Лекция А4
Регулярные языки, грамматики

Вадим Пузаренко

22 сентября 2023 г.

Теорема о накачке

Следующее утверждение является необходимым условием для регулярных языков и позволяет для достаточно богатого класса языков доказывать их нерегулярность.

Теорема о накачке

Следующее утверждение является необходимым условием для регулярных языков и позволяет для достаточно богатого класса языков доказывать их нерегулярность.

Теорема А4.1.

Пусть L — регулярный язык. Тогда существует натуральное число $n_0 \geqslant 1$, удовлетворяющее следующему условию: для любого слова $\lambda \in L$, $\mathrm{lh}(\lambda) \geqslant n_0$, существуют α, β, γ такие, что $\beta \neq \varepsilon$, $\mathrm{lh}(\alpha\hat{\ }\beta) \leqslant n_0$ и $\lambda = \alpha\hat{\ }\beta\hat{\ }\gamma$, для которых выполняется соотношение $\alpha\hat{\ }\beta^{n}\hat{\ }\gamma \in L$ для всех $n \in \omega$.

Теорема о накачке

Следующее утверждение является необходимым условием для регулярных языков и позволяет для достаточно богатого класса языков доказывать их нерегулярность.

Теорема А4.1.

Пусть L — регулярный язык. Тогда существует натуральное число $n_0 \geqslant 1$, удовлетворяющее следующему условию: для любого слова $\lambda \in L$, $\mathrm{lh}(\lambda) \geqslant n_0$, существуют α, β, γ такие, что $\beta \neq \varepsilon$, $\mathrm{lh}(\alpha\hat{}\beta) \leqslant n_0$ и $\lambda = \alpha\hat{}\beta\hat{}\gamma$, для которых выполняется соотношение $\alpha\hat{}\beta^n\hat{}\gamma \in L$ для всех $n \in \omega$.

Доказательство.

Так как L — регулярный язык, существует Д.К.А. $\mathfrak{A} = (Q; \Sigma; \delta, q_0, F)$ такой, что $\mathrm{L}(\mathfrak{A}) = L$. Положим $n_0 = n(Q)$; возьмём $\lambda \in L$ такое, что $\mathrm{lh}(\lambda) \geqslant n_0$. Пусть $\varepsilon = \lambda_0, \lambda_1, \ldots, \lambda_{n_0}$ — начальные подслова слова λ, $(\mathrm{lh}(\lambda_i) = i,\ 0 \leqslant i \leqslant n_0)$.

Доказательство (окончание).

По принципу Дирихле, найдётся состояние $q \in Q$, для которого имеет место $q = \delta^*(q_0, \lambda_i) = \delta^*(q_0, \lambda_j)$. Будем считать, что $i < j$ и, к тому же, $q \neq \delta^*(q_0, \lambda_p)$ для всех $i < p < j$. Положим $\alpha = \lambda_i$, а β и γ выберем так, чтобы $\lambda_i\hat{\ }\beta = \lambda_j$, $\lambda_j\hat{\ }\gamma = \lambda$. Тогда $\mathrm{lh}(\alpha\hat{\ }\beta) = \mathrm{lh}(\lambda_j) = j \leqslant n_0$. По условию, $\delta^*(q_0, \lambda) \in F$; далее, $\delta^*(q_0, \alpha\hat{\ }\beta^n\hat{\ }\gamma) = \delta^*(\delta^*(q_0, \alpha), \beta^n\hat{\ }\gamma) = \delta^*(q, \beta^n\hat{\ }\gamma) = \delta^*(\delta^*(q, \beta), \beta^{n-1}\hat{\ }\gamma) = \delta^*(q, \beta^{n-1}\hat{\ }\gamma) = \ldots = \delta^*(q, \gamma) \in F$ $(n \in \omega)$.
□

Доказательство (окончание).

По принципу Дирихле, найдётся состояние $q \in Q$, для которого имеет место $q = \delta^*(q_0, \lambda_i) = \delta^*(q_0, \lambda_j)$. Будем считать, что $i < j$ и, к тому же, $q \neq \delta^*(q_0, \lambda_p)$ для всех $i < p < j$. Положим $\alpha = \lambda_i$, а β и γ выберем так, чтобы $\lambda_i \hat{}\, \beta = \lambda_j$, $\lambda_j \hat{}\, \gamma = \lambda$. Тогда $\mathrm{lh}(\alpha \hat{}\, \beta) = \mathrm{lh}(\lambda_j) = j \leqslant n_0$. По условию, $\delta^*(q_0, \lambda) \in F$; далее, $\delta^*(q_0, \alpha \hat{}\, \beta^n \hat{}\, \gamma) = \delta^*(\delta^*(q_0, \alpha), \beta^n \hat{}\, \gamma) = \delta^*(q, \beta^n \hat{}\, \gamma) = \delta^*(\delta^*(q, \beta), \beta^{n-1} \hat{}\, \gamma) = \delta^*(q, \beta^{n-1} \hat{}\, \gamma) = \ldots = \delta^*(q, \gamma) \in F$ $(n \in \omega)$.
□

Теорема о накачке может восприниматься как игра, в которой играют двое ($\exists$ и $\forall$), при этом студент несет ответственность только за действия $\forall$-игрока. Сначала $\exists$-игрок предлагает натуральное число n_0; затем $\forall$-игрок предлагает слово λ $(\mathrm{lh}(\lambda) \geqslant n_0)$, за представление которого в виде $\alpha \hat{}\, \beta \hat{}\, \gamma$ отвечает $\exists$-игрок; и, наконец, $\forall$-игрок находит число n, для которого выполняется соотношение $\alpha \hat{}\, \beta^n \hat{}\, \gamma \notin L$.

