Теория автоматов и формальных языков

Контекстно-свободные языки

Лектор: Екатерина Вербицкая

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»

11 октября 2016г.

В предыдущей серии

- Контекстно-свободные грамматики (все правила имеют вид $A \to \alpha$)
- КС языки и разрешимость проверки пустоты
- Нормальная форма Хомского
- Алгоритм CYK

В предыдущей серии: НФХ

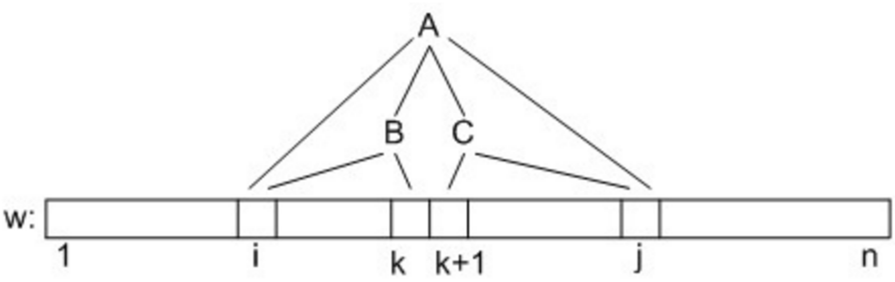
КС грамматика находится в **нормальной форме Хомского**, если все ее правила имеют вид:

- $A \to BC$, где $A, B, C \in V_N$
- $A \to a$, где $A \in V_N, a \in V_T$
- $S \to \varepsilon$, если в языке есть пустое слово, где S — стартовый нетерминал

1. Удалить стартовый нетерминал из правых частей правил
2. Избавиться от неодиночных терминалов в правых частях
3. Удалить длинные правила (длины больше 2)
4. Удалить непродуктивные правила (ε-правила)
5. Удалить цепные правила

В предыдущей серии: CYK

- Алгоритм синтаксического анализа, работающий с грамматиками в НФХ
- Динамическое программирование



В предыдущей серии: CYK

- Дано: строка ω длины n, грамматика $G = \langle V_T, V_N, P, S \rangle$ в НФХ
- Используем трехмерный массив d булевых значений размером $|V_N| \times n \times n$, $d[A][i][j] = true \Leftrightarrow A \overset{*}{\Rightarrow} \omega[i \dots j]$
- Инициализация: $i = j$
  - $d[A][i][i] = true$, если в грамматике есть правило $A \to \omega[i]$
  - $d[A][i][i] = false$, иначе
- Динамика. Предполагаем, d построен для всех нетерминалов и пар $\{(i', j') \mid j' - i' < m\}$
  - $d[A][i][j] = \bigvee_{A \to BC} \bigvee_{k=i}^{j-1} d[B][i][k] \wedge d[C][k+1][j]$
- В конце работы алгоритма в $d[S][0][n]$ записан ответ, выводится ли ω в данной грамматике

Восходящий синтаксический анализ

- Начинаем с символов входной строки, строим дерево вывода до стартового нетерминала
- CYK — один из примеров восходящего синтаксического анализа
- Контринтуитивен

Нисходящий синтаксический анализ

- Хотим построить левосторонний вывод строки
- Начинаем со стартового нетерминала, раскрываем нетерминалы до тех пор, пока не получим вывод строки
- Интуитивен

Нисходящий синтаксический анализ: функция *FIRST*

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

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

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

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.

$\forall (A, a), A \in V_N, a \in V_T, \exists$ самое большое 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* $S \overset{*}{\Rightarrow} \omega A\alpha$

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

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

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

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* $\Leftrightarrow \forall A \to \alpha_1 \,|\, \alpha_2 \,|\, \ldots \,|\, \alpha_n$ *верно:*

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