

Глава 4. Восходящий синтаксический анализ

4.2. Грамматики простого предшествования

4.2.4. Функции предшествования

Если грамматика состоит из n символов, то матрица предшествования имеет размер $n \times n$. Причем многие элементы будут иметь значение «пусто». Можно применить известные методы компактного хранения разреженных массивов, что обычно приводит к дополнительным временным затратам. Поэтому чаще вместо матриц предшествования синтаксические анализаторы используют так называемые *функции предшествования* f и g , отображающие символы грамматики в целые числа. При этом должны выполняться следующие соотношения:

$$f(X) = g(Y), \text{ если } X \doteq Y,$$

$$f(X) < g(Y), \text{ если } X < Y,$$

$$f(X) > g(Y), \text{ если } X > Y.$$

Однако не всякая матрица предшествования может иметь функции предшествования, хотя обычно на практике они существуют. Тогда вместо n^2 элементов достаточно хранить только $2n$ элементов.

Процедура построения функций предшествования по заданной матрице предшествования заключается в следующем.

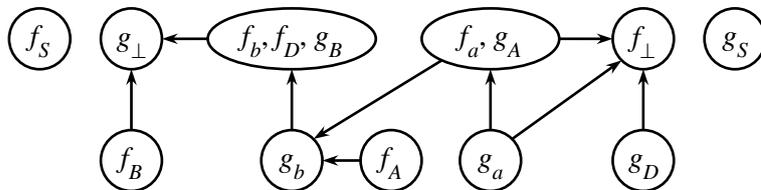
1. Создать символы f_X и g_X для каждого символа X грамматики, включая и символ \perp .

2. Сгруппировать созданные символы на как можно большее число групп таким образом, что если $X \doteq Y$, то f_X и g_Y должны входить в одну группу. Следует заметить, что в одну группу могут попасть символы, не связанные отношением \doteq . Например, если $a \doteq b$ и $c \doteq b$, то f_a и f_c должны находиться в одной и той же группе, поскольку оба находятся в той же группе, что и g_b . Если, кроме того, $c \doteq d$, то f_a и g_d находятся в одной группе, даже если не выполняется условие $a \doteq d$.

3. Создать ориентированный граф, вершины которого представляют собой определенные в п. 2 группы. Для всех X и Y , если $X < Y$, проводится дуга из группы, в которой находится g_Y , в группу с f_X . Если $X > Y$, дуга проводится из группы с f_X в группу с g_Y . Таким образом, дуга (путь) от группы с f_X к группе с g_Y означает, что $f(X)$ превосходит $g(Y)$; путь от группы с g_Y к группе с f_X означает, что $g(Y)$ должно превосходить $f(X)$.

4. Если построенный граф имеет циклы, то для заданной таблицы предшествования не существуют функции предшествования. Если циклов нет, то значением $f(X)$ является длина самого длинного пути, начинающегося в группе с f_X ; значение $g(X)$ равно длине самого длинного пути из группы с g_X .

В качестве примера определим функции предшествования для матрицы предшествования на рис. 4.4. Построенный граф показан на рис. 4.6. Поскольку $a \doteq A$, элементы f_a и g_A объединены в одну группу. Элементы f_b, f_D и g_B объединены в одну группу, т. к. $b \doteq B$ и $D \doteq B$, несмотря на то, что символы b и D не связаны никаким отношением предшествования.



	S	D	A	B	a	b	⊥
S							
D				=	<		
A						>	
B							>
a			=		<	>	
b				=		<	>
⊥		<	<		<		

Рис. 4.6. Граф отношений предшествования

Этот граф не имеет циклов, следовательно, функции предшествования существуют. Определим самые длинные пути от каждой вершины графа:

группы $\{f_S\}$, $\{g_{\perp}\}$, $\{f_{\perp}\}$ и $\{g_S\}$ не имеют исходящих дуг, т. е. длины путей равны 0,

$\{f_A\} \rightarrow \{g_b\} \rightarrow \{f_b, f_D, g_B\} \rightarrow \{g_{\perp}\}$ длины 3,

