

# Формальные языки, грамматики и автоматы

# Литература

1. Ахо А., Сети Р., Ульман Дж. Лам М. Компиляторы Принципы, технологии, инструменты. 2008г.
2. В.А.Серебряков, М.П.Галочкин Основы конструирования компиляторов.
3. Вирт Н. Построение компиляторов. 2010г.
4. Карпов Ю. Г. Основы построения трансляторов. 2005.
5. Хопкрофт, Джон Введение в теорию автоматов, языков и вычислений / Джон Хопкрофт , Раджив Мотвани , Джеффри Ульман. 2015.
6. Лаздин А. В. Формальные языки, грамматики, автоматы.

# От естественных языков к формальным

Исходный язык



Транслятор



Объектный язык

Из чего состоит язык?

Из букв? Алфавит.

Из слов? Лексика.

Из предложений!

Синтаксис определяется  
грамматикой языка.

# Виды предложений

Мама мыла раму.

однозначное

Ночь сменит день.

неоднозначное

стандарт отглагольного  
быстрое которая.

не является  
предложением  
русского языка

Делопроизводство  
трепещет синусоидами.

синтаксически верно,  
но бессмысленно

# Алфавиты и цепочки

- **Алфавит**  $\Sigma$  — это конечное множество символов.  
Например: кириллица, латинский алфавит, алфавит десятичных цифр.  $\Sigma = \{0,1\}$  — бинарный алфавит
- Всякая конечная (возможно пустая) последовательность символов из заданного алфавита называется **цепочкой**.
- Для алфавита  $\Sigma = \{a, b, c\}$  цепочками являются  $aba$ ,  $bbbaaa$ ,  $abc$ .
- Символом  $\varepsilon$  будем обозначать пустую цепочку.
- $\alpha = \varepsilon\alpha = \alpha\varepsilon = \varepsilon\alpha\varepsilon$ .

# Алфавиты и цепочки

- Степень алфавита.  $\Sigma = \{k, l, m\}$
- $\Sigma^1 = \{k, l, m\}$
- $\Sigma^2 = \{kk, kl, km, lk, ll, lm, mk, ml, mm\}$
- $\Sigma^n = \Sigma^{n-1} \times \Sigma$

- Усеченная итерация алфавита

$$\Sigma^+ = \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \dots \cup \Sigma^n \cup \dots$$

- Итерация алфавита  $\Sigma^* = \Sigma^0 \cup \Sigma^+$

$$\Sigma^0 = \{\varepsilon\}$$

# Языки и грамматики

- Правило подстановки (правило грамматики) – это упорядоченная пара  $(\alpha, \beta)$ , которая записывается следующим образом:  $\alpha ::= \beta$
- $\alpha$  – непустая конечная цепочка.
- $\beta$  – конечная цепочка, возможно пустая.
- Символ  $::=$  обозначает «есть по определению».

# Языки и грамматики

- Неформально грамматикой  $G(S)$  можно назвать конечное непустое множество правил.
- Все символы, которые встречаются в левых и правых частях правил образуют словарь грамматики  $V$ .
- Символ  $S$  –должен встретиться в левой части хотя бы одного правила.



# Пример грамматики

$\langle \text{целое\_число} \rangle ::= \langle \text{знак} \rangle \langle \text{число} \rangle$	(1)
$\langle \text{целое\_число} \rangle ::= \langle \text{число} \rangle$	(2)
$\langle \text{число} \rangle ::= \langle \text{цифра} \rangle$	(3)
$\langle \text{число} \rangle ::= \langle \text{цифра} \rangle \langle \text{число} \rangle$	(4)
$\langle \text{цифра} \rangle ::= 1$	(5)
$\langle \text{цифра} \rangle ::= 2$	(6)
$\langle \text{цифра} \rangle ::= 3$	(7)
$\langle \text{цифра} \rangle ::= 4$	(8)
$\langle \text{цифра} \rangle ::= 5$	(9)
$\langle \text{цифра} \rangle ::= 6$	(10)
$\langle \text{цифра} \rangle ::= 7$	(11)
$\langle \text{цифра} \rangle ::= 8$	(12)
$\langle \text{цифра} \rangle ::= 9$	(13)
$\langle \text{цифра} \rangle ::= 0$	(14)
$\langle \text{знак} \rangle ::= -$	(15)
$\langle \text{знак} \rangle ::= +$	(16)

