## 3. Крайни автомати. Регулярни езици. Теорема на Клини

Анотация: Детерминирани крайни автомати. Регулярни операции. Недетерминирани крайни автомати. Представяне на всеки недетерминиран краен автомат с детерминиран (с доказателство). Затвореност относно регулярните операции. Теорема на Клини (с доказателство). Лема за покачването (uvw) (с доказателство). Примери за регулярни и нерегулярни езици. Минимизация на състоянията. Теорема на Майхил-Нероуд (с доказателство). Алгоритъм за конструиране на минимален автомат, еквивалентен на даден детерминиран краен автомат.

**Краен детерминиран автомат (КДА)** наричаме петорката $\mathscr{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$, където: $Q$ е крайно множество от състояния; $\Sigma$ е крайна входна азбука; $\delta : Q \times \Sigma \rightharpoonup Q$ е частична функция на преходите, пресмятаща следващото състояние; $q_0 \in Q$ е началното състояние; $F \subseteq Q$ е множество от финалните състояния.

**Регулярни операции**. Именуваме следните символи:

$\varepsilon$ – празната буква;

$*$ – итерация (звезда на Клини);

• – конкатенация (символът за конкатенация може да се пропуска);

$\cup$ – обединение.

Нека $\Sigma$ е крайна азбука, за която $\Sigma \cap \{\varepsilon, *, \bullet, \cup, (, )\} = \emptyset$. Регулярен израз $\alpha$ над азбуката $\Sigma$ дефинираме индуктивно по следния начин:

а) $x$ за всяко $x \in \Sigma \cup \{\varepsilon\}$ е регулярен израз;

б) Ако $\alpha$ и $\beta$ са регулярни изрази, то $\alpha^*$, $\alpha \bullet \beta$ и $\alpha \cup \beta$ са регулярни изрази;

в) Няма други регулярни изрази, които да не следват от горните две правила.

Регулярен език $\mathscr{L}(\alpha)$ за регулярен израз $\alpha$ е функцията $\mathscr{L}$, която описва езика, който този регулярен израз $\alpha$ задава. Дефинираме $\mathscr{L}$ индуктивно по следния начин:

а) $\mathscr{L}(x) = \{x\}$, за всяко $x \in \Sigma \cup \{\varepsilon\}$ е регулярен език;

б) Ако $\mathscr{L}(\alpha)$ и $\mathscr{L}(\beta)$ са регулярни езици за регулярните изрази $\alpha$ и $\beta$, то $\mathscr{L}(\alpha^*) = \big(\mathscr{L}(\alpha)\big)^*$, $\mathscr{L}(\alpha \bullet \beta) = \mathscr{L}(\alpha) \bullet \mathscr{L}(\beta)$ и $\mathscr{L}(\alpha \cup \beta) = \mathscr{L}(\alpha) \cup \mathscr{L}(\beta)$ са рег. езици;

в) Няма други регулярни езици, които да не следват от горните две правила.

Нека $\Sigma = \{a, b\}$. Примери за регулярни изрази и съответните им езици:

| **Израз** | **Език** |
|---|---|
| $a \cup b$ | $\{a, b\}$ |
| $a(a \cup b)$ | $\{aa, ab\}$ |
| $a^*$ | $\{\varepsilon, a, a^2, \ldots, a^i, \ldots\}$ |
| $a^*(a \cup b)$ | $\{a, b, aa, ab, a^2a, a^2b, \ldots, a^ia, a^ib, \ldots\}$ |
| $(a \cup b)^*$ | $\Sigma^*$ |
| $a^*b \cup b^*a$ | $\{b, a, ab, ba, a^2b, b^2a, \ldots, a^ib, b^ia, \ldots\}$ |

**Краен недетерминиран автомат (КНА)** наричаме петорката $\mathscr{A} = \langle Q, \Sigma, \Delta, q_0, F \rangle$, където $Q$, $\Sigma$, $q_0$ и $F$ са дефинирани както при КДА, а $\Delta \subseteq Q \times \big(\Sigma \cup \{\varepsilon\}\big) \times Q$ е релация на преходите.

<u>Дефиниция (разширена функция на преходите $\tilde{\Delta}$)</u>. Нека $\mathscr{A} = \langle Q, \Sigma, \Delta, q_0, F \rangle$ е краен автомат (КА). Разширена функция на преходите $\tilde{\Delta}_{\mathscr{A}} : \underbrace{Q \times \Sigma^*}_{\text{траверсираме с цяла дума}} \rightarrow 2^Q$ на $\mathscr{A}$ дефинираме по следния начин:

$\tilde{\Delta}_{\mathscr{A}}(q, \varepsilon) = \{q\}, \forall q \in Q$ (предварително сме премахнали $\varepsilon$ преходите)

$$\tilde{\Delta}_{\mathscr{A}}(q, x) = \begin{cases} \underbrace{\Delta(q, x)}_{\text{множество}} \quad , & \forall q \in Q \text{ и } \forall x \in \Sigma, \text{ за които } !\Delta(q, x) \text{ (е дефинирана)} \\ \text{недефинирана}, & \text{ако } \neg\Delta(q, x) \text{ (не е дефинирана)} \end{cases}$$

$$\tilde{\Delta}_{\mathscr{A}}(q, \omega x) = \bigcup_{i=1}^{l} \Delta(q_{p_i}, x), \text{ ако } \tilde{\Delta}_{\mathscr{A}}(q, \omega) = \{q_{p_1}, q_{p_2}, \ldots, q_{p_l}\}$$

<u>Дефиниция</u>. Казваме, че КА $\mathscr{A}$ разпознава думата $\omega \in \Sigma^*$, ако $\tilde{\Delta}(q_0, \omega) \cap F \neq \emptyset$. Език разпознаван от КА $\mathscr{A}$ определяме като $\mathscr{L}(\mathscr{A}) = \{\omega \,|\, \omega \in \Sigma^*, \tilde{\Delta}(q_0, \omega) \cap F \neq \emptyset\}$.

