

Alfabet. Siruri

Alfabet (notat Σ sau T) – multime finita, nevida de simboluri.

Sir - o secventa finita de simboluri dintr-un alfabet.

- Sirul vid: **e**

$\Sigma_1 = \{0, 1\}$

Siruri: 0, 1, 00, 11, 01, 10, ...

$\Sigma_2 = \{a, b\}$

Siruri: a, b, aa, bb, ..., aabb, bbaa, ...

$\Sigma_3 = \{ (,) \}$

Siruri: (,), ((, (), ...

Multimi de siruri

Multimea sirurilor peste un alfabet Σ (notata Σ^*) (Kleene star)

Multimea Σ^* a şirurilor peste alfabetul Σ este definita:

• $e \in \Sigma^*$

• daca $a \in \Sigma$ si $w \in \Sigma^*$, atunci $aw \in \Sigma^*$

Operatii pe siruri:

- Lungimea unui sir
- Concatenarea a doua siruri
- Exponentiere
- Reverse

Operatii pe siruri

Lungimea unui sir:

$$||: \Sigma^* \rightarrow \mathbb{N}$$

$$|w| = \begin{cases} 0, & \text{daca } w = e \\ 1 + |w'|, & \text{daca } w = aw', a \in \Sigma \text{ si } w' \in \Sigma^* \end{cases}$$

Concatenarea a doua siruri:

$$\bullet: \Sigma_1^* \times \Sigma_2^* \rightarrow \Sigma^*$$

$$w_1 \bullet w_2 = w_1 w_2 = \begin{cases} w_2, & \text{dacă } w_1 = e \\ aw'_1 w_2, & \text{dacă } w_1 = aw'_1, a \in \Sigma \text{ si} \\ & w'_1, w_2 \in \Sigma^* \end{cases}$$

Operatii pe siruri

- Un sir $v \in \Sigma^*$ este un **subsir** al şirului $w \in \Sigma^*$ daca există sirurile $x, v, y \in \Sigma^*$ astfel incat $w = xvy$
- Daca $w = uv$, $v, w \in \Sigma^*$, oricare $u \in \Sigma^*$, atunci v este un **sufix** al lui w
- Daca $w = uv$, $u, w \in \Sigma^*$, oricare $v \in \Sigma^*$, atunci u este un **prefix** al lui w

Prefix: a, ab, abb, abba

abba

Sufix: a, ba, bba, abba

Subsir: b, bb, ...

Operatii pe siruri

- **e** este **subsir** pentru **orice şir** $w \in \Sigma^*$
- orice sir $w \in \Sigma^*$ este un **subsir propriu**
- **e** este **prefix** si **sufix** pentru orice $w \in \Sigma^*$
- orice sir $w \in \Sigma^*$ este **prefix propriu** si **sufix propriu**

Operatii pe siruri

Exponentierea unui sir peste un alfabet Σ :

$$E : \Sigma^* \times \mathbb{N} \rightarrow \Sigma^*$$

$$E(w, i) = w^i = \begin{cases} e, \text{ dacă } i = 0 \\ ww^{i-1}, \text{ în caz contrar} \end{cases}$$

$$(ab)^3 = ababab$$

$$(01)^2 = 0101$$

Operatii pe siruri

Revers-ul unui sir peste un alfabet Σ :

$$R : \Sigma^* \rightarrow \Sigma^*$$

$$R(w) = wR = \begin{cases} e, \text{ dac\u0103 } w = e \\ uRa, \text{ dac\u0103 } w = au, a \in \Sigma \text{ si } u \in \Sigma^* \end{cases}$$

$$abab^R = baba$$

$$1101^R = 1011$$

Σ^*, Σ^+

- Σ^k - mulțimea tuturor șirurilor de lungime k, formate din simboluri din alfabetul Σ

$$\Sigma = \{a, b\} \quad \Sigma^1 = \{a, b\} \quad \Sigma^2 = \{aa, ab, bb, ba\}$$

•Kleene star

$$\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \dots$$

•Kleene plus

$$\Sigma^+ = \Sigma^1 \cup \Sigma^2 \cup \dots$$

Limbaje

- **Limbaaj peste alfabetul Σ** - orice submulțime a mulțimii Σ^*
- Elementele limbajului L: **propozitii**

Exemplu: Limbaje peste alfabetul $\Sigma = \{a, b\}$

$$L_1 = \{w \in \{a, b\}^*\}$$

$$L_2 = \{w \in \{a, b\}^+\}$$

$$L_3 = \{w \in \{a, b\}^* \mid |w| = \text{para}\}$$

$$L_4 = \{w \in \{a, b\}^* \mid w = a^n b^n\}$$

$$L_5 = \{w \in \{a, b\}^* \mid w = a^n b^n a^n b^n\}$$

Limbaje

- **Reprezentarea limbajelor**
- **Pentru orice limbaj se poate gasi o reprezentare finita?**

Reprezentarea limbajelor

- Limbaj L peste alfabetul Σ
- A - alfabet folosit pentru reprezentarea limbajului L

Limbaj L peste alfabetul Σ	A - alfabet folosit pentru reprezentarea limbajului L
Numarul de limbaje peste Σ : $2^{ \Sigma^* }$	Numarul de siruri folosite pentru reprezentare: $ A^* $
2^{Σ^*} nenumarabila	A^* numarabila

- Nu putem avea o reprezentare finita pentru orice limbaj
- Ne intereseaza numai limbajele pentru care avem o reprezentare finita

