

Curs – sapt.3

- Gramatici. Clasificarea Chomsky
 - Gramatici regulare
- Limbaje regulare. Echivalente
 - Expresii regulare
- Proprietati ale limbajelor regulare

Gramatica

O gramatica este un cvadruplu $G = (N, \Sigma, P, S)$

- N este un alfabet de simboluri *neterminale*
- Σ este un alfabet de simboluri *terminale*
- $N \cap \Sigma = \emptyset$
- $P \subseteq (N \cup \Sigma)^* N (N \cup \Sigma)^* \times (N \cup \Sigma)^*$
 P multime finită (multimea regulilor de productie)
- $S \in N$ (simbolul de start - simbolul initial)

Notatie:

$(\alpha, \beta) \in P$ se noteaza: $\alpha \rightarrow \beta$

(α se înlocuieste cu β)

Notatii

- la nivel abstract (exemple matematice, specificari)
 - Σ : a,b,... litere mici de la inceputul alfabetului
 - N : A,B,.. litere mari de la inceputul alfabetului
 - Σ sau N : X,Y,...litere mari de la sfarsitul alfabetului
 - Σ^* : x,y,... litere mici de la sfarsitul alfabetului
 - $(\Sigma \cup N)^*$: α, β, \dots litere grecesti
- nu se folosesc spatii cand avem nevoie de mai multe caractere pentru a specifica un simbol (terminal sau neterminal)

Relatii de derivare

relatii binare peste $(\Sigma \cup N)^*$ adica $(\Sigma \cup N)^* \times (\Sigma \cup N)^*$

- derivare directa

$$\gamma \Rightarrow \delta \iff \exists \gamma_1, \gamma_2, \alpha, \beta \in (N \cup \Sigma)^*$$

$$\text{a.i. } \gamma = \gamma_1 \alpha \gamma_2, \delta = \gamma_1 \beta \gamma_2, \text{ iar } (\alpha \rightarrow \beta) \in P$$

- k -derivare $\stackrel{k}{=}$ \Rightarrow
(o succesiune de k derivări directe)

- $+$ derivare $\stackrel{+}{=}$ \Rightarrow
dacă $\exists k > 0$ a.i. cele 2 secvente să fie într-o relatie de " k derivare"

- $*$ derivare $\stackrel{*}{=}$ \Rightarrow

dacă fie cele 2 secvente sunt egale, fie între ele exista o relatie de $+$ derivare

Limбай generat de o gramatica

- Limбай generat gramatica $G=(N,\Sigma,P,S)$

$$L(G) = \{w \in \Sigma^* \mid S \xRightarrow{*} w\}$$

- Forma propozitionala

$$- \alpha \in (N \cup \Sigma)^* \text{ a.i. } S \xRightarrow{*} \alpha$$

- Propozitie (cuvant)

– un element din $L(G)$

- Gramatici echivalente

daca genereaza acelasi limбай

Tipuri de gramatici

- Gramatica monotona

- $\forall \alpha \rightarrow \beta \in P: |\alpha| \leq |\beta| \quad \alpha, \beta \in (N \cup \Sigma)^*$
- caz special: $S \rightarrow \epsilon$ poate sa apartina lui P. In acest caz S nu apare în membrul drept al nici unei reguli de productie.

- Gramatica dependenta de context

reguli de productie sunt de forma:

$$\alpha A \beta \rightarrow \alpha \gamma \beta$$

$$A \in N$$

$$\alpha, \beta, \gamma \in (N \cup \Sigma)^*, \gamma \neq \epsilon$$

- caz special: $S \rightarrow \epsilon$ poate sa apartina lui P. In acest caz S nu apare în membrul drept al nici unei reguli de productie.

