Представления регулярных языков. Критерии регулярности

Теория формальных языков

2021 г.

Лемма о накачке

Пусть n — мощность множества нетерминалов регулярной грамматики G.

- Каждое применение правила — +1 к длине порождаемого слова.
- Правила могут применяться независимо от контекста.

Лемма о накачке

Пусть n — мощность множества нетерминалов регулярной грамматики G.

- Каждое применение правила — +1 к длине порождаемого слова.
- Правила могут применяться независимо от контекста.

Рассмотрим слово $w \in L(G)$, $|w| \geqslant n+1$. Оно получается применением не меньше, чем $n+1$ правил $\Rightarrow$ после применения хотя бы двух из них в сентенциальной форме справа будет стоять один и тот же нетерминал A.

Лемма о накачке

Рассмотрим слово $w \in \mathrm{L}(G)$, $|w| \geqslant n+1$. Оно получается применением не меньше, чем $n+1$ правил $\Rightarrow$ после применения хотя бы двух из них в сентенциальной форме справа будет стоять один и тот же нетерминал A.

$$\underbrace{S \rightarrow \cdots \rightarrow \Phi\ A \rightarrow \cdots \rightarrow \Phi\ \Psi\ A}_{\text{не больше } n \text{ шагов}} \rightarrow \cdots \rightarrow \Phi\ \Psi\ \Theta$$

$$\Downarrow$$

$$|\Phi| + |\Psi| \leqslant n$$

Лемма о накачке

Рассмотрим слово $w \in L(G)$, $|w| \geqslant n+1$. Оно получается применением не меньше, чем $n+1$ правил $\Rightarrow$ после применения хотя бы двух из них в сентенциальной форме справа будет стоять один и тот же нетерминал A.

Известно, что $|\Phi| + |\Psi| \leqslant n$.

$$\underbrace{S \to \dots \to \Phi\, A}_{\rho_1:\,\text{вывод } \Phi A \text{ из } S} \overbrace{ \to \dots \to \Phi\, \Psi}^{\rho_2:\,\text{вывод } \Psi A \text{ из } A}\ \underbrace{A \to \dots \to \Phi\, \Psi\, \Theta}_{\rho_3:\,\text{вывод } \Theta \text{ из } A}$$

Все выводы вида $\rho_1 \, (\rho_2)^* \, \rho_3$ допустимы в $G \Rightarrow$ $\forall k(\Phi \, \Psi^k \, \Theta \in L(G))$.

Лемма о накачке

Утверждение

Если G — регулярная, то существует такое $n \in \mathbb{N}$, что $\forall w(w \in L(G)\ \&\ |w| > n \Rightarrow \exists w_1, w_2, w_3(|w_2| > 0\ \&$ $|w_1| + |w_2| \leqslant n\ \&\ w = w_1\ w_2\ w_3\ \&\ \forall k(k \geqslant 0 \Rightarrow$ $w_1\ {w_2}^k\ w_3 \in L(G)))).$

Известно, что $|\Phi| + |\Psi| \leqslant n$.

$$\underbrace{S \to \cdots \to \Phi}_{\rho_1:\ \text{вывод}\ \Phi A\ \text{из}\ S}\ \overbrace{A \to \cdots \to \Phi\ \Psi}^{\rho_2:\ \text{вывод}\ \Psi A\ \text{из}\ A}\ \underbrace{A \to \cdots \to \Phi\ \Psi}_{\rho_3:\ \text{вывод}\ \Theta\ \text{из}\ A}\ \Theta$$

Все выводы вида $\rho_1\ (\rho_2)^*\ \rho_3$ допустимы в $G \Rightarrow$ $\forall k(\Phi\ \Psi^k\ \Theta \in L(G))$.

Примеры применения леммы о накачке

Обозначим обращение (reversal) слова w как w^R.
Рассмотрим язык $L = \{w\,w^R \mid w \in \Sigma^+\}$.

Пусть длина накачки — n. Рассмотрим слово $b^{n+1} a\, a\, b^{n+1} \in L$. Поскольку $|\Phi| + |\Psi| \leqslant n$, то $\Psi = b^k$, $k \geqslant 1$. Но $b^m a\, a\, b^n \notin L$, если $m \neq n$. Поэтому L — не регулярный.

Примеры применения леммы о накачке

Обозначим обращение (reversal) слова w как w^R. Рассмотрим язык $L = \{w\,w^R \,|\, w \in \Sigma^+\}$.

Пусть длина накачки — n. Рассмотрим слово $b^{n+1} a\, a\, b^{n+1} \in L$. Поскольку $|\Phi| + |\Psi| \leqslant n$, то $\Psi = b^k$, $k \geqslant 1$. Но $b^m a\, a\, b^n \notin L$, если $m \neq n$. Поэтому L — не регулярный.