Нерегулярность языков

Пример А4.1.

Язык $\{0^n1^n \mid n \in \omega\}$ *нерегулярный.* $\exists$-Игрок предлагает число n_0; студент в лице $\forall$-игрока находит слово $0^{n_0}1^{n_0}$. Далее, $\exists$-игрок произвольным образом представляет данное слово в виде конкатенации $\alpha\hat{}\beta\hat{}\gamma$ с условием $\mathrm{lh}(\alpha\hat{}\beta) \leqslant n_0$ (следовательно, $\alpha = 0^k$, $\beta = 0^l$, $\gamma = 0^m1^{n_0}$, причём $n_0 = k + l + m$, $l > 0$). Остаётся только заметить, что $\alpha\hat{}\beta^0\hat{}\gamma = 0^{k+m}1^{n_0} \notin L$.

Нерегулярность языков

Пример А4.1.

Язык $\{0^n 1^n \mid n \in \omega\}$ *нерегулярный.* $\exists$-Игрок предлагает число n_0; студент в лице $\forall$-игрока находит слово $0^{n_0}1^{n_0}$. Далее, $\exists$-игрок произвольным образом представляет данное слово в виде конкатенации $\alpha\hat{}\beta\hat{}\gamma$ с условием $\mathrm{lh}(\alpha\hat{}\beta) \leqslant n_0$ (следовательно, $\alpha = 0^k$, $\beta = 0^l$, $\gamma = 0^m 1^{n_0}$, причём $n_0 = k + l + m$, $l > 0$). Остаётся только заметить, что $\alpha\hat{}\beta^0\hat{}\gamma = 0^{k+m}1^{n_0} \notin L$.

Пример А4.2.

Язык $\{a^p \mid p$ — *простое*$\}$ *нерегулярный.* $\exists$-Игрок предлагает число n_0; затем $\forall$-игрок находит $p \geqslant n_0 + 2$ (это возможно, поскольку простых чисел бесконечно много); $\exists$-игрок предлагает представление $a^k\hat{}a^l\hat{}a^m$, причём $0 < l \leqslant k + l \leqslant n_0$. Тогда $a^k\hat{}a^{l\cdot(k+m)}\hat{}a^m \notin L$, поскольку количество букв равняется $k + l(k+m) + m = (k+m)(l+1)$ и $l + 1 \geqslant 1 + 1 = 2$, $k + m = p - l \geqslant n_0 + 2 - n_0 = 2$.

ДКА: алфавит $\{0\}$

Теорема А4.2.

Пусть $\Sigma = \{0\}$; тогда для языка $L \subseteq \Sigma^*$ следующие утверждения эквивалентны:

1. L регулярный;
2. существует конечное множество $\{k_1(n), \ldots, k_m(n)\}$ $(m \in \omega)$ арифметических прогрессий, принимающих натуральные числа в качестве значений, для которого выполняется равенство $L = \{0^{k_i(n)} \mid n \in \omega,\ 1 \leqslant i \leqslant m\}$.

ДКА: алфавит $\{0\}$

Теорема А4.2.

Пусть $\Sigma = \{0\}$; тогда для языка $L \subseteq \Sigma^*$ следующие утверждения эквивалентны:

1. L регулярный;
2. существует конечное множество $\{k_1(n), \ldots, k_m(n)\}$ $(m \in \omega)$ арифметических прогрессий, принимающих натуральные числа в качестве значений, для которого выполняется равенство $L = \{0^{k_i(n)} \mid n \in \omega,\ 1 \leqslant i \leqslant m\}$.

Доказательство.

$(2 \Rightarrow 1)$ Пусть $k(n) = m + d(n-1)$ — арифметическая прогрессия; тогда регулярное выражение

$$\begin{cases} \varepsilon = \varnothing^*, & \text{если } m = d = 0; \\ 0^m, & \text{если } m > 0 \text{ и } d = 0; \\ (0^d)^*, & \text{если } m = 0 \text{ и } d > 0; \\ 0^m(0^d)^*, & \text{если } md > 0; \end{cases}$$

ДКА: алфавит $\{0\}$

Доказательство (продолжение).

(правильно расставить символ $\cdot$ и скобки!!!) представляет язык $\{0^{k(n)} \mid n \in \omega\}$. Для завершения следует воспользоваться тем, что регулярные языки замкнуты относительно операции объединения, а пустой язык (задаваемый пустой совокупностью арифметических прогрессий) регулярен.

ДКА: алфавит $\{0\}$

Доказательство (продолжение).

(правильно расставить символ $\cdot$ и скобки!!!) представляет язык $\{0^{k(n)} \mid n \in \omega\}$. Для завершения следует воспользоваться тем, что регулярные языки замкнуты относительно операции объединения, а пустой язык (задаваемый пустой совокупностью арифметических прогрессий) регулярен.