Transformarea gramaticilor monotone in gramatici dependente de context

Ideea:

Pentru fiecare regula de productie de forma: $X_1 \dots X_m \rightarrow Y_1 \dots Y_n$,(stim ca $m \leq n$,)
introducem:

$$\begin{array}{lcl}
 X_1 X_2 \dots X_{m-1} X_m & \rightarrow & Z_1 X_2 \dots X_{m-1} X_m \\
 Z_1 X_2 \dots X_{m-1} X_m & \rightarrow & Z_1 Z_2 \dots X_{m-1} X_m \\
 & \dots & \\
 Z_1 Z_2 \dots X_{m-1} X_m & \rightarrow & Z_1 Z_2 \dots Z_{m-1} X_m \\
 Z_1 Z_2 \dots Z_{m-1} X_m & \rightarrow & Z_1 Z_2 \dots Z_{m-1} Z_m Y_{m+1} \dots Y_n \\
 & & \\
 Z_1 Z_2 \dots Z_{m-1} Z_m Y_{m+1} \dots Y_n & \rightarrow & Y_1 Z_2 \dots Z_{m-1} Z_m Y_{m+1} \dots Y_n \\
 Y_1 Z_2 \dots Z_{m-1} Z_m Y_{m+1} \dots Y_n & \rightarrow & Y_1 Y_2 \dots Z_{m-1} Z_m Y_{m+1} \dots Y_n \\
 & \dots & \\
 Y_1 Y_2 \dots Z_{m-1} Z_m Y_{m+1} \dots Y_n & \rightarrow & Y_1 Y_2 \dots Y_{m-1} Z_m Y_{m+1} \dots Y_n \\
 Y_1 Y_2 \dots Y_{m-1} Z_m Y_{m+1} \dots Y_n & \rightarrow & Y_1 Y_2 \dots Y_{m-1} Y_m Y_{m+1} \dots Y_n
 \end{array}$$

- unde Z_i – sunt neterminale nou introduse, distincte pentru fiecare regula de productie

Tipuri de gramatici

- Gramatica regulara:

reg. prod. sunt de forma

- $A \rightarrow aB$

- $A \rightarrow b$

unde $A, B \in N$ si $a, b \in \Sigma$

caz special: $S \rightarrow \epsilon$ poate $\in P$ In acest caz S nu apare în membrul drept al nici unei reguli de productie.

- Gramatica independenta de context:

reg. productie sunt de forma $A \rightarrow \alpha$, $A \in N$, $\alpha \in (N \cup \Sigma)^*$

Clasificarea Chomsky

- Gramatici de tip 0
nici o restricție (*suplimentară*) referitoare la forma regulilor de producție
- Gramaticile de tip 1
dependente de context \Leftrightarrow *gramatici monotone*
(*monotonic, non-contracting*)
- Gramaticile de tip 2
gramatici independente de context
 \Rightarrow Limbaje independente de context
- Gramaticile de tip 3
gramatici regulare
 \Rightarrow Limbaje regulare

Ierarhia Chomsky

Fie

~ 1959-1963

- \mathcal{L}_0 - multimea limbajelor generate de gram. de tip 0
- \mathcal{L}_1 - multimea limbajelor generate de gram. de tip 1
- \mathcal{L}_2 - multimea limbajelor generate de gram. de tip 2
- \mathcal{L}_3 - multimea limbajelor generate de gram. de tip 3

Are loc:

$$\mathcal{L}_0 \supset \mathcal{L}_1 \supset \mathcal{L}_2 \supset \mathcal{L}_3$$



Ierarhia Chomsky: observatii

Teorema:

Fiecare dintre familiile de limbaje:

$$\mathcal{L}_0, \mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3$$

este inchisa fata de operatia de reuniune

Limbaje regulate. Echivalente

- Limbaj regular
= limbaj generat de o gramatica regulara
- putere de exprimare
AF: $AFN \Leftrightarrow AFD$
 $AF \Leftrightarrow \text{gr.regulare}$
 $AF \Leftrightarrow (\text{m.regulare} \Leftrightarrow \text{expr.reg.})$

Multimi regulate

Fie Σ un alfabet.

Multimile regulate peste Σ se definesc recursiv astfel:

1. Φ este o m. reg. peste Σ
2. $\{\epsilon\}$
3. $\{a\}$ daca: $a \in \Sigma$
4. $RU S$ daca R, S – multimii regulate peste Σ +
5. RS daca R, S – multimii regulate peste Σ
6. R^* daca R – multime regulara peste Σ
7. Orice alta multime regulara se obtine aplicand de un numar finit de ori reg. 1-6

Multimi regulate si expresii regulate

- Expresii regulate

1.	Φ	expr. reg. coresp. m.reg.	Φ
2.	ε		$\{\varepsilon\}$
3.	a	daca: $a \in \Sigma$	$\{a\}$
4.	$r+s$	daca r,s – expresii regulate	$R \cup S$
5.	rs	daca r,s – expresii regulate	RS
6.	r^*	daca r – expresie regulara	R^*
7.	Orice alta expr. reg. se obtine aplicand de un numar finit de ori reg. 1-6		