Рассмотрим язык $L' = \{a^n b^m \,|\, n \neq m\}$.

Пусть длина накачки — n. Рассмотрим множество слов $a^n b^{n+n!} \in L'$. Поскольку $|\Phi| + |\Psi| \leqslant n$, то $\Psi = a^k$, $k \geqslant 1$. Но для всех $k \leqslant n$ $\exists v (n + k * v = n + n!)$. Поэтому слово вида $a^{n+n!} b^{n+n!} \in L$, что абсурдно. Следовательно, L' не является регулярным.

Нерегулярные языки

Пусть $L = \{w \mid |w|_a = |w|_b\}$. Все слова вида $a^k b^k$ принадлежат L. Пусть длина накачки равна n. Рассмотрим слово $a^n b^n$. Поскольку $|\Phi| + |\Psi| \leqslant n$, то $\Psi = a^k$, $k > 0$. Но слова $a^{n+k*i} b^n$ не принадлежат L.

Совпадает ли L с языком правильных скобочных последовательностей P (язык Дика)? Если да, доказать. Если нет, исследовать язык $L \setminus P$. Регулярен ли он?

Анализ на достаточность

Гипотеза

G — регулярная $\overset{???}{\Longleftrightarrow}$ существует такое $n \in \mathbb{N}$, что $\forall w(w \in L(G) \;\&\; |w| > n \Rightarrow \exists w_1, w_2, w_3(|w_2| > 0 \;\&\; |w_1| + |w_2| \leqslant n \;\&\; w = w_1\ w_2\ w_3 \;\&\; \forall k(k \geqslant 0 \Rightarrow w_1\ {w_2}^k\ w_3 \in L(G))))$.

Анализ на достаточность

Гипотеза

G — регулярная $\overset{???}{\Longleftrightarrow}$ существует такое $n \in \mathbb{N}$, что $\forall w(w \in L(G)\ \&\ |w| > n \Rightarrow \exists w_1, w_2, w_3(|w_2| > 0\ \&\ |w_1| + |w_2| \leqslant n\ \&\ w = w_1\ w_2\ w_3\ \&\ \forall k(k \geqslant 0 \Rightarrow w_1\ {w_2}^k\ w_3 \in L(G)))).$

Рассмотрим язык $L = \{w\,w^R\,z | w \in \Sigma^+\ \&\ z \in \Sigma^+\}$ и $n = 4$.

- Если $|w| = 1$, тогда можно разбить слово $w\,w^R\,z$ так: $\Phi = w\,w^R$, $\Psi = z[1]$, $\Theta = z[2..|z|]$. Тогда для всех k $\Phi\,\Psi^k\,\Theta \in L$.
- Если $|w| \geqslant 2$, тогда разбиваем так: $\Phi = \varepsilon$, $\Psi = w[1]$, $\Theta = w[2..|w|]\,w^R\,z$. Слова $w[2..|w|]\,w^R\,z$ и $w[1]^k\,w[2..|w|]\,w^R\,z$ при $k \geqslant 2$ также принадлежат L.

Смысл леммы о накачке

Структура доказательства указывает, что длина накачки n регулярного языка L не больше (возможно, меньше) числа нетерминалов в минимальной грамматике для L.

Рассмотрим $L = a|b|(a\{a|b\}^*a)|(b\{a|b\}^*b)$. Если выбрать $n = 2$, то в качестве Ψ можно взять вторую букву слова из L. Пусть G имеет два нетерминала S, T и распознаёт L. Если G содержит правила $S \to aT$ и $S \to bT$ (или $S \to aS$, $S \to bS$), то для некоторого непустого z слова вида az и bz будут либо оба принадлежать L, либо нет, чего не может быть. Значит, G содержит либо пару $S \to aT$, $S \to bS$, либо пару $S \to bT$, $S \to aS$. Рассмотрим первый случай. Тогда для некоторого непустого z имеем $az \in L \Leftrightarrow b^+az \in L$, что абсурдно.

Недетерминированные КА

Определение

Недетерминированный конечный автомат (NFA) — это пятёрка $\mathscr{A} = \langle Q, \Sigma, q_0, F, \delta \rangle$, где:

- Q — множество состояний;
- Σ — алфавит терминалов;
- δ — множество правил перехода вида $\langle q_i, (a_i|\varepsilon), M_i \rangle$, где $q_i \in Q$, $a_i \in \Sigma$, $M_i \in 2^Q$;
- $q_0 \in Q$ — начальное состояние;
- $F \subseteq Q$ — множество конечных состояний.

Сокращаем: $\langle q_1, a, q_2 \rangle \in \delta \Leftrightarrow \langle q_1, a, M \rangle \in \delta \;\&\; q_2 \in M$.