(1 $\Rightarrow$ 2) Пусть Д.К.А. $\mathfrak{A} = (Q; \Sigma; \delta, q_0, F)$ таков, что $\mathrm{L}(\mathfrak{A}) = L$. Рассматривая слово $\alpha \in \{0\}^*$ с условием $\mathrm{lh}(\alpha) \geqslant n(Q)$, приходим к тому, что найдутся $\alpha_1, \alpha_2 \sqsubseteq_{\mathrm{beg}} \alpha$ $(\alpha_1 \neq \alpha_2)$ такие, что $\delta^*(q_0, \alpha_1) = \delta^*(q_0, \alpha_2)$. Возьмём слово α_0 наименьшей длины (скажем, $l_0 = \mathrm{lh}(\alpha_0)$) такое, что $q = \delta^*(q_0, \alpha_0) = \delta^*(q_0, \alpha_1)$ для некоторого $\alpha_1 \in \{0\}^*$ $(\mathrm{lh}(\alpha_0) > l_0)$.

Пусть $d \in \omega$ таково, что $\delta^*(q, a^d) = q$ и d — наименьшее с таким свойством. Далее, пусть $M_0 \subseteq \omega$ содержит все числа $m < n(Q)$, для которых $\delta^*(q_0, a^m) \in F$. Тогда язык L задается множеством арифметических прогрессий

$\{m + d(n-1) \mid m \in M_0,\ m \geqslant l_0\} \cup \{m \mid m \in M_0,\ m < l_0\}$:

$$L = \{0^m \mid m \in M_0,\ m < l_0\} \cup \{0^{m+d(n-1)} \mid m \in M_0,\ m \geqslant l_0,\ n \in \omega\}. \qquad (1)$$

ДКА: алфавит $\{0\}$

Доказательство (окончание).

($\subseteq$) Обозначим правую часть равенства (1) через S. Пусть $\alpha \in L$; тогда $\delta^*(q_0, \alpha) \in F$. Если $\mathrm{lh}(\alpha) < l_0$, то $\alpha = 0^m$, $m \in M_0$, и, следовательно, $\alpha \in S$. Если же $\mathrm{lh}(\alpha) \geqslant l_0$, то существуют и притом единственные n и $0 \leqslant r < d$ такие, что $\mathrm{lh}(\alpha) - l_0 = d \cdot n + r$ (теорема о делении с остатком); следовательно, $\delta^*(q_0, 0^{l_0+r+d \cdot n}) = \delta^*(q_0, 0^{l_0+r}) \in F$ и $l_0 + r < n(Q)$, т.е. $l_0 + r \in M_0$. Таким образом, $\alpha = 0^{(l_0+r)+d \cdot n}$ и $\alpha \in S$.

ДКА: алфавит $\{0\}$

Доказательство (окончание).

($\subseteq$) Обозначим правую часть равенства (1) через S. Пусть $\alpha \in L$; тогда $\delta^*(q_0, \alpha) \in F$. Если $\mathrm{lh}(\alpha) < l_0$, то $\alpha = 0^m$, $m \in M_0$, и, следовательно, $\alpha \in S$. Если же $\mathrm{lh}(\alpha) \geqslant l_0$, то существуют и притом единственные n и $0 \leqslant r < d$ такие, что $\mathrm{lh}(\alpha) - l_0 = d \cdot n + r$ (теорема о делении с остатком); следовательно, $\delta^*(q_0, 0^{l_0+r+d\cdot n}) = \delta^*(q_0, 0^{l_0+r}) \in F$ и $l_0 + r < n(Q)$, т.е. $l_0 + r \in M_0$. Таким образом, $\alpha = 0^{(l_0+r)+d\cdot n}$ и $\alpha \in S$.

($\supseteq$) Пусть сначала $\alpha = 0^m$ ($m \in M_0$, $m < l_0$). Следовательно, $\delta^*(q_0, 0^m) \in F$; тем самым, $\alpha \in L$. Пусть теперь $\alpha = 0^{m+dn}$ ($m \in M_0$, $m \geqslant l_0$, $n \in \omega$); тогда $\delta^*(q_0, 0^m) \in F$ и $\delta^*(q_0, 0^{m+dn}) = \delta^*(q_0, 0^m)$; таким образом, $\alpha \in L$. $\square$

ДКА: алфавит $\{0; 1\}$

Теорема А4.3.

Для каждого языка L в алфавите Σ $(\mathrm{card}(\Sigma) \geqslant 2)$ существует язык L' в алфавите $\{0; 1\}$, для которого выполняется следующее соотношение:
L — регулярный язык, если и только если L' — регулярный язык.
Более того, данная трансформация инъективна.

ДКА: алфавит $\{0;1\}$

Теорема А4.3.

Для каждого языка L в алфавите Σ $(\mathrm{card}(\Sigma) \geqslant 2)$ существует язык L' в алфавите $\{0;1\}$, для которого выполняется следующее соотношение:
L — регулярный язык, если и только если L' — регулярный язык.
Более того, данная трансформация инъективна.

Доказательство.

Пусть $\Sigma = \{a_1, a_2, \ldots, a_k\}$; определим $a_i \mapsto 01^i0$, $1 \leqslant i \leqslant k$.
Далее, индукцией по построению регулярного выражения, представляющего L, задаётся результат подстановки для регулярного выражения, представляющего L'. Свойство инъективности следует из того, что каждая буква задаётся двумя нулями, между которыми определенное количество единиц (аналог открывающей и закрывающей скобок). $\square$

Грамматики: определение

Определение A4.1.