$r \mid s$

- Expresii regulate echivalente:

- mult. regulate reprezentate de acestea sunt egale

Expresii regulate

- expresiile regulate – secv. obtinute prin concatenarea de simb. din $\Sigma \cup \{\Phi, \varepsilon, +, *, (,)\}$ (... prioritate ...)
- multimile regulate asociate expresiilor regulate sunt limbaje regulate

Deci: *Orice expresie regulara peste Σ descrie un limbaj regular peste Σ*

Proprietati de inchidere ale limbajelor regulate

Teorema:

Daca

L_1, L_2 sunt limbaje regulate peste alfabetul Σ

atunci:

$L_1 \cup L_2, L_1 \cap L_2, L_1 L_2, L_1^*, \text{complement}(L_1)$
sunt limbaje regulate peste alfabetul Σ

Lema de pompare pt. limbaje regulate

- Daca L este un limbaj regular,
- atunci $\exists p \in \mathbf{N}^*$ (fix pt. un limbaj dat)
(oricat de mare)
- astfel incat:
 $\forall w \in L$ de lungime cel putin p
exista o descompunere de forma $w=xyz$,
unde $0 < |y| \leq p$
cu proprietatea ca: $xy^iz \in L, \forall i \in \mathbf{N}$

Lema de pompare pt. limbaje regulate

$$\begin{aligned} &(\forall L \subseteq \Sigma^*) \\ &(\text{regular}(L) \Rightarrow \\ &((\exists p \geq 1)((\forall w \in L)((|w| \geq p) \Rightarrow \\ &((\exists x, y, z \in \Sigma^*)(w = xyz \wedge (|y| \geq 1 \wedge |y| \leq p \wedge (\forall n \geq 0)(xy^n z \in L)))))))))) \end{aligned}$$

(enunt formal al teoremei)

Lema de pompare pt. limbaje regulate

(o alta versiune, mai “puternica”)

Daca L este un limbaj regular,

- atunci $\exists p \in \mathbf{N}^*$ (fix pt. un limbaj dat)
(oricat de mare)
- astfel incat:
 $\forall w \in L$ de lungime cel putin p
exista o descompunere de forma $w=xyz$ astfel incat
$$0 < |y|$$
$$|xy| \leq p$$
$$xy^iz \in L, \forall i \in \mathbf{N}$$

Lema de pompare pt. limbaje regulate

Observatii:

- Lema da o conditie necesara dar nu suficienta
- daca un limbaj satisface conditiile lemei nu inseamna ca este regular
- folosim negatia lemei de pompare pt. a dem. ca un limbaj nu este regular

Lema de pompare pt. limbaje regulate

De ce se intampla asa:

- Daca L – limb. reg.
 \Rightarrow exista G – gram. reg. a.i. $L(G) = L$ (def.)
 \Rightarrow exista M – AF a.i. $L(M) = L$ (teorema)
- Fie p – nr. de stari ale lui M
- daca $|w| \geq p$ si w – acceptat
 $\Rightarrow \exists$ un drum in graful asociat lui M a.i. etichetele arcelor sunt simb. din w
 \Rightarrow drumul e de lung. p ; adica trece prin $p + 1$ noduri din graf
 $\Rightarrow \exists$ un nod prin care se trece de cel putin 2 ori
 \Rightarrow ciclu/bucla – care se poate repeta de oricate ori !!
 \Rightarrow se poate repeta sirul etichetelor arcelor din bucla !!
(de 0 sau mai multe ori)

Exemplu:

Fie L - limbajul regular corespunzator expresiei regulate:

aa^*b^*

1) fie $w = ab$;

Puteti identifica o descompunere $w=xyz$ a.i. xy^iz in L ?

