~ Seminar 3 ~

➢ Închiderea limbajelor regulate la operații (complement; intersecție, diferență, reuniune) [**în seminar 4**: *reuniune, concatenare, stelare, plus*]

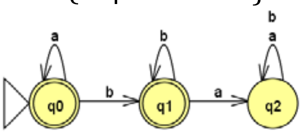
Obs: Dacă limbajul regulat L este acceptat de un AFD complet definit (fără tranziții lipsă) $AFD(L) = (Q, \Sigma, \delta, q_0, \boldsymbol{F})$, atunci putem construi un AFD care să accepte complementul lui L ($\Sigma^* \backslash L$) prin *interschimbarea stărilor finale cu cele nefinale* $AFD(\overline{L}) = (Q, \Sigma, \delta, q_0, \boldsymbol{Q \backslash F})$.

Obs: Dacă limbajele regulate L1 și L2 sunt acceptate de 2 automate AFD complet definite $AFD(L_1) = (Q_1 = \{q_0, q_1, \dots\}, \Sigma, \delta_1, \boldsymbol{q_0}, F_1)$ și $AFD(L_2) = (Q_2 = \{r_0, r_1, \dots\}, \Sigma, \delta_2, \boldsymbol{r_0}, F_2)$, atunci putem construi un AFD cu stări obținute prin *produs cartezian între mulțimile de stări* ale celor 2 automate: $AFD(L) = (Q, \Sigma, \delta, (\boldsymbol{q_0}, \boldsymbol{r_0}), F)$ având
- **stările** $Q = Q_1 \times Q_2 = \{(\boldsymbol{q_i}, \boldsymbol{r_j}) \mid q_i \in Q_1 \text{ și } r_j \in Q_2\}$,
- **tranzițiile** $\delta\left((q_i, r_j), x\right) = \left(\boldsymbol{\delta_1}(q_i, x), \boldsymbol{\delta_2}(r_j, x)\right), \forall (q_i, r_j) \in Q, \forall x \in \Sigma$,
- **starea inițială** $(\boldsymbol{q_0}, \boldsymbol{r_0})$ (perechea formată din cele două stări inițiale),
- **stările finale** ***F*** depind dacă automatul acceptă limbajul:

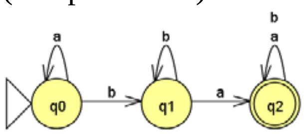
✓ *(intersecție)* $L = L_1 \cap L_2 = \{w \in \Sigma^* \mid w \in L_1 \text{ și } w \in L_2\} \Rightarrow F = F_1 \times F_2$
✓ *(diferență)* $L = L_1 \backslash L_2 = \{w \in \Sigma^* \mid w \in L_1 \text{ și } w \notin L_2\} \Rightarrow F = F_1 \times (Q_2 \backslash F_2)$
sau $L = L_2 \backslash L_1 = \{w \in \Sigma^* \mid w \notin L_1 \text{ și } w \in L_2\} \Rightarrow F = (Q_1 \backslash F_1) \times F_2$
✓ *(reuniune)* $L = L_1 \cup L_2 = \{w \in \Sigma^* \mid w \in L_1 \text{ sau } w \in L_2\} \Rightarrow F = (F_1 \times Q_2) \cup (Q_1 \times F_2)$

- ***Exemplu:***

$L = \{w \mid w \in a^* b^*\}$ => (complementul) $\overline{L} = \Sigma^* \backslash L = \{a, b\}^* \setminus \{a^* b^*\}$

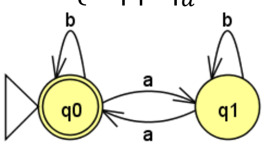


=>

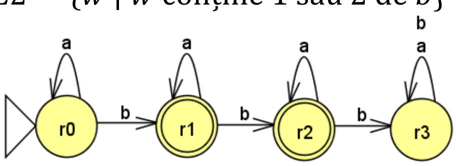


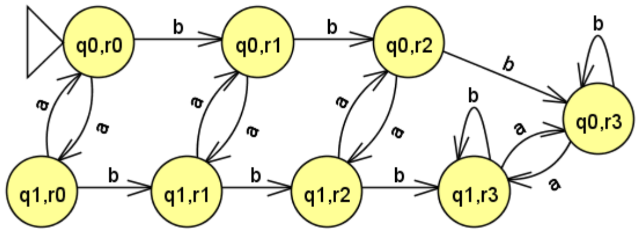
- ***Exemplu:***

$L1 = \{w \mid |w|_a \text{ este par}\}$



$L2 = \{w \mid w \text{ conține 1 sau 2 de } b\}$





Stările finale vor fi:
- pt. intersecție $L_1 \cap L_2 \Rightarrow F = \{q_0 r_1,\ q_0 r_2\}$
- pt. diferență $L_1 \backslash L_2 \Rightarrow F = \{q_0 r_0,\ q_0 r_3\}$
- pt. diferență $L_2 \backslash L_1 \Rightarrow F = \{q_1 r_1,\ q_1 r_2\}$
- pt. reuniune $L_1 \cup L_2 \Rightarrow$
$F = \{q_0 r_0,\ q_0 r_1,\ q_0 r_2,\ q_0 r_3,\ q_1 r_1,\ q_1 r_2\}$

