Глава 3. Нисходящий синтаксический анализ

3.4. *LL*(1)-грамматики

3.4.3. LL(1)-грамматики

LL(1)-грамматика является обобщением *q*-грамматик, принцип обобщения позволяет строить нисходящие детерминированные анализаторы.

Пусть $G = (V_T, V_N, P, S)$ – произвольная КС-грамматика. Определим функции *First* и *Follow*.

$$First(X) = \{a \mid X \overset{*}{\Rightarrow} a\beta\},\ X \in (V_T \cup V_N),\ a \in V_T,\ \beta \in (V_T \cup V_N)^*,$$

т. е. функция $First(X)$ определяет множество терминалов, с которых может начинаться строка, выводимая из символа X.

Расширим понятие функции *First* на строку символов $\alpha = X_1X_2...X_n$, где $X_i \in V_T \cup V_N$, $1 \le i \le n$:

$$First(X_1X_2...X_n) = \bigcup_{i=1}^{n} First(X_i) \mid X_1X_2...X_{i-1} \overset{*}{\Rightarrow} \varepsilon,$$

т. е. определяет множество терминалов, с которых может начинаться строка, выводимая из строки $\alpha = X_1X_2...X_n$, учитывая наличие ε-порождающих нетерминалов. Таким образом, сначала в множество $First(\alpha) = First(X_1X_2...X_n)$, добавляется множество $First(X_1)$, затем, если X_1 не порождает ε, то процесс вычисления завершается, если же $X_1 \overset{*}{\Rightarrow} \varepsilon$ (т. е. X_1 – ε-порождающий нетерминал), в множество $First(\alpha)$ добавляются элементы множества $First(X_2)$ и т. д.

Очевидно, что если строка $\alpha = a\beta$, где $a \in V_T$, $\beta \in (V_T \cup V_N)^*$, т. е. строка α начинается с терминала, то $First(a\beta) = \{a\}$, а также, что $First(\varepsilon) = \varnothing$. Из определения следует, что для нетерминала A с A-продукциями $A \to \alpha_1 | \alpha_2 | \ldots | \alpha_k$, $\alpha \in (V_T \cup V_N)^+$, $\alpha_i \in (V_T \cup V_N)^*$, $1 \le i \le k$

$First(A) = First(\alpha_1) \cup First(\alpha_2) \cup \ldots \cup First(\alpha_k)$.

Аргументом функции *Follow* является нетерминал $X \in V_N$:

$$Follow(X) = \{a \mid S \overset{*}{\Rightarrow} \alpha X a \beta\},\ a \in V_T,\ \alpha \in V_T^{*},\ \beta \in (V_T \cup V_N)^{*}.$$

Данная функция (как и в q-грамматике) определяет множество терминалов, которые могут следовать непосредственно за нетерминалом X в какой-либо сентенциальной форме, выводимой из начального нетерминала S. Из определения следует, что если существует продукция вида $A \to \alpha X \gamma$, $\alpha, \gamma \in (V_T \cup V_N)^{*}$, то

$$Follow(X) = First(\gamma) \cup Follow(A) \mid \gamma \overset{*}{\Rightarrow} \varepsilon.$$

Тогда множество DS направляющих символов продукций $LL(1)$-грамматики определяется следующим образом:

$DS(A \to \alpha) = First(\alpha) \cup Follow(A)$, если $\alpha \overset{*}{\Rightarrow} \varepsilon$,

$DS(A \to \alpha) = First(\alpha)$ в противном случае.

Другими словами, если правая часть продукции может генерировать пустую строку, то к элементам множества $First(\alpha)$ необходимо добавить элементы множества $Follow(A)$ для нетерминала из левой части продукции.

КС-грамматика $G = (V_T, V_N, P, S)$ называется $LL(1)$-*грамматикой* (*обладает* $LL(1)$-*свойствами*) тогда и только тогда, когда для любой пары несовпадающих альтернативных A-продукций вида $A \to \alpha$ и $A \to \beta$ справедливо утверждение, что

$$DS(A \to \alpha) \cap DS(A \to \beta) = \varnothing,$$

т. е. множества направляющих символов альтернативных продукций не пересекаются.

Из определения следует, что леворекурсивная грамматика не может быть $LL(1)$-грамматикой. Другим важным свойством $LL(1)$-грамматики является ее однозначность.

Существуют формальные алгоритмы определения, относится ли заданная КС-грамматика к классу $LL(1)$-грамматик, основанные на вычислении соответствующих функций и определения множеств направляющих символов через построение бинарных отношений, их транзитивных замыканий и вычислении их произведений [7; 10]. Эти алгоритмы могут быть полезны при создании программного инструментария для работы с грамматиками.

Рассмотрим грамматику с продукциями (в предположении, что имеется продукция $S' \to S\perp$ с маркером конца ввода $\perp$):

$S \to AB \mid DEa$ $\quad$ $C \to bC \mid \varepsilon$

$A \to ac \mid f$ $\quad$ $D \to dD \mid \varepsilon$

$B \to bC$ $\quad$ $E \to eE \mid \varepsilon$

Определим сначала множество всех ε-порождающих нетерминалов $N_\varepsilon = \{C, D, E\}$.

