Теория автоматов и формальных языков

Контекстно-свободные языки: нисходящий анализ

Лектор: Екатерина Вербицкая

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»

23 ноября 2021

В предыдущей серии

- Нисходящий анализ
- Алгоритм синтаксического анализа LL(1)

Когда LL-анализ не возможен

- Леворекурсивные правила
- Когда при построении таблицы в одну ячейку нужно записать больше одной записи
  - FIRST-FIRST конфликт
    - $A \to \alpha \mid \beta, FIRST(\alpha) \cap FIRST(\beta) \neq \varnothing$
    - $E \to T + E \mid T * E$
  - FIRST-FOLLOW конфликт
    - $FIRST(A) \cap FOLLOW(A) \neq \varnothing$
    - $S \to Aab, A \to a \mid \varepsilon$
- Как с этим бороться?
  - Избавиться от левой рекурсии
  - Избавиться от недетерминизма
  - Факторизовать грамматику
  - Использовать аннотации (если есть)
  - Переписать грамматику
  - Использовать более одного символа предпросмотра

Леворекурсивные правила грамматики

- Явная (непосредственная) левая рекурсия
  - $A \to A\beta$
- Неявная левая рекурсия
  - $A \to \alpha A\beta,\ \alpha \overset{*}{\Rightarrow} \varepsilon$
- Взаимная рекурсия
  - $A \to \alpha B\beta,\ B \to \gamma A\delta,\ \alpha \overset{*}{\Rightarrow} \varepsilon, \gamma \overset{*}{\Rightarrow} \varepsilon$

Избавление от левой рекурсии

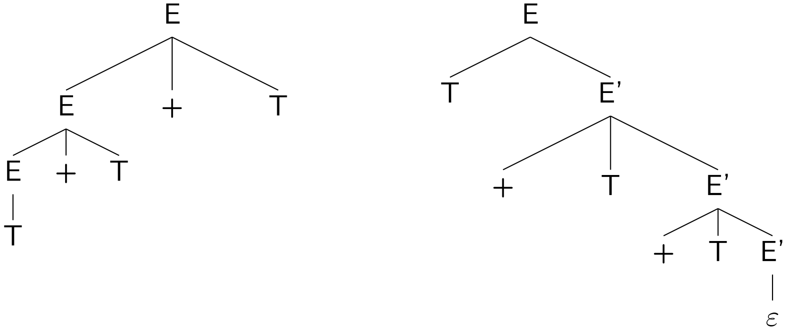
- $A \to A\alpha \mid \beta \Leftrightarrow A \to \beta A'$, $A' \to \varepsilon \mid \alpha A'$

Избавление от левой рекурсии

- $A \to A\alpha \mid \beta \Leftrightarrow A \to \beta A',\ A' \to \varepsilon \mid \alpha A'$
- $E \to E + T \mid T \Leftrightarrow E \to TE',\ E' \to \varepsilon \mid +TE'$

Избавление от левой рекурсии

- $A \to A\alpha \mid \beta \Leftrightarrow A \to \beta A',\ A' \to \varepsilon \mid \alpha A'$
- $E \to E + T \mid T \Leftrightarrow E \to TE',\ E' \to \varepsilon \mid +TE'$



Избавление от левой рекурсии: более общий случай

- $A \to A\alpha_1 \mid A\alpha_2 \mid \cdots \mid A\alpha_n \mid \beta_1 \mid \beta_2 \mid \cdots \mid \beta_k$
- $A \to \beta_1 A' \mid \beta_2 A' \mid \cdots \mid \beta_k A'$
- $A' \to \varepsilon \mid \alpha_1 A' \mid \alpha_2 A' \mid \cdots \mid \alpha_n A'$

