Лекция 14

Эквивалентность РВ, ДКА и НКА. Лемма о накачке. Теорема Майхилла – Нероуда. Замкнутость регулярных языков

Утв Регулярные выражения, детерминированные конечные автоматы и недетерминированные конечные автоматы задают одни и те же сущности.

□ Доказательство будет конструктивным. Будем строить по РВ → НКА; НКА → ДКА; ДКА → РВ ■

Построение НКА по РВ

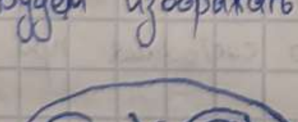
РВ строятся индуктивно. Будем строить НКА по РВ в том же порядке, соблюдая следующие ограничения:

1) НКА имеет ровно одно принимающее состояние

2) В начальное состояние НКА не ведёт ни один переход

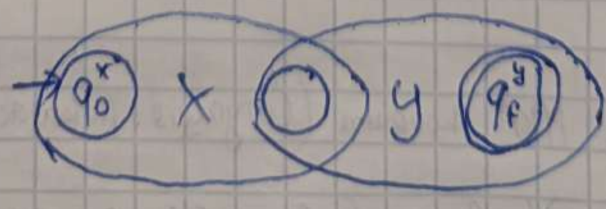
3) Из принимающего состояния НКА нет ни одного перехода

Будем изображать НКА языка X схематично так:

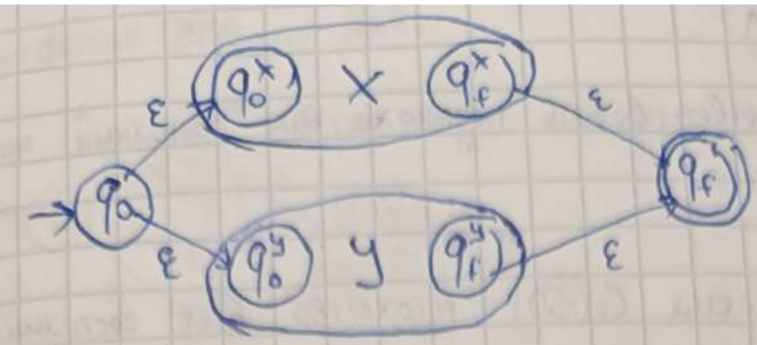


Построим НКА для операций:

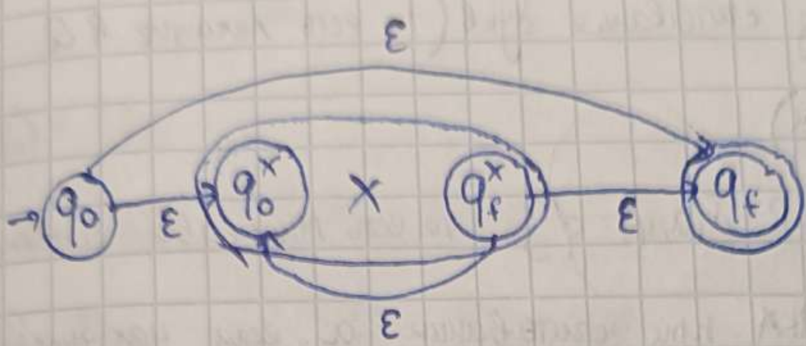
1) XY:



г) $X+Y$:



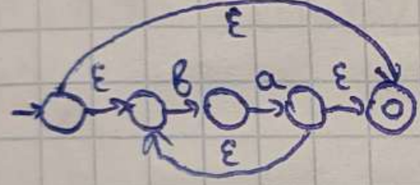
з) X^*



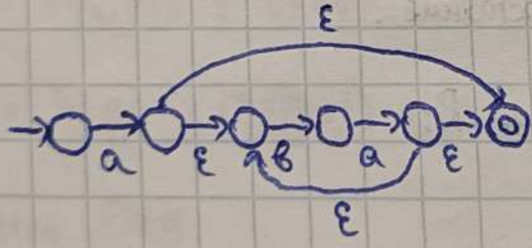
Пример. Построим НКА по $a(ba)^* + ba^*$

(ba):

$(ba)^*$:

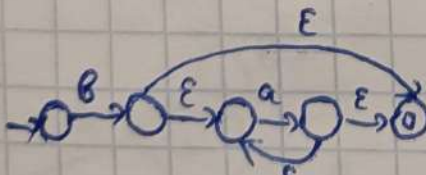


$a(ba)^*$:



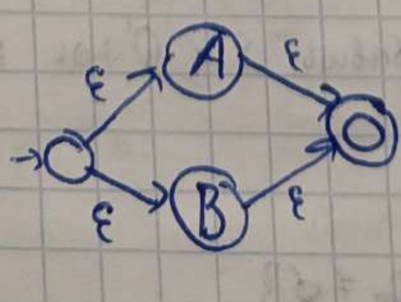
→ назовём это A

ba^*:



→ назовём это B

$a(ba)^* + ba^*$:



Построение ДКА по НКА:

Состояния ДКА будут соответствовать подмножествам состояний исходного НКА.

