Тема:
Языки и грамматики

Сергей Витальевич Рыбин
svrybin@etu.ru

СПбГЭТУ «ЛЭТИ», кафедра «Алгоритмической математики»

20 июня 2023 г.

Основные понятия

Основные понятия

1. **Порождающая грамматика** G — четверка объектов (V_T, V_N, P, S):

Основные понятия

1. **Порождающая грамматика** G — четверка объектов (V_T, V_N, P, S):
   - ✓ V_T — алфавит (словарь) терминальных символов (терминалов или примитивов);

Основные понятия

1. **Порождающая грамматика** G — четверка объектов (V_T, V_N, P, S):
   - ✓ V_T — алфавит (словарь) терминальных символов (терминалов или примитивов);
   - ✓ V_N — алфавит (словарь) нетерминальных символов (нетерминалов), предполагается, что $V_T \cap V_N = \emptyset$;

Основные понятия

1. **Порождающая грамматика** G — четверка объектов (V_T, V_N, P, S):
   - ✓ V_T — алфавит (словарь) терминальных символов (терминалов или примитивов);
   - ✓ V_N — алфавит (словарь) нетерминальных символов (нетерминалов), предполагается, что $V_T \cap V_N = \emptyset$;
   - ✓ множество P — конечное подмножество декартова произведения

     $$(V_T \cup V_N)^+ \times (V_T \cup V_N)^*.$$

     Элементы (α, β) множества P называют правилами вывода (продукционными правилами или продукциями) и записывают в виде $\alpha \to \beta$;

Основные понятия

1. **Порождающая грамматика** G — четверка объектов $(V_T\,,\, V_N\,,\, P\,,\, S)$:
   - ✓ V_T — алфавит (словарь) терминальных символов (терминалов или примитивов);
   - ✓ V_N — алфавит (словарь) нетерминальных символов (нетерминалов), предполагается, что $V_T \cap V_N = \emptyset$;
   - ✓ множество P — конечное подмножество декартова произведения

     $$(V_T \cup V_N)^+ \times (V_T \cup V_N)^*\,.$$

     Элементы $(\alpha\,,\, \beta)$ множества P называют правилами вывода (продукционными правилами или продукциями) и записывают в виде $\alpha \to \beta$;
   - ✓ S — начальный символ (цель или аксиома) грамматики, $S \in V_N$.

Основные понятия

1. **Порождающая грамматика** G — четверка объектов (V_T, V_N, P, S):
   - ✓ V_T — алфавит (словарь) терминальных символов (терминалов или примитивов);
   - ✓ V_N — алфавит (словарь) нетерминальных символов (нетерминалов), предполагается, что $V_T \cap V_N = \emptyset$;
   - ✓ множество P — конечное подмножество декартова произведения
     $$(V_T \cup V_N)^+ \times (V_T \cup V_N)^*.$$
     Элементы (α, β) множества P называют правилами вывода (продукционными правилами или продукциями) и записывают в виде $\alpha \to \beta$;
   - ✓ S — начальный символ (цель или аксиома) грамматики, $S \in V_N$.
2. Для того чтобы различать терминальные и нетерминальные символы, принято обозначать терминальные символы строчными, а нетерминальные символы прописными буквами латинского алфавита.

Основные понятия

1. **Порождающая грамматика** G — четверка объектов (V_T, V_N, P, S):
   - ✓ V_T — алфавит (словарь) терминальных символов (терминалов или примитивов);
   - ✓ V_N — алфавит (словарь) нетерминальных символов (нетерминалов), предполагается, что $V_T \cap V_N = \emptyset$;
   - ✓ множество P — конечное подмножество декартова произведения

     $$(V_T \cup V_N)^+ \times (V_T \cup V_N)^*.$$

     Элементы (α, β) множества P называют правилами вывода (продукционными правилами или продукциями) и записывают в виде $\alpha \to \beta$;
   - ✓ S — начальный символ (цель или аксиома) грамматики, $S \in V_N$.
2. Для того чтобы различать терминальные и нетерминальные символы, принято обозначать терминальные символы строчными, а нетерминальные символы прописными буквами латинского алфавита.
3. Для записи n правил с одинаковыми левыми частями

