

# Теория автоматов и формальных языков

## Формальные грамматики

**Лектор:** Екатерина Вербицкая

НИУ-ВШЭ

5 сентября 2022

- 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?

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

# 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  $\alpha \rightarrow \beta$ 
    - ★  $\alpha \in V^* V_N V^*$
    - ★  $\beta \in V^*$
  - ▶  $S$  — начальный нетерминал грамматики,
    - ★  $S \in V_N$

## Пример: язык чисел в двоичной системе счисления

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

$S$	$\rightarrow$	$0$
$S$	$\rightarrow$	$N$
$S$	$\rightarrow$	$-N$
$N$	$\rightarrow$	$1A$
$A$	$\rightarrow$	$0A$
$A$	$\rightarrow$	$1A$
$A$	$\rightarrow$	$\varepsilon$

## Пример: язык чисел в двоичной системе счисления

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

$S$	$\rightarrow$	$0$
$S$	$\rightarrow$	$N$
$S$	$\rightarrow$	$-N$
$N$	$\rightarrow$	$1A$
$A$	$\rightarrow$	$0A$
$A$	$\rightarrow$	$1A$
$A$	$\rightarrow$	$\varepsilon$

$S$	$\rightarrow$	$0 \mid N \mid -N$
$N$	$\rightarrow$	$1A$
$A$	$\rightarrow$	$0A \mid 1A \mid \varepsilon$

## Пример: язык чисел в двоичной системе счисления

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

$S$	$\rightarrow$	$0$
$S$	$\rightarrow$	$N$
$S$	$\rightarrow$	$-N$
$N$	$\rightarrow$	$1A$
$A$	$\rightarrow$	$0A$
$A$	$\rightarrow$	$1A$
$A$	$\rightarrow$	$\varepsilon$

$S$	$\rightarrow$	$0 \mid N \mid -N$
$N$	$\rightarrow$	$1A$
$A$	$\rightarrow$	$0A \mid 1A \mid \varepsilon$

$S$	$\rightarrow$	$0 \mid [-]N$
$N$	$\rightarrow$	$1A$
$A$	$\rightarrow$	$(0 \mid 1)A \mid \varepsilon$

# Отношение непосредственной выводимости

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



## Отношение непосредственной выводимости: пример

$$S \rightarrow 0 \mid N \mid -N$$

$$N \rightarrow 1A$$

$$A \rightarrow 0A \mid 1A \mid \varepsilon$$

$$S \Rightarrow -N$$

$$-N \Rightarrow -1A$$

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

Отношение выводимости является рефлексивно-транзитивным замыканием отношения непосредственной выводимости

- $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n \in V^*$
- $\alpha_0 \Rightarrow \alpha_1 \Rightarrow \alpha_2 \Rightarrow \dots \Rightarrow \alpha_n$
- $\alpha_0 \xRightarrow{*} \alpha_n$ :  $\alpha_n$  **выводится** из  $\alpha_0$

## Отношение выводимости: пример

$$S \rightarrow 0 \mid N \mid -N$$

$$N \rightarrow 1A$$

$$A \rightarrow 0A \mid 1A \mid \varepsilon$$

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

# Отношение выводимости: свойства

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

# Левосторонний вывод

На каждом шагу заменяем самый **левый** нетерминал

$$\begin{aligned} S &\rightarrow AA \mid s \\ A &\rightarrow AA \mid Bb \mid a \\ B &\rightarrow c \mid d \end{aligned}$$

$$\mathbf{S} \Rightarrow \mathbf{AA} \Rightarrow \mathbf{BbA} \Rightarrow \mathbf{cbA} \Rightarrow \mathbf{cbAA} \Rightarrow \mathbf{cbaA} \Rightarrow \mathbf{cbaa}$$

Аналогично определяется **правосторонний** вывод

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

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

Грамматика  $G_1$  и  $G_2$  эквивалентны, если  $L(G_1) = L(G_2)$

# Эквивалентность грамматик

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

$$\begin{aligned}V_T &= \{0, 1, -\} \\V_N &= \{S, N, A\}\end{aligned}$$

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

$$\begin{aligned}V_T &= \{0, 1, -\} \\V_N &= \{S, A\}\end{aligned}$$

$$\begin{aligned}S &\rightarrow 0 \mid 1A \mid -1A \\A &\rightarrow 0A \mid 1A \mid \varepsilon\end{aligned}$$



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

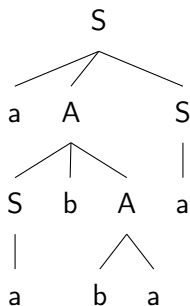
Дерево является **деревом вывода** для  $G = \langle V_N, V_T, P, S \rangle$ , если:

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

## Пример дерева вывода

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

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



## Теорема

Пусть  $G = \langle V_N, V_T, P, S \rangle$  — КС-грамматика

Вывод  $S \xRightarrow{*} \alpha$ , где  $\alpha \in V^*$ ,  $\alpha \neq \varepsilon$  существует  $\Leftrightarrow$  существует дерево вывода в грамматике  $G$  с результатом  $\alpha$

Упражнение: доказать теорему