Избавление от взаимной левой рекурсии

- Избавляемся от ε-продукций
- Упорядочиваем правила по индексу нетерминала
- Добиваемся того, чтобы не было правил вида $A_i \to A_j\alpha, j \leq i$
  - Перебираем все A_i
  - Перебираем все $A_j, 1 \leq j < i$
  - Для каждого правила $p : A_i \to A_j\gamma$
    - Удалить правило p
    - Для каждого правила $A_j \to \delta_1 \mid \cdots \mid \delta_k$ Добавить правила $A_i \to \delta_l$
  - Устранить непосредственную левую рекурсию для A_i

Левая факторизация грамматики

- Выделяем наибольший общий префикс продукций
  $A \to \alpha\beta \mid \alpha\gamma \Rightarrow A \to \alpha A',\ A' \to \beta \mid \gamma$

Пример

$$
\begin{aligned}
S &\rightarrow aSSbS \\
&\mid aSaSb \\
&\mid abb \\
&\mid b
\end{aligned}
$$

Пример

$$
\begin{array}{rcl}
S & \rightarrow & aSSbS \\
 & | & aSaSb \\
 & | & abb \\
 & | & b
\end{array}
$$

$$
\begin{array}{rcl}
S & \rightarrow & aS' \\
 & | & b \\
S' & \rightarrow & SSbS \\
 & | & SaSb \\
 & | & bb
\end{array}
$$

Пример

$$\begin{array}{rcl} S & \to & aSSbS \\ & | & aSaSb \\ & | & abb \\ & | & b \end{array}$$

$$\begin{array}{rcl} S & \to & aS' \\ & | & b \\ S' & \to & SSbS \\ & | & SaSb \\ & | & bb \end{array}$$

$$\begin{array}{rcl} S & \to & aS' \mid b \\ S' & \to & SS'' \mid bb \\ S'' & \to & SbS \mid aSb \end{array}$$

LL(k)-анализ

- Можно использовать более одного символа предпросмотра
- Все равно применимо не ко всем КС-грамматикам

LL(k): функция *FIRST*

- Функция $FIRST_k^G(\alpha) =$ $\{\omega \in V_T^* \mid \text{либо } |\omega| < k \text{ и } \alpha \overset{*}{\Rightarrow} \omega, \text{ либо } |\omega| = k \text{ и } \alpha \overset{*}{\Rightarrow} \omega\gamma, \gamma \in V_T^*\}$
  - По сути: первые k символов, встречающиеся в выводе из α
- Пример
  - $S \rightarrow SS \mid aSb \mid \varepsilon$
  - $FIRST_3^G(aSb) = \{ab, aab, aaa\}$
  - $aba \notin FIRST_3^G(aSb)$!

LL(k): функция FOLLOW

$$FOLLOW_k^G(\beta) = \{\omega \in V_T^* \mid S \overset{*}{\Rightarrow} \gamma\beta\alpha, \omega \in FIRST_k^G(\alpha)\}, k \geq 0$$

Пример: $S \to SS \mid aSb \mid \varepsilon$

- $FOLLOW_3^G(aa) = \{abb, aab, aaa, aba, baa, bab, bb, bba, \dots\}$
- $\varepsilon, b \notin FOLLOW_3^G$!

Нисходящий синтаксический анализ: LL-грамматики

Фундаментальное свойство: по сентенциальной форме $a_1 a_2 \dots a_j A\beta, a_i \in V_T, A \in V_N, \beta \in (V_T \cup V_N)^*$ однозначно определяется, какое правило нужно применять дальше, чтобы разобрать всю строку

Нисходящий синтаксический анализ: LL-грамматики

Фундаментальное свойство: по сентенциальной форме $a_1 a_2 \dots a_j A\beta, a_i \in V_T, A \in V_N, \beta \in (V_T \cup V_N)^*$ однозначно определяется, какое правило нужно применять дальше, чтобы разобрать всю строку

КС грамматика G является **LL(k)-грамматикой** для некоторого k, если для любых двух левосторонних выводов вида

