LFA - CURS 4

În cele ce urmează vom nota cu $\mathcal{L}_{REG}$ familia limbajelor descrise de expresiile regulate, care este, așa cum rezultă din echivalențele între AFD, AFN, AFNλ, totuna cu familiile de limbaje recunoscute de aceste tipuri de automate.

Teoremă $\mathcal{L}_{REG}$ este închisă la reuniune, intersecție, diferență, complementară, concatenare, iterație Kleene (*). (Aceasta înseamnă că $\forall L_1, L_2 \subseteq \mathcal{L}_{REG}$, $L_1, L_2 \subseteq \Sigma^*$, atunci $L_1 \cup L_2$, $L_1 \cap L_2$, $L_1 - L_2$, $\overline{L_1} = \Sigma^* - L_1$, $L_1 \cdot L_2$, $L_1^* \in \mathcal{L}_{REG}$)

Demonstrație Fie $L_1, L_2 \in \mathcal{L}_{REG}$, $L_1, L_2 \subseteq \Sigma^*$. Dintr-o teoremă anterioară știm că $L_1 \cup L_2$, $L_1 \cap L_2$, $L_1 - L_2$, $\overline{L_1} \in \mathcal{L}_{REG}$ (Vezi Cursul 2).

Fie r_1, r_2 expresii regulate peste Σ care descriu L_1, L_2, adică $L_1 = L(r_1)$, $L_2 = L(r_2)$, în conformitate cu teorema Kleene.

Atunci $L_1 L_2$ poate fi descris de $r_1 r_2$, iar L_1^* de r_1^*, deci $L_1 \cdot L_2$, $L_1^* \in \mathcal{L}_{REG}$

SUBSTITUȚII DE LIMBAJE

Definiție 1 Fie Σ_1, Σ_2 două alfabete. O *substituție* este o funcție $\sigma : \Sigma_1^* \to 2^{\Sigma_2^*}$ cu proprietățile:

1. $\sigma(\lambda) = \lambda$
2. $\sigma(x \cdot y) = \sigma(x) \cdot \sigma(y)$

Observații

- Prin $2^{\Sigma_2^*}$ înțelegem mulțimea submulțimilor lui Σ_2
- Definirea lui σ pe literele din Σ_1 definește complet pe σ, datorită condiției 2.
- Extensia lui σ la un limbaj $L \subseteq \Sigma_1^*$

$$\sigma(L) = \bigcup_{x \in L} \sigma(x)$$

Exemplu $\Sigma_1 = \{a, b\}$, $\Sigma_2 = \{a, b, c\}$, $\sigma \in \Sigma_1^* \to 2^{\Sigma_2^*}$

$\sigma(a) = \{ab, ac, b\}$, $\sigma(b) = \{b, ba\}$

$$\sigma(ba) = \sigma(b)\sigma(a) = \{b, ba\}\{ab, ac, b\} =$$
$$= \{bab, bac, bb, baab, baac\}$$

Definiție 2 O substituție $f : \Sigma_1^* \to 2^{\Sigma_2^*}$ se numește *substituție regulată* dacă $\forall a \in \Sigma_1$, $f(a) \in \mathcal{L}_{REG}$

Substituția f se numește *morfism* dacă $\forall a \in \Sigma_1$, $|f(a)| = 1$ (fiecare literă din Σ_1 are asociat un limbaj format dintr-un singur cuvânt).

Definiție 3 Morfisme inverse. Fie $h : \Sigma_1^* \to \Sigma_2^*$ un morfism. Definim morfismul invers $h^{-1} : \Sigma_2^* \to 2^{\Sigma_1^*}$ prin $h^{-1}(w) = \{x \mid x \in \Sigma_1^*, h(x) = w\}$, iar pentru $L \subseteq \Sigma_2^*$, $h^{-1}(L) = \bigcup_{x \in L} h^{-1}(x)$

Teorema 2 Familia limbajelor regulate, $\mathcal{L}_{REG}$, este închisă la:

1. substituții regulate
2. morfisme
3. morfisme inverse

Demonstrație $\boxed{1.}$ Fie $\sigma : \Sigma_1^* \to 2^{\Sigma_2^*}$ astfel încât $\forall a \in \Sigma_1$, $\sigma(a) \in \mathcal{L}_{REG}$. Fie $L_1 \in \mathcal{L}_{REG}$, $L_1 \subseteq \Sigma_1^*$.

