LFA CURS 3

EXPRESII REGULATE. ECHIVALENȚA CU AUTOMATELE FINITE

Definiție 1 Fie Σ un alfabet. Definim expresiile regulate peste Σ și limbajele descrise de acestea, într-un mod recursiv, astfel:

i) ϕ este expresie regulată peste Σ care descrie limbajul vid, ϕ.
ii) λ este expresie regulată peste Σ care descrie limbajul $\{\lambda\}$
iii) $\forall a \in \Sigma$, a este expresie regulată peste care descrie limbajul $\{a\}$

Fie r și s sunt expresii regulate peste Σ care descriu, respectiv, limbajele $R, S \subseteq \Sigma^*$. Atunci:

iv) $r+s$ (sau uneori $r|s$), rs și r^* sunt expresii regulate peste Σ care descriu, respectiv, limbajele: $R \cup S$, $R \cdot S$ și R^*
v) (r) este expresie regulată peste Σ care descrie limbajul R.

Definiție 2 Limbajul descris de o expresie regulată r se notează cu $L(r)$.

Observații

1) Prioritățile celor trei operatori ce apar în expresiile regulate sunt, în ordine crescătoare: $+$, $\cdot$, $*$
2) Operatorul $*$ se numește iterație Kleene
3) Parantezele sunt folosite pentru a schimba

prioritățile unor operatori.

EXEMPLE

1) $(0+1)^*$ descrie toate șirurile peste alfabetul $\{0,1\}$

2) $(0+1)^*00(0+1)^*$ descrie toate șirurile peste $\{0,1\}$ care conțin două '0'-uri consecutive

3) $(0^*1^*)^* = (0+1)^*$ - aceste expresii descriu același limbaj, cel al șirurilor peste $\{0,1\}$

4) $(a+b)^*ab(a+b)^* + b^*a^* = (a+b)^*$ este adevărată pentru că primul termen al expresiei stângi conține șirurile peste $\{a,b\}$ ce conțin 'ab' ca subșir, pe când cel de-al doilea termen, b^*a^*, descrie toate șirurile peste $\{a,b\}$ ce nu conțin 'ab' ca subșir.

5) Limbajul șirurilor din $\{a,b\}^*$ cu număr impar de 'a', să îl notăm cu L5:

a) $b^*ab^*(ab^*ab^*)^*$ sau

b) $b^*a(b^*ab^*a)^*b^*$ sau

c) $b^*a(b+ab^*a)^*$

Expresiile $b^*ab^*(ab^*a)^*b^* = r_1$ și $b^*a(b^*ab^*ab^*) = r_2$ nu sunt corecte, deoarece nu descriu toate șirurile din L_5.

Într-adevăr, a abababa $\notin L(r_1)$, dar este în L_5, iar a bababaa $\notin L(r_2)$, dar este în L_5.

6) Limbajul șirurilor peste $\{a,b\}$ în care numărul de 'a' și numărul de 'b' sunt pare

$$(aa+bb+(ab+ba)(aa+bb)^*(ab+ba))^*$$

7) Limbajul șirurilor peste $\{a,b\}$ care se termină în 'b' și nu conțin 'aa':

$$(b+ab)^*(b+ab)$$

<u>Teoremă</u> (Kleene) Un limbaj este recunoscut de un automat finit dacă și numai dacă poate fi descris de o expresie regulată

<u>Dem</u>. Vom demonstra Teorema în 2 pași.

<u>Lema 1</u> Fie r o expresie regulată. Atunci există un AFNλ care acceptă limbajul descris de r.

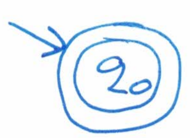
<u>Dem Lema 1</u> Arătăm prin inducție după numărul operatorilor $+, \cdot, *$ care apar în r, $n = Nop(r)$ este acceptat de un AFNλ cu o singură stare finală din care nu pleacă nicio tranziție.

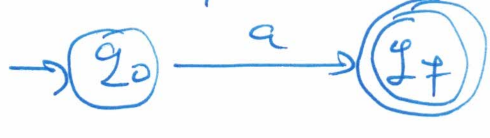
Fie Σ alfabetul de bază al expresiei r.