2) fie $w = aa$;

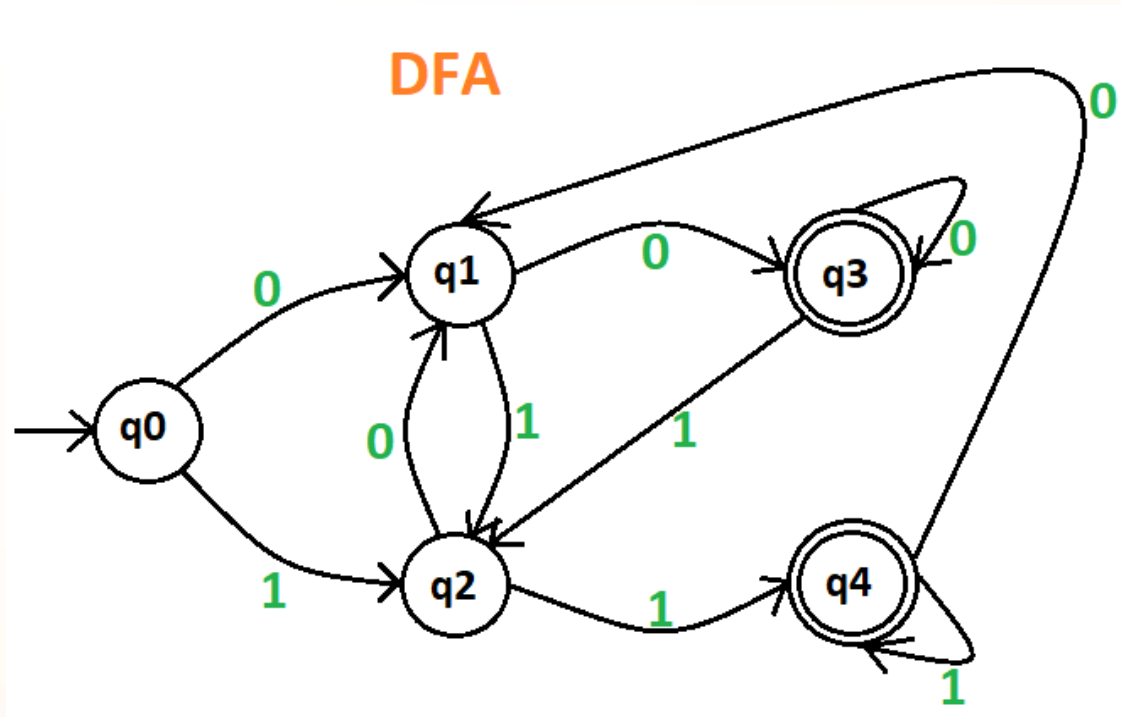
Puteti identifica o descompunere $w=xyz$ a.i. xy^iz in L ?