Недетерминированные КА

Определение

Недетерминированный конечный автомат (NFA) — это пятёрка $\mathscr{A} = \langle Q, \Sigma, q_0, F, \delta \rangle$.

- $q \xrightarrow{\varepsilon} q' \Leftrightarrow (q = q') \vee \exists p_1, \ldots, p_k(\langle q, \varepsilon, p_1 \rangle \in \delta \mathbin{\&} \langle p_k, \varepsilon, q' \rangle \in \delta \mathbin{\&} \forall i, 1 \leqslant i < k \langle p_i, \varepsilon, p_{i+1} \rangle \in \delta)$.
- $q \xrightarrow{a} q' \Leftrightarrow \exists p, p'(q \xrightarrow{\varepsilon} p \mathbin{\&} \langle p, a, p' \rangle \in \delta \mathbin{\&} p' \xrightarrow{\varepsilon} q')$.
- $q \xrightarrow{a_1 \ldots a_k} q' \Leftrightarrow \exists p_1, \ldots, p_{k-1}(q \xrightarrow{a_1} p_1 \mathbin{\&} p_{k-1} \xrightarrow{a_k} q' \mathbin{\&} \forall i, 1 \leqslant i < k-1 (p_i \xrightarrow{a_{i+1}} p_{i+1}))$.

Недетерминированные КА

Определение

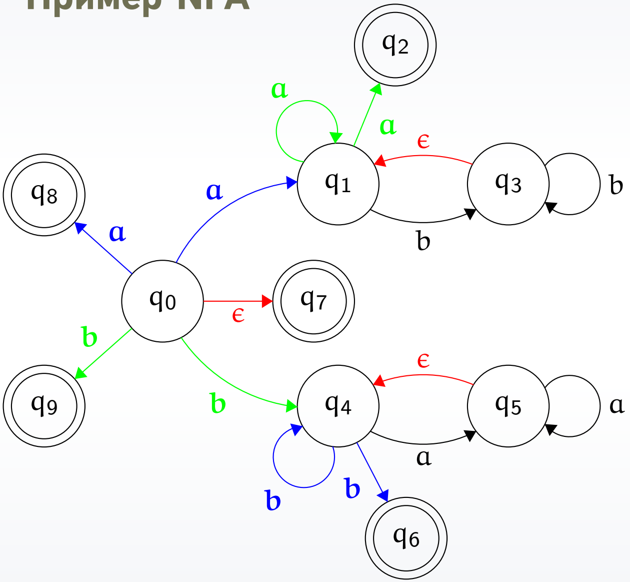
Недетерминированный конечный автомат (NFA) — это пятёрка $\mathscr{A} = \langle Q, \Sigma, q_0, F, \delta \rangle$.

- $q \xrightarrow{\varepsilon} q' \Leftrightarrow (q = q') \vee \exists p_1, \dots, p_k(\langle q, \varepsilon, p_1 \rangle \in \delta \;\&\; \langle p_k, \varepsilon, q' \rangle \in \delta \;\&\; \forall i, 1 \leqslant i < k \langle p_i, \varepsilon, p_{i+1} \rangle \in \delta)$.
- $q \xrightarrow{a} q' \Leftrightarrow \exists p, p'(q \xrightarrow{\varepsilon} p \;\&\; \langle p, a, p' \rangle \in \delta \;\&\; p' \xrightarrow{\varepsilon} q')$.
- $q \xrightarrow{a_1 \dots a_k} q' \Leftrightarrow \exists p_1, \dots, p_{k-1}(q \xrightarrow{a_1} p_1 \;\&\; p_{k-1} \xrightarrow{a_k} q' \;\&\; \forall i, 1 \leqslant i < k-1(p_i \xrightarrow{a_{i+1}} p_{i+1}))$.

Определение

Язык L, распознаваемый NFA $\mathscr{A}$ — это множество слов $\{w \mid \exists q \in F(q_0 \xrightarrow{w} q)\}$.

Пример NFA



Детерминированный КА

Определение

Детерминированный конечный автомат (DFA) — это пятёрка $\mathscr{A} = \langle Q, \Sigma, q_0, F, \delta \rangle$, где:

- Q — множество состояний;
- Σ — алфавит терминалов;
- δ — множество правил перехода вида $\langle q_i, a_i, q_j \rangle$, где $q_i, q_j \in Q$, $a_i \in \Sigma$, причём $\forall q_i, a_i \exists q_j (\langle q_i, a_i, q_j \rangle \in \delta \,\&\, \forall q_k (\langle q_i, a_i, q_k \rangle \in \delta \Rightarrow q_k = q_j))$;
- $q_0 \in Q$ — начальное состояние;
- $F \subseteq Q$ — множество конечных состояний.

ε-переходов нет $\Rightarrow q \xrightarrow{a} q' \Leftrightarrow \langle q, a, q' \rangle \in \delta$.

