символ (*заголовок*) основы сентенциальной формы, а X не входит в основу. В этом случае говорят, что символ Y *предшествует* символу X (поскольку символ Y должен быть свернут раньше символа X), и записывают в виде $X < Y$.

2. X и Y входят в одну и ту же основу. В этом случае говорят, что X и Y имеют *равное предшествование* (поскольку сворачиваются одновременно), и записывают в виде $X \doteq Y$.

3. X – последний символ (*окончание*) основы, а Y не входит в основу. В этом случае говорят, что символ X *предшествует* символу Y (поскольку символ X должен быть свернут раньше символа Y), и записывают в виде $X > Y$.

Отношения \prec , \doteq , \succ называются *отношениями предшествования*. Следует отметить, что, хотя эти отношения похожи на арифметические отношения $<$, $=$, $>$, они имеют совершенно иные свойства. В частности, они не обладают свойствами коммутативности и ассоциативности. Из отношения $X \succ Y$, не следует, что существует отношение $Y \prec X$. Эти отношения не являются симметричными, из отношения $X \doteq Y$ не следует $Y \doteq X$. Для одной и той же грамматики может быть так, что $X \prec Y$ и $X \succ Y$, или для некоторых пар символов не выполняется ни одно из отношений предшествования.

Формально эти отношения для символов $X, Y \in V_T \cup V_N$ грамматики определяются следующим образом:

1. $X \doteq Y$, если существует некоторая продукция $A \rightarrow \alpha XY \beta$, $A \in V_N$, $\alpha, \beta \in (V_T \cup V_N)^*$. Это значит, что в правосторонней сентенциальной форме X и Y входят в одну и ту же основу.

2. $X \triangleleft Y$, если существует некоторая продукция $A \rightarrow \alpha XB \beta$, $A, B \in V_N$, $\alpha, \beta \in (V_T \cup V_N)^*$, такая, что $B \xrightarrow{+} Y \delta$, $\delta \in (V_T \cup V_N)^*$. Это значит, что в правосторонней сентенциальной форме основа начинается с символа Y (Y является заголовком основы).

3. $X \triangleright Y$, если существует некоторая продукция $A \rightarrow \alpha BZ \beta$, $A, B \in V_N$, $Z \in V_T \cup V_N$, $\alpha, \beta \in (V_T \cup V_N)^*$, такая, что $B \xrightarrow{+} \gamma X$ и $Z \xrightarrow{*} Y \delta$, $\gamma, \delta \in (V_T \cup V_N)^*$. Это значит, что в правосторонней сентенциальной форме основа завершается символом X (X является окончанием основы). Следует отметить, что в правосторонней сентенциальной форме справа от основы может быть только терминальная строка. Поэтому в данном случае символ Y может быть только терминалом, т. е. $Y \in V_T$, и отношение \triangleright определяется на множестве $(V_T \cup V_N) \times V_T$. Кроме того, что если $Y \delta$ выводится из Z за нуль шагов, то $Z = Y$.

КС-грамматика $G = (V_T, V_N, P, S)$ называется *грамматикой простого предшествования*:

- 1) если не содержит ε -продукций;
- 2) никакие две продукции грамматики не имеют совпадающих правых частей (грамматики, в которых нет двух продукций с одинаковыми правыми частями, называются *обратимыми*);
- 3) любые два символа, составляющие элемент множества $(V_T \cup V_N) \times (V_T \cup V_N)$, связаны одним и тем же отношением предшествования.

Отношения предшествования обычно записывают в виде *матрицы предшествования*, строки и столбцы которой соответствуют символам грамматики. На пересечении i -й строки и j -го столбца записывается отношение предшествования между соответствующими символами грамматики. Элементами матрицы являются знаки $<$, \doteq , $>$ или «пусто». Последний случай означает, что соответствующие символы не могут стоять рядом ни в одной сентенциальной форме.

4.2.2. Вычисление отношений предшествования

Формальный процесс вычисления отношений предшествования для символов $X, Y \in V_T \cup V_N$ заданной КС-грамматики можно представить такой последовательностью действий (в описании действий строки $\alpha, \beta \in (V_T \cup V_N)^*, A \in V_N$):

1. Определить для каждого нетерминала X грамматики множество $L(X) = \{Y \mid X \xRightarrow{+} Y\alpha\}$, т. е. множество символов грамматики (как терминалов, так и нетерминалов), с которых могут начинаться строки, выводимые из нетерминала X . Для этого необходимо построить отношение $\langle \text{LEFT} \rangle$, определяемое следующим образом: $X \langle \text{LEFT} \rangle Y$, если в грамматике существует продукция вида $X \rightarrow Y\beta$. Затем вычислить отношение $\langle \text{LEFT} \rangle^+$ как транзитивное замыкание отношения $\langle \text{LEFT} \rangle$. Тогда $L(X)$ есть множество символов Y , для которых выполняется отношение $X \langle \text{LEFT} \rangle^+ Y$.

2. Вычислить отношение $\langle \text{LEFT} \rangle^*$ как рефлексивно-транзитивное замыкание отношения $\langle \text{LEFT} \rangle$. Очевидно, что отношение $\langle \text{LEFT} \rangle^*$ легко вычисляется по отношению $\langle \text{LEFT} \rangle^+$, поскольку имеет место соотношение $\langle \text{LEFT} \rangle^* = \langle \text{LEFT} \rangle^+ \cup I$, где I – отношение тождественности. Отношение $\langle \text{LEFT} \rangle^*$ понадобится для вычисления отношения \rangle .