$$\alpha \to \beta_1, \dots, \alpha \to \beta_n$$

часто используют сокращенную запись

$$\alpha \to \beta_1 \mid \dots \mid \beta_n.$$

Основные понятия

1. **Порождающая грамматика** G — четверка объектов (V_T, V_N, P, S):
   - ✓ V_T — алфавит (словарь) терминальных символов (терминалов или примитивов);
   - ✓ V_N — алфавит (словарь) нетерминальных символов (нетерминалов), предполагается, что $V_T \cap V_N = \emptyset$;
   - ✓ множество P — конечное подмножество декартова произведения

     $$(V_T \cup V_N)^+ \times (V_T \cup V_N)^*.$$

     Элементы (α, β) множества P называют правилами вывода (продукционными правилами или продукциями) и записывают в виде $\alpha \to \beta$;
   - ✓ S — начальный символ (цель или аксиома) грамматики, $S \in V_N$.
2. Для того чтобы различать терминальные и нетерминальные символы, принято обозначать терминальные символы строчными, а нетерминальные символы прописными буквами латинского алфавита.
3. Для записи n правил с одинаковыми левыми частями

$$\alpha \to \beta_1, \ldots, \alpha \to \beta_n$$

часто используют сокращенную запись

$$\alpha \to \beta_1 \mid \ldots \mid \beta_n.$$

4. Иногда при описании грамматики словари терминальных и нетерминальных символов не указывают.

В таком случае обычно предполагается, что грамматика содержит только те терминальные и нетерминальные символы, которые встречаются в правилах вывода.

Вывод в грамматике

Вывод в грамматике

1. Цепочка $\beta \in (V_T \cup V_N)^*$ **непосредственно выводима** из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \to \beta$), если

$$\alpha = \xi_1 \gamma \xi_2,\ \beta = \xi_1 \delta \xi_2, \qquad \xi_1, \xi_2, \delta \in (V_T \cup V_N)^*,\ \gamma \in (V_T \cup V_N)^+$$

и правило вывода $\gamma \to \delta$ содержится в P.

Вывод в грамматике

1. Цепочка $\beta \in (V_T \cup V_N)^*$ **непосредственно выводима** из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \to \beta$), если

$$\alpha = \xi_1 \gamma \xi_2, \ \beta = \xi_1 \delta \xi_2, \qquad \xi_1, \xi_2, \delta \in (V_T \cup V_N)^*, \ \gamma \in (V_T \cup V_N)^+$$

и правило вывода $\gamma \to \delta$ содержится в P.

2. Цепочка $\beta \in (V_T \cup V_N)^*$ **выводима** из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \Rightarrow \beta$), если существует последовательность цепочек $\gamma_0, \ldots, \gamma_n$ такая, что

$$\alpha = \gamma_0 \to \ldots \to \gamma_n = \beta.$$

Вывод в грамматике

1. Цепочка $\beta \in (V_T \cup V_N)^*$ **непосредственно выводима** из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \to \beta$), если

$$\alpha = \xi_1 \gamma \xi_2, \ \beta = \xi_1 \delta \xi_2, \qquad \xi_1, \xi_2, \delta \in (V_T \cup V_N)^*, \ \gamma \in (V_T \cup V_N)^+$$

и правило вывода $\gamma \to \delta$ содержится в P.

2. Цепочка $\beta \in (V_T \cup V_N)^*$ **выводима** из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \Rightarrow \beta$), если существует последовательность цепочек $\gamma_0, \dots, \gamma_n$ такая, что

$$\alpha = \gamma_0 \to \dots \to \gamma_n = \beta.$$

3. Последовательность $\gamma_0, \dots, \gamma_n$ называют **выводом** длины n.