Детерминированный КА

Определение

Детерминированный конечный автомат (DFA) — это пятёрка $\mathscr{A} = \langle Q, \Sigma, q_0, F, \delta \rangle$, где:

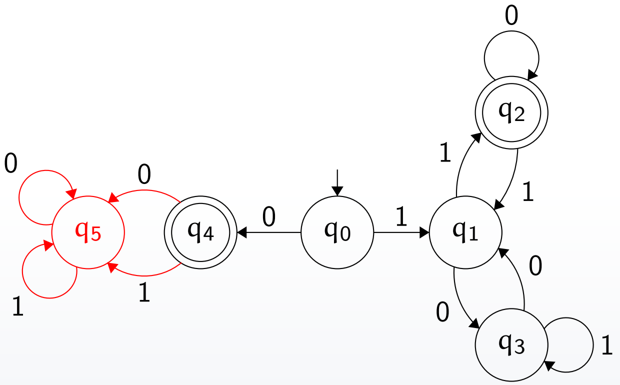
- Q — множество состояний;
- Σ — алфавит терминалов;
- δ — множество правил перехода вида $\langle q_i, a_i, q_j \rangle$, где $q_i, q_j \in Q$, $a_i \in \Sigma$, причём $\forall q_i, a_i \exists q_j (\langle q_i, a_i, q_j \rangle \in \delta \mathbin{\&} \forall q_k (\langle q_i, a_i, q_k \rangle \in \delta \Rightarrow q_k = q_j))$;
- $q_0 \in Q$ — начальное состояние;
- $F \subseteq Q$ — множество конечных состояний.

ε-переходов нет $\Rightarrow q \xrightarrow{a} q' \Leftrightarrow \langle q, a, q' \rangle \in \delta$.

Язык L, распознаваемый $\mathscr{A}$ — это множество слов $\{w \mid \exists q \in F (q_0 \xrightarrow{w} q)\}$.

Sink/trap state (состояние–ловушка)

«Ловушка» — не конечное состояние с переходами лишь в себя. Нужны для корректного задания DFA, но иногда по умолчанию не описываются.



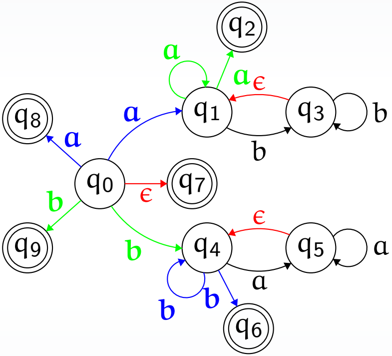
Детерминизация NFA

От $\mathscr{A}$ к $\mathrm{D}(\mathscr{A})$

Состояния DFA $\mathrm{D}(\mathscr{A})$ — это состояния $m_i \in 2^Q$, где Q — состояния NFA $\mathscr{A}$.

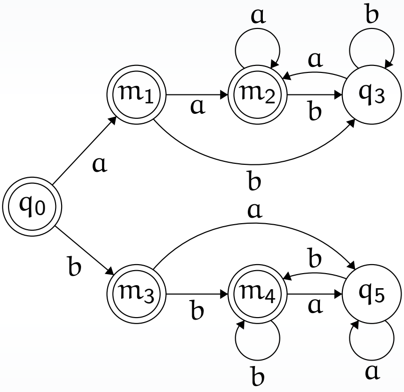
- $m_0 = \{q_i \mid q_0 \xrightarrow{\varepsilon} q_i\}$;
- $m_i \in F_D \Leftrightarrow \exists q_i, q_j \{q_i \in m_i \ \& \ q_j \in F(\mathscr{A}) \ \& \ q_i \xrightarrow{\varepsilon} q_j\}$;
- $\langle m, a, m' \rangle \in \delta_D \Leftrightarrow m' = \{q_i \mid \exists q_j \in m(q_j \xrightarrow{a} q_i)\}$.

Пример детерминизации



- $\{q_0\} \xrightarrow{a} \{q_1, q_8\}$, $\{q_0\} \xrightarrow{b} \{q_4, q_9\}$;
- $\{q_1, q_8\} \xrightarrow{a} \{q_1, q_2\}$,
  $\{q_1, q_8\} \xrightarrow{b} \{q_3\}$; $\{q_1, q_8\} \sim m_1$.
- $\{q_1, q_2\} \xrightarrow{a} \{q_1, q_2\}$,
  $\{q_1, q_2\} \xrightarrow{b} \{q_3\}$; $\{q_1, q_2\} \sim m_2$.
- $\{q_3\} \xrightarrow{a} \{q_1, q_2\}$, $\{q_3\} \xrightarrow{b} \{q_3\}$;
- $\{q_4, q_9\} \xrightarrow{b} \{q_4, q_6\}$,
  $\{q_4, q_9\} \xrightarrow{a} \{q_5\}$; $\{q_4, q_9\} \sim m_3$;
- $\{q_4, q_6\} \xrightarrow{b} \{q_4, q_6\}$,
  $\{q_4, q_6\} \xrightarrow{a} \{q_5\}$; $\{q_4, q_6\} \sim m_4$.
- $\{q_5\} \xrightarrow{b} \{q_4, q_6\}$, $\{q_5\} \xrightarrow{a} \{q_5\}$.