Грамматикой называется структура $\mathfrak{G} = (N, \Sigma, P, S)$, удовлетворяющая следующим условиям:

1. N — конечное множество символов, называемых **нетерминалами**;
2. Σ — конечный алфавит **терминалов** $(\Sigma \cap N = \varnothing)$;
3. $P \subseteq (N \cup \Sigma)^* N (N \cup \Sigma)^* \times (N \cup \Sigma)^*$ — множество **продукций**;
4. $S \in N$.

Грамматики: определение

Определение А4.1.

Грамматикой называется структура $\mathfrak{G} = (N, \Sigma, P, S)$, удовлетворяющая следующим условиям:

1. N — конечное множество символов, называемых **нетерминалами**;
2. Σ — конечный алфавит **терминалов** ($\Sigma \cap N = \varnothing$);
3. $P \subseteq (N \cup \Sigma)^* N (N \cup \Sigma)^* \times (N \cup \Sigma)^*$ — множество **продукций**;
4. $S \in N$.

Определение А4.2.

Опишем один такт преобразования слов грамматикой $\mathfrak{G}$. Пусть $\alpha, \beta \in (N \cup \Sigma)^*$; будем говорить что слово α **преобразуется** в слово β под действием грамматики $\mathfrak{G}$ (и записывать как $\alpha \Rightarrow_{\mathfrak{G}} \beta$), если найдутся слова γ_1, γ_2, δ_1, $\delta_2 \in (N \cup \Sigma)^*$ такие, что $\alpha = \gamma_1 \hat{}\, \delta_1 \hat{}\, \gamma_2$, $\beta = \gamma_1 \hat{}\, \delta_2 \hat{}\, \gamma_2$ и $P(\delta_1, \delta_2)$.

Грамматики: определение

Определение А4.3.

Определим отношение $\Rightarrow^*_{\mathfrak{G}}$ на словах из $(N \cup \Sigma)^*$ как рефлексивное и транзитивное замыкания отношения $\Rightarrow_{\mathfrak{G}}$. А именно, $\alpha \Rightarrow^*_{\mathfrak{G}} \beta$, если выполняется одно из следующих условий:

1. $\alpha = \beta$;
2. найдётся слово $\alpha' \in (N \cup \Sigma)^*$ такое, что $\alpha \Rightarrow^*_{\mathfrak{G}} \alpha'$ и $\alpha' \Rightarrow_{\mathfrak{G}} \beta$.

Грамматики: определение

Определение A4.3.

Определим отношение $\Rightarrow^*_{\mathfrak{G}}$ на словах из $(N \cup \Sigma)^*$ как рефлексивное и транзитивное замыкания отношения $\Rightarrow_{\mathfrak{G}}$. А именно, $\alpha \Rightarrow^*_{\mathfrak{G}} \beta$, если выполняется одно из следующих условий:

1. $\alpha = \beta$;
2. найдётся слово $\alpha' \in (N \cup \Sigma)^*$ такое, что $\alpha \Rightarrow^*_{\mathfrak{G}} \alpha'$ и $\alpha' \Rightarrow_{\mathfrak{G}} \beta$.

Определение A4.4.

Множество слов $\{\alpha \in \Sigma^* \mid S \Rightarrow^*_{\mathfrak{G}} \alpha\}$ называется **языком, порождённым грамматикой** $\mathfrak{G}$, и обозначается как $\mathrm{L}(\mathfrak{G})$.

Грамматики: определение

Определение A4.3.

Определим отношение $\Rightarrow^*_{\mathfrak{G}}$ на словах из $(N \cup \Sigma)^*$ как рефлексивное и транзитивное замыкания отношения $\Rightarrow_{\mathfrak{G}}$. А именно, $\alpha \Rightarrow^*_{\mathfrak{G}} \beta$, если выполняется одно из следующих условий:

1. $\alpha = \beta$;
2. найдётся слово $\alpha' \in (N \cup \Sigma)^*$ такое, что $\alpha \Rightarrow^*_{\mathfrak{G}} \alpha'$ и $\alpha' \Rightarrow_{\mathfrak{G}} \beta$.

Определение A4.4.

Множество слов $\{\alpha \in \Sigma^* \mid S \Rightarrow^*_{\mathfrak{G}} \alpha\}$ называется **языком, порождённым грамматикой** $\mathfrak{G}$, и обозначается как $\mathrm{L}(\mathfrak{G})$.

Основным атрибутом грамматики $\mathfrak{G}$ является язык, порождённый ею, поэтому грамматика однозначно задаётся множеством продукций и начальным нетерминалом.

Пример A4.1.

Пусть $\mathfrak{G} = (\{S, A, B\}, \{a, b, c\}, P, S)$, где $P = \{\langle S, aBc\rangle, \langle BA, AB\rangle, \langle aB, ab\rangle, \langle bB, bb\rangle, \langle B, BABc\rangle, \langle aA, aa\rangle, \langle S, \varepsilon\rangle\}$. Покажем, что $\mathrm{L}(\mathfrak{G}) = \{a^n b^n c^n \mid n \in \omega\}$.