Analog pt.: $a(ba)^*$

si $w = aba$

Analog pt.: $L=\{a,b\}$ si $w = a$

AF - example



Proprietati de inchidere ale limbajelor regulate

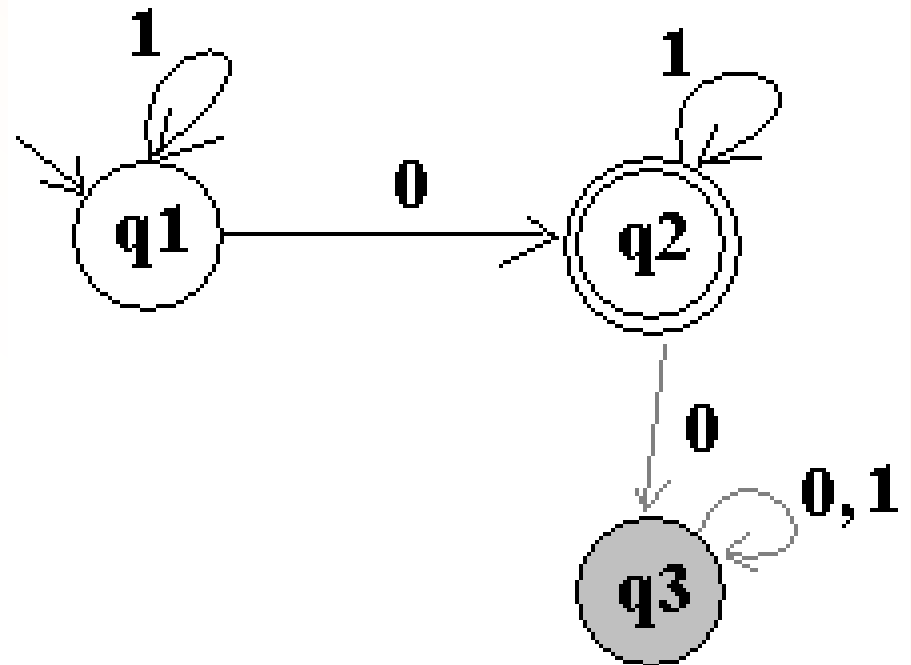
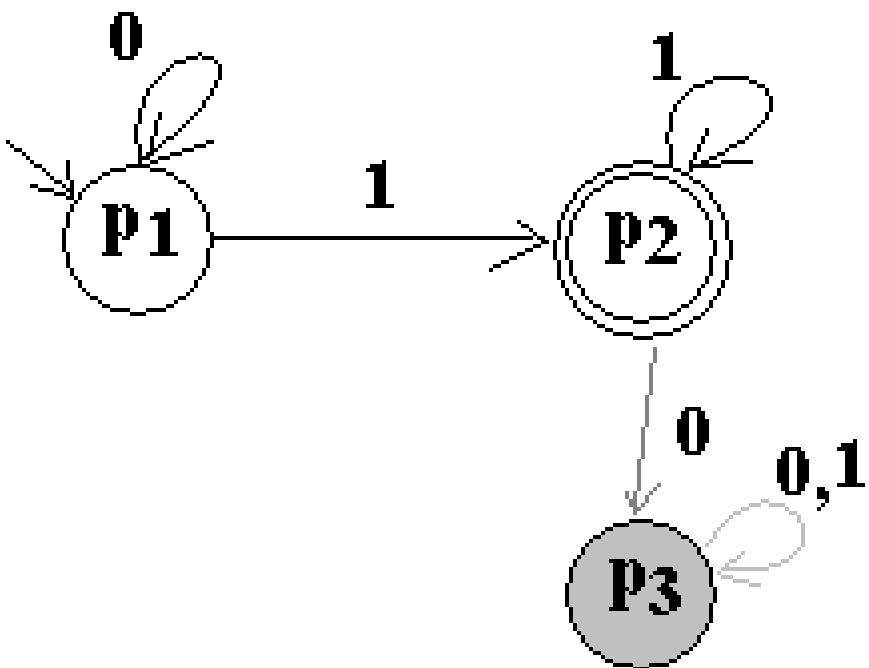
$$L_1 \cap L_2$$

- $M_1 = (Q_1, \Sigma, \delta_1, q_{01}, F_1)$
- $M_2 = (Q_2, \Sigma, \delta_2, q_{02}, F_2)$
- ? $M = (Q, \Sigma, \delta, q_0, F)$

PP. ca aut. M_1 si M_2 sunt deterministe, complet definite !

(alg. de constr. !!)

- $M = (Q_1 \times Q_2, \Sigma, \delta, (q_{01}, q_{02}), F_1 \times F_2)$
- $\delta((q_1, q_2), a) = (\delta_1(q_1, a), \delta_2(q_2, a))$



Proprietati de inchidere ale limbajelor regulate

$\text{complement}(L_1)$

- $M_1 = (Q_1, \Sigma, \delta_1, q_{01}, F_1)$
- ? $M = (Q, \Sigma, \delta, q_0, F)$

PP. ca aut. M_1 este determinist complet definit !

(alg. de constr.)

- $M_1 = (Q_1, \Sigma, \delta_1, q_{01}, Q_1 - F_1)$

In alte resurse : gram. regulare / liniare

Gramatica regulara la dreapta

- $A \rightarrow aB$
- $A \rightarrow b$
- $A \rightarrow \epsilon$ (accepted)

Alte resurse interzic ϵ -producțiile și presupun că secvența vidă nu este prezentă în limbaje.

Gramatica regulara la stanga

- $A \rightarrow B a$
- $A \rightarrow b \dots$

Pentru a avea un limbaj comun, în cadrul acestui curs vom respecta definițiile anterioare.

Gramatica liniara la dreapta (gr.regulara la dreapta extinsa)

- $A \rightarrow a_1 a_2 \dots a_n B$
 - $A \rightarrow b_1 b_2 \dots b_m \dots$
- $S \rightarrow abS \mid aba$

Gramatica liniara la stanga (gr.regulara la dreapta extinsa)

Gramatica liniara:

- are cel mult un neterminal in membrul drept al r.p.

ex.1: $S \rightarrow aSb$, $S \rightarrow ab$

ex.2: $S \rightarrow \{ A$, $A \rightarrow \} , A \rightarrow S \} , A \rightarrow \} S$