Пример детерминизации



- $\{q_0\} \xrightarrow{a} \{q_1, q_8\}$, $\{q_0\} \xrightarrow{b} \{q_4, q_9\}$;
- $\{q_1, q_8\} \xrightarrow{a} \{q_1, q_2\}$,
  $\{q_1, q_8\} \xrightarrow{b} \{q_3\}$; $\{q_1, q_8\} \sim m_1$.
- $\{q_1, q_2\} \xrightarrow{a} \{q_1, q_2\}$,
  $\{q_1, q_2\} \xrightarrow{b} \{q_3\}$; $\{q_1, q_2\} \sim m_2$.
- $\{q_3\} \xrightarrow{a} \{q_1, q_2\}$, $\{q_3\} \xrightarrow{b} \{q_3\}$;
- $\{q_4, q_9\} \xrightarrow{b} \{q_4, q_6\}$,
  $\{q_4, q_9\} \xrightarrow{a} \{q_5\}$; $\{q_4, q_9\} \sim m_3$;
- $\{q_4, q_6\} \xrightarrow{b} \{q_4, q_6\}$,
  $\{q_4, q_6\} \xrightarrow{a} \{q_5\}$; $\{q_4, q_6\} \sim m_4$.
- $\{q_5\} \xrightarrow{b} \{q_4, q_6\}$, $\{q_5\} \xrightarrow{a} \{q_5\}$.

Замыкания регулярных языков

Гомоморфизм над свободной полугруппой (множеством слов) полностью определяется значениями на буквах, поскольку по определению $h(a_1 \circ a_2 \circ \cdots \circ a_n) = h(a_1) \circ h(a_2) \circ \cdots \circ h(a_n)$. Здесь $\circ$ —конкатенация.

Утверждение

Пусть L — регулярный язык над Σ. Тогда регулярны:

- язык $\Sigma^* \setminus L$;
- для любого гомоморфизма h язык $\{h(w) \,|\, w \in L\}$;
- для любого гомоморфизма h язык $\{w \,|\, h(w) \in L\}$.

Замыкания регулярных языков

Утверждение

Пусть L — регулярный язык над Σ. Тогда регулярны:

- язык $\Sigma^* \setminus \mathrm{L}$;
- для любого гомоморфизма h язык $\{\mathrm{h}(w) \mid w \in \mathrm{L}\}$;
- для любого гомоморфизма h язык $\{w \mid \mathrm{h}(w) \in \mathrm{L}\}$.

Рассмотрим DFA $\mathscr{A} = \langle \mathrm{Q}, \Sigma, \mathrm{q}_0, \mathrm{F}, \delta \rangle$, распознающий L.
Построим $\mathscr{A}' = \langle \mathrm{Q}, \Sigma, \mathrm{q}_0, \mathrm{Q} \setminus \mathrm{F}, \delta \rangle$. Тогда
$w \notin \mathrm{L} \Leftrightarrow w \in \mathrm{L}(\mathscr{A}')$.

Замыкания регулярных языков

Утверждение

Пусть L — регулярный язык над Σ. Тогда регулярны:

- язык $\Sigma^* \setminus L$;
- для любого гомоморфизма h язык $\{h(w) \mid w \in L\}$;
- для любого гомоморфизма h язык $\{w \mid h(w) \in L\}$.

Рассмотрим регулярное выражение R такое, что $L(R) = L$. Заменим в нём все $a_i \in \Sigma$ на $h(a_i)$. Полученное таким образом выражение R' также регулярно, причём $L(R') = h(L)$.

Замыкания регулярных языков

Утверждение

Пусть L — регулярный язык над Σ. Тогда регулярны:

- язык $\Sigma^* \setminus L$;
- для любого гомоморфизма h язык $\{h(w) \mid w \in L\}$;
- для любого гомоморфизма h язык $\{w \mid h(w) \in L\}$.

Рассмотрим DFA $\mathscr{A} = \langle Q, \Sigma, q_0, F, \delta \rangle$, распознающий L. Построим $\mathscr{A}' = \langle Q, \Sigma, q_0, F, \delta' \rangle$ такой, что $\langle q_i, a, q_j \rangle \in \delta' \Leftrightarrow q_i \xrightarrow{h(a)} q_j$ в исходном автомате $\mathscr{A}$.