$(\supseteq)$ Доказывать будем индукцией по n. Если $n = 0$, то $S \Rightarrow^*_{\mathfrak{G}} \varepsilon$. Далее, индукцией по k докажем, что $S \Rightarrow^*_{\mathfrak{G}} aA^k B^{k+1} c^{k+1}$; если $k = 0$, то $S \Rightarrow^*_{\mathfrak{G}} aBc$. Предположим, что $S \Rightarrow^*_{\mathfrak{G}} aA^k B^k Bc^{k+1}$; тогда

$$S \Rightarrow^*_{\mathfrak{G}} aA^k B^k Bc^{k+1} \Rightarrow_{\mathfrak{G}} aA^k B^{k+1} ABc^{k+2} \Rightarrow_{\mathfrak{G}} aA^k B^k ABBc^{k+2} \Rightarrow_{\mathfrak{G}} \\ \Rightarrow_{\mathfrak{G}} aA^k B^{k-1} ABBBc^{k+2} \Rightarrow_{\mathfrak{G}} \ldots \Rightarrow_{\mathfrak{G}} aA^{k+1} B^{k+2} c^{k+2}.$$

Далее, имеем $S \Rightarrow^*_{\mathfrak{G}} aBc \Rightarrow_{\mathfrak{G}} abc$,

$$S \Rightarrow^*_{\mathfrak{G}} aA^n B^{n+1} c^{n+1} \Rightarrow_{\mathfrak{G}} aaA^{n-1} B^{n+1} c^{n+1} \Rightarrow_{\mathfrak{G}} \ldots \Rightarrow_{\mathfrak{G}} \\ \Rightarrow_{\mathfrak{G}} a^{n+1} B^{n+1} c^{n+1} \Rightarrow_{\mathfrak{G}} a^{n+1} bB^n c^{n+1} \Rightarrow_{\mathfrak{G}} \\ \Rightarrow_{\mathfrak{G}} a^{n+1} bbB^{n-1} c^{n+1} \Rightarrow_{\mathfrak{G}} \ldots \Rightarrow_{\mathfrak{G}} a^{n+1} b^{n+1} c^{n+1}.$$

Пример A4.1 (продолжение).

$(\subseteq)$ Сначала заметим, что если $S \Rightarrow^*_{\mathfrak{G}} \alpha(\in (N \cup \Sigma)^*)$, то суммарные количества вхождений букв соответственно a, A и b, B совпадают и равны количеству вхождений букв c (продукции $P(S, \varepsilon)$, $P(S, aBc)$ и $P(B, BABc)$ удовлетворяют данному условию; остальные сохраняют суммарные количества соответствующих букв).
Далее, используя продукции $P(S, aBc)$ и $P(aA, aa)$, нетрудно показать, что все вхождения букв 'a' должны находиться в начале слова $\alpha \in \mathrm{L}(\mathfrak{G})$.
Для завершения рассмотрений следует проверить, что буква 'c' не может находиться слева от буквы 'b' в слове $\alpha \in \mathrm{L}(\mathfrak{G})$. Действительно, если это не так, то выберем последнее такое вхождение буквы 'c'. Тогда после этой буквы стоит буква 'b', однако отсутствует продукция в такой ситуации, в которой буква 'B' преобразуется в букву 'b'.

Регулярные грамматики: основные сведения

Соглашение.

Пусть P — продукция некоторой грамматики; тогда вместо $\langle\alpha;\beta\rangle(\in P)$ будем иногда использовать обозначение $\alpha \longrightarrow \beta$.

Регулярные грамматики: основные сведения

Соглашение.

Пусть P — продукция некоторой грамматики; тогда вместо $\langle\alpha;\beta\rangle(\in P)$ будем иногда использовать обозначение $\alpha \longrightarrow \beta$.

Определение А4.5.

Грамматика $\mathfrak{G} = (N, \Sigma, P, S)$ называется **регулярной** или **праволинейной**, если каждая её продукция имеет один из следующих видов:

$$\begin{array}{ll} \langle A, aB\rangle; & \langle A, a\rangle; \\ \langle A, B\rangle; & \langle A, \varepsilon\rangle; \end{array}$$

где $A, B \in N$ и $a \in \Sigma$.

Регулярные грамматики: основные сведения

Соглашение.

Пусть P — продукция некоторой грамматики; тогда вместо $\langle\alpha;\beta\rangle(\in P)$ будем иногда использовать обозначение $\alpha \longrightarrow \beta$.

Определение A4.5.

Грамматика $\mathfrak{G} = (N, \Sigma, P, S)$ называется **регулярной** или **праволинейной**, если каждая её продукция имеет один из следующих видов:

$$\begin{array}{ll} \langle A, aB\rangle; & \langle A, a\rangle; \\ \langle A, B\rangle; & \langle A, \varepsilon\rangle; \end{array}$$

где $A, B \in N$ и $a \in \Sigma$.

Основной целью наших дальнейших действий является проверка того, что регулярные языки порождаются регулярными грамматиками, и наоборот.

Регулярные языки и регулярные грамматики

Теорема A4.4.

Язык L регулярный, если и только если он порождается некоторой регулярной грамматикой.

Регулярные языки и регулярные грамматики

Теорема А4.4.

Язык L регулярный, если и только если он порождается некоторой регулярной грамматикой.