Вывод в грамматике

1. Цепочка $\beta \in (V_T \cup V_N)^*$ **непосредственно выводима** из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \to \beta$), если

$$\alpha = \xi_1 \gamma \xi_2, \ \beta = \xi_1 \delta \xi_2, \qquad \xi_1, \xi_2, \delta \in (V_T \cup V_N)^*, \ \gamma \in (V_T \cup V_N)^+$$

и правило вывода $\gamma \to \delta$ содержится в P.

2. Цепочка $\beta \in (V_T \cup V_N)^*$ **выводима** из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \Rightarrow \beta$), если существует последовательность цепочек $\gamma_0, \dots, \gamma_n$ такая, что

$$\alpha = \gamma_0 \to \dots \to \gamma_n = \beta.$$

3. Последовательность $\gamma_0, \dots, \gamma_n$ называют **выводом** длины n.

4. **Языком**, порождаемым грамматикой G, называют множество $L(G) = \{\alpha \in V_T^* | S \Rightarrow \alpha\}$, т. е. $L(G)$ — все цепочки в алфавите V_T, которые выводимы из начального символа S с помощью правил вывода P.

Вывод в грамматике

1. Цепочка $\beta \in (V_T \cup V_N)^*$ **непосредственно выводима** из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \to \beta$), если

$$\alpha = \xi_1 \gamma \xi_2, \ \beta = \xi_1 \delta \xi_2, \qquad \xi_1, \xi_2, \delta \in (V_T \cup V_N)^*, \ \gamma \in (V_T \cup V_N)^+$$

и правило вывода $\gamma \to \delta$ содержится в P.

2. Цепочка $\beta \in (V_T \cup V_N)^*$ **выводима** из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \Rightarrow \beta$), если существует последовательность цепочек $\gamma_0, \dots, \gamma_n$ такая, что

$$\alpha = \gamma_0 \to \dots \to \gamma_n = \beta.$$

3. Последовательность $\gamma_0, \dots, \gamma_n$ называют **выводом** длины n.

4. **Языком**, порождаемым грамматикой G, называют множество $L(G) = \{\alpha \in V_T^* | S \Rightarrow \alpha\}$, т. е. $L(G)$ — все цепочки в алфавите V_T, которые выводимы из начального символа S с помощью правил вывода P.

Пример 1. Рассмотрим грамматику $G_1 = (\{a, b\}, \{A, S\}, P_1, S)$, где множество P_1 состоит из правил

$$S \to aAb \quad (1), \quad aA \to aaAb \quad (2), \quad A \to \epsilon \quad (3)$$

Вывод в грамматике

1. Цепочка $\beta \in (V_T \cup V_N)^*$ **непосредственно выводима** из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \to \beta$), если

$$\alpha = \xi_1 \gamma \xi_2, \ \beta = \xi_1 \delta \xi_2, \qquad \xi_1, \xi_2, \delta \in (V_T \cup V_N)^*, \ \gamma \in (V_T \cup V_N)^+$$

и правило вывода $\gamma \to \delta$ содержится в P.

2. Цепочка $\beta \in (V_T \cup V_N)^*$ **выводима** из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \Rightarrow \beta$), если существует последовательность цепочек $\gamma_0, \dots, \gamma_n$ такая, что

$$\alpha = \gamma_0 \to \dots \to \gamma_n = \beta.$$

3. Последовательность $\gamma_0, \dots, \gamma_n$ называют **выводом** длины n.

4. **Языком**, порождаемым грамматикой G, называют множество $L(G) = \{\alpha \in V_T^* | S \Rightarrow \alpha\}$, т. е. $L(G)$ — все цепочки в алфавите V_T, которые выводимы из начального символа S с помощью правил вывода P.

Пример 1. Рассмотрим грамматику $G_1 = (\{a, b\}, \{A, S\}, P_1, S)$, где множество P_1 состоит из правил

$$S \to aAb \quad (1), \quad aA \to aaAb \quad (2), \quad A \to \epsilon \quad (3)$$

Цепочка $aaAbb$ непосредственно выводима из цепочки aAb.