$\boxed{n=0}$ Expresia r nu are niciun operator $+, \cdot, *$.

Atunci avem cazurile

• $r=\emptyset$: AFNλ corespondent: 

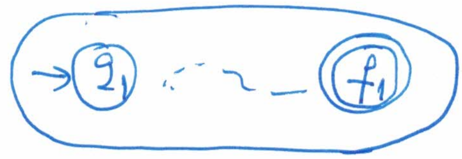
• $r=\lambda$: 

$r = a,\ a \in \Sigma$: 

$\boxed{n \to n+1}$ Presupunem că pentru orice expresie regulată e peste Σ cu $Nop(e) \le n$ există un AFN_λ care recunoaște $L(e)$.

Fie r expresie regulată cu $n+1$ operatori.

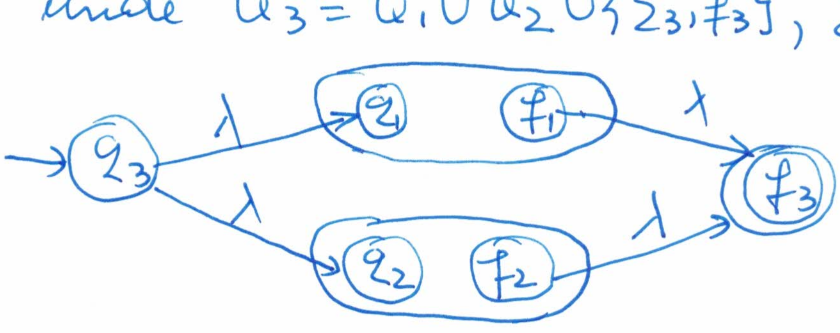
Atunci r poate fi de forma $r_1 + r_2$, $r_1 \cdot r_2$, r_1^* unde $Nop(r_1) \le n$, $Nop(r_2) \le n$. În conformitate cu ipoteza de inducție, există AFN_λ A_1 și A_2 pentru care $L(A_1) = L(r_1)$, $L(A_2) = L(r_2)$,

$A_1 = (Q_1, \Sigma, \delta_1, q_1, \{f_1\})$ 

$A_2 = (Q_2, \Sigma, \delta_2, q_2, \{f_2\})$ 

Presupunem că $Q_1 \cap Q_2 = \emptyset$ (putem redenumi stările lui Q_2 astfel încât $Q_1 \cap Q_2 = \emptyset$)

- Dacă $r = r_1 + r_2$, construim $A_3 = (Q_3, \Sigma, \delta_3, q_3, \{f_3\})$, unde $Q_3 = Q_1 \cup Q_2 \cup \{q_3, f_3\}$, δ_3 dată de:



$\delta_3(q_3, \lambda) = \{q_1, q_2\}$, $\delta_3(f_1, \lambda) = \{f_3\}$, $\delta_3(f_2, \lambda) = \{f_3\}$

$\delta_3(q, a) = \delta_1(q, a)$, $\forall\, q \in Q_1$, $\forall\, a \in \Sigma \cup \{\lambda\}$

$\delta_3(p, b) = \delta(p, b)$, $\forall\, p \in Q_2$, $\forall\, b \in \Sigma \cup \{\lambda\}$

$w \in L(A_3) \Leftrightarrow (q_3, w) \vdash^*_{A_3} (f_3, \lambda) \Leftrightarrow$

$(q_3, w) \vdash_{A_3} (q_1, w) \vdash^*_{A_3} (f_1, \lambda) \vdash_{A_3} (f_3, \lambda)$ sau

$(q_3, w) \vdash_{A_3} (q_2, w) \vdash^*_{A_3} (f_2, \lambda) \vdash_{A_3} (f_3, \lambda) \Leftrightarrow$

$(q_1, w) \vdash^*_{A_1} (f_1, \lambda)$ sau $(q_2, w) \vdash^*_{A_2} (f_2, \lambda) \Leftrightarrow$