Arătăm că $\sigma(L_1) \in \mathcal{L}_{REG}$.

Deoarece $L_1 \in \mathcal{L}_{REG}$, rezultă că există r_1 expresie regulată peste Σ_1 care descrie L_1, $L(r_1) = L_1$.

Deoarece $\forall a \in \Sigma$, $\sigma(a) \in \mathcal{L}_{REG}$, rezultă că există r_a expresie regulată peste Σ_2 astfel încât $L(r_a) = \sigma(a)$, $\forall a \in \Sigma_1$.

Construim expresia regulată r_2 peste Σ_2 pornind de la r_1, în care înlocuim fiecare simbol a din r_1 cu expresia r_a. Pentru că r_1 este Exp.Reg peste Σ_1 și r_a sunt Exp.Reg peste Σ_2, rezultă că și r_2 este Exp.Reg peste Σ_2, formată cu ajutorul operatorilor

$+, \cdot, *$ și al expresiilor regulate r_a.

Arătăm că $L(r_2) = \sigma(L_1)$ prin inducție după numărul operatorilor din r_1.

Baza r_1 are 0 operatori. Rezultă că $r_1 \in \{\emptyset, \lambda\} \cup \Sigma_1$.

Dacă $r_1 = \emptyset$ (adică $L_1 = \emptyset$), atunci luăm $r_2 = \emptyset$ și evident că $L(r_2) = \emptyset = \sigma(\emptyset) = \sigma(L(r_1)) = \sigma(L_1)$

Dacă $r_1 = \lambda$ (adică $L_1 = \{\lambda\}$), luăm $r_2 = \lambda$ și atunci $L(r_2) = L(\lambda) = \{\lambda\} = \sigma(\lambda) = \sigma(L(r_1)) = \sigma(L_1)$

Dacă $r_1 = a \in \Sigma_1$ (deci $L_1 = \{a\}$), luăm $r_2 = r_a$ și $L(r_2) = L(r_a) = \sigma(a) = \sigma(L_1)$

Ipoteza de inducție Presupunem că pentru orice limbaj L_1 astfel încât $L_1 = L(r_1)$, unde r_1 expresie regulată peste Σ_1 cu cel mult k operatori $+, \cdot, *$, există r_2 expresie regulată peste Σ_2 cu $L(r_2) = \sigma(L_1)$.

Saltul inductiv. Fie $L_1 \subseteq \Sigma_1^*$, $L_1 = L(r_1)$, unde r_1 expresie regulată cu $k+1$ operatori.

Avem cazurile:

a) $r_1 = r_1' + r_1''$ (sau $r_1 = r_1' \, r_1''$). Din ipoteza de inducție, există r_2', r_2'' expresii regulate peste Σ_2 astfel încât $L(r_2') = \sigma(L(r_1'))$, $L(r_2'') = \sigma(L(r_1''))$.

Luăm $r_2 = r_2' + r_2''$. Deoarece la substituire fiecare a cu r_a, rezultă că:

$$L(r_2)=L(r_2'+r_2'')=L(r_2')\cup L(r_2'')\underset{\substack{\text{ip. de}\\ \text{inducție}}}{=}\sigma(L(r_1'))\cup$$
$$\cup\,\sigma(L(r_2'))=\sigma(L(r_1')\cup L(r_2'))=\sigma(L(r_1'+r_2'))=\sigma(L(r_1)).$$

b) $r_1=r_1'\cdot r_2''$. Facem o construcție asemănătoare ca mai sus pentru r_2 și ținem cont de faptul că $\sigma(L_1\cdot L_2)=\sigma(L_1)\cdot\sigma(L_2)$.

c) $r_1=r_1'^{*}$. Raționăm ca mai sus, ținând cont de faptul că $\sigma(L_1^*)=\sigma(L_1)^*$.

$\boxed{2.}$ Demonstrația pentru morfisme rezultă din 1., deoarece morfismul este caz particular de substituție regulată: pentru fiecare $a\in\Sigma$, $\sigma(a)$ este un limbaj finit format dintr-un singur cuvânt, adică $\sigma(a)$ este regulat.

$\boxed{3.}$ Fie $h:\Sigma_1^*\to\Sigma_2^*$ morfism și $L\subseteq\Sigma_2^*$, $L\in\mathcal{L}_{REG}$.

Arătăm că $h^{-1}(L)\in\mathcal{L}_{REG}$.

Fie $A=(Q,\Sigma_2,\delta,q_0,F)$ un AFD astfel încât $L(A)=L$.

