

- 1 Недетерминированные конечные автоматы
- 2 Операции над автоматами языками
- 3 Регулярные языки
- 4 Автоматные (регулярные) фрагменты языков программирования
- 5 Пример нерегулярного языка

Язык, распознаваемый НКА

- Доопределим функцию переходов δ на $Q \times \Sigma^*$:

$$\delta(q, \varepsilon) = \{q\}, \quad \delta(q, ua) = \bigcup_{r \in \delta(q, u)} \delta(r, a).$$
- НКА $B = (Q, \Sigma, \delta, Q_0, F)$ распознает (или допускает) цепочку $w \in \Sigma^*$, если

$$\left(\bigcup_{q \in Q_0} \delta(q, w) \right) \cap F \neq \emptyset.$$
- Множество всех цепочек, допускаемых автоматом B , называется языком, распознаваемым автоматом B , и обозначается $L(B)$.

Теорема Рабина-Скотта

Теорема (Рабина-Скотта)
 Класс языков, распознаваемых НКА, совпадает с классом языков, распознаваемых ДКА.

Доказательство. ДКА является частным случаем НКА, поэтому класс языков, распознаваемых ДКА, содержится в классе языков, распознаваемых НКА. Докажем обратное включение.

- Пусть $B = (Q, \Sigma, \delta, Q_0, F)$ — произвольный НКА. Возьмём ДКА $A = (2^Q, \Sigma, \delta', Q_0, F')$, где

$$\delta'(P, a) = \bigcup_{q \in P} \delta(q, a), \quad F' = \{P \in 2^Q \mid P \cap F \neq \emptyset\},$$

и докажем, что $L(A) = L(B)$.

Лекции по теории формальных языков

Лекция 2.

- Недетерминированные конечные автоматы. Операции над автоматами языками.
- Регулярные языки.
- Автоматные фрагменты языков программирования

Александр Сергеевич Герасимов
<http://gas-teach.narod.ru>

Кафедра математических и информационных технологий
 Санкт-Петербургского академического университета
 Российской академии наук.
 Весенний семестр 2010/11 учебного года

18 февраля 2011 г.

Определение недетерминированного конечного автомата

Недетерминированным конечным автоматом (НКА) называется пятерка $B = (Q, \Sigma, \delta, Q_0, F)$, где

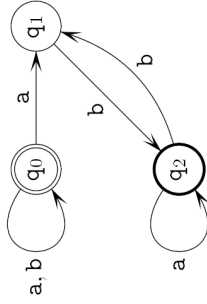
- Q — непустое конечное множество состояний,
- Σ — алфавит,
- $\delta : Q \times \Sigma \rightarrow 2^Q$ — функция переходов,
- $Q_0 \subseteq Q$ — множество начальных состояний,
- $F \subseteq Q$ — множество заключительных (или допускающих) состояний.

(δ может быть определено и как отношение $\delta \subseteq Q \times \Sigma \times Q$)

НКА, находящийся в состоянии q и обозревающий ячейку с символом a , может перейти в любое состояние из множества $\delta(q, a)$.

Задание НКА диаграммой переходов

НКА B тот же, что и на предыдущем слайде.



- $aab \in L(B)$: $\delta(q_0, a) = \{q_0, q_1\} = \delta(q_0, aa)$,
 $\delta(q_0, aab) = \{q_0, q_2\} \cap F \neq \emptyset$.
- $abba \notin L(B)$: $\delta(q_0, abba) = \{q_0, q_1\} \cap F = \emptyset$.