**Представяне на всеки недетерминиран краен автомат с детерминиран**

**Теорема**. За всеки КНА $\mathscr{A}$ съществува КДА $\mathscr{A}_1$ такъв, че $\mathscr{L}(\mathscr{A}) = \mathscr{L}(\mathscr{A}_1)$.

**Доказателство**

Нека $\mathscr{A} = \langle Q, \Sigma, \Delta, q_0, F \rangle$ е КНА. Ще построим КДА $\mathscr{A}_1 = \langle Q_1, \Sigma, \delta, t_0, F_1 \rangle$, който е еквивалентен на $\mathscr{A}$ (със същия език). $Q_1 = 2^Q$ и за краткост, ако множеството $\{q_{p_1}, q_{p_2}, \ldots, q_{p_l}\} \in Q_1$ ще го означаваме с $t_{[p_1, p_2, ..., p_l]}$.

При това означение определяме $t_0 = \{q_0\} = t_{[0]}$.

Нека $F_1 = \left\{ t_{[p_1, p_2, ..., p_l]} \,|\, \{q_{p_1}, q_{p_2}, \ldots, q_{p_l}\} \cap F \neq \emptyset \right\}$, а $\delta(t_{[p_1, p_2, ..., p_l]}, x) = t_{[r_1, r_2, ..., r_m]}, \Leftrightarrow$ $\bigcup_{i=1}^{l} \Delta(q_{p_i}, x) = \{q_{r_1}, q_{r_2}, \ldots, q_{r_m}\}$. Забележете, че не може да фиксираме в явен вид кои точно подмножества на $Q$ влизат в $Q_1$. Те се определят от изчислението на функцията $\delta$, започвайки от $\delta(t_0, x)$, $\forall x \in \Sigma$. С индукция по дължината на произволна дума $\omega \in \Sigma^*$ ще покажем, че:

$$\tilde{\Delta}_{\mathscr{A}_1}(t_{[0]}, \omega) = t_{[p_1, p_2, ..., p_l]} \Leftrightarrow \tilde{\Delta}_{\mathscr{A}}(q_0, \omega) = \{q_{p_1}, q_{p_2}, \ldots, q_{p_l}\}.$$

а) *Индукционна база*: Нека $|\omega| = 0 \Rightarrow \omega = \varepsilon$. Тогава $\tilde{\Delta}_{\mathscr{A}_1}(t_{[0]}, \varepsilon) = t_{[0]}$ и $\tilde{\Delta}_{\mathscr{A}}(q_0, \varepsilon) = \{q_0\}$ и твърдението е в сила директно от дефиницията за разширена функция на преходите;

б) *Индукционна хипотеза* (И.Х.): Допускаме, че твърдението е вярно за някоя дума $\omega$ с $|\omega| \geq 0$.

Ще покажем, че $\tilde{\Delta}_{\mathscr{A}_1}(t_{[0]}, \omega x) = t_{[r_1, r_2, ..., r_m]} \Leftrightarrow \tilde{\Delta}_{\mathscr{A}}(q_0, \omega x) = \{q_{r_1}, q_{r_2}, \ldots, q_{r_m}\}$, за произволно $x \in \Sigma$;

в) *Индукционнен преход*: От ляво на дясно ($\Rightarrow$). Нека $\tilde{\Delta}_{\mathscr{A}_1}(t_{[0]}, \omega x) = t_{[r_1, r_2, ..., r_m]}$, като $\tilde{\Delta}_{\mathscr{A}_1}(t_{[0]}, \omega) \overset{\text{И.Х.}}{=} t_{[p_1, p_2, ..., p_l]}$. Тогава $\delta(t_{[p_1, p_2, ..., p_l]}, x) = t_{[r_1, r_2, ..., r_m]}$ и съгласно построението на $\delta$, $\bigcup_{i=1}^{l} \Delta(q_{p_i}, x) = \{q_{r_1}, q_{r_2}, \ldots, q_{r_m}\}$.

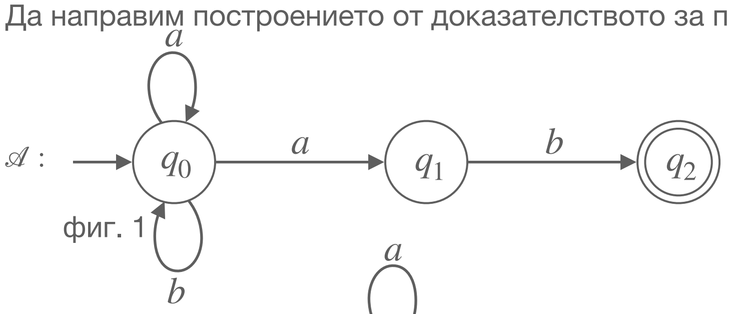
Но от допускането в И.Х. имаме, че $\tilde{\Delta}_{\mathscr{A}}(q_0, \omega) = \{q_{p_1}, q_{p_2}, \ldots, q_{p_l}\}$ и следователно $\tilde{\Delta}_{\mathscr{A}}(q_0, \omega x) = \bigcup_{i=1}^{l} \Delta(q_{p_i}, x) = \{q_{r_1}, q_{r_2}, \ldots, q_{r_m}\}$.

Обратно, от дясно на ляво ($\Leftarrow$). Нека $\widetilde{\Delta}_{\mathscr{A}}(q_0, \omega x) = \{q_{r_1}, q_{r_2}, \ldots, q_{r_m}\}$, като $\widetilde{\Delta}_{\mathscr{A}}(q_0, \omega) \overset{\text{И.Х.}}{=} \{q_{p_1}, q_{p_2}, \ldots, q_{p_l}\}$. Тогава $\bigcup_{i=1}^{l} \Delta(q_{p_i}, x) = \{q_{r_1}, q_{r_2}, \ldots, q_{r_m}\}$ и съгласно построението на $\delta$, $\delta(t_{[p_1,\, p_2,\, ...,\, p_l]}, x) = t_{[r_1,\, r_2,\, ...,\, r_m]}$.

