

4. Крайни автомати. Регулярни изрази. Теорема на Клини.

Дефиниция: Краен детерминиран автомат (КДА) наричаме петорката $A = \langle Q, \Sigma, \delta, s, F \rangle$, където:

- Q – крайно множество от състояния
- Σ – крайно множество от букви (символи)
- $\delta - Q \times \Sigma \rightarrow Q$ – функция на прехода
- $s \in Q$ – начално състояние
- $F \subseteq Q$ – множество от крайни състояния

Дефиниция: Дефинираме функцията $\hat{\delta}$ (разширение на δ), $\hat{\delta} : Q \times \Sigma^* \rightarrow Q$ по следния начин:

$$\hat{\delta}(q, \epsilon) = q, \hat{\delta}(q, aw) = \hat{\delta}(\delta(q, a), w), a \in \Sigma, w \in \Sigma^*$$

Дефиниция: Една дума $w \in \Sigma^*$ се разпознава от автомата A , ако $\hat{\delta}(s, w) = f \in F$. Така дефинираме, че A разпознава езика $L(A) = \{ w \in \Sigma^* \mid \hat{\delta}(s, w) \in F \}$.

Лема: $w \in L(A) \Leftrightarrow \exists$ път $S = sq_1q_2 \dots f = S \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \rightarrow \dots \xrightarrow{a_k} f$, където $w = a_1a_2\dots a_k$, $f \in F$.
Означаваме с $s \xRightarrow{w} f$.

Дефиниция: Текуща конфигурация на A ще наричаме текущото състояние, в което се намираме, и остатъка от думата, който не е прочетен: $(q, w) \in Q \times \Sigma^*$.

Дефиниция: Дефинираме релацията $\vdash_A \subseteq (Q \times \Sigma^*) \times (Q \times \Sigma^*)$ между конфигурации по следния начин: $(p, u) \vdash_A (q, v) \Leftrightarrow u = av$ и $\delta(p, a) = q$.

Дефинираме рефлексивно и транзитивно затваряне на $\vdash_A$:

$$\begin{aligned} (q, \epsilon) &\vdash_A^* (q, \epsilon) \\ (q, au) &\vdash_A^* (p, v) \Leftrightarrow (q, au) \vdash_A (r, u) \ \& \ (r, u) \vdash_A^* (p, v) \end{aligned}$$

Твърдение: $w \in L(A) \Leftrightarrow \hat{\delta}(s, w) \in F \Leftrightarrow s \xRightarrow{w} f \in F \Leftrightarrow (s, w) \vdash_A^* (f, \epsilon)$.

Дефиниция: Краен недетерминиран автомат (КНА) наричаме петорката $A = \langle Q, \Sigma, \delta, s, F \rangle$, където:

- Q – крайно множество от състояния
- Σ – крайно множество от букви (азбука)
- $\delta - Q \times \Sigma \rightarrow 2^Q$
- $s \in Q$ – начално състояние
- $F \subseteq Q$ – множество от крайни състояния

Дефиниция: Дефинираме функцията $\bar{\delta} : 2^Q \times \Sigma \rightarrow 2^Q : \bar{\delta}(M, a) = \bigcup_{p \in M} \delta(p, a)$.

Дефиниция: Дефинираме функцията $\hat{\delta} : 2^Q \times \Sigma^* \rightarrow 2^Q : \hat{\delta}(M, \epsilon) = M, \hat{\delta}(M, aw) = \hat{\delta}(\hat{\delta}(M, a), w)$.

Дефиниция: Езикът, който разпознава КНА A , дефиниран така:

$$L(A) = \{ w \in \Sigma^* : \hat{\delta}(s, w) \cap F \neq \emptyset \}$$

Дефиниция: Дефинираме аналогично $\vdash_A$ и $\vdash_A^* : (q, u) \vdash_A (p, v)$, ако $u = av$ и $p \in \delta(q, a)$.

Дефиниция: Автоматите A_1 и A_2 наричаме еквивалентни, ако $L(A_1) = L(A_2)$.

Теорема: За всеки недетерминиран автомат съществува еквивалентен детерминиран.