3. Определить для каждого нетерминала X грамматики множество $R(X) = \{Y \mid X \xRightarrow{+} \alpha Y\}$, т. е. множество символов грамматики, являющихся крайними справа в строках, выводимых из нетерминала X . Для этого необходимо построить отношение $\langle \text{RIGHT} \rangle$, определяемое следующим образом: $Y \langle \text{RIGHT} \rangle X$, если в грамматике существует продукция вида $X \rightarrow \beta Y$. Затем вычислить отношение $\langle \text{RIGHT} \rangle^+$ как транзитивное замыкание отношения $\langle \text{RIGHT} \rangle$. Тогда $R(X)$ есть множество символов Y , для которых выполняется отношение $Y \langle \text{RIGHT} \rangle^+ X$.

4. Построить для всех символов грамматики отношение \doteq по его определению, т. е. $X \doteq Y$, если в грамматике существует продукция вида $A \rightarrow \alpha XY\beta$.

5. Вычислить отношение \prec . Из его формального определения следует, что $X \prec Y$, если в грамматике имеется продукция вида $A \rightarrow \alpha XB\beta$, где $B \in V_N$, и $Y \in L(B)$. Таким образом, отношение \prec можно вычислить как произведение отношений \doteq и $\langle \text{LEFT} \rangle^+$, т. е. $(\prec) = (\doteq) (\langle \text{LEFT} \rangle^+)$.

6. Вычислить отношение $>$. Из его формального определения следует, что $X > Y$ (напомним, что отношение определено только для $Y \in V_T$), если существует продукция

а) вида $A \rightarrow \alpha B Y \beta$, где $B \in V_N$, $Y \in V_T$, и $X \in R(B)$;

б) вида $A \rightarrow \alpha B Z \beta$, где $B, Z \in V_N$, и $X \in R(B)$, $Y \in L_T(Z)$, где $L_T(Z) \subseteq L(Z)$ – подмножество терминалов множества $L(Z)$.

Таким образом, $(>) = (<\text{RIGHT}>^+) (\doteq) (<\text{LEFT}>^*)$. Поскольку Y может быть только терминалом, при вычислении произведения отношений следует рассматривать только те отношения $X <\text{LEFT}>^* Y$, где $Y \in V_T$.

7. Построить матрицу предшествования, объединив матрицы отношений \doteq , \leq и $>$ в одну и заменив единицы на соответствующие обозначения $(\doteq, \leq, >)$ отношений.

Пример процесса вычисления отношений предшествования для грамматики с продукциями

$$S \rightarrow AB$$

$$A \rightarrow aA \mid a$$

$$B \rightarrow bB \mid b$$

представлен на рис. 4.2.

$$S \rightarrow AB$$

$$A \rightarrow aA \mid a$$

$$B \rightarrow bB \mid b$$

Из отношений $\langle \text{LEFT} \rangle^+$
и $\langle \text{RIGHT} \rangle^+$ следует, что

$$L(S) = \{A, a\},$$

$$L(A) = \{a\},$$

$$L(B) = \{b\},$$

$$R(S) = \{B, b\},$$

$$R(A) = \{A, a\},$$

$$R(B) = \{B, b\}.$$

	$\langle \text{LEFT} \rangle$					
	S	A	B	a	b	
S		1				
A				1		
B					1	

	$\langle \text{LEFT} \rangle^+$					
	S	A	B	a	b	
S		1		1		
A				1		
B					1	

	$\langle \text{LEFT} \rangle^*$					
	S	A	B	a	b	
S	1	1		1		
A		1		1		
B			1		1	
a				1		
b					1	

	$\langle \text{RIGHT} \rangle$			
	S	A	B	
S				
A		1		
B	1		1	
a		1		
b			1	

	$\langle \text{RIGHT} \rangle^+$			
	S	A	B	
S				
A		1		
B	1		1	
a		1		
b	1		1	

	$\hat{=}$					
	S	A	B	a	b	
S						
A			1			
B						
a		1				
b			1			

	$\hat{<}$					
	S	A	B	a	b	
S						
A				1		
B						
a				1		
b					1	

	$\hat{>}$					
	S	A	B	a	b	
S						
A				1		
B						
a				1		
b						

Матрица предшествования						
	S	A	B	a	b	
S						
A		$\hat{=}$		$\hat{<}$	$\hat{>}$	
B						
a		$\hat{=}$		$\hat{<}$	$\hat{>}$	
b			$\hat{=}$		$\hat{<}$	

Рис. 4.2. Процесс вычисления отношений предшествования

Грамматика не является грамматикой простого предшествования, поскольку она не удовлетворяет третьему условию, требующему, чтобы любые два символа грамматики были связаны одним и тем же отношением предшествования. В нашем случае одновременно выполняются $A \prec b$ и $A \succ b$. Действительно $A \prec b$, так как имеется продукция $S \rightarrow AB$ и $b \in L(B)$, т. е. $B \xRightarrow{+} b$, а $A \succ b$, так как существует продукция $S \rightarrow AB$, такая, что $A \in R(A)$ и $b \in L(B)$, т. е. $A \xRightarrow{+} aA$ и $B \xRightarrow{*} b$.

Более кратко для вычисления вручную (пусть имеется продукция вида $A \rightarrow \alpha XY\beta$):

- 1) $X \doteq Y$ для любых символов (терминалов и нетерминалов),
- 2) если $Y \in V_N$, то X связано отношением \prec со всеми элементами из $L(Y)$,
- 3) если $X \in V_N$, то
 - а) если $Y \in V_T$, то все элементы из $R(X)$ связаны отношением \succ с элементом Y ,
 - б) если $Y \in V_N$, то все элементы из $R(X)$ связаны отношением \succ со всеми терминалами из $L(Y)$.