

Глава 3. Нисходящий синтаксический анализ

3.4. $LL(1)$ -грамматики

Прежде чем определить $LL(1)$ -грамматику, рассмотрим некоторые подклассы $LL(1)$ -грамматик и вопросы их разбора.

3.4.1. Разделенные грамматики (*s*-грамматики)

Разделенная, или *простая*, грамматика (*s*-грамматика) представляет собой грамматику, в которой правая часть каждой продукции начинается с терминального символа, за терминалом могут следовать нетерминалы и/или терминалы; правые части альтернативных продукций начинаются с разных терминалов.

Эти условия позволяют построить детерминированный нисходящий синтаксический анализатор, т. к. при выводе строки языка всегда можно сделать однозначный выбор между альтернативными продукциями для самого левого нетерминала в сентенциальной форме, предварительно исследовав один следующий символ входной строки.

Например, грамматика с продукциями

$$S \rightarrow pA\perp \mid qB\perp$$

$$A \rightarrow a \mid cAd$$

$$B \rightarrow b \mid cBg$$

представляет собой *s*-грамматику.

Рассмотрим разбор строки $pccadd\perp$. Разбор начинается с начального символа грамматики S . На первом шаге из двух альтернативных S -продукций выбираем продукцию $S \rightarrow pA\perp$, поскольку входным символом является символ p . В результате получим сентенциальную форму $pA\perp$. На втором шаге анализируется второй символ входной строки, поскольку это символ c , из A -продукций выбираем продукцию $A \rightarrow cAd$ и получаем сентенциальную форму $pcAd\perp$. Продолжая аналогичные действия, получим левостороннюю схему вывода

$$S \Rightarrow pA\perp \Rightarrow pcAd\perp \Rightarrow pccAdd\perp \Rightarrow pccadd\perp.$$

3.4.2. Слаборазделенные грамматики (*q*-грамматики)

Пусть дана следующая грамматика:

$$S \rightarrow pA\perp \mid qBA\perp$$

$$A \rightarrow a \mid bAd$$

$$B \rightarrow \varepsilon \mid cBg$$

Эта грамматика не принадлежит классу *s*-грамматик, поскольку содержит ε -продукцию $B \rightarrow \varepsilon$, правая часть которой не начинается с терминала. Выбрать при разборе продукцию по левому терминалу правой части не удастся.

Пусть $Follow(X)$ – множество терминалов, которые могут следовать непосредственно за нетерминалом X в какой-либо сентенциальной форме, выводимой из начального нетерминала. Например, для приведенной грамматики $Follow(B) = \{a, b, g\}$.

Критерием выбора между альтернативными продукциями является *множество направляющих символов (множество выбора) DS*, определяемое следующим образом:

Если продукция имеет вид $A \rightarrow b\alpha$, где $A \in V_N$, $b \in V_T$, $\alpha \in (V_T \cup V_N)^*$, то $DS(A \rightarrow b\alpha) = \{b\}$. Если продукция имеет вид $A \rightarrow \varepsilon$, то $DS(A \rightarrow \varepsilon) = Follow(A)$.

Контекстно-свободная грамматика называется *слаборазделенной (q-грамматикой)* при условии:

- правая часть каждой продукции либо начинается с терминала, либо представляет собой ε ;
- множества направляющих символов альтернативных продукций не пересекаются.

Таким образом, приведенная выше грамматика относится к классу *q*-грамматик, т. к. множества DS альтернативных продукций не пересекаются. Это позволяет при разборе строки детерминированно выбирать нужную продукцию из альтернативных.

Процесс вывода строки $qbad\perp$ соответствует следующей левосторонней схеме вывода:

$$S \Rightarrow qBA\perp \Rightarrow qA\perp \Rightarrow qbad\perp \Rightarrow qbad\perp.$$

На втором шаге вывода, поскольку $b \in DS(B \rightarrow \varepsilon) = \{a, b, g\}$, применена продукция $B \rightarrow \varepsilon$.