Примеры

Рассмотрим язык $L' = \{a^n b^m \mid n \neq m\}$.
Предположим, L' регулярен. Тогда $a^* b^* \setminus L' = \{a^n b^n\}$ также регулярен, а мы знаем, что это не так. $\perp$

Примеры

Рассмотрим язык $L' = \{a^n b^m \mid n \neq m\}$.

Предположим, L' регулярен. Тогда $a^* b^* \setminus L' = \{a^n b^n\}$ также регулярен, а мы знаем, что это не так. $\perp$

Рассмотрим язык $L^f = \{(abaabb)^n b^n\}$.

Попытка доказать его нерегулярность леммой о накачке породит перебор по накачиваемым строкам $(abaabb)^+$, $(abaabb)^* a$, $(abaabb)^* ab$,$(abaabb)^* aba$, $(abaabb)^* abaa$, $\ldots$. Рассмотрим гомоморфизм $h(a) = abaabb$, $h(b) = b$. $h^{-1}(L^f) = \{a^n b^n\}$, который был бы регулярен, если бы L^f был регулярен. $\perp$

Эквивалентность слов в DFA

Пусть дан DFA $\mathscr{A}$. Положим
$w_1 \equiv_{\mathscr{A}} w_2 \Leftrightarrow \exists q_i(q_0 \xrightarrow{w_1} q_i \,\&\, q_0 \xrightarrow{w_2} q_i)$.

Если $w_1 \equiv_{\mathscr{A}} w_2$, тогда $\forall z(w_1 z \in L(\mathscr{A}) \Leftrightarrow w_2 z \in L(\mathscr{A}))$.

Эквивалентность слов в DFA

Пусть дан DFA $\mathscr{A}$. Положим
$w_1 \equiv_{\mathscr{A}} w_2 \Leftrightarrow \exists q_i (q_0 \xrightarrow{w_1} q_i \; \& \; q_0 \xrightarrow{w_2} q_i)$.

Если $w_1 \equiv_{\mathscr{A}} w_2$, тогда $\forall z (w_1 z \in L(\mathscr{A}) \Leftrightarrow w_2 z \in L(\mathscr{A}))$.
Рассмотрим более общее отношение. Положим

$w_1 \equiv_L w_2 \Leftrightarrow \forall z (w_1 z \in L \Leftrightarrow w_2 z \in L)$. Это отношение разбивает L на классы эквивалентности.

Теорема Майхилла-Нероуда

Язык L регулярен тогда и только тогда, когда множество его классов эквивалентности по $\equiv_L$ конечно.

Критерий регулярности языка

Теорема Майхилла-Нероуда

Язык L регулярен тогда и только тогда, когда множество классов эквивалентности по $\equiv_{\mathrm{L}}$ конечно.

Критерий регулярности языка

Теорема Майхилла-Нероуда

Язык L регулярен тогда и только тогда, когда множество классов эквивалентности по $\equiv_L$ конечно.

$\Rightarrow$: Пусть L регулярен. Тогда он порождается некоторым DFA $\mathscr{A}$ с конечным числом состояний N. Значит, множество $\{q_i \mid q_0 \xrightarrow{w} q_i\}$ конечно, а для каждых двух w_1, w_2 таких, что $q_0 \xrightarrow{w_1} q_i$ и $q_0 \xrightarrow{w_2} q_i$, выполняется $w_1 \equiv_L w_2$.

Критерий регулярности языка

Теорема Майхилла-Нероуда

Язык L регулярен тогда и только тогда, когда множество классов эквивалентности по $\equiv_L$ конечно.

$\Leftarrow$: Пусть все слова в Σ^* принадлежат N классам эквивалентности $A_1, \dots, A_n$ по $\equiv_L$. Построим по ним DFA $\mathscr{A}$, распознающий L. Классы A_i объявим состояниями.

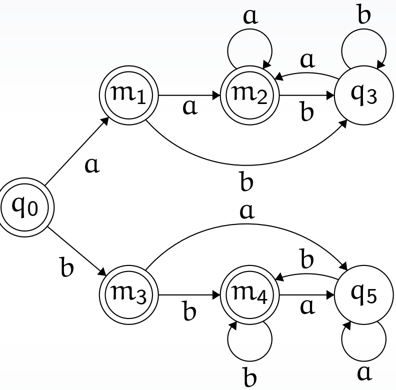
- Начальным состоянием объявим класс эквивалентности A_0 такой, что $\varepsilon \in A_0$.
- Конечными объявим такие A_j, что $\forall w \in A_j (w \in L)$.
- Если $w \in A_i$, $w\, a_k \in A_j$, тогда добавляем в δ правило $\langle A_i, a_k, A_j \rangle$. $\forall w_1, w_2 \in A_i$, $w_1 a_k$ и $w_2 a_k$ всегда принадлежат одному и тому же A_j.