$w \in L(A_1) \cup L(A_2) \Leftrightarrow$

$\Leftrightarrow w \in L(A_1)$ sau $w \in L(A_2) \Leftrightarrow$

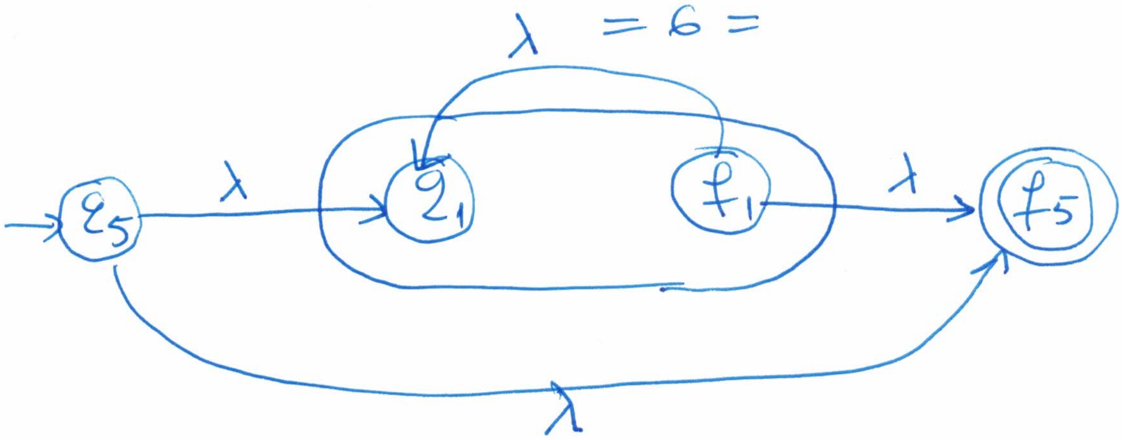
(ipoteza de inducție) $w \in L(r_1) \cup L(r_2) \Leftrightarrow w \in L(r_1 + r_2)$

- $r = r_1 \cdot r_2$. În acest caz construim AFNλ $A_4 = (Q_4, \Sigma, \delta_1, q_1, \{f_2\})$, unde $Q_4 = Q_1 \cup Q_2$, iar δ_4 descrisă de:

$w \in L(A_4) \Leftrightarrow (q_1, w) \vdash^*_{A_4} (f_2, \lambda) \Leftrightarrow \exists\, w_1, w_2 \in \Sigma^*$, $w = w_1 w_2$ astfel ca: $(q_1, w_1 w_2) \vdash^*_{A_1} (f_1, w_2) \vdash_{A_4} (q_2, w_2) \vdash^*_{A_2} (f_2, \lambda) \Leftrightarrow \exists\, w_1 \in L(A_1), w_2 \in L(A_2)$, $w = w_1 w_2 \Leftrightarrow w \in L(A_1) \cdot L(A_2) \Leftrightarrow w \in L(r_1) L(r_2) \Leftrightarrow$

$\Leftrightarrow w \in L(r_1 \cdot r_2)$

- $r = r_1^*$. Construim $A_5 = (Q_5, \Sigma, \delta_5, q_5, \{f_5\})$, unde $Q_5 = Q_1 \cup \{q_5, f_5\}$, iar δ_5 descrisă de:



$w \in L(A_5) \Leftrightarrow (q_5, w) \vdash_{A_5} (f_5, \lambda)$, $w = \lambda$, sau

$\exists\, w_1, \dots, w_n \in \Sigma^*$, $n \geq 1$, $w = w_1 \dots w_n$ pentru care

$(q_5, w) = (q_5, w_1 \dots w_n) \vdash_{A_5} (q_1, w_1 \dots w_n) \vdash^*_{A_1} (f_1, w_2 \dots w_n)$

$\vdash_{A_5} (q_1, w_2 \dots w_n) \vdash^*_{A_1} \dots \vdash_{A_5} (q_1, w_n) \vdash^*_{A_1} (f_1, \lambda) \vdash_{A_5} (f_5, \lambda)$

$\Leftrightarrow$ $w = \lambda$ sau $\exists\, w_1, \dots, w_n \in \Sigma^*$, $n \geq 1$, $w = w_1 \dots w_n$,

$w_i \in L(A_1)$, $i = 1, \dots, n \Leftrightarrow w \in L(A_1)^* = (L(r_1))^*$

Cu aceasta am demonstrat Lema 1

<u>Lema 2</u> Pentru orice AFD A există o expresie regulată care descrie limbajul recunoscut de A.