- 1 Недетерминированные конечные автоматы
- 2 Операции над автоматами языками
- 3 Регулярные языки
- 4 Автоматные (регулярные) фрагменты языков программирования
- 5 Пример нерегулярного языка

Задание НКА расширенной таблицей переходов

НКА B :

| | a | b | Q_0 | F |
|-------|------------|-------|-------|-----|
| q_0 | q_0, q_1 | q_0 | 1 | 0 |
| q_1 | | q_2 | 0 | 0 |
| q_2 | | q_1 | 0 | 1 |

Теорема Рабина-Скотта: продолжение доказательства

- Индукцией по длине произвольной $w \in \Sigma^*$ установим, что для любого $P \in 2^Q$

$$\delta'(P, w) = \bigcup_{q \in P} \delta(q, w). \quad (*)$$
- База индукции ($|w| = 0$) верна, поскольку
$$\delta'(P, \varepsilon) = P = \bigcup_{q \in P} \delta(q, \varepsilon).$$
- Индукционный переход. Рассмотрим $w = ua$.

$$\delta'(P, ua) = \delta'(\delta'(P, u), a) \stackrel{\text{предп.}}{=} \bigcup_{q \in P} \delta(q, u, a) \stackrel{\text{опр. } \delta'}{=} \bigcup_{q \in P} \delta(q, a) = \bigcup_{r \in \bigcup_{q \in P} \delta(q, u)} \delta(r, a) = \bigcup_{q \in P} \delta(q, ua).$$

- Наконец, пользуясь определениями и беря Q_0 в качестве P в $(*)$, получаем

$$L(A) = \{w \in \Sigma^* \mid \delta'(Q_0, w) \in F'\} = \{w \in \Sigma^* \mid \delta'(Q_0, w) \cap F \neq \emptyset\} = L(B).$$

□
Применённый в доказательстве теоремы способ получения ДКА по НКА называется «построением подмножеств».

Определение недетерминированного конечного автомата с ε -переходами

Определяемые здесь автоматы будут удобно строить по заданным языкам.

Недетерминированным конечным автоматом с ε -переходами

(ε -НКА) называется пятёрка $B = (Q, \Sigma, \delta, Q_0, F)$, где

- Q, Σ, Q_0, F — те же, что и в случае НКА,
- $\delta : Q \times (\Sigma \cup \{\varepsilon\}) \rightarrow 2^Q$ — функция переходов.

ε -НКА может не сдвигаться по входной цепочке по окончании некоторых тактов (на таком такте автомат прочитывает ε).

Диаграмма переходов ε -НКА может содержать дуги, помеченные ε .

ε -НКА распознаёт цепочку $w \Leftrightarrow$ в диаграмме переходов существует путь из начального состояния в заключительное, помеченный w .

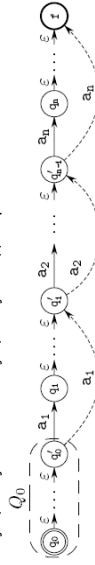
(Определение распознаваемой цепочки через обобщение функции переходов будет дано ниже.)

Теорема

Класс языков, распознаваемых ε -НКА, совпадает с классом языков, распознаваемых НКА.

Доказательство.

- Пусть $B = (Q, \Sigma, \delta, \{q_0\}, \{f\})$ — (нормальный) ε -НКА. Определим НКА $B' = (Q, \Sigma, \delta', \{q_0\}, \{f\})$, где $Q_0 = Clo(q_0)$, $\delta'(q, a) = \delta(q, a)$.
- Пути, помеченные цепочкой $w \in \Sigma^*$, в диаграммах переходов B и B' существуют или не существуют одновременно:



- Таким образом, множества цепочек, распознаваемых B и B' , совпадают.

□

Алгоритм нахождения ДКА по НКА (построением подмножеств).

Вход. НКА $B = (Q, \Sigma, \delta, Q_0, F)$.

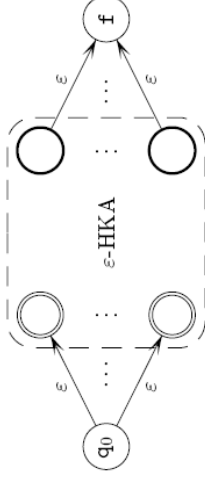
Выход. Эквивалентный ДКА $A = (Q', \Sigma, \delta', Q_0, F')$ без недостижимых состояний.