Доказательство.

$(\Leftarrow)$ Пусть $L = \mathrm{L}(\mathfrak{G})$ для некоторой регулярной грамматики $\mathfrak{G} = (N, \Sigma, P, S)$. Определим ε-НКА $\mathfrak{A} = (N \uplus \{*\}, \Sigma, \delta, \{S\}, \{*\})$ следующим образом $(A \in N,\ a \in \Sigma)$:

- $B \in \delta(A, a) \Leftrightarrow P(A, aB)$ $(B \in N)$;
- $B \in \delta(A, \varepsilon) \Leftrightarrow P(A, B)$ $(B \in N)$;
- $* \in \delta(A, a) \Leftrightarrow P(A, a)$;
- $* \in \delta(A, \varepsilon) \Leftrightarrow P(A, \varepsilon)$.

Отметим, что $\delta(*, \varepsilon) = \delta(*, a) = \varnothing$ для любого $a \in \Sigma$. Докажем теперь, что $L = \mathrm{L}(\mathfrak{A})$.

Регулярные языки и регулярные грамматики

Доказательство (продолжение).

Докажем сначала следующее соотношение ($\alpha = a_1 a_2 \ldots a_n \in \Sigma^*$, $A \in N$):

$$
\begin{aligned}
&[S \Rightarrow^*_{\mathfrak{G}} \alpha\hat{\ }A] \Leftrightarrow \\
&\Leftrightarrow [\text{ найдётся последовательность } (S =)A_0^0, A_0^1, \ldots, A_0^{k_0}, \\
&A_1^0, A_1^1, \ldots, A_1^{k_1}, \ldots, A_n^0, A_n^1, \ldots, A_n^{k_n}(= A) \in N \\
&\text{такая, что } A_i^{j+1} \in \delta(A_i^j, \varepsilon),\ 0 \leqslant j < k_i,\ 0 \leqslant i \leqslant n,\ \text{ и} \\
&A_{i+1}^0 \in \delta(A_i^{k_i}, a_{i+1}), 0 \leqslant i < n].
\end{aligned}
\qquad (1)
$$

(1⇒) Доказывать будем индукцией по отношению $\Rightarrow^*_{\mathfrak{G}}$. Если $S \Rightarrow^*_{\mathfrak{G}} S(= \varepsilon\hat{\ }S)$, то в качестве искомой последовательности возьмём $A_0^0 = S$.

Если $S \Rightarrow^*_{\mathfrak{G}} \beta$ и $\beta \Rightarrow_{\mathfrak{G}} \alpha\hat{\ }A$, где $\alpha = a_1 a_2 \ldots a_m$, то β может иметь один из следующих видов.

Регулярные языки и регулярные грамматики

Доказательство (продолжение).

1) $\beta = \alpha\hat{\ }B$ **для некоторого** $B \in N$**.** Тогда существует последовательность $\Omega = \{(S =)A_0^0,\ A_0^1,\ \ldots,\ A_0^{l_0},\ A_1^0,\ A_1^1,\ \ldots,\ A_1^{l_1},\ \ldots,\ A_m^0,\ A_m^1,\ \ldots,\ A_m^{l_m}(= B)\}$, удовлетворяющая условиям $A_i^{j+1} \in \delta(A_i^j, \varepsilon)$ $(0 \leqslant j < l_i,\ 0 \leqslant i \leqslant m)$ и $A_{i+1}^0 \in \delta(A_i^{l_i}, a_{i+1})$ $(0 \leqslant i < m)$; кроме того, $P(B, A)$, а следовательно, $A \in \delta(B, \varepsilon)$; в качестве искомой последовательности следует взять Ω, A.

2) $\beta = \alpha_1\hat{\ }B$**, где** $B \in N$ **и** $\alpha = \alpha_1\hat{\ }a_m$**.** Тогда существует последовательность $\Omega = \{(S =)A_0^0,\ A_0^1,\ \ldots,\ A_0^{l_0},\ A_1^0,\ A_1^1,\ \ldots,\ A_1^{l_1},\ \ldots,\ A_{m-1}^0,\ A_{m-1}^1,\ \ldots,\ A_{m-1}^{l_{m-1}}(= B)\}$, удовлетворяющая условиям $A_i^{j+1} \in \delta(A_i^j, \varepsilon)$ $(0 \leqslant j < l_i,\ 0 \leqslant i \leqslant m-1)$ и $A_{i+1}^0 \in \delta(A_i^{l_i}, a_{i+1})$ $(0 \leqslant i < m-1)$; кроме того, $P(B, a_m\hat{\ }A)$, а следовательно, $A \in \delta(B, a_m)$; в качестве искомой последовательности следует взять Ω, A.

3) **Других вариантов для** β **не существует** (почему?)

Регулярные языки и регулярные грамматики

Доказательство (продолжение).

(1⇐) Доказывать будем индукцией по длине последовательности. В самом деле, $S \Rightarrow^*_{\mathfrak{G}} S(=\varepsilon\hat{\ }S)$, поэтому база индукции выполняется.