<u>Dem</u> Fie AFD $A = (Q, \Sigma, \delta, q_1, F)$, $Q = \{q_1, q_2, \dots, q_n\}$.

Notăm cu $R^k_{ij} = \{ w \in \Sigma^* \mid \delta(q_i, w) = q_j$, stările intermediare prin care trece A atunci când este în starea q_i, primește la intrare w, iar la final ajunge în q_j, au, toate, indicele $\leq k \}$

<u>Obs</u>. Automatul determinist A este posibil să nu fie complet, adică $\exists\, p \in Q, \exists\, a \in \Sigma$, $\delta(p, a) = \emptyset$

Cu alte cuvinte:

$R_{ij}^{k} = \{x \in \Sigma^* \mid \delta(q_i, x) = q_j$, $\forall y$ prefix propriu al lui x, $y \neq \lambda$ și $y \neq x$, astfel încât $\delta(q_i, y) = q_l$, atunci $l \leq k\}$

Deoarece numărul stărilor lui A este n, atunci R_{ij}^{n} conține toate șirurile fac ca automatul A să treacă din starea q_i în starea q_j.

Definim inductiv, după k, R_{ij}^{k}:

$$R_{ij}^{0} = \begin{cases} \{a \in \Sigma \mid \delta(q_i, a) = q_j\}, & \text{dacă } i \neq j \\ \{a \in \Sigma \mid \delta(q_i, a) = q_i\} \cup \{\lambda\}, & \text{dacă } i = j \end{cases}$$

$$R_{ij}^{k} = R_{ij}^{k-1} \cup R_{ik}^{k-1} (R_{kk}^{k-1})^* R_{kj}^{k-1}, \quad k \geq 1$$

Aceasta înseamnă că:

- fie R_{ij}^{k} conține doar șiruri care fac ca automatul A să treacă din starea q_i în starea q_j trecând doar prin stări intermediare de indice $\leq k-1$ (R_{ij}^{k-1})
- fie R_{ij}^{k} conține șiruri de forma

$w = u w_1 w_2 \dots w_p v$, $p \geq 1$,

$\delta(q_i, u) = q_k$, $u \in R_{ik}^{k-1}$

$\delta(q_k, w_1) = q_k, \dots, \delta(q_k, w_p) = q_k$, $w_1, \dots, w_p \in R_{kk}^{k-1}$

$\delta(q_k, v) = q_j$, $v \in R_{kj}^{k-1}$

Arătăm că pentru orice i, j, k, R_{ij}^{k} este descrisă de o expresie regulată r_{ij}^{k}.

Pentru i, j fixați, $1 \le i, j \le n$, facem inducție după k, $0 \le k \le n$.

$\boxed{k=0}$ R_{ij}^{0} este un limbaj finit format din $a \in \Sigma$ pentru care $\delta(q_i, a) = q_j$, la care se adaugă λ dacă $i = j$. Atunci putem scrie:

$r_{ij}^{0} = a_1 + a_2 + \dots + a_m$, $i \ne j$, $\Sigma = \{a_1, \dots, a_m\}$, sau

$r_{ij}^{0} = a_1 + a_2 + \dots + a_m + \lambda$, dacă $i = j$.

$\boxed{k-1 \to k}$ Formula recursivă R_{ij}^{k} implică
$k \ge 1$
doar operatorii: reuniune $(+)$, concatenare $(\cdot)$ și iterație (închidere) Kleene $(*)$. Din ipoteza de inducție, pentru fiecare l, m, $1 \le l, m \le n$ există expresia regulată r_{lm}^{k-1} astfel încât $L(r_{lm}^{k-1}) = R_{lm}^{k-1}$. Atunci putem scrie

$$r_{ij}^{k} = r_{ij}^{k-1} + r_{ik}^{k-1} (r_{kk}^{k-1})^{*} r_{kj}^{k-1}$$

ceea ce completează inducția.