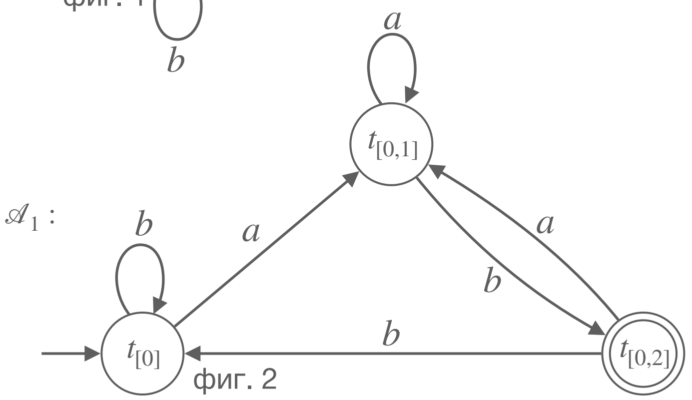
Но от допускането в И.Х. знаем, че $\widetilde{\Delta}_{\mathscr{A}_1}(t_{[0]}, \omega) = t_{[p_1,\, p_2,\, ...,\, p_l]}$ и следователно $\widetilde{\Delta}_{\mathscr{A}_1}(t_{[0]}, \omega x) = \delta(t_{[p_1, p_2,\, ...,\, p_l]}, x) = t_{[r_1,\, r_2,\, ...,\, r_m]}$.

$\square$

Да направим построението от доказателството за произволен КНА $\mathscr{A}$.



фиг. 1



фиг. 2

| от състояния | с буква | до състояния | състояние в КДА |
|---|---|---|---|
| $\{q_0\}$ | $\varepsilon$ | $\{q_0\}$ | $t_{[0]}$ – ново |
| $\{q_0\}$ | $a$ | $\{q_0, q_1\}$ | $t_{[0,1]}$ – ново |
| $\{q_0\}$ | $b$ | $\{q_0\}$ | $t_{[0]}$ |
| $\{q_0, q_1\}$ | $a$ | $\{q_0, q_1\}$ | $t_{[0,1]}$ |
| $\{q_0, q_1\}$ | $b$ | $\{q_0, q_2\}$ | $t_{[0,2]}$ – ново |
| $\{q_0, q_2\}$ | $a$ | $\{q_0, q_1\}$ | $t_{[0,1]}$ |
| $\{q_0, q_2\}$ | $b$ | $\{q_0\}$ | $t_{[0]}$ |

$t_{[0,2]}$ е финално, тъй като $\{q_0, q_2\} \cap F = \{q_2\} \neq \emptyset$.

**Затвореност относно регулярните операции**
Класът на езиците, разпознавани от краен автомат е затворен относно операциите:
а) Обединение; б) Конкатенация; в) Итерация. Като следствие от а), б) и в) получаваме и г) Допълнение; д) Сечение.

За всяка от операциите изброени по-горе, ще построим автомат, който приема език, зададен чрез прилагане на съответната операция върху езиците на крайните автомати $\mathscr{A}_1$ и $\mathscr{A}_2$ (или само върху езика на един от тях – при унарни оператори като итерация и допълнение). Равенството на езиците се доказва, но тук ще посочим само построенията.

а) **Обединение**. Построяваме автомат $\mathscr{A}_\cup$ с език равен на $\mathscr{L}(\mathscr{A}_1) \cup \mathscr{L}(\mathscr{A}_2)$. За $\mathscr{A}_\cup$:

1) Състояния: състоянията на $\mathscr{A}_1$ и $\mathscr{A}_2$ и ново състояние $q$;
2) Начално състояние: $q$;
3) Финални състояния: финалните състояния на $\mathscr{A}_1$ и $\mathscr{A}_2$ се запазват. $q$ е финално $\Leftrightarrow$ поне едно от началните състояния на $\mathscr{A}_1$ и $\mathscr{A}_2$ е финално;
4) Преходи: преходите на $\mathscr{A}_1$ и $\mathscr{A}_2$ остават. $q$ копира преходите на началните състояния $\mathscr{A}_1$ и $\mathscr{A}_2$.

$\mathscr{A}_\cup$ използва недетерминизъм, за да отгатне дали входът ще е в $\mathscr{L}(\mathscr{A}_1)$ или $\mathscr{L}(\mathscr{A}_2)$ след което имитира един от двата автомата и обработва подадената като вход дума, точно както съответния автомат би го направил.

б) **Конкатенация**. Построяваме автомат $\mathscr{A}_\bullet$ с език равен на $\mathscr{L}(\mathscr{A}_1) \bullet \mathscr{L}(\mathscr{A}_2)$. За $\mathscr{A}_\bullet$:

1) Състояния: състоянията на $\mathscr{A}_1$ и $\mathscr{A}_2$;
2) Начално състояние: началното състояние на левия автомат $\mathscr{A}_1$;
3) Финални състояния: финалните състояния на десния автомат $\mathscr{A}_2$. Добавяме финалните състояния на $\mathscr{A}_1$ тогава и само тогава, когато началното състояние на десния автомат е финално;
4) Преходи: преходите на $\mathscr{A}_1$ и $\mathscr{A}_2$ остават. Всяко финално състояние на левия автомат $\mathscr{A}_1$ копира преходите на началното състояние на десния автомат $\mathscr{A}_2$.

$\mathscr{A}_\bullet$ работи, като симулира $\mathscr{A}_1$ за известно време и след това „скача“ недетерминистично от крайно състояние на $\mathscr{A}_1$ до първоначалното състояние на $\mathscr{A}_2$.

в) **Итерация**. Построяваме автомат $\mathscr{A}^*$ с език равен на $\mathscr{L}(\mathscr{A}_1)^*$. За $\mathscr{A}^*$:

1) Състояния: състоянията на $\mathscr{A}_1$ и ново състояние $q$;
2) Начално състояние: $q$;
3) Финални състояния: финалните състояния на $\mathscr{A}_1$ и $q$ (за да може да разпознаем и $\varepsilon$);
4) Преходи: преходите в $\mathscr{A}_1$ остават. Всички финални състояния на $\mathscr{A}^*$ копират преходите на началното състояние на $\mathscr{A}_1$.