Construim AFD $M=(Q,\Sigma_1,\delta',q_0,F)$ astfel ca:

$$\delta'(q,a)=\delta(q,h(a)),\ \forall q\in Q,\ \forall a\in\Sigma$$

↑ extensia lui δ la cuvinte; $\delta(q,h(a))\in Q$

Arătăm că $\delta'(q,x)=\delta(q,h(x))\ \forall x\in\Sigma_1^*$.

(δ' și δ sunt extensiile pe cuvinte)

Facem inducție după $n=|x|$.

$\boxed{n=0}$ Rezultă $x=\lambda$. Avem $\delta'(q,\lambda)=q=\delta(q,\lambda)=$

$$=\delta(q,h(\lambda))$$

$\boxed{n\to n+1}$ Presupunem că $\delta'(q,x)=\delta(q,h(x))$ pentru orice cuvânt $x\in\Sigma_1^*$ de lungime n.

Fie acum $x=x'a$, $|x|=n+1$, $a\in\Sigma_1$, $|x'|=n$. Avem:

$$\delta'(q,x)=\delta'(q,x'a)\underset{\uparrow\atop\text{def. extensie }\delta'}{=}\delta'(\delta'(q,x'),a)\underset{\text{ip. de inducție}}{=}\delta'(\delta(q,h(x')),a)$$

$$\underset{\uparrow\atop\text{definiție }\delta'}{=}\delta(\delta(q,h(x')),h(a))\underset{\uparrow\atop\text{def. extensie }\delta}{=}\delta(q,h(x')h(a))=$$

$$=\delta(q,h(x'a))=\delta(q,h(x))$$

Rezultă că $\delta'(q_0,w)\in F\iff\delta(q,h(w))\in F$, adică $w\in L(M)\iff h(w)\in L$, deci $L(M)=h^{-1}(L)$, q.e.d.

LEMA DE POMPARE PENTRU LIMBAJE REGULATE

<u>Teoremă</u> Fie $A=(Q,\Sigma,\delta,q_0,F)$ un AFD și $n=|Q|$. Atunci pentru orice șir $x\in L(A)$ astfel încât $|x|\geq n$, există o descompunere a lui x, $x=uvw$, astfel încât:

(i) $|uv|\leq n$

(ii) $|v|\geq 1$

(iii) $\forall k\geq 0$, $uv^kw\in L(A)$

<u>Demonstrație</u>. Fie $x \in L(A)$, $|x| \geq n$, $x = a_1 \cdots a_m$, $m \geq n$, $a_1, \dots, a_m \in \Sigma$. Atunci în A există schimbările de configurații:

$$(q_0, a_1 \cdots a_m) \vdash (q_1, x_2 \cdots x_m) \vdash \cdots \vdash (q_{m-1}, a_m) \vdash (q_m, \lambda),$$

unde $q_i = \delta(q_{i-1}, a_i)$, $1 \leq i \leq m$, $q_m \in F$.

Avem $\{q_0, q_1, \dots, q_n\} \subseteq Q$. Deoarece $|Q| = n$, rezultă că există $i \neq j$, $0 \leq i < j \leq n$ astfel încât $q_i = q_j$.

Considerăm $u = a_1 \cdots a_i$, $v = a_{i+1} \cdots a_j$, $w = a_{j+1} \cdots a_m$, unde $(q_0, u) \vdash^* (q_i, \lambda)$, $(q_i, v) \vdash^* (q_j, \lambda) = (q_i, \lambda)$, $(q_j, w) \vdash^* (q_m, \lambda)$, $q_m \in F$. Rezultă că:

$$(q_i, v^k) \vdash^* (q_j, \lambda) = (q_i, \lambda), \forall k \geq 0.$$

Să observăm că $|uv| \leq n$, deoarece $j \leq n$, $|v| \geq 1$, deoarece $i < j$, iar $uv^k w \in L$. Într-adevăr

$$(q_0, uv^k w) \vdash^* (q_i, v^k w) \vdash^* (q_i, v^{k-1} w) \vdash^* \cdots \vdash^*$$
$$\vdash^* (q_i, w) = (q_j, w) \vdash^* (q_m, \lambda), \; q_m \in F, \text{ deci}$$
$$uv^k w \in L, \; \forall k \geq 0.$$

q.e.d.