În final, observăm că

$$L(A) = \bigcup_{q_j \in F} R_{1j}^{n}$$, deci putem scrie:

$r = r^{n}_{1j_1} + r^{n}_{1j_2} + \dots r^{n}_{1j_p}$, unde $F = \{q_{j_1}, \dots, q_{j_p}\}$.

Cu aceasta am încheiat demonstrația lemei 2 și totodată a Teoremei Kleene.

Exemple

1) Se dă expresia $r = 01^* + 1$. Să se transforme această expresie în AFNλ.

Avem descompunerile:

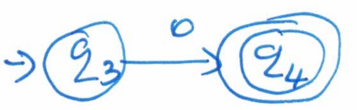
$r = r_1 + r_2$, $r_1 = 01^*$, $r_2 = 1$;

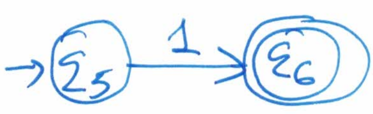
$r_1 = r_3 r_4$, $r_3 = 0$, $r_4 = 1^*$;

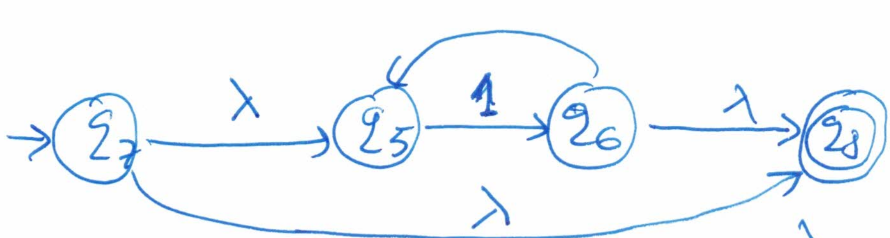
$r_4 = r_5^*$, $r_5 = 1$;