$\mathscr{A}^*$ обработва (прочита) дума от $\mathscr{L}(\mathscr{A}_1)$ или празната дума $\varepsilon$ (тъй като новото начално състояние е и финално състояние), след което се възобновява от първоначалното състояние на $\mathscr{A}_1$. Конструкцията е подобна на конкатенацията, но с тази разлика, че конкатенираме автомата със себе си.

г) **Допълнение**. Построяваме автомат $\overline{\mathscr{A}}$ с език равен на $\Sigma^* \backslash \mathscr{L}(\mathscr{A}_1)$.

1) Тоатализираме автомата (важна стъпка, за да е коректна конструкцията)
2) Състояния: състоянията на $\mathscr{A}_1$;
3) Начално състояние: началното състояние на $\mathscr{A}_1$;
4) Финални състояния: състоянията на $\mathscr{A}_1$, които не са финални ($Q \backslash F$);
5) Преходи: преходите на $\mathscr{A}_1$.

$\overline{\mathscr{A}}$ е идентичен с $\mathscr{A}_1$, с изключение на това, че финалните и нефиналните състояния се разменят.

д) **Сечение**. $\mathscr{L}(\mathscr{A}_1) \cap \mathscr{L}(\mathscr{A}_2) \overset{\text{де Морган}}{=} \overline{\overline{\mathscr{L}(\mathscr{A}_1)} \cup \overline{\mathscr{L}(\mathscr{A}_2)}} = \Sigma^* \backslash \big((\Sigma^* \backslash \mathscr{L}(\mathscr{A}_1) \cup (\Sigma^* \backslash \mathscr{L}(\mathscr{A}_2))\big)$, т.е. затвореността при операцията сечение следва директно от затвореността при операцията обединение и допълнение.

**Теорема на Клини**

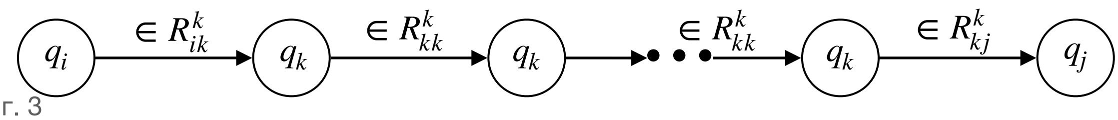
Един език е регулярен, тогава и само тогава, когато се разпознава от краен автомат. (Тоест множеството на регулярните езици и множеството на автоматните езици съвпадат.)

**Доказателство**

($\Rightarrow$) Ще покажем с индукция по дефиницията на регулярни езици, че всеки регулярен език е автоматен. Действително $\varnothing$ и $\{x\}$, $\forall x \in \Sigma$ са автоматни езици, тъй като са крайни. Ако допуснем, че регулярните езици $\mathscr{L}(\alpha)$ и $\mathscr{L}(\beta)$, съответни на регулярните изрази $\alpha$ и $\beta$ са автоматни, тогава $\mathscr{L}(\alpha) \cup \mathscr{L}(\beta)$, $\mathscr{L}(\alpha) \bullet \mathscr{L}(\beta)$ и $\mathscr{L}(\alpha)^*$ са автоматни, тъй като са съответно обединение, конкатенация и итерация на автоматни езици. Тъй като други регулярни езици няма, всеки регулярен език е автоматен.

($\Leftarrow$) Нека езикът $L$ е автоматен. Ще докажем, че $L$ е регулярен език. Съществува КДА $\mathscr{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$ такъв, че $L = \mathscr{L}(\mathscr{A})$. Нека автоматът $\mathscr{A}$ е представен с краен ориентиран мултиграф $G$. Нека състоянията на $\mathscr{A}$ са $Q = \{q_0, q_1, \ldots, q_n\}$ и $F = \{q_{p_1}, q_{p_2}, \ldots, q_{p_l}\}$. Да означим с $R_{ij}^k$ множеството от маршрутите в $G$ от връх $q_i$ до връх $q_j$, които не използват като вътрешни върхове такива с индекс по-голям или равен на $k$. Очевидно, всеки маршрут от $q_i$ до $q_j$ еднозначно определя дума $\omega \in \Sigma^*$, такава че $\tilde{\Delta}(q_i, \omega) = q_j$. Така на множеството от маршрути $R_{ij}^k$ може да гледаме като на множество от съответните думи от $\Sigma^*$, т.е. $R_{ij}^k$ е език над $\Sigma^*$ – *маршрутен език*. В автомата няма състояния с номера по-големи от $n$ и затова $R_{ij}^k$ е езика на всички маршрути от $q_i$ до $q_j$, когато $k > n$.

Да разгледаме маршрутите $R_{ij}^{k+1}$, не минаващи през $q_{\overline{k+1,\, n}}$. Разбиваме тези маршрути на две подмножества: $R_{ij}^k$ (такива, които не минават през $q_k$) и $\tilde{R}$ (такива, които минават през $q_k$). Получаваме, че $R_{ij}^{k+1} = R_{ij}^k \cup \tilde{R}$. Всеки маршрут от $q_i$ до $q_j$, минаващ през $q_k$ и не минаващ през $q_{\overline{k+1,\, n}}$ да разбием на части с краища срещанията на $q_k$ в маршрута.




фиг. 3

Всеки такъв маршрут започва с подмаршрут от $R_{ik}^k$ и завършва с подмаршрут от $R_{kj}^k$, между които може да има произволен брой подмаршрути от $R_{kk}^k$, тоест $\tilde{R} = R_{ik}^k \left(R_{kk}^k\right)^* R_{kj}^k$ и така $R_{ij}^{k+1} = R_{ij}^k \cup R_{ik}^k \left(R_{kk}^k\right)^* R_{kj}^k$.

