LFA CURS 7

LIMBAJE REGULATE - RECAPITULARE

Notații:

$\mathcal{L}_{AFD}$ - familia limbajelor <u>recunoscute</u> de automatele finite deterministe

$\mathcal{L}_{AFN}$ - familia limbajelor <u>recunoscute</u> de automatele finite nedeterministe

$\mathcal{L}_{AFN_\lambda}$ - familia limbajelor <u>recunoscute</u> de automatele finite nedeterministe cu λ-tranziții

$\mathcal{L}_{ExpReg}$ - familia limbajelor <u>descrise</u> de expresiile regulate

$\mathcal{L}_{REG}$ - familia limbajelor <u>generate</u> de gramatici regulate

<u>Teoremă</u>. Există următoarele egalități:

$$\mathcal{L}_{AFD} = \mathcal{L}_{AFN} = \mathcal{L}_{AFN_\lambda} = \mathcal{L}_{ExpReg} = \mathcal{L}_{Reg}$$

<u>Dem</u>. Rezultă din teoremele demonstrate în cursurile anterioare

<u>Obs</u>. Mai târziu vom folosi pentru $\mathcal{L}_{REG}$ notația $\mathcal{L}_3$

Aplicație pentru limbajele regulate

Lexicul unui limbaj de programare (format din următoarele mulțimi: cuvinte cheie, identificatori, constante de diferite tipuri, operatori, separatori, comentarii) poate fi descris cu ajutorul limbajelor regulate.

Fie Σ mulțimea caracterelor permise de un limbaj de programare oarecare. Presupunem că identificatorii încep cu o literă și sunt formați litere, cifre și _, iar constantele întregi din cifre. Putem scrie expresiile regulate următoare pentru a descrie lexicul unui astfel de limbaj, în care în loc de + vom folosi operatorul | (cu aceeași semnificație, de reuniune). Astfel:

letter → A|B|C|...|Z|a|b|...|z

digit → 0|1|2|3|4|5|6|7|8|9

id → (letter)(letter|digit|_)*

constint → digit (digit)*

În acest exemplu id și constint constituie tipuri de token-i. Recunoașterea token-ilor este realizată de un program (funcție subrutină) care se numește analizor lexical sau scanner și care este de fapt implementarea unei AFD (deoarece $\mathcal{L}_{ExpReg} = \mathcal{L}_{AFD}$)

GRAMATICI ȘI LIMBAJE LINIARE

Reamintim că o gramatică independentă de context are o structură $G=(N,\Sigma,S,P)$ unde producțiile din P sunt de forma

$$A \to \alpha, \ A\in N, \ \alpha \in (N\cup\Sigma)^* \text{ (vezi Curs 6)}$$

__Definiție__ Spunem că gramatica $G=(N,\Sigma,S,P)$ este liniară dacă producțiile sale sunt de forma:

$$A \to xBy \quad A,B\in N, \ x,y\in\Sigma^*$$

sau

$$A \to x \quad A\in N, \ x\in\Sigma^*$$

Observăm că o gramatică liniară este o gramatică independentă de context în care fiecare producție are în membrul drept cel mult un neterminal.

__Exemple__

1) G_1: $S \to aSb \mid ab$

$L(G_1) = \{a^n b^n \mid n\geq 0\}$

2) G_2 $S \to aSa \mid bSb \mid a \mid b \mid \lambda$

$L(G_2) = \{w\in\{a,b\}^* \mid w \text{ palindrom}\}$

3) G_3 $S \to aAb \mid bBa$

$A \to aAb \mid \lambda$

$B \to bbBaa \mid \lambda$

$L(G_3) = \{a^n b^n / n \geq 1\} \cup \{b^{2n+1} a^{2n+1} \mid n \geq 0\}$

GRAMATICI ȘI LIMBAJE SEMILINIARE

__Definiție__ Fie $G = (N, \Sigma, S, P)$ o gramatică liniară. Atunci:

a) Spunem că G este liniară la stânga dacă producțiile lui G au una dintre formele:

$A \to Bx$ sau $A \to y$, $x, y \in \Sigma^*$, $A, B \in N$

b). Spunem că G este liniară la dreapta dacă G are producții de una dintre formele:

$A \to xB$ sau $A \to y$, $x, y \in \Sigma^*$, $A, B \in N$

__Lema 1__ Orice gramatică liniară la dreapta este echivalentă cu o gramatică regulată.

__Demonstrație__. Fie $G = (N, \Sigma, S, P)$ gramatică liniară la dreapta. Vom construi gramatica regulată $G' = (N', \Sigma, S, P')$ unde P' conține toate regulile din

