

LIMBAJE FORMALE, AUTOMATE ȘI CALCULABILITATE

Bibliografie

1. Leon Livovschi, Horia Georgescu, Nicolae Țândăreanu, Constantin Popovici – *Bazele Informaticii*, Ed. Didactică și Pedagogică, 1981;
2. Cireșica Jalobeanu, Daniela Marinescu – *Bazele teoriei calculului*, Ed. Albastră, 2007;
3. Gabriel Orman – *Limbaje formale și acceptori*, Ed. Albastră, 2007;
4. Adrian Atanasiu – *Arhitectura calculatoarelor*, InfoData, Cluj, 2006;
5. John Hopcroft, Rajeev Motwani, Jeffrey Ullman – *Introduction to Automata Theory, Languages and Computation*, Addison Wesley, 2006;
6. Horia Georgescu, Radu Boriga, Ana Cristina Dăscălescu – *Limbaje formale, automate și calculabilitate*, Ed. Universității Titu Maiorescu, București, 2014.

Cerințe examen

Examen scris, cu subiecte de teorie și aplicații:	50%
Teme:	10%
Lucrări practice pe calculator:	30%
Prezența curs/ seminar/ laborator:	10%

Nota minimă de promovare este 5.

Introducere gramatici

1. Limbaje formale

1.1. Alfabet, cuvânt, mulțime de cuvinte

Se consideră cunoscute elementele uzuale de algebră și teoria mulțimilor.

Se notează totdeauna mulțimile cu litere latine mari – A, B, C, \dots , eventual indexate.

Definiția 1.1. Fie V o mulțime nevidă (de obicei finită), numită **alfabet** (sau **vocabular**).

Definiția 1.2. Elementele alfabetului V se numesc **caractere** (sau **simboluri**) și vor fi notate cu litere mici de la începutul alfabetului latin, eventual indexate (a, b, c_5, \dots).

Definiția 1.3. Șirul finit de simboluri se numește **cuvânt**.

Cuvântul nul este notat cu ε sau λ .

Definiția 1.4. Se definește **lungimea** unui cuvânt u : numărul simbolurilor sale.

Notăție: $|u|$, deci $|\varepsilon| = 0$.

Definiția 1.5. Fie V un alfabet. Se definește V^* - **mulțimea tuturor cuvintelor peste alfabetul V** , inclusiv ε .

Exemplu 1.1. $\{0,1\}^* = \{\varepsilon, 0, 1, 00, 01, 10, 11, 000, 001, \dots\}$.

Definiția 1.6. Fie V un alfabet. Se definește V^+ - **mulțimea tuturor cuvintelor nenule peste alfabetul V** .

Exemplu 1.2. $\{0,1\}^+ = \{0, 1, 00, 01, 10, 11, 000, 001, \dots\}$.

1.2. Operații cu cuvinte

Definiția 1.7. Se definește **concatenarea** a două cuvinte x, y : cuvântul $x \cdot y$ obținut din simbolurile lui x , în ordinea în care apar, urmate de cele ale lui y de asemenea în ordinea în care apar.

De exemplu: $x = 0100, y = 100$, atunci $x \cdot y = 0100100$ sau $x = 000, y = \varepsilon$, atunci $x \cdot y = 000$.

Observație. Concatenarea este asociativă.

Definiția 1.8. (V^*, \cdot) este monoid (ε este element neutru) și se numește **monoidul liber generat de V** .

Definiția 1.9. Fie V un alfabet. O submulțime $L \subseteq V^*$ este un **limbaj** (formal) **peste alfabetul V** (sau **V - limbaj**) dacă L are o descriere (matematică) finită.

O descriere poate fi:

- *neformală*:
 - ◆ mulțimea cuvintelor peste alfabetul $\{0, 1\}$ care conțin un număr par de 0;
 - ◆ $\{a^n b^n \mid n \in \mathbb{N}\}$.
- *formală* (descriere matematică):
 - ◆ o descriere inductivă a cuvintelor;
 - ◆ o descriere generativă a cuvintelor (gramatică generativă);
 - ◆ o descriere a unei metode de recunoaștere a cuvintelor din limbaj (automat finit, automat pushdown, etc.);
 - ◆ descriere algebrică (expresii regulate).

Exemplul 1.3.