➢ ***Lema de pompare** pentru limbaje regulate (REG)* *[vezi curs 5, pag 19 – 21]*

Fie L un limbaj regulat. Atunci $\exists p \in \mathbb{N}$ (număr natural) astfel încât pentru $\forall \alpha \in L$ cuvânt, cu $|\alpha| \geq p$, există o descompunere $\alpha = u \cdot v \cdot w$ cu proprietățile:

(1) $|u \cdot v| \leq p$

(2) $|v| \geq 1$

(3) $u \cdot v^i \cdot w \in L, \ \forall i \geq 0.$

- Vrem să demonstrăm că un limbaj NU este regulat.

Presupunem prin reducere la absurd că "limbajul este regulat" (predicatul P) și atunci rezultă că "afirmația din lemă este adevărată" (predicatul Q).

Obs: Știm de la logică faptul că $(P \to Q) \equiv (\neg Q \ \to \ \neg P)$. Așa că vom nega afirmația lemei $(\neg Q)$ și va rezulta că limbajul nu este regulat $(\neg P)$. Practic negarea constă în interschimbarea cuantificatorilor logici (∃ și ∀) între ei, iar la condiția (3) ∈ devine ∉.

→ ***Schema demonstrației***

- Vrem să demonstrăm că L **nu este** limbaj regulat (adică nu se poate construi niciun automat finit care să-l recunoască pe L), folosind lema de pompare ***negată***.

- Presupunem prin reducere la absurd că L **este** limbaj regulat. Atunci $\exists p \in \mathbb{N}$ (unde $p = |Q|$ numărul de stări ale unui automat finit care-l recunoaște pe L) și putem aplica lema de pompare. *(În continuare negăm afirmația lemei.)*

- Alegem (adică ∃) un cuvânt $\boldsymbol{\alpha}$ din limbajul L, care să respecte ipoteza lemei de pompare, adică să aibă lungimea cel puțin p, deci $|\alpha| \geq p, \forall p \in \mathbb{N}$.

- Pentru fiecare (adică ∀) posibilă descompunere a cuvântului $\alpha = u \cdot v \cdot w$ care respectă condițiile (1) și (2) din lemă ($|u \cdot v| \leq p$ și $|v| \geq 1$):

- **alegem** (adică ∃) convenabil câte un număr natural $\boldsymbol{i} \geq 0$ pentru care să obținem o contradicție a condiției (3), adică să rezulte că cuvântul $\beta = u \cdot v^i \cdot w \notin L$ și deci presupunerea făcută este falsă.

Observație: Demonstrația este completă și corectă doar dacă se obține contradicție ($\beta \notin L$) pentru toate descompunerile posibile ale cuvântului α.

- ***Exemplu:*** $\boldsymbol{L_0 = \{a^n b^n \mid n \geq 0\} \notin REG}$

Idee: Vrem să demonstrăm că L_0 **nu este limbaj regulat**. Observăm că L_0 conține cuvinte formate din *n* litere de "a" urmate tot de *n* litere de "b". În demonstrație trebuie să obținem un cuvânt β care să **nu** respecte această proprietate (adică să fie de forma $a^* b^*$, dar să aibă număr *diferit* de a-uri și b-uri).

Obs: Dacă aveam condiția $n \geq x$ (cu $x \in \mathbb{N}$ o constantă), atunci ar fi trebuit să alegem cuvântul $\alpha = a^{p+x}b^{p+x} \in L_0$ (pentru a fi sigur un cuvânt din limbaj $\forall p \in \mathbb{N}$).

Demonstrație: Presupunem prin reducere la absurd că L_0 este limbaj regulat. Atunci $\exists p \in \mathbb{N}$ și putem aplica lema de pompare. *(În continuare negăm afirmația lemei.)*

Alegem cuvântul $\boldsymbol{\alpha = a^p b^p \in L_0}$, cu $|\alpha| = 2p \geq p, \forall p \in \mathbb{N}$ *(deci lungimea cuvântului respectă ipoteza lemei)*. Conform lemei, cuvântul poate fi scris sub forma $\alpha = u \cdot v \cdot w$.

