Теория автоматов и формальных языков

Введение

Лектор: Екатерина Вербицкая

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»

14 сентября 2021

В предыдущей серии

- Формальные языки повсюду. Язык — множество строк над алфавитом
- Существует множество способов описать язык
- Задачи теории формальных языков
  - Как представить язык?
  - Какие есть характеристики у разных представлений языка?
  - Как определить, принадлежит ли строка данному языку?

Метаязык

- Язык, на котором дано описание языка
  - Естественный язык
  - Язык металингвистических формул Бэкуса (БНФ)
  - Синтаксические диаграммы
  - **Грамматики**
  - ...

Описание языка: формальная грамматика

- **Порождающая грамматика** G — это четверка $\langle V_T, V_N, P, S\rangle$
  - V_T — алфавит терминальных символов (терминалов)
  - V_N — алфавит нетерминальных символов (нетерминалов)
    - $V_T \cap V_N = \emptyset$
    - $V ::= V_T \cup V_N$
  - P — конечное множество правил вида $\alpha \to \beta$
    - $\alpha \in V^* V_N V^*$
    - $\beta \in V^*$
  - S — начальный нетерминал грамматики,
    - $S \in N$

Пример: язык чисел в двоичной системе счисления

$$V_T = \{0, 1, -\} V_N = \{S, N, A\}$$

$$\begin{array}{lcl} S & \rightarrow & 0 \\ S & \rightarrow & N \\ S & \rightarrow & -N \\ N & \rightarrow & 1A \\ A & \rightarrow & 0A \\ A & \rightarrow & 1A \\ A & \rightarrow & \varepsilon \end{array}$$

Пример: язык чисел в двоичной системе счисления

$$V_T = \{0, 1, -\} V_N = \{S, N, A\}$$

$$\begin{array}{rcl} S & \rightarrow & 0 \\ S & \rightarrow & N \\ S & \rightarrow & -N \\ N & \rightarrow & 1A \\ A & \rightarrow & 0A \\ A & \rightarrow & 1A \\ A & \rightarrow & \varepsilon \end{array}$$

$$\begin{array}{rcl} S & \rightarrow & 0 \mid N \mid -N \\ N & \rightarrow & 1A \\ A & \rightarrow & 0A \mid 1A \mid \varepsilon \end{array}$$

Пример: язык чисел в двоичной системе счисления

$$V_T = \{0, 1, -\} V_N = \{S, N, A\}$$

$$\begin{array}{lcl} S & \to & 0 \\ S & \to & N \\ S & \to & -N \\ N & \to & 1A \\ A & \to & 0A \\ A & \to & 1A \\ A & \to & \varepsilon \end{array}$$

$$\begin{array}{lcl} S & \to & 0 \mid N \mid -N \\ N & \to & 1A \\ A & \to & 0A \mid 1A \mid \varepsilon \end{array}$$

$$\begin{array}{lcl} S & \to & 0 \mid [-]N \\ N & \to & 1A \\ A & \to & (0 \mid 1)A \mid \varepsilon \end{array}$$

Отношение непосредственной выводимости

- $\alpha \to \beta \in P$
- $\gamma, \delta \in V^*$
- $\gamma\alpha\delta \Rightarrow \gamma\beta\delta$: $\gamma\beta\delta$ **непосредственно выводится** из $\gamma\alpha\delta$ при помощи правила $\alpha \to \beta$

Отношение непосредственной выводимости: пример

$$\begin{array}{rcl} S & \to & 0 \mid N \mid -N \\ N & \to & 1A \\ A & \to & 0A \mid 1A \mid \varepsilon \end{array}$$

$$S \Rightarrow -N$$

$$-N \Rightarrow -1A$$

$$-1A \Rightarrow -11A$$

Отношение выводимости

Отношение выводимости является рефлексивно-транзитивным замыканием отношения непосредственной выводимости

- $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_n \in V^*$
- $\alpha_0 \Rightarrow \alpha_1 \Rightarrow \alpha_2 \Rightarrow \cdots \Rightarrow \alpha_n$
- $\alpha_0 \overset{*}{\Rightarrow} \alpha_n$: α_n **выводится** из α_0