APLICAȚII ALE LEMEI DE POMPARE

Lema de pompare este utilizată pentru a arăta că un limbaj nu poate fi acceptat de un AFD (nu este regulat), dar nu poate fi utilizată pentru a arăta că un limbaj este acceptat de un AFD. Condiția din Lema de pompare este o condiție necesară, dar nu suficientă pentru ca un limbaj să fie regulat.

EXEMPLE

1) $L = \{a^i b^i \mid i \geq 0\}$ nu poate fi acceptat de un AFD.

Dem. Presupunem că $L = L(A)$, unde $A = (Q, \Sigma, \delta, q_0, F)$ este un AFD, $n = |Q|$. Fie $x = a^t b^t$ cu $|x| = 2t \geq n$.

Din Lema de pompare rezultă că există o descompunere $x = uvw$ astfel încât:

(i) $|uv| \leq n$

(ii) $|v| \geq 1$

(iii) $uv^k w \in L, \forall k \geq 0$

Avem cazurile:

a) $v = a^k, k \geq 1, k \leq t$

Dar atunci $uv^0 w = a^{t-k} b^k \notin L$

$\boxed{\text{Notație}}$ Pentru un șir $v \in \Sigma^*$, notăm:

$v^0 = \lambda$, $v^1 = v$, $\dots$, $v^i = v^{i-1}v$, $i \geq 1$.

b) $v = b^k$, $k \geq 1$.

$uv^0w = uw = a^t b^{t-k} \notin L$

c) $v = a^k b^l$, $k, l \geq 1$, $u = a^{t-k}$, $w = a^{t-l}$.

Atunci $uv^2w = a^t b^l a^k b^t \notin L$.

2) $L = \{a^{2^n} \mid n \geq 0\}$ nu poate fi recunoscut de un AFD.

Presupunem că există $A = (Q, \Sigma, \delta, q_0, F)$ un AFD

cu $L(A) = L$, $n = |Q|$.

Considerăm $x = a^{2^{n+1}}$, $|x| \geq n$

Din Lema de pompare, există u, v, w cu

$x = uvw$, $|uv| \leq n$, $|v| \geq 1$, $uv^iw \in L$, $\forall i \geq 0$.

Rezultă că v este de forma $v = a^k$, $1 \leq k \leq n$.

Observăm că $uv^0w = uw = a^{2^{n+1}-k}$, iar

$2^{n+1} > 2^{n+1} - k \geq 2^{n+1} - n > 2^{n+1} - 2^n = 2^n$, deci

$2^{n+1} > |uw| > 2^n$, adică $uw \notin L$.

ALTE APLICAȚII ALE LEMEI DE POMPARE — PROBLEME DE DECIZIE PENTRU LIMBAJELE REGULATE

<u>Teoremă</u> Fie $A = (Q, \Sigma, \delta, q_0, F)$ un AFD. Atunci $L(A)$ este infinit dacă și numai dacă $L(A)$ conține un cuvânt x cu $|Q| \leq |x| < |Q|$.

<u>Dem</u> „$\Rightarrow$" Presupunem că $L(A)$ este infinit. Fie $n = |Q|$. Presupunem că nu există $x \in L(A)$ astfel încât $n \leq |x| < n$. Deoarece $L(A)$ infinit, atunci există $x \in L(A)$ cu $|x| \geq 2n$. Din lema de pompare, $x = uvw$, $1 \leq |v| \leq n$ și $x' = uv^0w = uw \in L(A)$.

Observăm că $|x| > |x'| \geq |x| - n$.

Dacă $|x'| \geq 2n$, repetăm procedeul.

Altfel, avem $|x'| < 2n$, $|x| \geq 2n$, $|x'| \geq |x| - n \geq$ $\geq 2n - n = n$, adică $n \leq |x'| < 2n$.

„$\Leftarrow$" Aplicăm lema de pompare.

<u>Teoremă</u> Fie AFD $A = (Q \Sigma, \delta, q_0, F)$. Atunci $L(A) \neq \emptyset \Leftrightarrow$ $\Leftrightarrow \exists x \in \Sigma^*$, $|x| < n$.

<u>Dem</u> „$\Rightarrow$" Fie $x \in L(A)$ cu cea mai mică lungime posibilă. Presupunem $|x| \geq n$. Din lema de pompare, $x = uvw$, $|uv| \leq n$, $|v| \geq 1$, $uv^iw \in L$, $\forall i \geq 0$. Deci $uw \in L$, dar $|uw| < |uvw| = |x|$, contradicție.

„$\Leftarrow$" Evident.