- Fie $V = \{a, b\} \Rightarrow V^* = \{\varepsilon, a, b, ab, ba, bb, aba, bba, \dots, baba, \dots\}$. Putem defini un limbaj
 $L = \{a, ba, aba, baba\}$;
- V = alfabetul englez, iar L = cuvintele corecte din limba engleză;
- $V = \{\text{litere}, +, -, *, /, (,)\}$, iar L = mulțimea expresiilor aritmetice corecte formate cu aceste caractere;
- V = alfabetul ASCII, iar L = mulțimea programelor corecte din *Java* (un program poate fi considerat ca un cuvânt din V^*).

Fie V un alfabet, iar L un limbaj peste V , reprezentând în general cuvintele „corecte” formate cu litere din V . Cum se poate defini limbajul?

- Dacă L este finit, se poate forma o listă cu cuvintele sale;
- Dacă L este infinit, dar cuvintele sale au o anumită formă, de exemplu $\{a^n b^n c^n \mid n > 0\}$;
- Altfel, se pot preciza reguli de formare a cuvintelor sale, deci se va lucra cu **gramatici**!

1.3. Operații cu limbaje

Între limbaje (fiind mulțimi) au loc următoarele operații:

- **Operațiile cu mulțimi** (reuniune, intersecție, diferență, etc);
- **Produs de limbaje**: $L_1 \cdot L_2 = \{u v \mid u \in L_1, v \in L_2\}$.

Exemplul 1.4. Fie limbajele: $L_1 = \{a^n \mid n \geq 1\}$, $L_2 = \{b^n \mid n \geq 1\}$, atunci

$$L_1 \cdot L_2 = \{a^n b^m \mid n \geq 1, m \geq 1\}.$$

- **Iterația (produsul Kleene):** $L^* = \bigcup_{n \geq 0} L^n$, unde:

- ♦ $L^0 = \{\varepsilon\};$
- ♦ $L^{n+1} = L^n \cdot L.$

Exemplul 1.5. Fie limbajul: $L = \{a\}$, atunci

$$L^0 = \{\varepsilon\}, L^1 = L, L^2 = \{aa\}, \dots, L^n = \left\{ \underbrace{aa \dots a}_{\text{de } n \text{ ori}} \stackrel{\text{def}}{=} a^n \right\}, \text{ iar } L^* = \{a^n, n \geq 0\}.$$

Exemplul 1.6. Fie limbajul: $L = \{a^n b^n \mid n \geq 1\}$, atunci $L \cdot L = \{a^{n_1} b^{n_1} a^{n_2} b^{n_2} \mid n_i \geq 1\}$ și

$$L^* = \{a^{n_1} b^{n_1} a^{n_2} b^{n_2} \dots a^{n_k} b^{n_k} \mid n_i \geq 1, k \geq 0\}.$$

2. Mecanisme de generare a limbajelor: gramatici

2.1. Gramatici

Definiția 2.1. Se numește **gramatică** un cvadruplu $G = (N, T, S, P)$ în care T și N sunt alfabet nevide disjuncte, care sunt partiții ale unui alfabet finit și nevid notat prin V :

- N se numește **alfabetul simbolurilor neterminale**, ale cărui elemente se numesc *atribute* și se notează cu litere latine mari (A, B, C, ...);
- T se numește **alfabetul simbolurilor terminale**, ale cărui elemente se notează cu litere latine mici (a, b, c, ...);
- S se numește **simbolul inițial al gramaticii sau simbolul de start** ($S \in N$);
- P este o mulțime finită și nevidă numită **mulțimea producțiilor**, adică reguli de substituție de forma $u \rightarrow v$, unde $u \in (N \cup T)^* N (N \cup T)^*$, iar $v \in (N \cup T)^*$.

Definiția 2.2. Fie $G = (N, T, S, P)$ o gramatică și $u, v \in (N \cup T)^*$.

Se spune că v este **derivat direct (într-un pas) de la u** prin aplicarea regulii

$$x \rightarrow y \in P,$$

și se notează

$$u \Rightarrow v,$$

dacă $\exists p, q \in (N \cup T)^*$ a.î. $u = pxq$ și $v = pyq$.

Definiția 2.3. Dacă $u_1 \Rightarrow u_2 \dots \Rightarrow u_n, n > 1$, se spune că u_n este **derivat din u_1 în G** și notăm

$$u_1 \stackrel{+}{\Rightarrow} u_n \text{ (1 sau mai multe derivări).}$$

Scriem $u \stackrel{*}{\Rightarrow} v$ (0 sau mai multe derivări), dacă $u \stackrel{+}{\Rightarrow} v$ sau $u = v$.