Пусть теперь найдётся последовательность $(S=)A_0, A_1, \ldots, A_k, A_{k+1} \in N$, удовлетворяющая соотношению (1). По предположению индукции, $S \Rightarrow^*_{\mathfrak{G}} \alpha_k\hat{\ }A_k$ для соответствующего $\alpha_k \in \Sigma^*$. Далее, предположим, что $A_{k+1} \in \delta(A_k, a)$ **используется в соотношении (1), где** $a \in \Sigma \cup \{\varepsilon\}$. Тогда $P(A_k, a\hat{\ }A_{k+1})$ и, следовательно, $\alpha_k\hat{\ }A_k \Rightarrow_{\mathfrak{G}} \alpha_k\hat{\ }(a\hat{\ }A_{k+1})$; таким образом, имеем $S \Rightarrow^*_{\mathfrak{G}} (\alpha_k\hat{\ }a)\hat{\ }A_{k+1}$.

Теперь для завершения доказательства (⇐) достаточно проверить справедливость следующего условия $(\alpha = a_1 a_2 \ldots a_n \in \Sigma^*)$:

Регулярные языки и регулярные грамматики

Доказательство (продолжение).

$[S \Rightarrow^*_{\mathfrak{G}} \alpha] \Leftrightarrow$
$\Leftrightarrow$ [найдётся последовательность $(S =)A_0^0, A_0^1, \ldots, A_0^{k_0},$
$A_1^0, A_1^1, \ldots, A_1^{k_1}, \ldots, A_n^0, A_n^1, \ldots, A_n^{k_n}(= *)$ (2)
такая, что $A_i^{j+1} \in \delta(A_i^j, \varepsilon)$, $0 \leqslant j < k_i$, $0 \leqslant i \leqslant n$, и
$A_{i+1}^0 \in \delta(A_i^{k_i}, a_{i+1}), 0 \leqslant i < n$].

(2$\Rightarrow$) Доказывать будем индукцией по построению отношения $\Rightarrow^*_{\mathfrak{G}}$. Если $P(S, \varepsilon)$ (что равносильно тому, что $S \Rightarrow_G \varepsilon$), то $* \in \delta(S, \varepsilon)$ и, следовательно, $\varepsilon \in \mathrm{L}(\mathfrak{A})$. Пусть теперь $S \Rightarrow^*_{\mathfrak{G}} \beta(\in (N \cup \Sigma)^*)$ и $\beta \Rightarrow_{\mathfrak{G}} \alpha$. Возможны только два случая.
1) $\beta = \alpha\hat{\ }A$ **для подходящего** $A \in N$ **и** $P(A, \varepsilon)$**.** Из (1) следует существование последовательности $(S =)A_0$, A_1, ..., $A_k(= A)$, подтверждающей, что к этому моменту слово α считывается автоматом $\mathfrak{A}$. Далее, имеем $* \in \delta(A, \varepsilon)$, поэтому $\alpha \in \mathrm{L}(\mathfrak{A})$.

Лекция А4 Регулярные языки, грамматики

Вадим Пузаренко

Теорема о накачке

ДКА: алфавиты $\{0\}$ и $\{0;1\}$

Грамматики: общие сведения

Регулярные грамматики

Доказательство (продолжение).

2) $\alpha = \alpha_1\hat{\ }a$, $\beta = \alpha_1\hat{\ }A$ **для подходящих** $A \in N$, $a \in \Sigma$, $\alpha_1 \in \Sigma^*$ **и** $P(A, a)$. Из (1) следует существование последовательности $(S =)A_0, A_1, \ldots, A_k(= A)$, подтверждающей, что к этому моменту слово α_1 считывается автоматом $\mathfrak{A}$. Далее, имеем $* \in \delta(A, a)$, поэтому $\alpha = \alpha_1\hat{\ }a \in \mathrm{L}(\mathfrak{A})$.

Доказательство (продолжение).

2) $\alpha = \alpha_1\hat{\ }a$, $\beta = \alpha_1\hat{\ }A$ **для подходящих** $A \in N$, $a \in \Sigma$, $\alpha_1 \in \Sigma^*$ **и** $P(A, a)$. Из (1) следует существование последовательности $(S =)A_0, A_1, \ldots, A_k(= A)$, подтверждающей, что к этому моменту слово α_1 считывается автоматом $\mathfrak{A}$. Далее, имеем $* \in \delta(A, a)$, поэтому $\alpha = \alpha_1\hat{\ }a \in \mathrm{L}(\mathfrak{A})$.

(2⇐) Пусть последовательность состояний $(S =)A_0, A_1, \ldots, A_k, A_{k+1}(= *)$ удовлетворяет условию определения $\alpha \in \mathrm{L}(\mathfrak{A})$. Так как $\delta(*, a) = \varnothing$ для любого $a \in \Sigma \cup \{\varepsilon\}$, имеем $A_k \in N$ и $* \in \delta(A_k, b)$ для некоторого $b \in \Sigma \cup \{\varepsilon\}$ такого, что $\alpha = \alpha_1\hat{\ }b$, где α_1 таково, что $S \Rightarrow^*_{\mathfrak{G}} \alpha_1\hat{\ }A_k$, что следует из (1). Кроме того, $P(A_k, b)$; таким образом, $S \Rightarrow^*_{\mathfrak{G}} \alpha$.

Регулярные языки и регулярные грамматики

Доказательство (продолжение).