- $S \overset{*}{\Rightarrow} \omega A\alpha \Rightarrow \omega\beta\alpha \overset{*}{\Rightarrow} \omega\delta$
- $S \overset{*}{\Rightarrow} \omega A\alpha \Rightarrow \omega\gamma\alpha \overset{*}{\Rightarrow} \omega\eta$

в которых $FIRST_k^G(\delta) = FIRST_k^G(\eta)$, верно $\beta = \gamma$

КС грамматика G является **LL-грамматикой**, если она является $LL(k)$-грамматикой для некоторого $k \geq 0$

Пример LL(1)-грамматики

$S \to aBS \mid b$
$B \to a \mid bSB$
Надо показать: для любых левосторонних выводов

- $S \overset{*}{\Rightarrow} \omega A\alpha \Rightarrow \omega\beta\alpha \overset{*}{\Rightarrow} \omega\delta$
- $S \overset{*}{\Rightarrow} \omega A\alpha \Rightarrow \omega\gamma\alpha \overset{*}{\Rightarrow} \omega\eta$

если δ и η начинаются с одного символа, то $\beta = \gamma$
Рассматриваем выводы, где роль A выполняет S: $S \Rightarrow aBS, S \Rightarrow b$.
$\omega = \alpha = \varepsilon, \beta = aBS, \gamma = b$. Любая цепочка, выводимая из $\beta\alpha = aBS$ начинается на a; любая цепочка, выводимая из $\gamma\alpha = b$ начинается на b. Однозначно определяется, какой альтернативе следовать.
Аналогично с $A = B : S \Rightarrow aBS \Rightarrow aaS; S \Rightarrow aBS \Rightarrow abSBS$

Простая LL(1)-грамматика

КС-грамматика G называется **простой LL(1)-грамматикой**, если в ней нет ε-правил, и все альтернативы для каждого нертерминала начинаются с терминалов, и притом различных.

$\forall (A, a), A \in V_N, a \in V_T, \exists$ максимум 1 альтернатива вида $A \to a\alpha$

LL-грамматика: необходимое и достаточное условие

Теорема

КС грамматика $G = \langle V_N, V_T, P, S\rangle$ *является* $LL(k)$*-грамматикой*
$\Leftrightarrow FIRST_k^G(\beta\alpha) \cap FIRST_k^G(\gamma\alpha) = \varnothing$, *для всех таких*
$\alpha, \beta, \gamma : A \to \beta, A \to \gamma \in P, \beta \neq \gamma, \exists$ *вывод* $S \overset{*}{\Rightarrow} \omega A\alpha$

LL(1)-грамматика: необходимое и достаточное условие

Теорема

КС-грамматика $G = \langle V_N, V_T, P, S \rangle$ *является* $LL(1)$*-грамматикой* $\Leftrightarrow FIRST_1^G(\beta FOLLOW_1^G(A)) \cap FIRST_1^G(\gamma FOLLOW_1^G(A)) = \varnothing, \forall A \in V_N, \beta, \gamma \in (V_N \cup V_T)^*, A \to \gamma, A \to \beta \in P, \beta \neq \gamma$

LL(1)-грамматика: необходимое и достаточное условие: другая формулировка

Теорема

КС-грамматика $G = \langle V_N, V_T, P, S \rangle$ *является* $LL(1)$*-грамматикой* $\Leftrightarrow \forall A \to \alpha_1 \mid \alpha_2 \mid \cdots \mid \alpha_n$ *верно:*

- $FIRST_1^G(\alpha_i) \cap FIRST_1^G(\alpha_j) = \varnothing, i \neq j, 1 \leq i, j \leq n$
- *если* $\alpha_i \overset{*}{\Rightarrow} \varepsilon$, *то* $FIRST_1^G(\alpha_j) \cap FOLLOW_1^G(A) = \varnothing, 1 \leq j \leq n, i \neq j$

Леворекурсивность

Теорема

Если КС-грамматика $G = \langle V_N, V_T, P, S \rangle$ *леворекурсивна, то она не является* $LL(k)$*-грамматикой ни при каком* k