Минимизация DFA

1. Построим таблицу всех двухэлементных множеств $\{q_i, q_j\}$, $q_i, q_j \in Q$.
2. Пометим все множества $\{q_i, q_j\}$ такие, что одно из q_i, q_j из F, а второе нет.
3. Пометим все множества $\{q_i, q_j\}$ такие, что $\exists a(q_i \xrightarrow{a} q_1' \;\&\; q_j \xrightarrow{a} q_2' \;\&\; \{q_1', q_2'\}$ — помеченная пара).
4. Продолжаем шаг 3, пока не будет появляться новых помеченных пар.

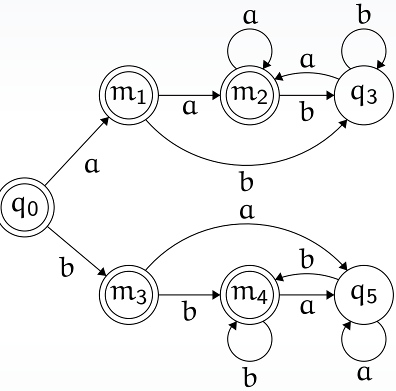
Пары, оставшиеся непомеченными, можно объединить.

Пример детерминизации



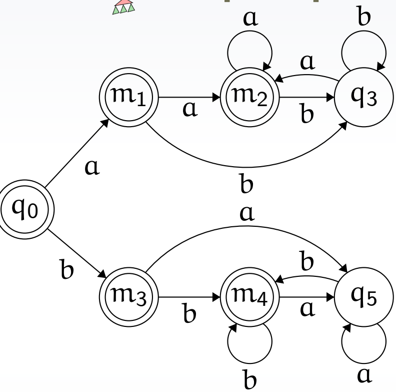
m_1						
m_2						
q_3						
m_3						
m_4						
q_5						
	q_0	m_1	m_2	q_3	m_3	m_4

Пример детерминизации



m_1						
m_2						
q_3	✓	✓	✓			
m_3				✓		
m_4				✓		
q_5	✓	✓	✓		✓	✓
	q_0	m_1	m_2	q_3	m_3	m_4

Пример детерминизации

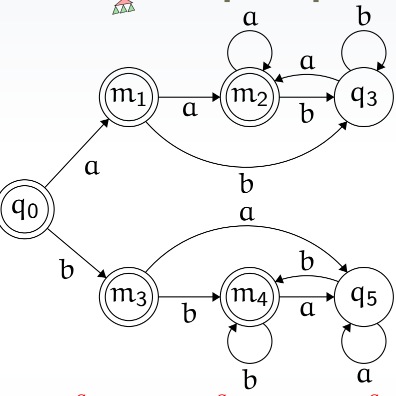


m_1						
m_2						
q_3	✓	✓	✓			
m_3				✓		
m_4				✓		
q_5	✓	✓	✓		✓	✓
	q_0	m_1	m_2	q_3	m_3	m_4

$q_0 \xrightarrow{a} m_1, m_1 \xrightarrow{a} m_2$ $\quad$ $q_0 \xrightarrow{b} m_3, m_1 \xrightarrow{b} q_3$ $\quad$ $\{m_1, m_2\} \xrightarrow{a} m_2$ $\quad$ $\{m_1, m_2\} \xrightarrow{b} q_3$

$q_0 \xrightarrow{a} m_1, m_2 \xrightarrow{a} m_2$ $\quad$ $q_0 \xrightarrow{b} m_3, m_2 \xrightarrow{b} q_3$

Пример детерминизации

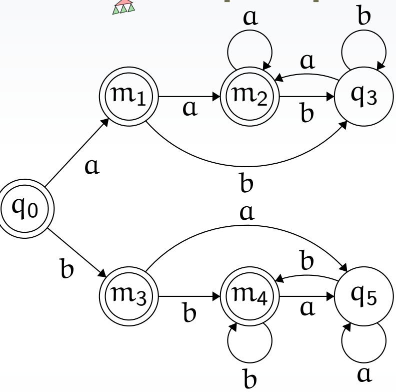


m_1	✓					
m_2	✓					
q_3	✓	✓	✓			
m_3				✓		
m_4				✓		
q_5	✓	✓	✓		✓	✓
	q_0	m_1	m_2	q_3	m_3	m_4

$q_0 \xrightarrow{a} m_1, m_3 \xrightarrow{a} q_5$ $\quad$ $q_0 \xrightarrow{a} m_1, m_4 \xrightarrow{a} q_5$ $\quad$ $m_1 \xrightarrow{a} m_2, m_3 \xrightarrow{a} q_5$