1. для каждого $P \subseteq Q$ $label(P) := 0$;
2. $Q' := \{Q_0\}$;
3. пока $(\exists P \in Q' : label(P) = 0)$ повторять
4. для каждого $a \in \Sigma$
5. $\delta'(P, a) := \bigcup_{q \in P} \delta(q, a)$;
6. $Q' := Q' \cup \{\delta'(P, a)\}$;
7. $label(P) := 1$;
8. $F' := \{P \in Q' \mid P \cap F \neq \emptyset\}$

Все состояния автомата A достижимы, поскольку каждое состояние в Q' , кроме начального, получено переходом в него из ранее построенного состояния.

Нормальные ε -НКА

- Произвольный ε -НКА эквивалентен ε -НКА с единственным начальным и единственным заключительным состоянием: добавим следующим образом к исходному автомату состояния q_0 и f



и объявим $\{q_0\}$ множеством начальных состояний, а $\{f\}$ — множеством заключительных состояний нового автомата.

- ε -НКА вида $(Q, \Sigma, \delta, \{q_0\}, \{f\})$ будем называть нормальными.
- В дальнейшем мы рассматриваем только нормальные ε -НКА.

Алгоритм нахождения ДКА по ε -НКА

Вход. ε -НКА $B = (Q, \Sigma, \delta, \{q_0\}, F)$.

Выход. Эквивалентный ДКА $A = (Q', \Sigma, \delta', Q_0, F')$ без недостижимых состояний.

1. для каждого $P \subseteq Q$ $label(P) := 0$;
2. $Q_0 := Clo(q_0)$; $Q' := \{Q_0\}$;
3. пока $(\exists P \in Q' : label(P) = 0)$ повторять
4. для каждого $a \in \Sigma$
5. $\delta'(P, a) := \bigcup_{q \in P} \delta(q, a)$;
6. $Q' := Q' \cup \{\delta'(P, a)\}$;
7. $label(P) := 1$;
8. $F' := \{P \in Q' \mid P \cap F \neq \emptyset\}$

Это видоизменение алгоритма, описанного на слайде 11. Замыкание $Clo(q)$ вычисляется поиском из состояния q .

Пример построения ДКА по НКА

| | a | b | Q_0 | F |
|-------|------------|-------|-------|---|
| q_0 | q_0, q_1 | q_0 | 1 | 0 |
| q_1 | | q_2 | 0 | 0 |
| q_2 | | q_1 | 0 | 1 |

| | a | b | F |
|---------------------|---------------------|---------------------|---|
| $\{q_0\}$ | $\{q_0, q_1\}$ | $\{q_0\}$ | 0 |
| $\{q_0, q_1\}$ | $\{q_0, q_1\}$ | $\{q_0, q_2\}$ | 0 |
| $\{q_0, q_2\}$ | $\{q_0, q_1, q_2\}$ | $\{q_0, q_1\}$ | 1 |
| $\{q_0, q_1, q_2\}$ | $\{q_0, q_1, q_2\}$ | $\{q_0, q_1, q_2\}$ | 1 |

Цепочки, распознаваемые (нормальным) ε -НКА

Пусть $B = (Q, \Sigma, \delta, \{q_0\}, \{f\})$ — (нормальный) ε -НКА.

- Определим отношение $\rho \subseteq Q^2$: $(q, p) \in \rho \Leftrightarrow p \in \delta(q, \varepsilon)$.
- $(q, p) \in \rho^*$ (рефлексивно-транзитивное замыкание отношения ρ) \Leftrightarrow автомат может перейти из состояния q в состояние p , не сдвигаясь по входной цепочке.
- Замыкание состояния q : $Clo(q) = \{p \in Q \mid (q, p) \in \rho^*\}$.
- Замыкание множества состояний $P \subseteq Q$: $Clo(P) = \bigcup_{q \in P} Clo(q)$.
- Обобщённая функция переходов $\bar{\delta} : Q \times \Sigma^* \rightarrow 2^Q$,
 $\bar{\delta}(q, \varepsilon) = Clo(q)$, $\bar{\delta}(q, ua) = \bigcup_{x \in \bar{\delta}(q, u)} Clo(x, a)$.

- $x \in \bar{\delta}(q, w) \Leftrightarrow$ в диаграмме переходов автомата B существует путь из q в x , помеченный w .

- Цепочка w распознаётся автоматом B , если $f \in \bar{\delta}(q_0, w)$.