2.2. Limbaj generat

Definiția 2.4. Limbajul generat de gramatica G este:

$$L(G) = \{w \in T^* \mid S \stackrel{+}{\Rightarrow} w\}.$$

Definiția 2.5. Două gramatici G_1 și G_2 sunt **echivalente** dacă $L(G_1) = L(G_2)$.

Exemplul 2.1. Fie gramatica $G = (N, T, S, P)$, $N = \{S, S_1, X\}$, $T = \{a, b, c\}$, iar mulțimea producățiilor P constă din:

$$S \rightarrow abc \quad (1)$$

$$S \rightarrow aS_1Xc \quad (2)$$

$$S_1 \rightarrow abc \quad (3)$$

$$cX \rightarrow Xc \quad (4)$$

$$bX \rightarrow bb \quad (5).$$

Gramatica are trei terminale (a, b, c), trei neterminale (S, S_1, X) și cinci producății.

Cuvintele $abc, a^2b^2c^2 \in L(G)$, deoarece se pot obține printr-o derivare din simbolul inițial S astfel:

$$\begin{array}{c} * \\ S \Rightarrow abc \in L(G) \\ (1) \end{array}$$

și

$$\begin{array}{ccccccc} * & & * & & * & & * \\ S \Rightarrow aS_1Xc \Rightarrow a(abc)Xc \Rightarrow aabcXc \Rightarrow aabXcc \Rightarrow aabbcc = a^2b^2c^2 \in L(G). \\ (2) & & (3) & & (4) & & (5) \end{array}$$

Se observă că $L(G) = \{abc, a^2b^2c^2\}$.

Exemplul 2.2. Limbajul $L = \{a^n b^n \mid n \geq 1\}$ are următoarele definiții:

- Definiția *inductivă*:
 - ♦ $ab \in L$;
 - ♦ Dacă $X \in L$, atunci $aXb \in L$;
 - ♦ Nici un alt cuvânt nu face parte din L .
- Definiția *generativă*:
 - ♦ $G = (\{X\}, \{a, b\}, X, P)$, unde $P = \{X \rightarrow aXb, X \rightarrow ab\}$;

♦ Derivarea cuvântului a^3b^3 :

$$X \Rightarrow aXb \Rightarrow a(aXb)b \Rightarrow aa(ab)bb = a^3b^3 \in L(G).$$

Exemplul 2.3. Limbajul generat de gramatica $G = (N, T, S, P)$, $N = \{S, X\}$, $T = \{a, b, c\}$, P constând din:

$$S \rightarrow abc \quad (1)$$

$$S \rightarrow aSXc \quad (2)$$

$$cX \rightarrow Xc \quad (3)$$

$$bX \rightarrow bb \quad (4)$$

este $L = \{a^n b^n c^n \mid n \geq 1\}$.

Într-adevăr,

$$\begin{array}{ccccccc} & * & & * & & * & & * \\ S & \Rightarrow & aSXc & \Rightarrow & aabcXc & \Rightarrow & aabXcc & \Rightarrow & aabbcc = a^2b^2c^2 \in L, \\ (2) & & (1) & & (3) & & (4) \end{array}$$

$$\begin{array}{ccccccccccc} & * & & * & & * & & * & & * & & * \\ S & \Rightarrow & aSXc & \Rightarrow & aaSXcXc & \Rightarrow & aaabcXcXc & \Rightarrow & aaabXccXc & \Rightarrow & aaabbccXc & \Rightarrow & aaabbXcc & \Rightarrow & aaabbXccc \\ (2) & & (2) & & (1) & & (3) & & (4) & & (3) & & (3) \end{array}$$

$$\begin{array}{ccc} & * & \\ & \Rightarrow & aaabbbccc = a^3b^3c^3 \in L \\ & (4) \end{array}$$

și se demonstrează că $L = \{a^n b^n c^n \mid n \geq 1\}$ astfel:

se pornește de la simbolul de start S și până când S “dispare”, cuvântul curent are forma $a^n S \alpha$, unde în α există n de X și c .

După dispariția lui S :

- toate c - urile sunt deplasate la dreapta după X -uri, conform (3);
- toate X - urile care urmează lui b trec în b , conform (4),
și se obține cuvântul $a^n b^n c^n$, $n \geq 1$, format numai din terminale.

Temă:

1. Fie $V = \{a, b, c\}$, $W = \{x, y, z, u\}$. Să se scrie $V \cdot W$, $W \cdot V$, $W \cdot W$.
2. Fie A o mulțime nevidă. A este finită dacă și numai dacă $A^* = \{\varepsilon\}$.