Сега остана да докажем, че за всеки $q_i, q_j \in Q : R_{ij}^k$, за $k = \overline{0, n+1}$ е регулярен език.
Доказателството ще го осъществим с помощта на индукция по $k$.

а) *Индукционна база*: Нека $k = 0$. Ако $i = j$, $R_{ii}^0$ са всички маршрути от $q_i$ до $q_i$, не минаващи през никой друг връх. Ако мултиграфът на автомата няма примки в $q_i$, единствен такъв маршрут е празният, без нито едно ребро и съответната му дума е $\varepsilon$. Тогава $R_{ii}^0 = \{\varepsilon\}$ е регулярен език. Другата възможност е в мултиграфа да има примки от $q_i$ до $q_i$, надписани с буквите $x_{i_s}$, за $s = \overline{1, j}$. Тогава $R_{ii}^0 = \{\varepsilon, x_{i_1}, x_{i_2}, \ldots, x_{i_j}\}$ е регулярен език, тъй като се представя с регулярния израз

$$\varepsilon \cup x_{i_1} \cup x_{i_2} \cup \ldots \cup x_{i_j} = \bigcup_{s=1}^{j} x_{i_s} \cup \varepsilon.$$

Ако $i \neq j$ тогава $R_{ij}^0$ са всички маршрути от $q_i$ до $q_j$ не минаващи през които и да е от останалите върхове, т.е. ребрата от $q_i$ до $q_j$. Ако няма такива ребра $R_{ij}^0 = \emptyset$ е регулярен език, а ако $x_{i_s}$, за $s = \overline{1, k}$ са буквите по ребрата от $q_i$ до $q_j$, тогава $R_{ij}^0 = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$ и $R_{ij}^0$ е регулярен език, защото се представя с регулярния израз $\bigcup_{s=1}^{k} x_{i_s}$.

б) Индукционна хипотеза: Да допуснем, че за някое $k \geq 0$, $R_{ij}^k$ е регулярен език $\forall q_i, q_j \in Q$, тоест съществува регулярен израз $\alpha_{ij}^k$ за него.

в) Езикът $R_{ij}^{k+1}$ също е регулярен, защото може да се представи по следния начин: $R_{ij}^{k+1} = R_{ij}^{k} \cup R_{ik}^{k} \left(R_{kk}^{k}\right)^{*} R_{kj}^{k} = \alpha_{ij}^{k} \cup \alpha_{ij}^{k} \left(\alpha_{kk}^{k}\right)^{*} a_{kj}^{k}$, като израза съдържа само обединение, конкатенация и итерация на регулярни изрази. Но ние вече знаем за затвореността на регулярните изрази и езици спрямо тези операции, откъдето следва желаният резултат.

Окончателно, от дефиницията на език, разпознаван от КДА получаваме, че $\mathscr{L}(\mathscr{A}) = R_{0p_1}^{n+1} \cup R_{0p_2}^{n+1} \cup \ldots \cup R_{0p_l}^{n+1}$, тъй като езика на $\mathscr{A}$ се състои от всички маршрути (разбирай съответните им думи), довеждащи автомата от началното състояние $q_0$ до някое от крайните $q_{p_1}, q_{p_2}, \ldots, q_{p_l}$. Следователно $\mathscr{L}(\mathscr{A})$ е регулярен език, защото е крайно обединение на регулярни езици. $\square$

Доказателството на тази Теорема ни дава и алгоритъм, по който може да построяваме регулярни изрази на езици, зададени с автомат. Нека построим регулярен израз за езика, който се задава от автомата от фиг. 2.
Нека $q_0 = t_{[0]}$, $q_1 = t_{[0,1]}$, $q_2 = t_{[0,2]}$. Получаваме, че $\mathscr{L}(\mathscr{A}) = R_{02}^{3} = R_{02}^{2} \cup R_{02}^{2}(R_{22}^{2})^{*}R_{22}^{2}$, но

$$R_{02}^{2} = R_{02}^{1} \cup R_{01}^{1}(R_{11}^{1})^{*}R_{12}^{1} = \varnothing \cup b^{*}a(a)^{*}b = b^{*}aa^{*}b,$$
$$R_{22}^{2} = R_{22}^{1} \cup R_{21}^{1}(R_{11}^{1})^{*}R_{12}^{1} = \varnothing \cup (a \cup bb^{*}a)(a)^{*}b = (\varepsilon \cup b^{+})aa^{*}b = b^{*}aa^{*}b.$$

Сега, $R_{02}^{2} = R_{22}^{2} = R$ и
$$\mathscr{L}(\mathscr{A}) = R \cup RR^{*}R = R(\varepsilon \cup R^{*}R) = R(\varepsilon \cup R^{+}) = R^{*}R = (b^{*}aa^{*}b)^{*}b^{*}aa^{*}b =$$
$$= (b^{*}a^{*}ab)^{*}b^{*}a^{*}ab\,.$$
Ще покажем, че $(b^{*}a^{*}ab)^{*}b^{*}a^{*} = (a \cup b)^{*}$, т.е. съответният му език съдържа всяка дума на азбуката $\{a, b\}$. Очевидно $\varepsilon$ е от езика на израза $(b^{*}a^{*}ab)^{*}b^{*}a^{*}$. Затова да разгледаме произволна дума $\omega \neq \varepsilon$. Да намерим в $\omega$ всички места, в които се среща комбинацията от букви $ab$. Очевидно е, че между всеки две двойки $ab$ може да стои само дума от вида $b^{*}a^{*}$, а след последното срещане на $ab$ – дума от същия вид. Следователно, всяка дума от $(a \cup b)^{*}$, различна от $\varepsilon$ се представя във вида $(b^{*}a^{*}ab)^{*}b^{*}a^{*}$. И така получаваме $\mathscr{L}(\mathscr{A}) = (a \cup b)^{*}ab$ – факт, който се вижда лесно в недетерминирания автомат за същия език (виж фиг. 1).