План

1. Недетерминированные конечные автоматы
2. Операции над автоматными языками
3. Регулярные языки
4. Автоматные (регулярные) фрагменты языков программирования
5. Пример нерегулярного языка

Определение автоматного языка. Теорема о замкнутости класса автоматных языков относительно некоторых операций

Назовём язык **автоматным**, если он распознаётся некоторым конечным автоматом (каким?). Обозначим через \mathbb{A} класс всех автоматных языков.

Теорема

Класс \mathbb{A} замкнут относительно объединения, пересечения, дополнения, произведения и итерации.

Доказательство.

- Замкнутость относительно объединения. Пусть $B_1 = (Q_1, \Sigma, \delta_1, \{q_1\}, \{f_1\})$ и $B_2 = (Q_2, \Sigma, \delta_2, \{q_2\}, \{f_2\})$ — нормальные ε -НКА, $Q_1 \cap Q_2 = \emptyset$. Тогда язык $L(B_1) \cup L(B_2)$ распознаётся следующим нормальным ε -НКА $(Q_1 \cup Q_2 \cup \{q_0, f\}, \Sigma, \delta_3, \{q_0\}, \{f\})$:

Автоматность конечных языков

Предложение

Любой конечный язык является автоматным.

Доказательство.

- Языки $\emptyset, \{\varepsilon\}, \{a\}$ (где a — символ) распознаются следующими автоматами:



- Любой язык, состоящий из одной цепочки, либо совпадает с $\{\varepsilon\}$, либо является произведением конечного числа языков из одного символа, поэтому по предыдущей теореме такой язык является автоматным.
- Любой конечный язык либо совпадает с \emptyset , либо является объединением конечного числа языков из одной цепочки, поэтому по предыдущей теореме такой язык является автоматным.

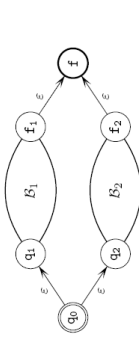
Теорема Клини: продолжение доказательства

База индукции.

- Функция переходов δ нигде не определена. Тогда $L(A) = \{\varepsilon\}$, если $q_0 \in F$, и $L(A) = \emptyset$ иначе.
- Существует ровно один переход $\delta(q, a) = x$. Если $q \neq q_0$, то имеем: $L(A) = \{\varepsilon\}$ при $q_0 \in F$, а иначе $L(A) = \emptyset$. Разберём все возможные случаи, если $q = q_0$:

| $x = q_0$ | $q \in F$ | $x \in F$ | $L(A)$ |
|-----------|-----------|-----------|----------------------|
| нет | нет | нет | \emptyset |
| нет | нет | да | $\{a\}$ |
| нет | да | нет | $\{\varepsilon\}$ |
| нет | да | да | $\{\varepsilon, a\}$ |
| да | нет | нет | \emptyset |
| да | да | да | $\{a\}^*$ |

Теорема: продолжение доказательства



- Замкнутость относительно произведения. $L(B_1)L(B_2)$
- Замкнутость относительно итерации. $L(B_1)^*$



План

- Недетерминированные конечные автоматы
- Операции над автоматными языками
- Регулярные языки
- Автоматные (регулярные) фрагменты языков программирования
- Пример нерегулярного языка

Теорема Клини: продолжение доказательства

Индукционный переход. Пусть автомат $A = (Q, \Sigma, \delta, q_0, F)$ имеет $k > 1$ переходов.

- Зафиксируем один переход $\delta(q, a) = x$. Через δ' обозначим (частичную) функцию переходов на $Q \times \Sigma$ такую, что значение δ' не определено на паре (q, a) и совпадает со значением δ на всех остальных парах.
- Рассмотрим автоматы $A_0 = (Q, \Sigma, \delta', q_0, F)$, $A_1 = (Q, \Sigma, \delta', q_0, \{q\})$, $A_2 = (Q, \Sigma, \delta', x, \{q\})$, $A_3 = (Q, \Sigma, \delta', x, F)$, каждый из которых имеет $(k-1)$ переход. По индукционному предположению языки $L(A_0), L(A_1), L(A_2), L(A_3)$ регулярны.