Доказателство: Нека $A = \langle Q, \Sigma, \delta, s, F \rangle$ е КНА. Образоваме $A' = \langle Q', \Sigma, \bar{\delta}, s, F' \rangle$, $Q' \subseteq 2^Q$, $F' = \{ M \subseteq Q \mid M \cap F \neq \emptyset \}$, $\bar{\delta}(M, a) = \bigcup_{p \in M} \delta(p, a)$. A' е КДА: новите състояния представляват множество от старите състояния, така че новополученият A' да е детерминиран (всяко състояние има най-много един преход с буква от Σ).

$$L(A) = \{ w \in \Sigma^* : \hat{\delta}(s, w) \cap F \neq \emptyset \} \stackrel{\text{def}}{=} \{ w \in \Sigma^* : \hat{\delta}(s, w) \in F' \} = L(A')$$

Следователно A и A' са еквивалентни.

Дефиниция: Нека Σ е крайна азбука. Множеството от регулярни изрази в Σ^* е :

1. $\emptyset, \epsilon, a$, където $a \in \Sigma$
2. Ако α и β са регулярни изрази $\Rightarrow (\alpha \cdot \beta)$ също е регулярен израз
3. Ако α и β са регулярни изрази $\Rightarrow (\alpha \cup \beta)$ също е регулярен израз
4. Ако α е регулярен израз $\Rightarrow \alpha^*$ също е регулярен израз
5. Няма други регулярни изрази

Дефиниция: Ако имаме даден регулярен израз α , **регулярен език** е множеството $L(\alpha)$, където L е функция, дефинирана по следния начин: (α, β са регулярни изрази)

1. $L(\emptyset) = \emptyset$ и $L(a) = \{a\}$, $a \in \Sigma$
2. $L(\alpha\beta) = L(\alpha)L(\beta)$
3. $L(\alpha \cup \beta) = L(\alpha) \cup L(\beta)$
4. $L(\alpha^*) = L(\alpha)^*$

Теорема: Всеки регулярен език е автоматен.

Теорема: Автоматните езици са затворени относно :

- Обединение
- Конкатенация
- Звезда на Клини
- Допълнение
- Сечение

Теорема: (Клини) Всеки автоматен език е регулярен.

Доказателство: Нека $Q = \{ 1 \dots n \}$ са състояния на КДА $A = \langle Q, \Sigma, \delta, s, F \rangle$.
Дефинираме $L_{ij} = L(\langle \{ 1 \dots n \}, \Sigma, \delta, i, \{ j \} \rangle)$. Дефинираме $L_f = \{ w \in \Sigma^* : \hat{\delta}(s, w) = f \}, \forall f \in F$.
Очевидно $L_{sf} = L_f$. $L(A) = \bigcup_{f \in F} L_f$. Ще покажем, че за всяко L_f съществува регулярен израз, описващ този език.

Дефинираме L_{ij}^m , $m \in \mathbb{N}_0$, индуктивно :

$$L_{ij}^0 = \{ a \in \Sigma : j \in \delta(i, a) \}, i \neq j$$

$$L_{ij}^0 = \{ a \in \Sigma : j \in \delta(i, a) \} \cup \{ \epsilon \}, i = j$$

$L_{ij}^m = \{ w \in \Sigma^* : \exists \text{ работен път } i \xRightarrow{w} j = iPj \text{ и } P \in \{ 1 \dots m \}^* \}$, т.е. iPj е преход от i до j с междинни състояния $\leq m$.

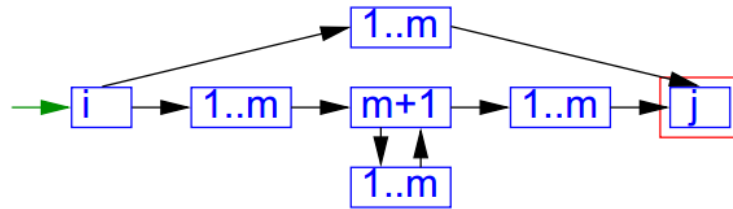
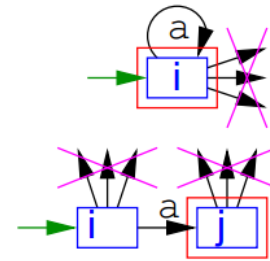
Забелязваме, че $L_{ij}^n = L_{ij}$. Построяваме регулярен израз за L_{ij}^m индуктивно :

$$\text{Ако } m = 0, i = j: \alpha_{ii}^0 = \bigcup_{a \in \Sigma: \delta(i,a)=i} a \cup \epsilon$$

$$\text{Ако } m = 0, i \neq j: \alpha_{ij}^0 = \bigcup_{a \in \Sigma: \delta(i,a)=j} a$$