Опр ε-замыкание (обозначается $C_\varepsilon(Q)$) - множество всех состояний, достижимых из всех $q \in Q$ без считывания букв (то есть лежащие в Q или достижимые по ε-переходам)

Опр $move(R; a) = \{q' \mid \exists q \in R : \delta(a; q) = q'\}$ - то есть множество состояний, в которые может перейти НКА при считывании a, если находился в одном из состояний $R \subseteq Q$

Алгоритм построения ДКА по НКА:

Сначала Q' и δ' пусты.

1) Определить $q_0' = C_\varepsilon(q_0)$

2) Добавить q_0' в Q' как непомеченное состояние.

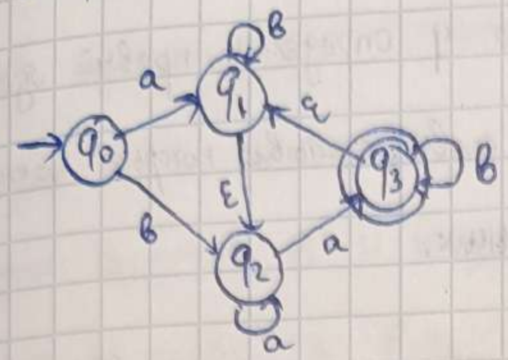
3) Пока в Q' есть непомеченное состояние R:

- Пометить R
- Для каждого символа $a \in \Sigma$:
  - $S = C_\varepsilon(move(R; a))$
  - Если $S \neq \varnothing$ и $S \notin Q'$, то добавить S в Q' как непомеченное состояние и определить $\delta'(a; R) = S$

4) Определить $Q_F' = \{S \mid S \in Q', S \cap Q_F \neq \varnothing\}$

Пример построения ДКА по НКА:

НКА:



$Q_0 = C_\varepsilon(q_0) = \{q_0\}$

$move(Q_0; a) = \{q_1\} \quad C_\varepsilon(\{q_1\}) = \{q_1; q_2\} = Q_1$

$move(Q_0; b) = \{q_2\} \quad C_\varepsilon(\{q_2\}) = \{q_2\} = Q_2$

$move(Q_1; a) = \{q_2; q_3\} \quad C_\varepsilon(\{q_2; q_3\}) = \{q_1; q_2; q_3\} = Q_3$

$move(Q_1; b) = \{q_1\} \quad C_\varepsilon(\{q_1\} = \{q_1; q_2\} = Q_1$

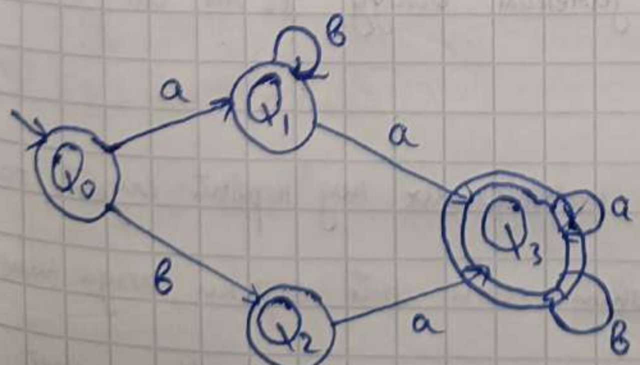
$move(Q_2; a) = \{q_2; q_3\} \quad C_3(\{q_2; q_3\}) = \{q_1; q_2; q_3\} = Q_3$

$move(Q_2; b) = \varnothing$

$move(Q_3; a) = \{q_2; q_3\} \; ; \; C_\varepsilon(\{q_2; q_3\}) = Q_3$

$move(Q_3; b) = \{q_1; q_3\} \quad C_3(\{q_1; q_3\}) = \{q_1; q_2; q_3\} = Q_3$

Получаем ДКА:



Построение РВ по ДКА

Опр По заданному НКА/ДКА для каждого q определим правый язык $R_q = \{x \mid q \xrightarrow{x} q_f;\ q_f \in Q_f\}$, то есть все слова, считывая которые можно попасть из состояния q в одно из финальных.