$m_2 \xrightarrow{a} m_2, m_3 \xrightarrow{a} q_5$ $\quad$ $m_1 \xrightarrow{a} m_2, m_4 \xrightarrow{a} q_5$ $\quad$ $m_2 \xrightarrow{a} m_2, m_4 \xrightarrow{a} q_5$

Пример детерминизации

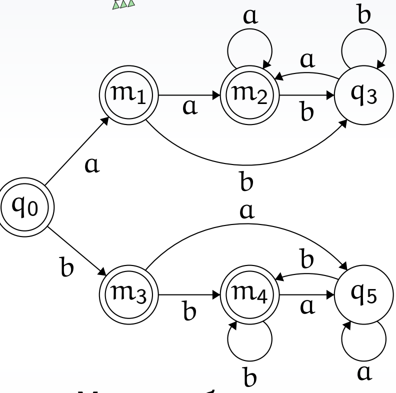


m_1	✓					
m_2	✓					
q_3	✓	✓	✓			
m_3	✓	✓	✓	✓		
m_4	✓	✓	✓	✓		
q_5	✓	✓	✓		✓	✓
	q_0	m_1	m_2	q_3	m_3	m_4

$\{m_3, m_4\} \xrightarrow{a} q_5$ $\{m_3, m_4\} \xrightarrow{b} m_4$

$q_3 \xrightarrow{a} m_2, q_5 \xrightarrow{a} m_4$

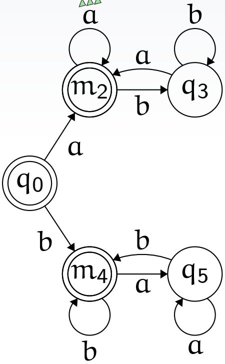
Пример детерминизации



m_1	✓					
m_2	✓					
q_3	✓	✓	✓			
m_3	✓	✓	✓	✓		
m_4	✓	✓	✓	✓		
q_5	✓	✓	✓	✓	✓	✓
	q_0	m_1	m_2	q_3	m_3	m_4

Можно объединить состояния m_1 и m_2 и состояния m_3 и m_4.

Пример детерминизации



m_1	✓					
m_2	✓					
q_3	✓	✓	✓			
m_3	✓	✓	✓	✓		
m_4	✓	✓	✓	✓		
q_5	✓	✓	✓	✓	✓	✓
	q_0	m_1	m_2	q_3	m_3	m_4

Меньше чем пятью состояниями не обойтись. Рассмотрим слова ε, a, b, ab, ba. Каждые два из них различаются по $\equiv_L$ при выборе одного из трёх z: ε, a или b.

Применение теоремы М.– N.

Задача

Дан язык L. Показать, что он не регулярен, пользуясь теоремой Майхилла–Нероуда.

Стандартный подход

1. Подобрать бесконечную последовательность префиксов $w_1, \ldots, w_n, \ldots$.
2. Подобрать бесконечную последовательность суффиксов $z_1, \ldots, z_n, \ldots$, такую, что $w_i +\!\!+\, z_i \in L$.
3. Доказать, что в таблице конкатенаций все строки различны (значит, $\forall i, j\, \exists k (w_i z_k \in L \,\&\, w_j z_k \notin L)$).

Диагональная конструкция (условие $w_i +\!\!+\, z_i \in L$) — одна из многих возможных, обычно она довольно удобна.

Диагональная конструкция

Рассмотрим язык $L = \{a^n b^n\}$. Положим $w_i = a^i$, $z_i = b^i$. Тогда таблица конкатенаций w_i, z_j будет выглядеть следующим образом. Здесь $+$ — это то же, что «$\in L$», $-$ читаем как «$\notin L$».

	$z_1 = b$	$z_2 = b^2$	$z_3 = b^3$	$\dots$	$z_n = b^n$	$\dots$
$w_1 = a$	$+$	$-$	$-$		$-$	
$w_2 = a^2$	$-$	$+$	$-$		$-$	
$w_3 = a^3$	$-$	$-$	$+$		$-$	
$\dots$			$\dots$			
$w^n = a^n$	$-$	$-$	$-$		$+$	
$\dots$						

Доказательство минимальности

Так же можно обосновывать минимальность DFA. Рассмотрим минимальный автомат из примера выше. Его язык — слова в $\{a, b\}^*$, начинающиеся и заканчивающиеся одной и той же буквой. Построим таблицу классов эквивалентности по $w_i \in \{\varepsilon, a, b, ab, ba\}$.

	ε	a	b
ε	+	+	+
a	+	+	−
b	+	−	+
ab	−	+	−
ba	−	−	+

В этой таблице все строчки различны, значит, выбранные w_i действительно лежат в различных классах эквивалентности, и DFA, распознающий язык L, не может иметь меньше пяти состояний.

При доказательстве минимальности DFA достаточно подобрать $[\log_2 n] + 1$ различающих суффиксов z_i, где n — число состояний автомата.