Din condiția (1) avem $|u \cdot v| \leq p$. Rezultă că cuvântul $u \cdot v$ conține doar litere de ”a” (pentru că $u \cdot v$ este un prefix al primelor p caractere din α).

Atunci notăm $\boldsymbol{v = a^k}$. Din condițiile (1) și (2) avem $1 \leq |v| \leq p$ *(pentru că se poate ca $u = \lambda$, adică $|u| = 0$)*. Deci $1 \leq |a^k| \leq p$, adică $1 \leq k \leq p$ (*).

De asemenea, condiția (3) spune că $u \cdot v^i \cdot w \in L_0,\ \forall i \geq 0$. Alegem $\boldsymbol{i = 2}$ și avem cuvântul $\boldsymbol{\beta} = u \cdot v^2 \cdot w = \boldsymbol{a^{p+k} b^p \notin L_0}$ (pentru că în relația (*) avem $k \geq 1$, deci numărul de ”a”-uri este strict mai mare decât numărul de ”b-uri” din cuvântul β). Am obținut o contradicție pentru (3), deci presupunerea făcută este falsă și L_0 nu este limbaj regulat. ■

- ***Exemplu:*** $\boldsymbol{L_1 = \{a^m b^n \mid m > n \geq 0\} \notin REG}$

Idee: Vrem să demonstrăm că L_1 **nu este limbaj regulat**. Observăm că L_1 conține cuvinte formate din litere de ”a” urmate de strict mai puține litere de ”b”. În demonstrație trebuie să obținem un cuvânt β care să **nu** respecte această proprietate (adică să fie de forma a^*b^*, dar să aibă $|\beta|_a \leq |\beta|_b$).

Obs: Ca să putem ajunge la contradicție, trebuie să alegem cuvântul α care respectă la limită inegalitatea dintre a-uri și b-uri (adică *m* exact cu o unitate mai mare decât *n*). De asemenea, b-urile trebuie să existe (deci alegem $n \neq 0$).

Demonstrație: Presupunem prin reducere la absurd că L_1 este limbaj regulat. Atunci $\exists p \in \mathbb{N}$ și putem aplica lema de pompare. *(În continuare negăm afirmația lemei.)*

Alegem cuvântul $\boldsymbol{\alpha = a^{p+1} b^p \in L_1}$, cu $|\alpha| = 2p + 1 \geq p, \forall p \in \mathbb{N}$ *(deci lungimea cuvântului respectă ipoteza lemei)*. Conform lemei, cuvântul poate fi scris sub forma $\alpha = u \cdot v \cdot w$.

Din condiția (1) avem $|u \cdot v| \leq p$. Rezultă că cuvântul $u \cdot v$ conține doar litere de ”a” (pentru că $u \cdot v$ este un prefix al primelor p caractere din α).

Atunci notăm $\boldsymbol{v = a^k}$. Din condițiile (1) și (2) avem $1 \leq |v| \leq p$ *(pentru că se poate ca $u = \lambda$, adică $|u| = 0$)*. Deci $1 \leq |a^k| \leq p$, adică $1 \leq k \leq p$ (*).

De asemenea, condiția (3) spune că $u \cdot v^i \cdot w \in L_1,\ \forall i \geq 0$. Alegem $\boldsymbol{i = 0}$ și avem cuvântul $\boldsymbol{\beta} = u \cdot v^0 \cdot w = u \cdot w = \boldsymbol{a^{p+1-k} b^p}$. Conform (*) avem $1 \leq k \leq p$ (înmulțim cu -1) <=> $-p \leq -k \leq -1$ (adunăm p+1) <=> $1 \leq p + 1 - k \leq p$. Avem $|\beta|_a \leq |\beta|_b$ => $\boldsymbol{\beta \notin L_1}$**.** Am obținut o contradicție pentru (3), deci presupunerea făcută este falsă și L_1 nu este limbaj regulat. ■

- ***Exemplu:*** $\boldsymbol{L2} = \{\boldsymbol{a^m b^{3n} c^{n+3}} \mid \boldsymbol{m} \geq \mathbf{5}, \boldsymbol{n} \geq \mathbf{1}\} \notin \boldsymbol{REG}$

Obs: Limbajul L_2 nu este regulat pentru că există o corelație între numărul de apariții ale simbolurilor *b* și *c* (ambele depind de valoare lui *n*). De aceea, când alegem cuvântul α aparițiile *b*-urilor și *c*-urilor vor depinde de numărul *p* din lemă, iar pentru *m* vom alege valoarea care <u>respectă la limită inegalitatea</u> $m \geq 5$.