Ако $m \rightsquigarrow m+1$:

$$\alpha_{ij}^{m+1} = \alpha_{ij}^m \cup \alpha_{i,m+1}^m \cdot (\alpha_{m+1,m+1}^m)^* \cdot \alpha_{m+1,j}^m$$



Т.е. за $\forall m$ съществува регулярен израз α , $L(\alpha) = L_{ij}^m \rightarrow L_{ij}^n \rightarrow L_{ij} \rightarrow L_{sf} \rightarrow L_f \rightarrow \bigcup_{f \in F} L_f = L$.

Теорема: (Pumping lemma, Лема за покачването)

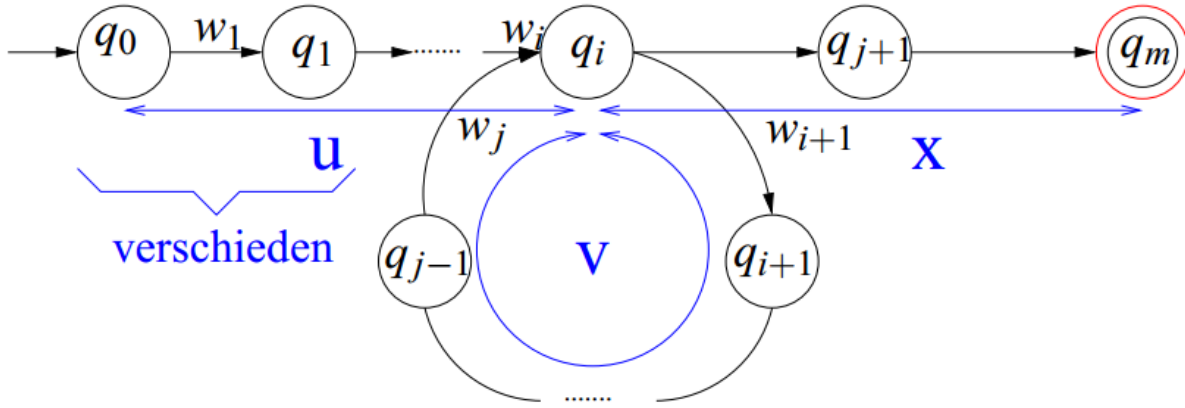
Нека L е регулярен език. Тогава $\exists n \in \mathbb{N}, \forall w \in L, |w| \geq n, \exists u,v,x : w = uvx$ и

1. $v \neq \epsilon$
2. $|uv| \leq n$
3. $uv^i x \in L, \forall i \in \mathbb{N}_0$

Доказателство:

Нека $A = \langle Q, \Sigma, \delta, q_0, F \rangle$. Нека $|Q| = n$ и $w = w_1 \dots w_m, |w| = m \geq n, w \in L. (\forall i = 1 \dots m, w_i \in \Sigma)$.

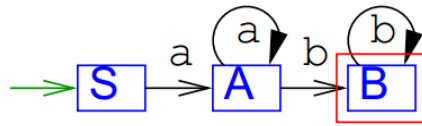
Нека $q_i = \delta(q_{i-1}, w_i) \forall i = 1 \dots m$.



Т.к. думата се състои от поне n букви, то необходимите са поне $(n + 1)$ състояния в свидетелстващия път на w . Тогава от принципа на Дирихле $\exists q_i = q_j, i \neq j, i < j \leq n$.