**Лема за покачването** $(xyz)$

**За всеки** регулярен език $L$, **съществува** цяло положително число $n$, такова че **за всяка** дума $\omega \in L$ с дължина $|\omega| > n$, **съществуват** думи $x$, $y$ и $z$, за които $\omega = xyz$, $y \neq \varepsilon$ и $|xy| \leq n$, такива че **за всяко** $i \geq 0$: $xy^{i}z \in L$.

**Доказателство**

За регулярния език $L$ съществува КДА $\mathscr{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$, който го разпознава. Избираме $n = |Q|$. Ще покажем, че това е търсеното цяло число. Нека $\omega \in L$ и $\omega = a_{j_1}a_{j_2}\ldots a_{j_k}$, $|\omega| = k > n$. Тъй като $\mathscr{A}$ разпознава $\omega$, то $\Delta(q_0, \omega) \in F$,където за улеснение сме означили разширената функция на преходите на $\mathscr{A}$ да е $\Delta$ вместо $\tilde{\Delta}$.
Тогава $\delta(q_0, a_{j_1}) = q_{j_1}$, $\delta(q_{j_1}, a_{j_2}) = q_{j_2}$ , ... , $\delta(q_{j_{k-1}}, a_{j_k}) = q_{j_k} \in F$. Ако разгледаме редицата от състояния $q_0, q_{j_1}, \ldots, q_{j_k}$, $k > n$, през които автоматът преминава при работа върху думата $\omega$, то от принципа на Дирихле получаваме, че в редицата има поне една двойка съвпадащи състояния. Избираме най-ляво разположената двойка $q_{j_m} \equiv q_{j_p}$, $m < p$, т.е. вляво от $q_{j_m}$ няма друга такава двойка. Разбиваме $\omega$ на три поддуми $x$, $y$ и $z$, такива че

$\Delta(q_0, x) = q_{j_m}$, $\Delta(q_{j_m}, y) = q_{j_p}$, $\Delta(q_{j_p}, z) = q_{j_k} \in F$. Ще покажем, че тези три думи удовлетворяват твърденията на теоремата.

1) Тъй като $m < p \Rightarrow y \neq \varepsilon$
2) Тъй като $q_{j_m}$ и $q_{j_p}$ е най-ляво разположената двойка, то $|xy| \leq n$. В противен случай отново щяхме да можем да приложим принципа на Дирихле и да намерим друга двойка съвпадащи състояния в редицата $q_0, q_{j_1}, \ldots, q_{j_{p-1}}$.
3) С индукция по $i$ доказваме, че $\Delta(q_0, xy^i) = q_{j_m} = q_{j_p}$. Действително, $\Delta(q_0, xy^0) = $ $= \Delta(q_0, x) = q_{j_m} = q_{j_p}$. Допускаме, че твърдението е вярно и за някое $i \geq 0$ и да разгледаме $\Delta(q_0, xy^{i+1}) = \Delta(\Delta(q_0, xy^i), y) = \Delta(q_{j_m}, y) = q_{j_p} = q_{j_m}$.

Сега за всяко цяло число $i \geq 0$ имаме $\Delta(q_0, xy^iz) = \Delta(\Delta(q_0, xy^i), z) = \Delta(q_{i_p}, z) = q_{j_k} \in F$ и следователно $xy^iz \in L$, $\forall i = 0, 1, 2, \ldots$.

**Примери за регулярни и нерегулярни езици**

1.1. Езикът $L = \{a^m b^m \,|\, m \geq 1\}$ не е регулярен.

Доказателство. Допуснеме, че $L$ е регулярен. Лемата за покачването ни дава число $n \geq 1$, за което $\omega = a^n b^n$ е такава, че $\omega \in L$ и $|\omega| = 2n > n$. Следователно $\omega = xyz$, за които $|xy| \leq n$, $y \neq \varepsilon$. Това означава, че $y = a^j$ за някое $j > 0$. Но според лемата $xy^iz \in L$, за всяко $i \geq 0$ и ако вземем $i = 0$ получаваме, че $xz = a^{n-j}b^n \in L$, което не е изпълнено и следователно допускането, че $L$ е регулярен е грешно.

1.2. Езикът $L = \{a^m \,|\, m$ е просто$\}$ не е регулярен.

Доказателство. Допуснеме, че $L$ е регулярен. Лемата за покачването ни дава число $n \geq 1$, за което избираме просто число $p > n$ и следователно $\omega = a^p \in L$ и $|\omega| > n$. Тогава $\omega = xyz$, за които $|xy| \leq n$, $y \neq \varepsilon$ и $xy^iz \in L$, за всяко $i = 0, 1, 2, \ldots$. Нека $x = a^q$, $y = a^r$, $z = a^s$ и $p = q + r + s$ за $q, s \geq 0$ и $r > 0$. Тогава $q + ir + s$ е просто число за всяко $i$. Но ако вземем $i = q + 2r + s + 2$ ще получим, че $q + ir + s = q + (q + 2r + s + 2)r + s = q + (q + 2r + s)r + 2r + s =$ $= (r + 1)(q + 2r + s)$, което е съставно и следователно допускането, че $L$ е регулярен е грешно.

1.3. Езикът $L = \{\omega \in \{a, b\}^* \,|\, \omega = \omega^R\}$ не е регулярен.

Доказателство. Допускаме, че $L$ е регулярен. Лемата за покачването ни дава число $n \geq 1$. Нека $\omega = a^n b a^n$ (взимаме някакъв палиндром), за която $|\omega| = 2n + 1 > n$. Следователно $\omega = xyz$, за които $|xy| \leq n$, $y \neq \varepsilon$. Това означава, че $x$ и $y$ са съставени само от буквите $a$. Нека тогава $x = a^k$, $y = a^m$, за $k \geq 0$ и $m > 0$. Следователно $z = a^{n-k-m}ba^n$. Но според лемата $xy^iz \in L$, за всяко $i \geq 0$. Тоест $k + im + n - k - m = n$. Нека $i = 2$. Получаваме, че $m = 0$, което е противоречие с лемата и следователно допускането, че $L$ е регулярен е грешно.