Опр Назовём

$$\begin{cases} X_1 = \alpha_{11}X_1 + \alpha_{12}X_2 + \ldots + \alpha_{1n}X_n + \beta_1 \\ \vdots \\ X_n = \alpha_{n1}X_1 + \alpha_{n2}X_2 + \ldots + \alpha_{nn}X_n + \beta_n \end{cases}$$

системой уравнений с регулярными коэффициентами, где X_i - переменные-множества, $\alpha_{ij}; \beta_i$ - регулярные языки, задаваемые регулярными выражениями

Опр Решением системы с языками-переменными $X_1; X_2 \ldots X_n$ будем называть такой набор языков $A_1; A_2 \ldots A_n$, удовлетворяющий системе, что для $\forall$ набора языков $B_1; B_2 \ldots B_n$, удовлетворяющего системе, выполнено $A_i \subseteq B_i\ \forall i \in \overline{1..n}$

Алгоритм решения:

Случай 1: Если в системе есть строка вида:

$X_i = a_{ii}X_i + \gamma$, где $a_{ii} \neq \varnothing$, то заменим всюду X_i на $a_{ii}^*\gamma$, а саму строку удалим

Случай 2: Если в системе нет строк, подходящих под первый случай, то переменная X_n не встречается в правой части n-й строки, тогда после замены всюду X_n на правую часть n-ой строки переменная X_n исчезнет

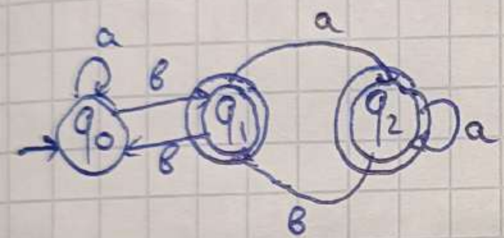
Алгоритм построения РВ по ДКА:

В качестве переменных при составлении системы будем использовать правые языки всех состояний.

Для каждого состояния q_i и каждого символа a: если из состояния q_i есть переход по символу a в q_j, то к правой части строки R_{q_i} добавим $+aR_{q_j}$, к правым частям финальных состояний добавим $+\varepsilon$

Решим систему ~~с заменой уравнений подставим все переменные~~. Ответом будет R_0.

Пример построения РВ по ДКА:

ДКА: 

Составим систему:

$$\begin{cases} R_0 = aR_0 + bR_1 \\ R_1 = aR_2 + bR_0 + \varepsilon \\ R_2 = aR_2 + bR_1 + \varepsilon \end{cases} \Leftrightarrow \begin{cases} R_0 = a^*bR_1 \\ R_1 = ba^*bR_1 + aR_2 + \varepsilon \\ R_2 = bR_1 + aR_2 + \varepsilon \end{cases} \Leftrightarrow \begin{cases} R_0 = a^*bR_1 \\ R_1 = (ba^*b)^*(aR_2 + \varepsilon) \\ R_2 = b(ba^*b)^*(aR_2 + \varepsilon) + aR_2 + \varepsilon \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} R_0 = a^*bR_1 \\ R_1 = (ba^*b)^*(aR_2 + \varepsilon) \\ R_2 = (b(ba^*b)^*a + a)R_2 + b(ba^*b)^* + \varepsilon \end{cases}$$

Подставим R_2 и потом R_1 и найдём R_0.

Теорема (лемма о накачке/разрастании)

Пусть $L \subset \Sigma^*$ - регулярен. Тогда $\exists$ такое $k \in \mathbb{N}$: $\forall w \in L$: $|w| > k$ $\exists$ слова $x; y; z$: $w = xyz$ и:

1) $|y| > 0$

2) $|xy| \leq k$

3) $xy^n z \in L$ $\forall n \in \mathbb{N} \cup \{0\}$

Зам Это не критерий. Пример нерегулярного языка, удовлетворяющего условиям леммы:

$L = \{a^i b^j c^k \mid i \neq 1; j \geq 0; k \geq 0\} \cup \{ab^i c^i \mid i \geq 1\}$

Теорема (Теорема Майхилла-Нерода)

Пусть дан $L \subset \Sigma^*$. Определим отношение Майхилла-Нерода ($\equiv_L$) на Σ^* следующим образом: $x \equiv_L y \Leftrightarrow \forall z \in \Sigma^* \hookrightarrow xz \in L \Leftrightarrow yz \in L$

Тогда:

Язык L регулярен $\Leftrightarrow$ число классов эквивалентности по отношению Майхилла-Нерода конечно.

□ $\Rightarrow$ Язык L - регулярен $\Rightarrow \exists$ min ПДКА, соответствующий этому языку. Тогда слова x и y эквивалентны, если при запуске min ПДКА на словах x и y он останавливается в одном и том же состоянии. Т.к. состояний конечно, то и классов эквивалентности конечно

$\Leftarrow$ Т.к. конечное число классов, то можно для каждого класса завести вершину в ПДКА. Для слова x из каждого класса проведём рёбра со всеми символами в те вершины, в которых лежат xa, где a тот символ. Тогда можно отметить тот класс, в котором есть ε за начальное состояние. А за финальные - те, в которых есть слова из L ■

Замкнутость регулярных языков:

Пусть $A, B \in REG$ над Σ^*. Тогда:

1) $A+B$; AB; $A^* \in REG$

2) $\overline{A} = \Sigma^* \setminus A \in REG$

3) $A \cap B \in REG$

4) $A \setminus B \in REG$

5) Можно получить регулярный язык, сделав гомоморфизм и реверс.