Фиксираме $u = w_1 \dots w_i; v = w_{i+1} \dots w_j; w = w_{j+1} \dots w_m$. Т.к. $i \neq j \Rightarrow v \neq \varepsilon; i < j \leq n \Rightarrow |uv| \leq n$
 $(q_0, w) \vdash_A^* (q_m, \varepsilon) \Rightarrow (q_0, w) \vdash_A^* (q_i, w_{i+1} \dots w_j \dots w_m) \vdash_A^* (q_j, w_{j+1} \dots w_m)$
 $\Rightarrow (q_0, w_1 \dots w_i) \vdash_A^* (q_i, \varepsilon)$ и $(q_i, w_{i+1} \dots w_j) \vdash_A^* (q_j, \varepsilon)$ и $q_i = q_j$
 $\Rightarrow (q_0, w_1 \dots w_i w_{j+1} \dots w_m) \vdash_A^* (q_m, \varepsilon)$ и
 $(q_0, uv^k x) \vdash_A^* (q_i, v^k x) \vdash_A^* (q_i, v^{k-1}, x) \vdash_A^* \dots \vdash_A^* (q_j, x) \vdash_A^* (q_m, \varepsilon)$

Пример: Езикът $L = \{ a^n b^m \mid n > 0, m > 0 \}$ е регулярен.



Пример: Езикът $L = \{ a^n b^n \mid n \geq 0 \}$ не е регулярен.

Допускаме, че L е регулярен. Тогава от $PL \Rightarrow \exists n \in \mathbb{N}$ и

$w = a^n b^n = uvx$ ($|w| = 2n > n$), $|uv| \leq n, v \neq \varepsilon$

Следователно $v = a^i, n \geq i > 0 \Rightarrow ux = a^{n-i} b^n \in L$ – противоречие.

Дефиниция: Двуместната релация $R \subseteq X \times X$ наричаме **релация на еквивалентност**, ако :

- 1) е рефлексивна ($\forall x (x, x) \in R$)
- 2) е симетрична ($\forall x, y (x, y) \in R \Rightarrow (y, x) \in R$)
- 3) е транзитивна ($\forall x, y, z (x, y) \in R, (y, z) \in R \Rightarrow (x, z) \in R$)

Дефиниция: **Индекс** – броят на класовете на еквивалентност в дадена релация на еквивалентност. Означаваме с $|R|$. Класовете на еквивалентност означаваме с $[x] = \{ y \mid yRx \}$.

Дефиниция: Нека R и R' са релации на еквивалентност. Казваме, че R **прецизира** R' , ако $R \subseteq R'$.

Свойство: Ако R прецизира R' , то $|R| \geq |R'|$.

Доказателство: $y \in [x]_R \Leftrightarrow (y, x) \in R \Rightarrow (y, x) \in R' \Leftrightarrow y \in [x]_{R'}$. Т.е. $\forall [x]_R : [x]_R \subseteq [x]_{R'}$.

Дефиниция: За езика L дефинираме **релация на Нероуд** :

$$R_L := \{ (x, y) \in \Sigma^* \times \Sigma^* \mid \forall z \in \Sigma^* : xz \in L \Leftrightarrow yz \in L \}$$

Свойство: R_L е релация на еквивалентност.

Дефиниция: КДА $M = \langle Q, \Sigma, \delta, s, F \rangle$. Дефинираме $R_M = \{ (x, y) \in \Sigma^* \times \Sigma^* \mid \hat{\delta}(s, x) = \hat{\delta}(s, y) \}$

Свойство: R_M е релация на еквивалентност - за всеки клас на еквивалентност съответства едно състояние.

Твърдение 1: R_M прецизира R_L .

Доказателство: Нека $(x, y) \in R_M \Rightarrow \hat{\delta}(s, x) = \hat{\delta}(s, y) \Rightarrow \forall z \in \Sigma^* : \hat{\delta}(s, xz) = \hat{\delta}(s, yz)$, т.е. $\forall z \in \Sigma^* : xz \in L \Leftrightarrow yz \in L \Rightarrow (x, y) \in R_L$

Твърдение 2: Ако L е регулярен, то $|R_L| < \infty$.

Доказателство: Т.к. L е регулярен, то $\exists$ КДА $M = \langle Q, \Sigma, \delta, s, F \rangle$, $L(M) = L$, $|Q| < \infty$. Имаме, че $|Q| \geq |R_M| \geq (\text{Тв. 1}) |R_L|$. Ако допуснем, че $|R_L| = \infty$, то $|Q| = \infty$ - противоречие. Следователно $|R_L| < \infty$.