Вывод в грамматике

1. Цепочка $\beta \in (V_T \cup V_N)^*$ **непосредственно выводима** из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \to \beta$), если

$$\alpha = \xi_1 \gamma \xi_2, \ \beta = \xi_1 \delta \xi_2, \qquad \xi_1, \xi_2, \delta \in (V_T \cup V_N)^*, \ \gamma \in (V_T \cup V_N)^+$$

и правило вывода $\gamma \to \delta$ содержится в P.

2. Цепочка $\beta \in (V_T \cup V_N)^*$ **выводима** из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \Rightarrow \beta$), если существует последовательность цепочек $\gamma_0, \dots, \gamma_n$ такая, что

$$\alpha = \gamma_0 \to \dots \to \gamma_n = \beta.$$

3. Последовательность $\gamma_0, \dots, \gamma_n$ называют **выводом** длины n.

4. **Языком**, порождаемым грамматикой G, называют множество $L(G) = \{\alpha \in V_T^* | S \Rightarrow \alpha\}$, т. е. $L(G)$ — все цепочки в алфавите V_T, которые выводимы из начального символа S с помощью правил вывода P.

Пример 1. Рассмотрим грамматику $G_1 = (\{a, b\}, \{A, S\}, P_1, S)$, где множество P_1 состоит из правил

$$S \to aAb \quad (1), \quad aA \to aaAb \quad (2), \quad A \to \epsilon \quad (3)$$

Цепочка $aaAbb$ непосредственно выводима из цепочки aAb.

Пример 2. В грамматике G_1 из примера 1 $S \Rightarrow aaabbb$, так как существует вывод

$$S \xrightarrow{1} aAb \xrightarrow{2} aaAbb \xrightarrow{2} aaaAbbb \xrightarrow{3} aaabbb.$$

При этом длина вывода равна 4.

Вывод в грамматике

1. Цепочка $\beta \in (V_T \cup V_N)^*$ **непосредственно выводима** из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \to \beta$), если

$$\alpha = \xi_1 \gamma \xi_2, \;\; \beta = \xi_1 \delta \xi_2, \qquad \xi_1, \xi_2, \delta \in (V_T \cup V_N)^*, \;\; \gamma \in (V_T \cup V_N)^+$$

и правило вывода $\gamma \to \delta$ содержится в P.

2. Цепочка $\beta \in (V_T \cup V_N)^*$ **выводима** из цепочки $\alpha \in (V_T \cup V_N)^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначают $\alpha \Rightarrow \beta$), если существует последовательность цепочек $\gamma_0, \ldots, \gamma_n$ такая, что

$$\alpha = \gamma_0 \to \ldots \to \gamma_n = \beta.$$

3. Последовательность $\gamma_0, \ldots, \gamma_n$ называют **выводом** длины n.

4. **Языком**, порождаемым грамматикой G, называют множество $L(G) = \{\alpha \in V_T^* | S \Rightarrow \alpha\}$, т. е. $L(G)$ — все цепочки в алфавите V_T, которые выводимы из начального символа S с помощью правил вывода P.

Пример 1. Рассмотрим грамматику $G_1 = (\{a, b\}, \{A, S\}, P_1, S)$, где множество P_1 состоит из правил

$$S \to aAb \quad (1), \quad aA \to aaAb \quad (2), \quad A \to \epsilon \quad (3)$$

Цепочка $aaAbb$ непосредственно выводима из цепочки aAb.

Пример 2. В грамматике G_1 из примера 1 $S \Rightarrow aaabbb$, так как существует вывод

$$S \xrightarrow{1} aAb \xrightarrow{2} aaAbb \xrightarrow{2} aaaAbbb \xrightarrow{3} aaabbb.$$

При этом длина вывода равна 4.

Для грамматики G_1 из предыдущих примеров $L(G_1) = \{a^n b^n \mid n > 0\} = \{ab, aabb, aaabbb, \ldots\}$.

Примеры

Примеры

Пример 3

1. Построить грамматику G, порождающую язык
$$L(G) = \{a^n bc^m \mid n, m > 0\}.$$

Примеры

Пример 3

1. Построить грамматику G, порождающую язык
$$L(G) = \{a^n bc^m \mid n, m > 0\}.$$

2. Структура любой цепочки языка имеет вид $\alpha b \beta$, где часть α состоит только из символов a, а часть β — только из символов c.