P de forma $A \to xB$ sau $A \to y$, cu $x, y \in \Sigma \cup \{\lambda\}$, iar fiecare producție

$$A \to a_1 a_2 \dots a_n B, \; n \geq 2, \; a_1, \dots, a_n \in \Sigma$$

va fi înlocuită cu producțiile:

$$A \to a_1 A_1, \; A_1 \to a_2 A_2, \dots, A_{n-1} \to a_n B$$

unde $A_1, A_2, \dots, A_n$ sunt neterminali noi, asociați în mod unic cu producția $A \to a_1 a_2 \dots a_n B$.

De asemenea, producțiile de forma

$$A \to b_1 b_2 \dots b_m, \; m \geq 2, \; b_1, \dots, b_m \in \Sigma$$

vor fi înlocuite cu producțiile

$$A \to b_1 B_1, \; B_1 \to a_2 B_2, \dots, B_{m-1} \to b_m$$

unde $B_1, \dots, B_{m-1}$ sunt neterminali noi asociați în mod unic cu $A \to b_1 \dots b_m$.

Evident, N' va conține ~~toți~~ neterminalii din N la care se adaugă toți neterminalii noi introduși ca mai sus.

Se arată prin inducție după lungimile derivărilor că

$$S \overset{*}{\underset{G}{\Rightarrow}} w \iff S \overset{*}{\underset{G'}{\Rightarrow}} w$$

de unde rezultă $L(G) = L(G')$, G' gramatică regulată.

Lema 2 Orice gramatică liniară la stânga este echivalentă cu o gramatică regulată

Dem Exercițiu

DERIVĂRI STÂNGI ȘI DREPTE ÎN GRAMATICILE INDEPENDENTE DE CONTEXT. ARBORI DE DERIVARE

Exemplu Fie G_4 gramatica cu producțiile $S \to aSbS \mid bSaS \mid \lambda$

Se poate arăta că $L(G_4) = \{w \in \{a,b\}^* \mid |w|_a = |w|_b\}$

Să considerăm $w = abab$

Pentru acest șir putem avea derivările:

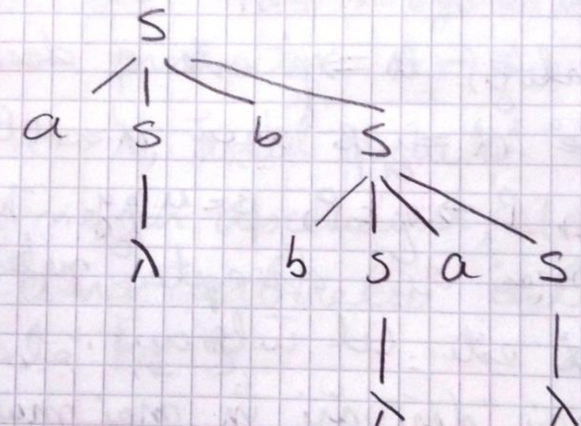
$$\underline{S} \Rightarrow a\underline{S}bS \Rightarrow abS \Rightarrow abaSbS \Rightarrow \begin{cases} \Rightarrow ababS \Rightarrow abab & (1) \\ \Rightarrow abaSb \Rightarrow abab & (2) \end{cases}$$

$$S \Rightarrow aSbS \Rightarrow a\underline{S}b \Rightarrow a(bSaS)b \Rightarrow \begin{cases} abaSb \Rightarrow abab & (3) \\ abSab \Rightarrow abab & (4) \end{cases}$$

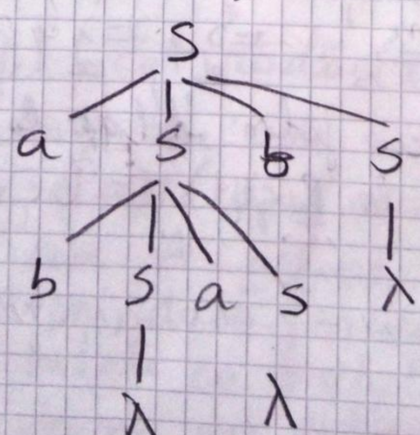
$S \Rightarrow aSbS \Rightarrow abSaSbS \Rightarrow$

$$\Rightarrow \begin{cases} abaSbS \Rightarrow \begin{cases} ababS \Rightarrow abab & (5) \\ abaSb \Rightarrow abab & (6) \end{cases} \\ abSabS \Rightarrow \begin{cases} ababS \Rightarrow abab & (7) \\ ababS \Rightarrow abab & (8) \end{cases} \\ abSaSb \Rightarrow \begin{cases} abaSb \Rightarrow abab & (9) \\ abSab \Rightarrow abab & (10) \end{cases} \end{cases}$$