<u>Demonstrație:</u> Presupunem prin reducere la absurd că L_2 este limbaj regulat. Atunci $\exists p \in \mathbb{N}$ și putem aplica lema de pompare. *(În continuare negăm afirmația lemei.)*

Alegem cuvântul $\boldsymbol{\alpha} = \boldsymbol{a^5 b^{3p} c^{p+3}} \in \boldsymbol{L_2}$, cu $|\alpha| = 4p + 8 \geq p, \forall p \in \mathbb{N}$ *(deci lungimea cuvântului respectă ipoteza lemei)*. Conform lemei, cuvântul poate fi scris sub forma $\alpha = u \cdot v \cdot w$.
Din condiția (1) avem $|u \cdot v| \leq p$. Rezultă că cuvântul $u \cdot v$ poate conține doar litere de ”a” și/sau ”b” (pentru că $u \cdot v$ este un prefix al primelor *p* caractere din α).

- Caz I: Fie $\boldsymbol{v} = \boldsymbol{a^k}$, $1 \leq k \leq 5$. Din condițiile (1) și (2) avem $1 \leq |v| \leq p$. Deci $1 \leq |a^k| \leq p$, adică $1 \leq k \leq p$ (*).
De asemenea, condiția (3) spune că $u \cdot v^i \cdot w \in L_2,\ \forall i \geq 0$. Alegem $\boldsymbol{i} = \mathbf{0}$ și avem cuvântul $\boldsymbol{\beta} = u \cdot v^0 \cdot w = u \cdot w = \boldsymbol{a^{5-k} b^{3p} c^{p+3}} \in L_2 <=> |\beta|_a \geq 5 <=> 5 - k \geq 5 <=> k \leq 0$, contradicție cu (*). **[1]**

- Caz II: Fie $\boldsymbol{v} = \boldsymbol{a^k b^t}$, $0 \leq k \leq 5$ și $\mathbf{1} \leq \boldsymbol{t} \leq p - 5$. Din condițiile (1) și (2) avem $1 \leq |v| \leq p$. Deci $1 \leq |a^k b^t| \leq p$, adică $1 \leq k + t \leq p$ (**).
De asemenea, condiția (3) spune că $u \cdot v^i \cdot w \in L_2,\ \forall i \geq 0$. Alegem $\boldsymbol{i} = \mathbf{0}$ și avem cuvântul $\boldsymbol{\beta} = u \cdot v^0 \cdot w = u \cdot w = \boldsymbol{a^{5-k} b^{3p-t} c^{p+3}} \in L_2 <=> |\beta|_b = 3 * (|\beta|_c - 3)$
$<=> 3p - t = 3 * (p + 3 - 3) <=> t = 0$, contradicție cu alegerea lui $t \geq 1$ și cu (**). **[2]**

Din relațiile [1] și [2] (contradicțiile obținute pentru fiecare descompunere posibilă a lui α) rezultă că presupunerea făcută este falsă și L_2 nu este limbaj regulat. ■

- ***Exemplu:*** $\boldsymbol{L3} = \left\{\boldsymbol{a^{n^2}} \mid \boldsymbol{n} \geq \mathbf{1}\right\} = \{a^1, a^4, a^9, a^{16}, a^{25}, a^{36}, \dots\} \notin REG$

Obs: Proprietatea cuvintelor din limbajul L_3 este aceea că sunt formate doar din litere de ”a” și lungimea lor este un număr pătrat perfect. Deci pentru a obține contradicția ($\beta \notin L_3$) trebuie să arătăm că lungimea lui β **nu** poate fi un pătrat perfect.

<u>Demonstrație:</u> Presupunem prin reducere la absurd că L_3 este limbaj regulat. Atunci $\exists p \in \mathbb{N}$ și putem aplica lema de pompare. *(În continuare negăm afirmația lemei.)*

Alegem cuvântul $\boldsymbol{\alpha} = \boldsymbol{a^{p^2}} \in \boldsymbol{L_3}$, cu $|\alpha| = p^2 \geq p, \forall p \in \mathbb{N}$ *(deci lungimea cuvântului respectă ipoteza lemei)*. Conform lemei, cuvântul poate fi scris sub forma $\alpha = u \cdot v \cdot w$.

Observăm că toate cuvintele din L_3 sunt formate doar din litere de ”a”. Atunci notăm $\boldsymbol{v} = \boldsymbol{a^k}$.
Din condițiile (1) și (2) ale lemei avem $1 \leq |v| \leq p$. Deci $1 \leq |a^k| \leq p$, adică $1 \leq k \leq p$ (*).