Имайки тези нерегулярни езици и знаейки, че регулярните езици са затворени спрямо операциите споменати по-рано в темата, може да конструираме лесно още езици, които не са регулярни. Например:

1.4. Езикът $L = \{\omega \in \{a, b\}^* \,|\, \omega$ не е палиндром$\}$ не е регулярен. Това е така, тъй като този език е допълнението на езика от 1.3., който вече доказахме, че не е регулярен.

1.5. Езикът $L = \{\omega \in \{a, b\}^* \,|\, \omega$ има равен брой $a$-та и $b$-та$\}$ не е регулярен. $L$ не е регулярен, защото ако беше, то тогава и езика $L \cap a^*b^*$ щеше да е регулярен. Но този език е точно $\{a^m b^m : m \geq 0\}$, който вече доказахме, че не е регулярен в 1.1.

2.1. $L = \{\omega \in \{a\}^* \,|\, \omega$ има поне едно $a\}$ е рег. език, т.к. $L = a^*a$. Кратък запис $L = a^+$.

2.2. $L = \{\omega \in \{0, 1\}^* \,|\, \exists k \geq 0$ и $\omega$ е двоичен запис на $2^k + 1\}$ е регулярен език, тъй като $L = 10 \cup 10^*1 = 1(0 \cup 0^*1)$.

**Минимизация на състоянията**

Един автоматен (регулярен) език $L$ може да се разпознава от повече от един автомат. Крайните детерминирани автомати $\mathscr{A}_1$ и $\mathscr{A}_2$ наричаме еквивалентни, ако $\mathscr{L}(\mathscr{A}_1) = \mathscr{L}(\mathscr{A}_2)$. Възниква въпроса: „Кой е най-простият автомат, който разпознава езика $L$?“

Дефиниция (минимален автомат). КДА $\mathscr{A}_\mu$, разпознаващ автоматния език $L$, наричаме минимален (за езика $L$), ако за всеки друг автомат $\mathscr{A}$, разпознаващ $L$ е изпълнено $|Q_\mu| \leq |Q|$, където $Q_\mu$ и $Q$ са множествата от състояния съответно на $\mathscr{A}_\mu$ и $\mathscr{A}$.

Задачата за намиране на минималния автомат за автоматния език $L$ по зададен КДА $\mathscr{A}$, разпознаващ $L$ наричаме **минимизация** (на състоянията).

**Теорема на Майхил-Нероуд**

Дефиниция. Релацията $R \subseteq \Sigma^* \times \Sigma^*$, където $\Sigma$ е произволна крайна азбука, наричаме дясно-инвариантна, ако $\forall \gamma \in \Sigma^*$ е в сила $(\alpha, \beta) \in R \Rightarrow (\alpha\gamma, \beta\gamma) \in R$.

Дефиниция ($R_\mathscr{A}$). Нека $\mathscr{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$ е КДА с разширена функция на преходите $\Delta$. Релацията $R_\mathscr{A} \subseteq \Sigma^* \times \Sigma^*$ определяме с $R_\mathscr{A} = \{(\alpha, \beta) \,|\, \Delta(q_0, \alpha) = \Delta(q_0, \beta)\}$.

Дефиниция ($R_L$). Нека $L \subseteq \Sigma^*$. Релацията $R_L \subseteq \Sigma^* \times \Sigma^*$ се състои от такива двойки $(\alpha, \beta)$, за които $\forall \gamma \in \Sigma^*$ $\alpha\gamma$ и $\beta\gamma$ едновременно са или не са в $L$.

Лема. Релацията $R_L$ за езика $L \subset \Sigma^*$ е релация на еквивалентност и е дясно-инвариантна.

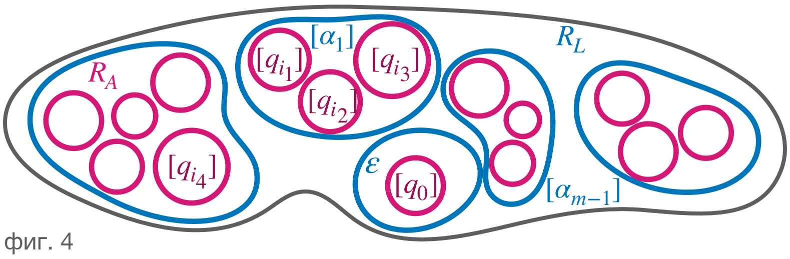
Доказателство на лемата.

1) Релация на еквивалентност.
   a) Рефлексивност. $(\alpha, \alpha) \in R_L$, $\forall \alpha \in \Sigma^*$, защото $\alpha\gamma$ и $\alpha\gamma$ едновременно са или не са в $L$, $\forall \gamma \in \Sigma^*$.
   b) Симетричност. Нека $(\alpha, \beta) \in R_L$, т.е. $\alpha\gamma$ и $\beta\gamma$ едновременно са или не са в $L$, $\forall \gamma \in \Sigma^*$. Тогава $\beta\gamma$ и $\alpha\gamma$ едновременно са или не са в $L$, $\forall \gamma \in \Sigma^*$ и следователно $(\beta, \alpha) \in R_L$ (равенството запазва симетричността).
   c) Транзитивност. Нека $(\alpha, \beta) \in R_L$ и $(\beta, \phi) \in R_L$. Следователно, $\forall \gamma \in \Sigma^*$, $\alpha\gamma$ и $\beta\gamma$ едновременно са или не са в $L$ и $\beta\gamma$ и $\phi\gamma$ едновременно са или не са в $L$. Тогава $\alpha\gamma$ и $\phi\gamma$ едновременно са или не са в $L$, $\forall \gamma \in \Sigma^*$ и $(\alpha, \phi) \in R_L$.
   Следователно $R_L$ е релация на еквивалентност.
2) Дясно-инвариантна.
   Нека $(\alpha, \beta) \in R_L$, т.е. $\forall \gamma \in \Sigma^*$ $\alpha\gamma$ и $\beta\gamma$ едновременно са или не са в $L$. Да разбием $\gamma$ по произволен начин на $\gamma = \gamma'\gamma''$, т.е. $\forall \gamma'' \in \Sigma^*$ $(\alpha\gamma')\gamma''$ и $(\beta\gamma')\gamma''$ едновременно са или не са в $L$. Следователно $\forall \gamma' \in \Sigma^*$ $(\alpha\gamma', \beta\gamma') \in R_L$. Следователно $R_L$ е дясно-инвариантна.

