

Глава 4. Восходящий синтаксический анализ

4.2. Грамматики простого предшествования

4.2.3. Синтаксический анализ

Если грамматика является грамматикой простого предшествования, то отношения предшествования позволяют легко определить границы основы правосторонней сентенциальной формы. Для этого достаточно проанализировать сентенциальную форму слева направо и найти самую левую пару символов X_j и X_{j+1} , таких, что $X_j > X_{j+1}$, т. е. X_j – окончание основы, а отношение $>$ отмечает ее правую границу. Затем сентенциальная форма просматривается справа налево, начиная с символа X_j , до тех пор, пока не будет найдена пара символов X_{i-1} и X_i , таких, что $X_{i-1} < X_i$, т. е. X_i – заголовок основы, а отношение $<$ отмечает ее левую границу. Ясно, что между всеми соседними символами внутри основы выполняется отношение \doteq . В результате будет найдена основа, имеющая вид $\beta = X_i \dots X_j$, и должна существовать продукция вида $A \rightarrow \beta$, т. е. основа может быть свернута к нетерминалу A . Например, если сентенциальная форма с отношениями предшествования имеет вид

$$X_1 < X_2 \doteq X_3 \doteq X_4 > X_5,$$

то $X_2X_3X_4$ является ее основой (в грамматике должна быть продукция вида $A \rightarrow X_2X_3X_4$), а в результате свертки получится сентенциальная форма X_1AX_5 .

Возникает трудность, если заголовок основы является первым символом или окончание основы является последним символом сентенциальной формы. Для преодоления этой трудности вводится специальный символ \perp , принадлежащий множеству терминалов, который используется для обозначения начала и конца сентенциальной формы. При этом предполагается, что $\perp \prec X$ и $X \succ \perp$ для всех $X \in V_T \cup V_N$, и результатом последовательности сверток входной строки $\perp x \perp$, где $x \in L(G)$, будет строка вида $\perp S \perp$. Другой подход основан на том, что отношения предшествования вычисляются в предположении, что в грамматике имеется продукция вида $S' \rightarrow \perp S \perp$. Тогда $\perp \prec X$ для всех X , для которых $S \xrightarrow{\pm} X\delta$, т. е. для всех $X \in L(S)$, а $Y \succ \perp$ для всех Y , для которых $S \xrightarrow{\pm} \alpha Y$, т. е. для всех $Y \in R(S)$. Это позволяет выявлять ошибки на более ранних этапах синтаксического анализа.

Как уже отмечалось, для реализации свертки удобным средством является стек. В начале анализа входной строки в вершину стека записывается символ \perp . Затем осуществляется поиск окончания самой левой основы анализируемой строки. Для этого по матрице предшествования проверяются отношения между символом X , находящимся в вершине стека, и очередным входным символом Y . Если условие $X \succ Y$ не выполняется, то очередной входной символ заносится в стек и процесс продолжается до тех пор, пока не будет выполнено условие $X \succ Y$, т. е. не будет найдено окончание самой левой основы. Таким образом, если между символом в вершине стека и очередным входным символом выполняется отношение \prec или \doteq , то синтаксический анализатор выполняет операцию переноса входного символа в стек (окончание основы еще не найдено). После обнаружения окончания основы (выполняется отношение \succ) в верхней части стека будет содержаться основа сентенциальной формы.

Затем, чтобы выполнить свертку, производится поиск заголовка основы. Для этого из стека исключаются все символы, имеющие равное предшествование \doteq (начиная с вершины стека), до тех пор, пока между символом в вершине стека и последним исключенным из стека символом не выполнится отношение $<$, т. е. последний исключенный из стека символ будет являться заголовком основы. Завершается свертка поиском продукции с соответствующей правой частью и записью в стек нетерминала из левой части этой продукции. Если такая продукция не найдена (т. е. подстрока не является основой), то фиксируется синтаксическая ошибка.

Разбор входной строки успешно завершен, если в результате работы стек будет содержать строку $\perp S$ (содержимое стека представляется строкой, в которой самый правый символ находится в вершине стека), а входная строка будет содержать символ \perp . Если основа не будет найдена или между парами соседних символов не выполняется ни одно из отношений предшествования, то фиксируется синтаксическая ошибка.

Детали реализации алгоритма синтаксического анализа предлагаются в качестве упражнения.

В качестве примера рассмотрим грамматику простого предшествования с продукциями

$$S \rightarrow DB$$

$$D \rightarrow A$$

$$A \rightarrow aA \mid a$$

$$B \rightarrow bB \mid b$$

матрица предшествования для которой приведена на рис. 4.4. Отношения предшествования вычислены в предположении, что в грамматике имеется продукция $S' \rightarrow \perp S \perp$.

	S	D	A	B	a	b	\perp
S							
D				\doteq	$<$		
A						$>$	
B							$>$
a			\doteq		$<$	$>$	
b				\doteq	$<$	$>$	
\perp	$<$	$<$		$<$			

Рис. 4.4. Матрица предшествования

Краткое правило для продукции вида $A \rightarrow \alpha XY\beta$:

1) $X \doteq Y$,

2) если $Y \in V_N$, то $X < L(Y)$,

3) если $X \in V_N$, то

а) если $Y \in V_T$, то $R(X) > Y$,

б) если $Y \in V_N$, то $R(X) >$ со всеми терминалами из $L(Y)$.

Рассмотрим работу анализатора на примере разбора строки $aabb\perp$, которая выводится в соответствии со следующей правосторонней схемой (символ \perp начала строки опущен)

$$S\perp \Rightarrow DB\perp \Rightarrow DbB\perp \Rightarrow Dbb\perp \Rightarrow Abb\perp \Rightarrow aAbb\perp \Rightarrow aabb\perp.$$

Процесс разбора показан на рис. 4.5. Содержимое стека представлено строкой, в которой для наглядности между каждой парой соседних символов указано отношение предшествования (реально стек содержит только символы грамматики).

	S	D	A	B	a	b	\perp
S							
D			\doteq		$<$		
A					$>$		
B						$>$	
a			\doteq		$<$	$>$	
b			\doteq		$<$	$>$	
\perp	$<$	$<$	$<$				

Входной буфер	Содержимое стека	Основа	Выполняемое действие
$aabb\perp$	\perp		Перенос a в стек, т. к. $\perp < a$
$abb\perp$	$\perp < a$		Перенос a в стек, т. к. $a < a$
$bb\perp$	$\perp < a < a$	a	Свертка для $A \rightarrow a$, т. к. $a > b$
$bb\perp$	$\perp < a \doteq A$	aA	Свертка для $A \rightarrow aA$, т. к. $A > b$
$bb\perp$	$\perp < A$	A	Свертка для $D \rightarrow A$, т. к. $A > b$
$bb\perp$	$\perp < D$		Перенос b в стек, т. к. $D < b$
$b\perp$	$\perp < D < b$		Перенос b в стек, т. к. $b < b$
\perp	$\perp < D < b < b$	b	Свертка для $B \rightarrow b$, т. к. $b > \perp$
\perp	$\perp < D < b \doteq B$	bB	Свертка для $B \rightarrow bB$, т. к. $B > \perp$
\perp	$\perp < D \doteq B$	DB	Свертка для $S \rightarrow DB$, т. к. $B > \perp$
\perp	$\perp \doteq S$		Разбор успешно завершен

Рис. 4.5. Процесс разбора строки $aabb\perp$