De asemenea, condiția (3) din lema spune că $u \cdot v^i \cdot w \in L_3,\ \forall i \geq 0$.
Alegem $\boldsymbol{i} = \mathbf{2}$ și avem cuvântul $\boldsymbol{\beta} = u \cdot v^2 \cdot w$ de lungime
$|\boldsymbol{\beta}| = |u \cdot v^2 \cdot w| = |u \cdot v \cdot w| + |v| = |\alpha| + |v| = p^2 + |v| = \boldsymbol{p^2} + \boldsymbol{k}$.

Conform (*), avem $1 \leq k \leq p$ (adunăm peste tot p^2) $<=> p^2 + 1 \leq p^2 + k \leq p^2 + p$.

Dar $p^2 < p^2 + 1$ și $p^2 + p < (p+1)^2$. Rezultă că $p^2 < p^2 + k < (p+1)^2$, adică $p^2 < |\beta| < (p+1)^2$. Deci $|\beta|$ nu poate fi pătrat perfect (pentru că este inclus strict între două pătrate perfecte consecutive) => $\beta \notin L_3$, contradicție cu condiția (3) din lema), deci presupunerea făcută este falsă și L_3 nu este limbaj regulat. ■

- ***Exemple:***

$L4 = \{x \cdot x^R \mid x \in \{a, b\}^*\} \notin REG$ *(discutat alegerea cuvântului α)*
$L5 = \{x \cdot x \mid x \in \{a, b\}^*\} \notin REG$ *(discutat alegerea cuvântului α)*

Obs: (R = reversed) Cuvântul $\boldsymbol{w^R}$ este oglindirea cuvântului $\boldsymbol{w}$, de exemplu $(abb)^R = bba$.

→ Pentru L4:

-- Cuvântul $\boldsymbol{\alpha = a^p a^p} = a^{2p} = (aa)^p$ NU este o alegere bună (pentru acest α putem desena un AFD având un circuit de 2 de a).

- Dacă $\boldsymbol{v = a^{2k}}$ => $|\beta| = |u \cdot v^i \cdot w| = |u \cdot v \cdot w| + |v|^{(i-1)} =$
$= |a^{2p+(2k)*(i-1)}| = 2 * (p + k * (i-1))$ este **par** (adică $\beta \in L_4$), $\forall \boldsymbol{i \geq 0}$ (NU avem contradicție cu condiția 3 din lemă).
- Dacă $\boldsymbol{v = a^{2k+1}}$ => $|\beta| = |a^{2p+(2k+1)*(i-1)}| = 2 * \left(p + k * (i-1)\right) + (i-1)$ este **impar**, adică $\beta \notin L_4$ (contradicție cu condiția 3 din lemă) <=> i este par (de exemplu alegem $i = 0$).

Concluzie: Dacă există descompuneri ale lui α (forme ale lui v) pentru care NU putem obține contradicție, înseamnă că demonstrația nu este corectă și trebuie ales un alt α.

-- Cuvântul $\boldsymbol{\alpha = a^p bba^p}$ este o alegere bună (pentru acest α nu putem desena un AFD).

- Dacă $\boldsymbol{v = a^k}$ alegem $i = 2$ => $\beta = a^{p+k}bba^p \notin L_4$ pentru că $1 \leq k \leq p$ (din primele două condiții din lemă).

Concluzie: Avem un singur caz de descompunere a lui α (o singură formă a lui v), am obținut contradicție, deci demonstrația este corectă.

→ Pentru L5:

-- Cuvântul $\boldsymbol{\alpha = a^p ba^p b}$ este o alegere bună (pentru acest α nu putem desena un AFD). Avem un singur caz de descompunere a lui α.

- Dacă $\boldsymbol{v = a^k}$ alegem $i = 2$ => $\beta = a^{p+k}ba^p b \notin L_5$ pentru că $1 \leq k \leq p$ (din primele două condiții din lemă). Deci avem contradicție și demonstrația este corectă.

-- Cuvântul $\boldsymbol{\alpha = ba^p ba^p}$ este tot o alegere bună (pentru acest α nu putem desena un AFD), dar avem mai multe cazuri de descompunere a lui α.

- Dacă $\boldsymbol{v = ba^k}$ (cu $0 \leq k \leq p-1$) alegem $i = 0$ => $\beta = a^{p-k}ba^p \notin L_5$.
- Dacă $\boldsymbol{v = a^k}$ (cu $1 \leq k \leq p$) alegem $i = 2$ => $\beta = ba^{p+k}ba^p \notin L_5$.

Concluzie: Avem contradicție pe fiecare caz posibil de descompunere, deci demonstrația este corectă.