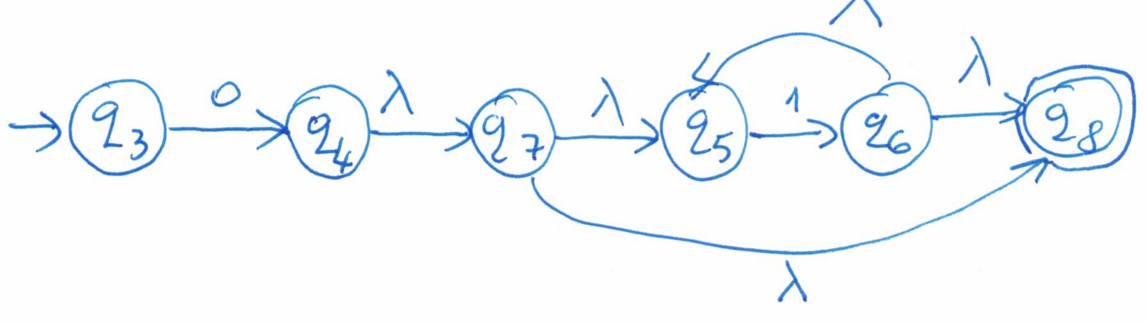
$r = r_3 r_5^* + r_2$

r_2

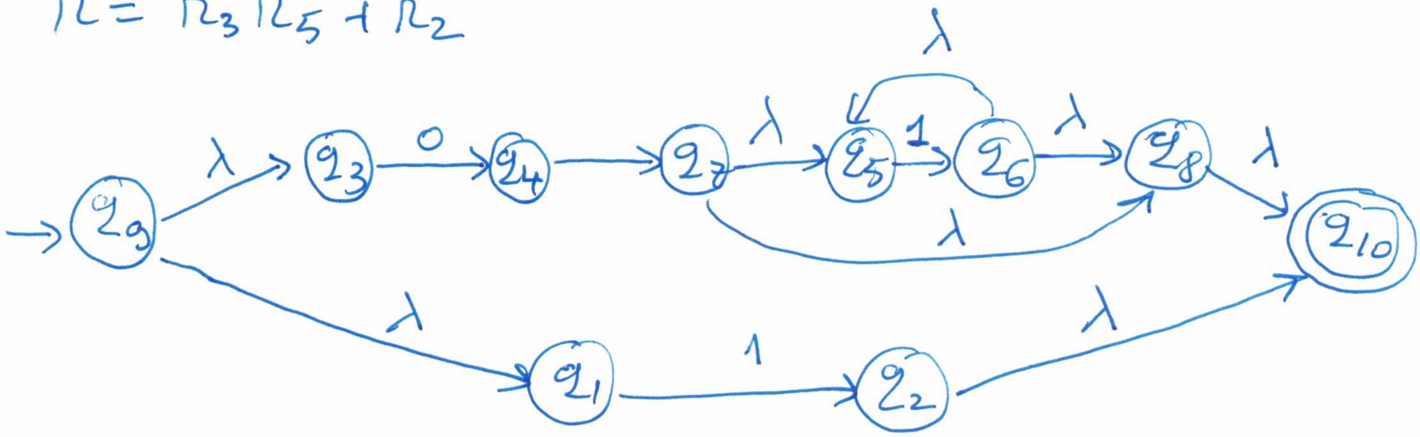

r_3


r_5


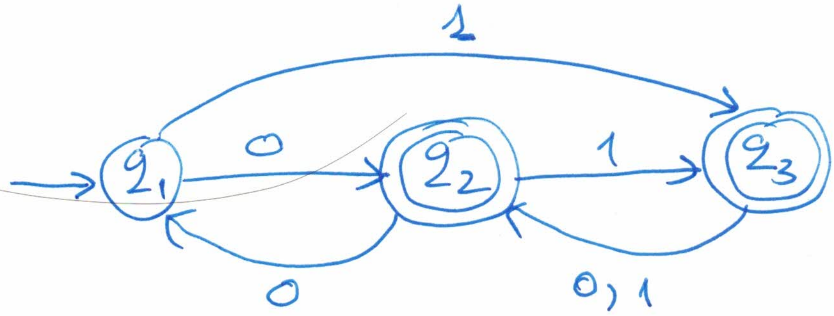
$r_4 = r_5^*$


$r_3 r_5^*$


$r = r_3 r_5^* + r_2$



2) Fie automatul următor, A:



În cele ce urmează vom ține cont de următoarele relații valide pentru expresiile regulate:

$(r+s)t = rt + st$

$(\lambda + r)^* = r^*$

$(\lambda)^* = \lambda$

$t(r+s) = tr + ts$

	$k=0$	$k=1$	$k=2$
r_{11}^k	λ	λ	$(00)^*$
r_{12}^k	0	0	$0(00)^*$
r_{13}^k	1	1	0^*1
r_{21}^k	0	0	$0(00)^*$
r_{22}^k	λ	$\lambda+00$	$(00)^*$
r_{23}^k	1	$1+01$	0^*1
r_{31}^k	ϕ	ϕ	$(0+1)(00)^*0$
r_{32}^k	$0+1$	$0+1$	$(0+1)(00)^*$
r_{33}^k	λ	λ	$\lambda+(0+1)0^*1$

$$L(A) = r_{12}^{3} + r_{13}^{3}$$

$$r_{12}^{3} = r_{12}^{2} + r_{13}^{2}\,(r_{33}^{2})^{*}\,r_{32}^{2} =$$

$$= 0(00)^{*} + 0^{*}1\,(\lambda + (0+1)^{*}0^{*}1)^{*}(0+1)\,(00)^{*}$$

$$= 0(00)^{*} + 0^{*}1\,((0+1)^{*}0^{*}1)^{*}(0+1)\,(00)^{*}$$

$$r_{13}^{3} = r_{13}^{2} + r_{13}^{2}\,(r_{33}^{2})^{*}\,r_{33}^{2} =$$

$$= 0^{*}1 + 0^{*}1\,(\lambda + (0+1)^{*}0^{*}1)^{*}(\lambda + (0+1)0^{*}1)$$

$$= 0^{*}1\,((0+1)0^{*}1)^{*}$$