$\{f_B\} \rightarrow \{g_{\perp}\}$ длины 1,

$\{f_b, f_D, g_B\} \rightarrow \{g_{\perp}\}$ длины 1,

$\{f_a, g_A\} \rightarrow \{g_b\} \rightarrow \{f_b, f_D, g_B\} \rightarrow \{g_{\perp}\}$ длины 3,

$\{g_D\} \rightarrow \{f_{\perp}\}$ длины 1,

$\{g_a\} \rightarrow \{f_a, g_A\} \rightarrow \{g_b\} \rightarrow \{f_b, f_D, g_B\} \rightarrow \{g_{\perp}\}$ длины 4,

$\{g_b\} \rightarrow \{f_b, f_D, g_B\} \rightarrow \{g_{\perp}\}$ длины 2.

Полученные длины путей и будут являться значениями соответствующих функций предшествования. Например, поскольку группы $\{f_S\}$, $\{g_\perp\}$, $\{f_\perp\}$ и $\{g_S\}$ не имеют исходящих дуг (длины путей равны 0), $f(S) = g(\perp) = f(\perp) = g(S) = 0$. Самый длинный путь из $\{f_a, g_A\}$ имеет длину 3, поэтому $f(a) = g(A) = 3$. Аналогично определяются остальные значения функций. Результат построения функций предшествования выглядит следующим образом:

	S	D	A	B	a	b	\perp
f	0	1	3	1	3	1	0
g	0	1	3	1	4	2	0

Недостатком использования функций предшествования является то, что теряется информация о несуществующих отношениях, которые могут быть использованы для обнаружения синтаксических ошибок. На практике потеря возможности обнаружения ошибок не считается достаточно серьезной, поскольку в большинстве случаев они могут быть обнаружены при попытке выполнения свертки для необнаруженной основы, но на более поздних этапах разбора.

В качестве примера обнаружения ошибки рассмотрим разбор строки $aba\perp$, не принадлежащей языку, порождаемому грамматикой. При использовании матрицы предшествования (рис. 4.7) ошибка фиксируется на более раннем этапе разбора по несуществующему отношению для символов b и a . При использовании функций предшествования (рис. 4.8) ошибка обнаруживается на более позднем этапе при попытке выполнения операции свертки подстроки bD , поскольку в грамматике нет продукции с данной правой частью, т. е. основа не обнаружена.

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

Входной буфер	Содержимое стека	Основа	Выполняемое действие
$aba\perp$	\perp		Перенос a в стек, т. к. $\perp < a$
$ba\perp$	$\perp < a$	a	Свертка для $A \rightarrow a$, т. к. $a > b$
$ba\perp$	$\perp < A$	A	Свертка для $D \rightarrow A$, т. к. $A > b$
$ba\perp$	$\perp < D$		Перенос b в стек, т. к. $D < b$
$a\perp$	$\perp < D < b$		Синтаксическая ошибка, т. к. для символов b и a не существует отношение предшествования

Рис. 4.7. Процесс разбора строки $aba\perp$ по матрице предшествования

	S	D	A	B	a	b	\perp
f	0	1	3	1	3	1	0
g	0	1	3	1	4	2	0

Входной буфер	Содержимое стека	Основа	Выполняемое действие
$aba\perp$	\perp		Перенос a в стек, т. к. $f(\perp) < g(a)$
$ba\perp$	$\perp < a$	a	Свертка для $A \rightarrow a$, т. к. $f(a) > g(b)$
$ba\perp$	$\perp < A$	A	Свертка для $D \rightarrow A$, т. к. $f(A) > g(b)$
$ba\perp$	$\perp < D$		Перенос b в стек, т. к. $f(D) < g(b)$
$a\perp$	$\perp < D < b$		Перенос a в стек, т. к. $f(b) < g(a)$
\perp	$\perp < D < b < a$	a	Свертка для $A \rightarrow a$, т. к. $f(a) > g(\perp)$
\perp	$\perp < D < b < A$	A	Свертка для $D \rightarrow A$, т. к. $f(A) > g(\perp)$
\perp	$\perp < D < b = D$	bD	Попытка свертки подстроки bD , т. к. $f(D) > g(\perp)$. Синтаксическая ошибка, поскольку нет продукции с правой частью bD

Рис. 4.8. Процесс разбора строки $aba\perp$ по функциям предшествования