**Теорема (Майхил-Нероуд)**. Нека $L \subseteq \Sigma^*$. Релацията $R_L \subseteq \Sigma^* \times \Sigma^*$ има краен индекс $\Leftrightarrow L$ е автоматен.

**Доказателство**.

($\Leftarrow$) Нека $L$ е автоматен език. Тогава съществува КДА $\mathscr{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$, който разпознава $L$. Без ограничение на общността можем да считаме, че сме се освободили от недостижимите състояния на $\mathscr{A}$, което не се отразява на разпознавания език. Образуваме релацията $R_{\mathscr{A}}$. Нека $(\alpha, \beta) \in R_{\mathscr{A}}$. Ще покажем, че $(\alpha, \beta) \in R_L$. От $(\alpha, \beta) \in R_{\mathscr{A}}$ следва $\Delta(q_0, \alpha\gamma) = \Delta(\Delta(q_0, \alpha), \gamma) = \Delta(\Delta(q_0, \beta), \gamma) = \Delta(q_0, \beta\gamma) = q$. Или $q \in F$ и тогава $\alpha\gamma \in L$ и $\beta\gamma \in L$, или $q \notin F$ и тогава $\alpha\gamma \notin L$ и $\beta\gamma \notin L$. Следователно $(\alpha, \beta) \in R_L$. От тук получаваме, че всеки клас на еквивалентност на $R_{\mathscr{A}}$ се съдържа в клас на еквивалентност на $R_L$ (виж фиг. 4). Това означава, че $R_{\mathscr{A}}$ е изфиняване на $R_L$ и следоватлно $IX(R_L) \leq IX(R_{\mathscr{A}}) = |Q_{\mathscr{A}}|$. Следователно $IX(R_L)$ е краен.




фиг. 4

($\Rightarrow$) Нека $IX(R_L) = m < \infty$(т.е. е краен). Да означим с $[\alpha]$ класа на еквивалентност на $R_L$, съдържащ думата $\alpha \in \Sigma^*$. Да означим с $Q = \{[\varepsilon], [\alpha_1], \ldots, [\alpha_{m-1}]\}$ множеството от класовете на еквивалентност на $R_L$, където $\alpha_0 = \varepsilon, \alpha_1, \ldots, \alpha_{m-1}$ са представители на съответните класове. Образуваме КДА $\mathscr{A} = \langle Q, \Sigma, \delta, [\varepsilon], F \rangle$, където $F = \{[\alpha_{p_1}], [\alpha_{p_2}], \ldots, [\alpha_{p_r}]\}$ са всички класове на еквивалентност на $R$, съдържащи само думи от езика $L$. Ясно е от дефиницията на $R_L$, че всички думи от един и същ клас на $R_L$ едновременно са или не са в $L$, така че дефиницията на $F$ е коректна. Функцията на преходите $\delta$ дефинираме по следния начин: $\delta([\alpha], x) = [\alpha x]$. Тази дефиниция също е коректна, защото $\forall \beta \in [\alpha]$, $[\beta x] = [\alpha x]$, заради дясната инвариантност на $R_L$.

Ще докажем, че КДА $\mathscr{A}$ разпознава точно езика $L$.

a) Нека $\omega \in L_{\mathscr{A}}$, т.е. $\Delta([\varepsilon], \omega) \in F$. Но $\Delta([\varepsilon], \omega) = [\varepsilon\omega] = [\omega]$, т.е. съществува $p_j$, $[\omega] = [\alpha_{p_j}]$. Но тогава $\omega \in L$, защото всички думи на $[\alpha_{p_j}]$ са от $L$ според дефиницията на $[\alpha_{p_j}]$.

b) Нека $\omega \in L$. Тогава $[\omega] = [\alpha_{p_j}]$ за някое $p_j$ и $\Delta([\varepsilon], \omega) = [\omega] = [\alpha_{p_j}] \in F$. Следователно $\omega \in L_{\mathscr{A}}$.

Следователно езикът $L$ е автоматен. $\square$

**Алгоритъм за конструиране на минимален автомат, еквивалентен на даден детерминиран краен автомат**

**Дадено:** КДА $\mathscr{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$ с език $L_{\mathscr{A}}$

**Резултат:** КДА $\mathscr{A}_\mu$ – минимален за езика $L_{\mathscr{A}}$

**Процедура:**

1. Образуваме разбиването $\mathscr{Q}^0 = \{Q_1^0, Q_2^0\}$ на $Q$, където $Q_1^0 = Q \backslash F$, $Q_2^0 = F$. Нека $i = 0$.
2. Нека сме построили разбиването $\mathscr{Q}^i = \{Q_1^i, Q_2^i, \ldots, Q_{l_i}^i\}$ на $Q$ за някое $i$. Всяко $Q_j^i$, $j = 1, 2, \ldots, l_i$, разбиваме на такива подмножества, елементите на които не се проявяват като нееквивалентни с теста на едната буква, за никое $x \in \Sigma$. Обединяваме получените разбивания и нека резултатът е разбиването $\mathscr{Q}^{i+1} = \{Q_1^{i+1}, Q_2^{i+1}, \ldots, Q_{l_{i+1}}^{i+1}\}$.
3. Ако $\mathscr{Q}^i$ и $\mathscr{Q}^{i+1}$ са едно и също разбиване – край на процедурата. В противен случай, $i = i + 1$ и преминаваме към 2.