

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 $Follow$.

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

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

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

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

т. е. определяет множество терминалов, с которых может начинаться строка, выводимая из строки $\alpha = X_1X_2\dots X_n$, учитывая наличие ε -порождающих нетерминалов. Таким образом, сначала в множество $First(\alpha) = First(X_1X_2\dots X_n)$, добавляется множество $First(X_1)$, затем, если X_1 не порождает ε , то процесс вычисления завершается, если же $X_1 \xRightarrow{*} \varepsilon$ (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 $First(a\beta) = \{a\}$. Очевидно также, что $First(\varepsilon) = \emptyset$. Из определения следует, что для нетерминала A с A -продукциями $A \rightarrow \alpha_1 | \alpha_2 | \dots | \alpha_k$, $\alpha \in (V_T \cup V_N)^+$, $\alpha_i \in (V_T \cup V_N)^*$, $1 \leq i \leq k$

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

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

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

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 $A \rightarrow \alpha X \gamma$, $\alpha, \gamma \in (V_T \cup V_N)^*$, то

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

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

$$DS(A \rightarrow \alpha) = First(\alpha) \cup Follow(A), \text{ если } \alpha \xrightarrow{*} \varepsilon,$$

$$DS(A \rightarrow \alpha) = First(\alpha) \text{ в противном случае.}$$

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.

Контекстно-свободная 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 $A \rightarrow \alpha$ и $A \rightarrow \beta$ справедливо утверждение, что

$$DS(A \rightarrow \alpha) \cap DS(A \rightarrow \beta) = \emptyset,$$

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 вычисления 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 $S' \rightarrow S\perp$ с маркером конца ввода \perp):

$$\begin{array}{ll} S \rightarrow AB \mid DEa & C \rightarrow bC \mid \varepsilon \\ A \rightarrow ac \mid f & D \rightarrow dD \mid \varepsilon \\ B \rightarrow bC & E \rightarrow eE \mid \varepsilon \end{array}$$

Определим сначала множество всех ε -порождающих нетерминалов $N_\varepsilon = \{C, D, E\}$ (см. разд. 1.6.6).

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

$$\begin{array}{ll} First(S) = \{a, d, e, f\}, & First(C) = \{b\}, \\ First(A) = \{a, f\}, & First(D) = \{d\}, \\ First(B) = \{b\}, & First(E) = \{e\}. \end{array}$$

Рассмотрим подробнее вычисление $First(S)$, остальные значения вычисляются аналогично. В соответствии с продукциями $S \rightarrow AB \mid DEa$ 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 $A \rightarrow ac \mid f$

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

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 $D \rightarrow dD \mid \varepsilon$

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

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

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

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 заметить, 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 $A \rightarrow \alpha X \gamma$, $\alpha, \gamma \in (V_T \cup V_N)^*$, то

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

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

$$Follow(A) = \{b\}, \text{ поскольку } A \text{ встречается только в правой части продукции } S \rightarrow AB, \text{ а } First(B) = \{b\},$$

причем $B \notin N_\varepsilon$.

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

$Follow(C) = \{\perp\}$, поскольку C встречается в правых частях продукций $B \rightarrow bC$ и $C \rightarrow bC$, а $Follow(B) = \{\perp\}$, праворекурсивная продукция $C \rightarrow bC$ 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 $S \rightarrow DEa$ (из тех же соображений, что и при вычислении $Follow(C)$, не учитываем праворекурсивную продукцию $D \rightarrow dD$), 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 $S \rightarrow DEa$ (из тех же соображений, что и при вычислении $Follow(C)$, не учитываем праворекурсивную продукцию $E \rightarrow eE$), а $First(a) = \{a\}$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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