Отношение выводимости: пример

$$\begin{array}{rcl} S & \to & 0 \mid N \mid -N \\ N & \to & 1A \\ A & \to & 0A \mid 1A \mid \varepsilon \end{array}$$

$$S \Rightarrow -N \Rightarrow -1A \Rightarrow -11A \overset{*}{\Rightarrow} -1101A \Rightarrow -1101$$

Отношение выводимости: свойства

- Транзитивность:
  $\forall \alpha, \beta, \gamma \in V^* : \ \alpha \overset{*}{\Rightarrow} \beta, \beta \overset{*}{\Rightarrow} \gamma$ следовательно $\alpha \overset{*}{\Rightarrow} \gamma$
- Рефлексивность: $\forall \alpha \in V^* : \ \alpha \overset{*}{\Rightarrow} \alpha$
- $\alpha_0 \overset{+}{\Rightarrow} \alpha_n$: вывод использует хотя бы одно правило грамматики
- $\alpha_0 \overset{k}{\Rightarrow} \alpha_n$: вывод происходит за k шагов

Левосторонний вывод

На каждом шагу заменяем самый **левый** нетерминал

$$
\begin{array}{lcl}
S & \to & AA \mid s \\
A & \to & AA \mid Bb \mid a \\
B & \to & c \mid d
\end{array}
$$

$$\boldsymbol{S} \Rightarrow \boldsymbol{A}A \Rightarrow \boldsymbol{B}bA \Rightarrow cb\boldsymbol{A} \Rightarrow cb\boldsymbol{A}A \Rightarrow cba\boldsymbol{A} \Rightarrow cbaa$$

Аналогично определяется **правосторонний** вывод

Язык, порождаемый грамматикой $G = \langle V_T, V_N, P, S \rangle$

$$L(G) = \{\omega \in V_T^* \mid S \overset{*}{\Rightarrow} \omega\}$$

Эквивалентность грамматик

Грамматики G_1 и G_2 **эквивалентны**, если $L(G_1) = L(G_2)$

Грамматики G_1 и G_2 **эквивалентны**, если $L(G_1) = L(G_2)$

$$\begin{array}{rcl} V_T & = & \{0, 1, -\} \\ V_N & = & \{S, N, A\} \\ \\ S & \to & 0 \mid N \mid -N \\ N & \to & 1A \\ A & \to & 0A \mid 1A \mid \varepsilon \end{array}$$

$$\begin{array}{rcl} V_T & = & \{0, 1, -\} \\ V_N & = & \{S, A\} \\ \\ S & \to & 0 \mid 1A \mid -1A \\ A & \to & 0A \mid 1A \mid \varepsilon \end{array}$$

Контекстно-свободная грамматика

Контекстно-свободная грамматика — грамматика, все правила которой имеют вид $A \to \alpha, A \in V_N, \alpha \in V^*$

Дерево вывода

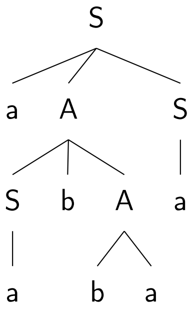
Дерево является **деревом вывода** для $G = \langle V_N, V_T, P, S \rangle$, если:

- Каждый узел помечен символом из алфавита V
- Метка корня — S
- Листья помечены терминалами, остальные узлы — нетерминалами
- Если узлы $n_0, \ldots, n_k$ — прямые потомки узла n, перечисленные слева направо, с метками $A_0, \ldots, A_k$; метка n — A, то $A \to A_0 \ldots A_k \in P$

Пример дерева вывода

$$G = \langle \{S, A\}, \{a, b\}, \{S \to aAS \mid a, A \to SbA \mid ba \mid SS\}, S \rangle$$

$$S \Rightarrow aAS \Rightarrow aSbAS \Rightarrow aabAS \Rightarrow aabbaS \Rightarrow aabbaa$$



Теорема

Пусть $G = \langle V_N, V_T, P, S \rangle$ *— КС-грамматика*
Вывод $S \overset{*}{\Rightarrow} \alpha$*, где* $\alpha \in V^*, \alpha \neq \varepsilon$ *существует* $\Leftrightarrow$ *существует дерево вывода в грамматике* G *с результатом* α

Упражнение: доказать теорему