# Пример (продолжение)

$V = \{ \langle \text{целое\_число} \rangle, \langle \text{число} \rangle, \langle \text{знак} \rangle, \langle \text{цифра} \rangle, 1, 2, 3, 4, 5, 6, 7, 8, 9, 0, -, + \}$

Начальным символом  $S$  для этой грамматики является символ  $\langle \text{целое\_число} \rangle$

Множество нетерминальных символов грамматики обозначается  $VN$ ; множество терминальных символов грамматики обозначается  $VT$ . Очевидно, что  $VN \cup VT = V$ , при этом  $VN \cap VT = \emptyset$

# Пример (продолжение)

Исходная цепочка	Полученная цепочка	№ правила
<целое_число>	<знак><число>	1
<знак><число>	<знак><цифра><число>	4
<знак><цифра><число>	<знак> <b>5</b> <число>	9
<знак>5<число>	<знак>5<цифра>	3
<знак>5<цифра>	+5<цифра>	16
+5<цифра>	+5 <b>3</b>	7

# Языки и грамматики

- *Порождающей грамматикой* называется четверка

$$G = (V_T, V_N, P, S), \text{ где}$$

- $V_T$  – множество терминальных символов;
- $V_N$  – множество нетерминальных символов;
- $P$  – непустое конечное множество правил грамматики;
- $S$  – начальный(стартовый) символ грамматики,  $S \in V_N$ .

# Языки и грамматики

$\langle \text{целое\_число} \rangle \rightarrow \langle \text{знак} \rangle \langle \text{число} \rangle \mid \langle \text{число} \rangle$

$\langle \text{число} \rangle \rightarrow \langle \text{цифра} \rangle \mid \langle \text{цифра} \rangle \langle \text{число} \rangle$

$\langle \text{цифра} \rangle \rightarrow 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9 \mid 0$

$\langle \text{знак} \rangle \rightarrow - \mid +$

# Языки и грамматики

Пусть  $G$  грамматика, тогда говорят, что цепочка  $\alpha$  непосредственно порождает цепочку  $\beta$ , обозначается  $\alpha \Rightarrow \beta$

- $\alpha = \xi_1 \gamma \xi_2$  и  $\beta = \xi_1 \delta \xi_2$ , причем  $\xi_1, \xi_2 \in (V_T \cup V_N)^*$
- $\gamma \rightarrow \delta$  – правило из множества  $P$

Для любого правила грамматики  $\gamma \rightarrow \delta$  справедливо  $\gamma \Rightarrow \delta$

$\langle \text{знак} \rangle 5 \langle \text{число} \rangle \Rightarrow +5 \langle \text{число} \rangle$

# Языки и грамматики

Цепочка  $\alpha$  порождает цепочку  $\beta$ , если существует конечное количество непосредственных выводов

$$\alpha \Rightarrow \gamma_0 \Rightarrow \gamma_1 \Rightarrow \gamma_2 \dots \gamma_n \Rightarrow \beta$$

обозначается  $\alpha \Rightarrow^+ \beta$

Если в процессе вывода цепочки всякий раз раскрывается **самый левый** нетерминальный символ, это называется **левосторонним** выводом цепочки; если всякий раз раскрывается **самый правый** нетерминальный символ, то это **правосторонний** вывод.

# Языки и грамматики

Цепочка  $\beta$  **выводима** из цепочки  $\alpha$ ,  
обозначается  $\alpha \Rightarrow^* \beta$ ,  
если  $\alpha \Rightarrow \beta$  или  $\alpha \Rightarrow^+ \beta$

$\langle \text{знак} \rangle \langle \text{число} \rangle \Rightarrow^+ + 5 \langle \text{цифра} \rangle$

$\langle \text{целое\_число} \rangle \Rightarrow^* \langle \text{знак} \rangle 5 \langle \text{цифра} \rangle$



# Языки и грамматики

*Сентенциальной формой* грамматики называется цепочка  $\alpha$  такая, что  $S \Rightarrow^* \alpha$ .

*Предложение* – это сентенциальная форма, состоящая только из терминальных символов, если  $S \Rightarrow^* \gamma$ , и  $\gamma \in VT^*$ , то  $\gamma$  – предложение языка.

# Языки и грамматики

Неформально

Язык, порождаемый грамматикой  $G$ , есть подмножество всех цепочек из  $VT^*$ .