Теорема: окончание доказательства

- Замкнутость относительно дополнения. Пусть $A = (Q, \Sigma, \delta, q_0, F)$ — (полный) ДКА. Тогда язык $L(A)$ распознаётся ДКА $(Q, \Sigma, \delta, q_0, Q \setminus F)$.
- Замкнутость относительно пересечения. $K \cap L = \overline{\overline{K} \cup \overline{L}}$

□

Определение регулярного языка. Теорема Клини

Язык называется **регулярным**, если он может быть получен из конечных языков (или языков вида $\emptyset, \{\varepsilon\}, \{a\}$, где a — символ) в результате конечного числа операций объединения, произведения и итерации. Обозначим через \mathbb{R} класс всех регулярных языков.

Теорема (теорема Клини)

$\mathbb{R} = \mathbb{A}$.

Доказательство. $\mathbb{R} \subseteq \mathbb{A}$ по предыдущим теореме и предложению. Покажем, что $\mathbb{A} \subseteq \mathbb{R}$.

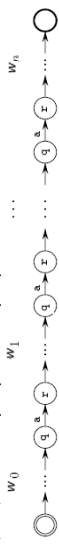
Пусть $A = (Q, \Sigma, \delta, q_0, F)$ — неполный ДКА. Индукцией по числу переходов автомата A докажем, что язык $L(A)$ регулярен.

Теорема Клини: окончание доказательства

- Для завершения доказательства достаточно показать, что

$$L(A) = (L(A_0) \cup L(A_1)\{a\}(L(A_2)\{a\})^*L(A_3))^*L(A_3). \quad (*)$$

- Пусть $w \in L(A)$. Если автомат A распознал w , не совершив перехода $\delta(q, a) = x$, то $w \in L(A_0)$.
- Если же переход $\delta(q, a) = x$ был совершён $n \geq 1$ раз, то $w = w_0 a w_1 a w_2 \dots w_{n-1} a w_n$ для некоторых $w_0 \in L(A_1), w_1, \dots, w_{n-1} \in L(A_2), w_n \in L(A_3)$:



Так что $w \in L(A_1)\{a\}(L(A_2)\{a\})^*L(A_3)$.

- Пусть w принадлежит правой части $(*)$. Если $w \in L(A_0)$, то по определению автомата A_0 верно $w \in L(A)$.
- Если $w \in L(A_1)\{a\}(L(A_2)\{a\})^*L(A_3)$, то $w = w_0 a w_1 a \dots w_{n-1} a w_n$ для некоторых $n \geq 1, w_0 \in L(A_1), w_1, \dots, w_{n-1} \in L(A_2), w_n \in L(A_3)$. Отсюда $w \in L(A)$.

Определение регулярных выражений

- Регулярные языки можно задавать так называемыми **регулярными выражениями**.
- Регулярные выражения** в алфавите (над алфавитом) Σ и языки, которые они обозначают, определяются следующим образом:
 - символы \emptyset, ε и a (где $a \in \Sigma$) являются регулярными выражениями в Σ , обозначающими языки $\emptyset, \{\varepsilon\}$ и $\{a\}$ соответственно;
 - если r и s — регулярные выражения в Σ , обозначающие языки R и S соответственно, то $(r|s)$, (rs) и r^* являются регулярными выражениями в Σ , обозначающими языки $R \cup S$, RS и R^* соответственно.
- Класс \mathbb{R} совпадает с классом всех языков, определяемых регулярными выражениями.
- Переносим соглашение о приоритетах операций над языками.
- Пример: $\Sigma = \{a, b\}$,
 $a^*b|bb^*a = \{a^*b^2 \mid m \geq 0\} \cup \{b^*a \mid n \geq 1\}$,
 $(ab)^*a = \{wa \mid w \in \Sigma^*\}$.