Am găsit 10 derivări ale șirului abab. Derivările (1) și (2) pot fi scrise și în felul următor:



Derivările (3) - (10) le putem reprezenta



Din aceste exemple deducem că derivările (1) și (2) și respectiv (3)–(10) au ceva în comun.

Def Fie $G=(N,\Sigma,S,P)$ gram. indep. de context.

Spunem că derivarea directă (într-un pas) $\alpha \Rightarrow \beta$ este o derivare stângă într-un pas dacă $\alpha = xA\delta$, $x\in\Sigma^*$, $A\to\gamma\in P$, $\beta = x\gamma\delta$, cu alte cuvinte neterminalul cel mai din stânga al lui α este cel înlocuit.

Vom scrie $\alpha \Rightarrow_s \beta$

Analog, $\alpha \Rightarrow \beta$ este o derivare dreaptă, notată $\alpha \Rightarrow_d \beta$ dacă $\alpha = uBy$, $y\in\Sigma^*$, $B\in N$, $B\to\eta\in P$, $\beta = u\eta y$. În acest caz, cel mai din dreapta neterminal al lui α este cel înlocuit.

Pentru derivări în mai mulți pași vom folosi notațiile:

$\overset{*}{\Rightarrow}_s$, $\overset{+}{\Rightarrow}_s$, $\overset{n}{\Rightarrow}_s$, $\overset{*}{\Rightarrow}_d$, $\overset{+}{\Rightarrow}_d$, $\overset{n}{\Rightarrow}_d$

sau uneori putem folosi 'l' (de la Left) în loc de 's', și 'r' (de la Right) în loc de 'd'.

<u>Def</u> Fie $G = (N, \Sigma, S, P)$ gram. independ. de context. Numim arbore de derivare sau arbore sintactic în G un arbore în care:
- frunzele sunt etichetate cu λ sau cu un simbol $x \in \Sigma$.
- nodurile interne sunt etichetate cu neterminali.
- Dacă un nod intern este etichetat cu neterminalul A și descendenții săi direcți sunt etichetați cu $Z_1 Z_2 \dots Z_m$, $m \geq 1$, $Z_1, \dots, Z_m \in N \cup \Sigma$ de la stânga la dreapta, at. în G avem producția $A \to Z_1 \dots Z_m$.

  Pentru $A \to \lambda$, nodul etichetat cu A va avea un singur descendent, etichetat cu λ
- Rădăcina arborelui este etichetată cu S.

<u>Obs</u> Putem extinde definiția de mai sus astfel încât rădăcina să poată fi etichetată cu un neterminal oarecare, $A \in N$. Vom numi acest arbore un A-arbore
- De asemeni, dacă permitem ca frunzele să fie etichetate cu simbolurii $X \in N \cup \Sigma \cup \{\lambda\}$, vorbim despre un arbore de derivare parțial.

În exemplul de mai sus, pt. -10-
gramatica cu producțiile

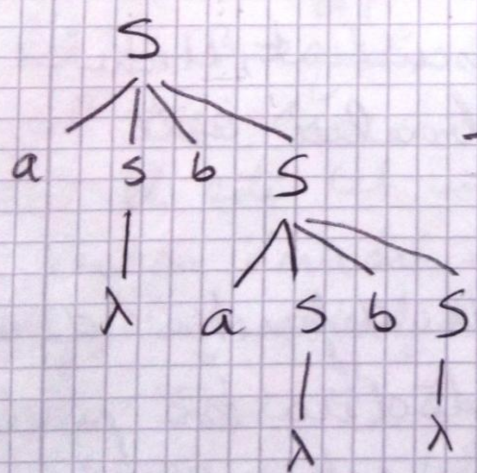
$S \to aSbS \mid bSaS \mid \lambda$

și șirul $abab = w \in L(G)$, găsim două derivări stângi:

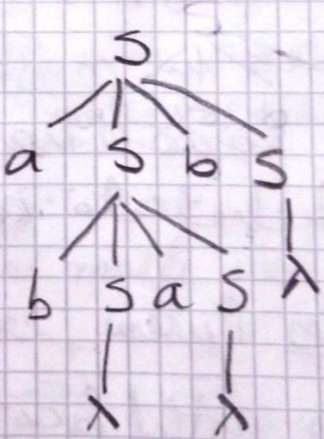
$$S \underset{s}{\Rightarrow} aSbS \underset{s}{\Rightarrow} abS \underset{s}{\Rightarrow} abaSbS \underset{s}{\Rightarrow} ababS \underset{s}{\Rightarrow} abab,$$ și respectiv

$$S \underset{s}{\Rightarrow} aSbS \underset{s}{\Rightarrow} abSaSbS \underset{s}{\Rightarrow} abaSbS \underset{s}{\Rightarrow} ababS \underset{s}{\Rightarrow} abab$$ și arborii

:



T_2:



Frontierele celor 2 arbori, de la stânga la dreapta, sunt reprezentate de șirul $w = abab$ (λ nu apare ca atare, deoarece este elementul neutru pentru concatenare.

Obs Deducem din exemplul anterior că unei derivări stângi (sau drepte) îi corespunde un unic arbore de derivare și reciproc, unui arbore de derivare în G îi corespunde o unică derivare stângă sau dreaptă.

Dar, la modul general, un arbore de derivare poate să corespundă mai multor derivări, în care avem pași de derivare și stângă și dreaptă.

T_1 corespunde la derivările 1) - 2)

T_2 — " — 3) - 10)

Exercițiu. Se dă gramatica cu producțiile:

$E \to E+T \mid T$

$T \to T*F \mid F$

$F \to a$

1) Să se găsească toate derivările pentru $a+a$

2) Să se găsească derivările stângi și cele drepte pt $a*a$

=12=

Teoremă Următoarele 3 afirmații sunt echivalente pt o gram. indep de context G și $x \in L(G)$.

1) x are cel puțin doi arbori de derivare distincți
2) x are cel puțin 2 derivări stângi distincte
3) x are cel puțin 2 derivări drepte distincte.

Dem Fie $G=(N,\Sigma,S,P)$, $x \in L(G)$, $x \in \Sigma^*$.

Arătăm că 1) $\Leftrightarrow$ 2).

1) $\Rightarrow$ 2) Dacă x are 2 arbori de derivare distincți, atunci fiecare arbore de derivare corespunde unei derivări stângi. Cele 2 derivări vor fi diferite, altfel acestea ar corespunde la 2 arbori de derivare diferiți, ceea ce este imposibil pentru o derivare stângă sau dreaptă.

2) $\Rightarrow$ 1) Dacă x are 2 derivări stângi distincte, fie T_1 și T_2 arborii de derivare corespondenți.

Punem în evidență primul pas în care cele 2 derivări stângi diferă.

$$S \underset{s}{\overset{*}{\Rightarrow}} xA\beta \underset{s}{\Rightarrow} x\alpha_1\beta$$

$$\alpha_1 \neq \alpha_2, \; x \in \Sigma^*$$

$$S \underset{s}{\overset{*}{\Rightarrow}} xA\beta \underset{s}{\Rightarrow} x\alpha_2\beta$$

Atât în T_1 cât și în T_2 există nodul intern etichetat cu A și arborii de derivare parțiali pentru $S \overset{*}{\underset{s}{\Rightarrow}} xA\alpha$ sunt identice.

Însă deoarece în T_1 nodul etichetat cu A are ca descendenți sb. din α_1 iar în T_2 nodul corespondent are ca descendenți simbolurile lui $\alpha_2 \neq \alpha_1$, rezultă că $T_1 \neq T_2$.

Analog rezultă $(1) \Leftrightarrow (3)$

Def Spunem că gram. indep. de context G este ambiguă dacă există $x \in L(G)$ astfel încât x are cel puțin 2 arbori de derivare distincți (sau, echivalent, 2 derivări stângi distincte sau 2 derivări drepte distincte).

Altfel, spunem că G este neambiguă

Exemple G_4: $S \to aSbS \mid bSaS \mid \lambda$

G_5: $E \to E+T \mid T$; $T \to T*F \mid F$; $F \to a$

G_6: $S \to iEtS \mid iEtSeS \mid b$; $E \to a$

Am văzut că G_4 este ambiguă

G_6 formalizează cele 2 cazuri pt instrucțiunea if.

Exercițiu. Arătați că G_5, G_6 sunt ambigue