Твърдение 3: Минималният автомат, разпознаващ $L(M)$, съдържа $|R_L|$ на брой състояния. (Очевидно от $|Q| \geq |R_M| \geq |R_L|$). Всеки автомат съдържа поне $|R_L|$ състояния. Неговите състояния съответстват на класове на еквивалентност на $R_L : [w_1] \dots [w_k]$. Дефинираме

$$M_{\equiv} := \langle \{ [w_1] \dots [w_k] \}, \Sigma, \delta_{\equiv}, [\epsilon], F_{\equiv} \rangle$$

$$\delta_{\equiv}([w], a) = [wa]$$

$$F_{\equiv} = \{ [w] \mid w \in L \}$$

1) $\delta_{\equiv}$ е добре дефинирана

$$xR_L y \Rightarrow \forall z \in \Sigma^* \quad (xz) \in L \Leftrightarrow yz \in L$$

$$\Rightarrow \forall az \in \Sigma^* \quad x(az) \in L \Leftrightarrow y(az) \in L$$

$$\Rightarrow \forall a \in \Sigma^* \forall z \in \Sigma^* \quad (xa)z \in L \Leftrightarrow (ya)z \in L$$

$$\Rightarrow \forall a \in \Sigma^* : \quad xa R_L ya$$

(т.е. $\delta_{\equiv}$ е еднозначна за всяка една буква)

$$2) \hat{\delta}_{\equiv}([x], y) = [xy]$$

Индуктивно по $|y|$:

$$2.1. \hat{\delta}_{\equiv}([x], \epsilon) = [x]$$

$$2.2. \hat{\delta}_{\equiv}([x], aw) \stackrel{\text{def}}{=} \hat{\delta}_{\equiv}(\delta_{\equiv}([x], a), w) \stackrel{\text{def}}{=} \hat{\delta}_{\equiv}([xa], w) = [xaw]$$

$$3) L(M_{\equiv}) = L$$

$$w \in L(M_{\equiv}) \stackrel{\text{def}}{=} \hat{\delta}_{\equiv}([\epsilon], w) \in \{ [w] : w \in L \} \Leftrightarrow w \in L$$

$$(\hat{\delta}_{\equiv}([\epsilon], w) = [w])$$

Теорема: (Майхил-Нероуд) Езикът L е регулярен $\Leftrightarrow |R_L| < \infty$.

Доказателство:

$\Rightarrow$) от тв. 2

$\Leftarrow$) $|R_L| < \infty$. Тогава построяваме автомата $M_{\equiv}$ и от тв. 3 $L(M_{\equiv}) = L$.

Следствие: Всички минимални автомати за L са изоморфни на $M_{\equiv}$.

Алгоритъм за конструиране на минимален автомат :

Дефинираме релацията $\equiv \subseteq Q \times Q$ за $A = \langle Q, \Sigma, \delta, s, F \rangle$.

$$p \equiv q \Leftrightarrow \forall z \in \Sigma^* : \hat{\delta}(p, z) \in F \Leftrightarrow \hat{\delta}(q, z) \in F$$

За да намерим тези класове на еквивалентност на $\equiv$ (тези класове на еквивалентност са всъщност състоянията на минималния автомат, разпознаващ $L(A)$), ще пресметнем последователно $\equiv_0, \equiv_1, \dots$, където $p \equiv_n q : \forall z \in \Sigma^*, |z| \leq n, \hat{\delta}(p, z) \in F \Leftrightarrow \hat{\delta}(q, z) \in F$.

1. За $\equiv_0$, класовете на еквивалентност са F и $Q \setminus F$
2. Нека сме намерили класове на еквивалентност на $\equiv_n$
Тогава за $\forall p, q \in Q, p \equiv_{n+1} q \Leftrightarrow p \equiv_n q \ \& \ \forall a \in \Sigma : \delta(p, a) \equiv_n \delta(q, a)$. По този начин разбираме дали 2 състояния са в един и същ клас на еквивалентност.
3. Продължаваме, докато не стигнем до стъпка, при която класовете на еквивалентност на $\equiv_n$ съвпадат с класовете на еквивалентност на $\equiv_{n-1}$.