($\Rightarrow$) Пусть $L = \mathrm{L}(\mathfrak{A})$, где $\mathfrak{A} = (Q, \Sigma, \delta, q_0, F)$ — ДКА. Определим регулярную грамматику $\mathfrak{G} = (Q, \Sigma, P, q_0)$ так, что $P = \{\langle q, aq'\rangle \mid q \in Q,\ a \in \Sigma,\ q' = \delta(q, a)\} \cup \{\langle q, \varepsilon\rangle \mid q \in F\}$.
Докажем, что $L = \mathrm{L}(\mathfrak{G})$. Сначала индукцией по $\mathrm{lh}(\alpha)$ докажем следующее соотношение ($q \in Q$):

$$q_0 \Rightarrow^*_{\mathfrak{G}} \alpha\hat{\ }q \Leftrightarrow \delta^*(q_0, \alpha) = q. \qquad (3)$$

Если $\alpha = \varepsilon$, то справедливы соотношения $q_0 \Rightarrow^*_{\mathfrak{G}} q_0 (= \varepsilon\hat{\ }q_0)$ и $\delta^*(q_0, \varepsilon) = q_0$.
Пусть теперь $\alpha = \alpha_1\hat{\ }a$; по индукционному предположению, $q_0 \Rightarrow^*_{\mathfrak{G}} \alpha_1\hat{\ }q_1 \Leftrightarrow \delta^*(q_0, \alpha_1) = q_1$. Далее, имеем $P(q_1, aq) \Leftrightarrow q = \delta(q_1, a)$.
(3$\Leftarrow$) Пусть теперь $q = \delta^*(q_0, \alpha) = \delta(\delta^*(q_0, \alpha_1), a)$; тогда $q_0 \Rightarrow^*_{\mathfrak{G}} \alpha_1\hat{\ }q_1$ и $P(q_1, aq)$, а следовательно, $\alpha_1\hat{\ }q_1 \Rightarrow_{\mathfrak{G}} \alpha_1\hat{\ }aq (= \alpha\hat{\ }q)$; таким образом, $q_0 \Rightarrow^*_{\mathfrak{G}} \alpha\hat{\ }q$.

Доказательство (продолжение).

(3⇒) В обратную сторону, пусть имеет место $q_0 \Rightarrow^*_{\mathfrak{G}} \alpha\hat{\ }q$; тогда выполняются соотношения $q_0 \Rightarrow^*_{\mathfrak{G}} \beta$ и $\beta \Rightarrow_{\mathfrak{G}} \alpha\hat{\ }q$; из определения грамматики $\mathfrak{G}$ вытекает, что $\beta = \alpha_1\hat{\ }q_1$ и $P(q_1, aq)$. Следовательно, $q_1 = \delta^*(q_0, \alpha_1)$ и $q = \delta(q_1, a)$. Таким образом, $q = \delta(q_1, a) = \delta(\delta^*(q_0, \alpha_1), a) = \delta^*(q_0, \alpha_1\hat{\ }a) = \delta^*(q_0, \alpha)$.
Тем самым, соотношение (3) выполняется для всех $\alpha \in \Sigma^*$ и $q \in Q$. Для завершения доказательства достаточно проверить следующее ($\alpha \in \Sigma^*$):

$$(\alpha \in \mathrm{L}(\mathfrak{G}) \Leftrightarrow)[q_0 \Rightarrow^*_{\mathfrak{G}} \alpha] \Leftrightarrow \delta^*(q_0, \alpha) \in F(\Leftrightarrow \alpha \in \mathrm{L}(\mathfrak{A})). \qquad (4)$$

(4⇐) Пусть $\delta^*(q_0, \alpha) = q \in F$; тогда из (3) следует, что $q_0 \Rightarrow^*_{\mathfrak{G}} \alpha\hat{\ }q$, а из определения грамматики $\mathfrak{G}$ — $P(q, \varepsilon)$; следовательно, $\alpha\hat{\ }q \Rightarrow_{\mathfrak{G}} \alpha(= \alpha\hat{\ }\varepsilon)$; таким образом, $q_0 \Rightarrow^*_{\mathfrak{G}} \alpha$.

Регулярные языки и регулярные грамматики

Доказательство (окончание).

(4⇒) В обратную сторону, пусть $q_0 \Rightarrow^*_{\mathfrak{G}} \alpha$; из определения грамматики $\mathfrak{G}$ следует, что $q_0 \Rightarrow^*_{\mathfrak{G}} \alpha\hat{\ }q$ и $P(q,\varepsilon)$ для подходящего $q \in Q$. Следовательно, $q \in F$ и $\delta^*(q_0,\alpha) = q$. $\square$

Регулярные языки и регулярные грамматики

Доказательство (окончание).

(4⇒) В обратную сторону, пусть $q_0 \Rightarrow^*_{\mathfrak{G}} \alpha$; из определения грамматики $\mathfrak{G}$ следует, что $q_0 \Rightarrow^*_{\mathfrak{G}} \alpha\hat{\ }q$ и $P(q,\varepsilon)$ для подходящего $q \in Q$. Следовательно, $q \in F$ и $\delta^*(q_0,\alpha) = q$. □

Замечание А4.1.

Часто в литературе встречается вариант регулярной грамматики, когда продукции имеют вид только $P(A,a)$ и $P(A,a\hat{\ }B)$. В этом случае теорема А4.4 справедлива для языков, не содержащих ε (обосновать).

Спасибо за внимание.