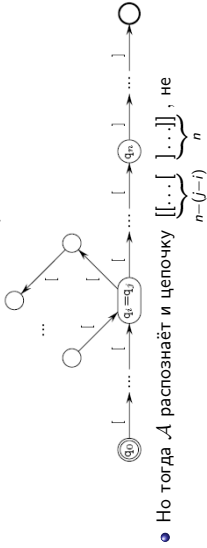
Регулярные выражения, задающие идентификаторы

и числа

- Имя (или идентификатор):
 $\langle \text{имя} \rangle = \langle \text{буква} \rangle (\langle \text{буква} \rangle | \langle \text{цифра} \rangle)^*$, где
 $\langle \text{буква} \rangle = a|b|\dots|z$,
 $\langle \text{цифра} \rangle = 0|1|\dots|9$
- Целое без знака:
 $\langle \text{целое_без_знака} \rangle = \langle \text{цифра} \rangle (\langle \text{цифра} \rangle)^*$
- Цисло без знака:
 $\langle \text{целое_без_знака} \rangle = \langle \text{целое_без_знака} \rangle (\langle \text{целое_без_знака} \rangle | \varepsilon)$
 $(\varepsilon | + | - | \varepsilon) \langle \text{целое_без_знака} \rangle | \varepsilon$

Нерегулярность скобочного языка LB

- Предположим, что язык LB распознаётся некоторым ДКА A .
- Пусть n — число состояний этого автомата. Автомат A распознаёт цепочку $w = \underbrace{[\dots [] \dots] }_n$.
- Из начального состояния q_0 автомат A при чтении n скобок $[$ последовательно переходит в состояния, обозначаемые q_1, \dots, q_n .
- Найдутся такие $i < j$, что $q_i = q_j$. Путь в автомате A , помеченный цепочкой w , можно представить так:

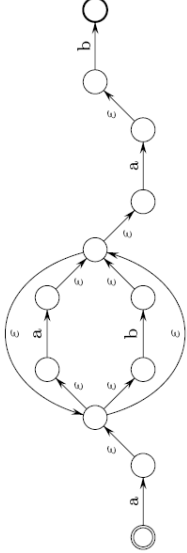


- Но тогда A распознаёт и цепочку $\underbrace{[\dots [] \dots] }_{n-(j-i)}$, не

принадлежащую LB. Противоречие.

Построение конечного автомата по регулярному выражению

Нормальный ε -НКА, распознающий язык $a(ba)^*ab$:



Регулярные выражения, задающие константы и простые типы

- Константа (в языке Паскаль):
 $\langle \text{константа} \rangle = ((+ | - | \varepsilon) \langle \text{число_без_знака} \rangle (\langle \text{имя} \rangle | '(\langle \text{буква} \rangle | \langle \text{цифра} \rangle)^*'))'$
- Простой тип (в языке Паскаль):
 $\langle \text{простой_тип} \rangle = (\langle \text{имя} \rangle | "(\langle \text{имя} \rangle | \langle \text{константа} \rangle)^*") | \langle \text{константа} \rangle$
- Арифметические выражения (со сбалансированными скобками) не задаются регулярными выражениями, иначе говоря, не распознаются конечными автоматами.

Литература

- Основная литература
 - Замятин А. П., Шур А. М. Языки, грамматики, распознаватели: Учебное пособие. Екатеринбург : Изд-во Урал. ун-та, 2007 (электронный вариант книги — на <http://elag.usu.ru>, поиск)
 - Дополнительная литература
 - Ахо А., Лам М., Сети Р., Ульман Дж. Компиляторы: принципы, технологии и инструментарий. М.: ООО "И.Д. Вильямс", 2008
 - Ахо А., Ульман Дж. Теория синтаксического анализа, перевода и компиляции. М.: Мир, 1978
 - Мартыненко Б. К. Языки и трансляции: Учеб. пособие. СПб.: Издательство С.-Петербургского университета, 2004 (электронный вариант книги — на <http://www.math.spbu.ru/user/mbk>)

План

- Недетерминированные конечные автоматы
- Операции над автоматными языками
- Регулярные языки
- Автоматные (регулярные) фрагменты языков программирования**
- Пример нерегулярного языка

План

- Недетерминированные конечные автоматы
- Операции над автоматными языками
- Регулярные языки
- Автоматные (регулярные) фрагменты языков программирования
- Пример нерегулярного языка**