Пример 3

1. Построить грамматику G, порождающую язык
$$L(G) = \{a^n b c^m \mid n, m > 0\}.$$
2. Структура любой цепочки языка имеет вид $\alpha b \beta$, где часть α состоит только из символов a, а часть β — только из символов c.
3. Обе части можно порождать независимо друг от друга. Получаем следующую грамматику:
$$S \to AbC; \quad A \to aA \mid a; \quad C \to cC \mid c.$$

Примеры

Пример 3

1. Построить грамматику G, порождающую язык
$$L(G) = \{a^n b c^m \mid n, m > 0\}.$$
2. Структура любой цепочки языка имеет вид $\alpha b \beta$, где часть α состоит только из символов a, а часть β — только из символов c.
3. Обе части можно порождать независимо друг от друга. Получаем следующую грамматику:
$$S \to AbC; \quad A \to aA \mid a; \quad C \to cC \mid c.$$

Пример 4

1. Построить грамматику G, порождающую язык
$$L(G) = \{\alpha \alpha^R \mid \alpha \in \{a, b\}^*\},$$

где α^R обозначает слово α, записанное в обратном порядке.

Примеры

Пример 3

1. Построить грамматику G, порождающую язык
$$L(G) = \{a^n b c^m \mid n, m > 0\}.$$
2. Структура любой цепочки языка имеет вид $\alpha b \beta$, где часть α состоит только из символов a, а часть β — только из символов c.
3. Обе части можно порождать независимо друг от друга. Получаем следующую грамматику:
$$S \to AbC; \quad A \to aA \mid a; \quad C \to cC \mid c.$$

Пример 4

1. Построить грамматику G, порождающую язык
$$L(G) = \{\alpha \alpha^R \mid \alpha \in \{a, b\}^*\},$$

где α^R обозначает слово α, записанное в обратном порядке.

2. Имеем
$$S \to aSa \mid bSb \mid \epsilon$$

Примеры

Пример 3

1. Построить грамматику G, порождающую язык
$$L(G) = \{a^n b c^m \mid n, m > 0\}.$$
2. Структура любой цепочки языка имеет вид $\alpha b \beta$, где часть α состоит только из символов a, а часть β — только из символов c.
3. Обе части можно порождать независимо друг от друга. Получаем следующую грамматику:
$$S \to AbC; \quad A \to aA \mid a; \quad C \to cC \mid c.$$

Пример 4

1. Построить грамматику G, порождающую язык
$$L(G) = \{\alpha \alpha^R \mid \alpha \in \{a, b\}^*\},$$

где α^R обозначает слово α, записанное в обратном порядке.

2. Имеем
$$S \to aSa \mid bSb \mid \epsilon$$
3. $L(G) = \{aa,\ bb,\ abba, \ldots\}$

Примеры (2)

Примеры (2)

Пример 5

1. Построить грамматику G, порождающую язык
$$L(G) = \{a^n \mid n > 0\}.$$

Примеры (2)

Пример 5

1. Построить грамматику G, порождающую язык
$$L(G) = \{a^n \mid n > 0\}.$$
2. Имеем: $S \to a\,S \mid a$.

Примеры (2)

Пример 5

1. Построить грамматику G, порождающую язык
$$L(G) = \{a^n \mid n > 0\}.$$
2. Имеем: $S \to aS \mid a$.

Пример 6

1. Усложним предыдущий пример. Построим грамматику G, порождающую язык
$$L(G) = \{a^{2n+1} \mid n \geqslant 0\}.$$

Примеры (2)

Пример 5

1. Построить грамматику G, порождающую язык
$$L(G) = \{a^n \mid n > 0\}.$$
2. Имеем: $S \to aS \mid a$.

Пример 6

1. Усложним предыдущий пример. Построим грамматику G, порождающую язык
$$L(G) = \{a^{2n+1} \mid n \geqslant 0\}.$$
2. Имеем: $S \to aB \mid a;\ B \to aS$.