Формально

$$L(G) = \{ \alpha \in VT^* \mid S \Rightarrow^* \alpha \}$$

# Примеры языков

G1(S)	$P = \{ S \rightarrow aB$ $B \rightarrow cD$ $D \rightarrow k \}$	$VT = \{a, c, k\}$ $VN = \{S, B, D\}$
G2(S)	$P = \{ S \rightarrow aB$ $B \rightarrow cB \mid k \}$	$VT = \{a, c, k\}$ $VN = \{S, B\}$
G3(S)	$P = \{ S \rightarrow aB$ $B \rightarrow cB \}$	$VT = \{a, c\}$ $VN = \{S, B\}$

# Определение языка как множества

- $L = \{\alpha \mid \text{информация об } \alpha\}$
- $L_1 = \{\alpha \mid \alpha \text{ содержит одинаковое число нулей и единиц}\}$
- $L_2 = \{1^n 0^n \mid n \geq 1\}$

$G = (VT = \{0,1\}, VN = \{S\}, P = \{S \rightarrow 1S0 \mid 10\}, S)$

- $L_3 = \{0^m 1^n \mid n \geq m \geq 0\}$
- $L_4 = \{\alpha \mid \alpha = a^+\}$

# Классификация грамматик

Говорят, что  $G$  — это грамматика *типа 0* или грамматика с фразовой структурой, если этой грамматике правила имеют вид:

$$\alpha \rightarrow \beta, \text{ где } \alpha \in V^+ \text{ и } \beta \in V^*$$

# Классификация грамматик

Грамматики типа 1, называемые также *неукорачивающимися* или *контекстно-зависимыми* грамматиками.

$\alpha \rightarrow \beta$ , где  $\alpha, \beta \in V^+$  и  $1 \leq |\alpha| \leq |\beta|$

или

$\alpha \rightarrow \beta$ , где

$\alpha = \xi_1 N \xi_2$ ,  $\beta = \xi_1 \gamma \xi_2$ ;

$N \in VN^+$ ,  $\gamma \in V^+$ ;

$\xi_1, \xi_2 \in (VT \cup VN)^*$

# Классификация грамматик

Грамматика *типа 2*, или *контекстно-свободная* (КС) – это грамматика все правила которой имеют следующий вид:

$A \rightarrow \alpha$ , где  $A \in VN$ ,  $\alpha \in V^+$ .

Если  $\alpha \in V^*$ , то такая грамматика является укорачивающейся КС грамматикой

# Классификация грамматик

Грамматика *типа 3*, также называемая автоматной или *регулярной* грамматикой *праволинейная РГ*

каждое правило грамматики имеет вид:

$A \rightarrow t$ , или  $A \rightarrow tB$ , где  $A, B \in VN$ ;  $t \in VT^*$

*леволинейная РГ*, если каждое правило грамматики имеет вид:

$A \rightarrow t$ , или  $A \rightarrow Bt$ , где  $A, B \in VN$ ;  $t \in VT^*$ .



# Классификация грамматик

$$1) S \rightarrow aaCFD;$$

$$2) AD \rightarrow D;$$

$$3) F \rightarrow AFB \mid AB;$$

$$4) Cb \rightarrow bC;$$

$$5) AB \rightarrow bBA;$$

$$6) CB \rightarrow C;$$

$$7) Ab \rightarrow bA;$$

$$8) bCD \rightarrow \varepsilon.$$

Грамматика типа 0

# Классификация грамматик

1)  $S \rightarrow aSBC \mid aBC;$       2)  $CB \rightarrow BC;$

3)  $aB \rightarrow ab;$       4)  $bB \rightarrow bb;$

5)  $bC \rightarrow bc;$       6)  $cC \rightarrow cc.$

Грамматика типа 1

Контекстно-зависимый язык

$$L(G) = \{a^n b^n c^n \mid n \geq 1\}$$

# Классификация грамматик

$$S \rightarrow aQb \mid acsb;$$
$$Q \rightarrow cSc.$$

Грамматика типа 2

Контекстно-свободный язык

$$L(G) = \{(ac)^n(cb)^n \mid n > 0\}$$

# Классификация грамматик

$$S \rightarrow A \mid B$$

$$A \rightarrow a \mid Ba$$

$$B \rightarrow b \mid Bb \mid Ab$$

Грамматика типа 3

Регулярный язык  $L(G) = \{\alpha \mid \alpha \in \{a, b\}^+, \text{ где нет двух рядом стоящих } a\}$