Вычислим значение функции $First(A)$ для всех нетерминалов $A \in V_N$ (для терминалов значениями функции являются сами терминалы):

$First(S) = \{a, d, e, f\}$, $\quad$ $First(C) = \{b\}$,

$First(A) = \{a, f\}$, $\quad$ $First(D) = \{d\}$,

$First(B) = \{b\}$, $\quad$ $First(E) = \{e\}$.

Рассмотрим подробнее вычисление $First(S)$, остальные значения вычисляются аналогично. В соответствии с продукциями $S \to AB \mid DEa$ имеем

$First(S) = First(AB) \cup First(DEa)$.

Поскольку $A \notin N_\varepsilon$ (нетерминал A не порождает пустую строку),

$First(AB) = First(A)$.

В соответствии с продукциями $A \to ac \mid f$

$First(A) = First(ac) \cup First(f) = \{a\} \cup \{f\} = \{a, f\}$.

Так как $D, E \in N_\varepsilon$ (являются ε-порождающими нетерминалами),

$First(DEa) = First(D) \cup First(E) \cup First(a)$.

В соответствии с продукциями $D \to dD \mid \varepsilon$

$First(D) = First(dD) \cup First(\varepsilon) = \{d\} \cup \varnothing = \{d\}$.

В соответствии с продукциями $E \to eE \mid \varepsilon$

$First(E) = First(eE) \cup First(\varepsilon) = \{e\} \cup \varnothing = \{e\}$.

Таким образом,

$First(DEa) = \{d\} \cup \{e\} \cup \{a\} = \{a, d, e\}$.

В итоге получаем

$First(S) = First(AB) \cup First(DEa) = \{a, f\} \cup \{a, d, e\} = \{a, d, e, f\}$.

Вычислим значение функции *Follow*. Следует отметить, что вычисление этой функции требуется только для ε-порождающих нетерминалов. Но для учебных целей вычислим для всех нетерминалов. Вычисление выполняется в соответствии со сформулированным выше правилом: если существует продукция вида $A \to \alpha X\gamma$, $\alpha, \gamma \in (V_T \cup V_N)^*$, то $Follow(X) = First(\gamma) \cup Follow(A) \mid \gamma \overset{*}{\Rightarrow} \varepsilon$.

$Follow(S) = \{\perp\}$, поскольку предполагается наличие продукции $S' \to S\perp$ и $First(\perp) = \{\perp\}$.

$Follow(A) = \{b\}$, поскольку A встречается только в правой части продукции $S \to AB$, а $First(B) = \{b\}$, причем $B \notin N_\varepsilon$.

$Follow(B) = \{\perp\}$, поскольку B встречается только в правой части продукции $S \to AB$, а $Follow(S) = \{\perp\}$.

$Follow(C) = \{\perp\}$, поскольку C встречается в правых частях продукций $B \to bC$ и $C \to bC$, а $Follow(B) = \{\perp\}$, праворекурсивная продукция $C \to bC$ не влияет на результат, поскольку для нее согласно правилу имеет место $Follow(C) = Follow(C)$.

$Follow(D) = \{a, e\}$, поскольку D встречается только в правой части продукции $S \to DEa$ (из тех же соображений, что и при вычислении $Follow(C)$, не учитываем праворекурсивную продукцию $D \to dD$), а $First(Ea) = First(E) \cup First(a) = \{a, e\}$, так как $E \in N_\varepsilon$, а строка Ea не порождает пустую строку.

$Follow(E) = \{a\}$, поскольку E встречается только в правой части продукции $S \to DEa$ (из тех же соображений, что и при вычислении $Follow(C)$, не учитываем праворекурсивную продукцию $E \to eE$), а $First(a) = \{a\}$.

Вычислим множества направляющих символов DS (в учебных целях вычислим для всех продукций, хотя по определению достаточно вычисления только для альтернативных продукций):

$DS(S \to AB) = First(AB) = First(A) = \{a, f\}$,

$DS(S \to DEa) = First(D) \cup First(E) \cup First(a) = \{a, d, e\}$,

$DS(A \to ac) = First(ac) = \{a\}$,

$DS(A \to f) = First(f) = \{f\}$,

$DS(B \to bC) = First(bC) = \{b\}$,

$DS(C \to bC) = First(bC) = \{b\}$,

$DS(C \to \varepsilon) = Follow(C) = \{\perp\}$,

$DS(D \to dD) = First(dD) = \{d\}$,

$DS(D \to \varepsilon) = Follow(D) = \{a, e\}$,

$DS(E \to eE) = First(eE) = \{e\}$,

$DS(E \to \varepsilon) = Follow(E) = \{a\}$.

Таким образом, грамматика не является $LL(1)$-грамматикой, поскольку для S-продукций $S \to AB$ и $S \to DEa$ множества направляющих символов пересекаются:

$DS(S \to AB) \cap DS(S \to DEa) = \{a, f\} \cap \{a, d, e\} = \{a\}$.