Operatii asupra limbajelor

- Reuniunea

- Intersectia

- Diferenta

- **Concatenarea** a două limbaje L_1 si L_2
(L_1 si L_2 sunt definite peste alfabetul Σ^*):

$$L_1 \cdot L_2 = L_1 L_2 = \{w_1 w_2 \mid \text{oricare } w_1 \in L_1 \text{ și } w_2 \in L_2\}$$

$$L_1 = \{ab, ba, aa, bb\}$$

$$L_2 = \{c, cc\}$$

$$L_1 L_2 = \{abc, bac, aac, bbc, abcc, bacc, aacc, bbcc\}$$

Operatii asupra limbajelor

- **Exponentierea** unui limbaj L (peste alfabetul Σ^*):

$$L^i = \begin{cases} \{e\}, & \text{dacă } i=0 \\ LL^{i-1}, & \text{altfel} \end{cases}$$

$$L_1 = \{ab, ba\}$$

$$L_1^2 = \{ba, ab\}\{ba, ab\} = \{baba, baab, abba, abab\}$$

- **Reversul** unui limbaj L (peste alfabetul Σ):

$$L^R = \{w^R \mid \text{oricare } w \in L\}$$

$$L_1 = \{ab, ba, aa, bb\}$$

$$L_1^R = \{ba, ab, aa, bb\}$$

$$L^*, L^+$$

- **Kleene star** pentru un limbaj L : L^*
 - $\epsilon \in L^*$
 - dacă $w \in L$ și $w' \in L^*$ atunci $ww' \in L^*$
- **Kleene plus** pentru un limbaj L : L^+
 - $L \subseteq L^+$
 - dacă $v \in L$ și $w \in L^+$ atunci $vw \in L^+$

Proprietatile limbajelor

1. $L\Phi = \Phi L = \Phi$

2. $L\{e\} = L$

3. $L_1(L_2L_3) = (L_1L_2)L_3$

4. $L_1(L_2 \cup L_3) = (L_1L_2) \cup (L_1L_3)$

5. $L^1 = L$

6. $L^{i+j} = L^iL^j$

7. $L^* = \bigcup L^i, i \geq 0$

8. $L^+ = \bigcup L^i, i \geq 1$

9. $L^+ = LL^*$

Definirea limbajelor

- **Generare**
- **gramatici**
- **Acceptoare**
- **Automate**
 - **Automate finite**
 - **Automate cu stiva**
 - **Masina Turing**

Gramatici

$$G = (N, \Sigma, P, S)$$

- N - mulțimea simbolurilor neterminali
- Σ - mulțimea simbolurilor terminali ($\Sigma \cap N = \emptyset$)
- P - o submulțime finită din
$$(N \cup \Sigma)^* N (N \cup \Sigma)^* \times (N \cup \Sigma)^*$$
- $p \in P, p = (\alpha, \beta)$ este notat cu $\alpha \rightarrow \beta$
 $p = \text{productie}$
- $S \in N$ - simbolul de start al gramaticii G .

Gramatici – Conventii de notare

- **simboluri terminale** - literele mici de la inceputul alfabetului latin (a, b, c, \dots)
- **siruri de simboluri terminale** - literele mici de la sfarsitul alfabetului latin (u, v, x, \dots)
- **simboluri neterminale** - literele mari de la inceputul alfabetului latin (A, B, C, \dots) si S
- **simboluri terminale sau neterminale** - literele mari de la sfirsitul alfabetului latin (U, V, X, \dots) - elemente din $N \cup \Sigma$
- **siruri din $(N \cup \Sigma)^*$** - literele alfabetului grecesc (α, β, \dots)

Gramatici

Forma propozitională în gramatica G :

- orice sir din $(N \cup \Sigma)^*$ obținut prin:
 - S este o formă propozitională
 - dacă $\alpha\beta\gamma$ este o formă propozitională și există o producție $\beta \rightarrow \delta$ atunci $\alpha\delta\gamma$ este o formă propozitională.

Relația de derivare într-o gramatică $G \Rightarrow_G$

- definită asupra formelor propozitionale
 - α și β două forme propoziționale
 - $\alpha \Rightarrow \beta$ dacă și numai dacă
există w_1, w_2 și $\gamma \rightarrow \delta \in P$ astfel încât
 $\alpha = w_1\gamma w_2$ și $\beta = w_1\delta w_2$.

Gramatici

- **Derivarea in k pasi, $\alpha \Rightarrow^k \beta$**

- daca exista $\alpha_0, \alpha_1, \dots, \alpha_k$ forme propositionale astfel incat:

$$\alpha = \alpha_0 \Rightarrow \alpha_1 \Rightarrow \dots \Rightarrow \alpha_{i-1} \Rightarrow \alpha_i \dots \Rightarrow \alpha_k = \beta$$

- **inchiderea tranzitiva si reflexiva a relatiei \Rightarrow^***

$$\alpha \Rightarrow^* \beta \iff \alpha \Rightarrow^k \beta \text{ pentru orice } k \geq 0$$

- **inchiderea tranzitiva \Rightarrow^+**

$$\alpha \Rightarrow^+ \beta \iff \alpha \Rightarrow^k \beta \text{ pentru orice } k \geq 1$$

Gramatici

Propozitie - o forma propozitionala intr-o gramatică **G**, care contine numai simboluri terminale.

Limbajul generat de o gramatica **G**, notat cu **$L(G)$** , este reprezentat de mulțimea tuturor propozitiilor generate de gramatica **G**:

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

Doua gramatici **G1**, **G2** se numesc **echivalente** daca genereaza acelasi limbaj adica, **$L(G1) = L(G2)$** .

Clasificarea gramaticilor

Ierarhia Chomsky

- **gramatici de tip 0 (gramatici fără restricții - GFR)**

$\alpha \rightarrow \beta$ cu $\alpha \in (N \cup \Sigma)^* N (N \cup \Sigma)^*$, $\beta \in (N \cup \Sigma)^*$

- **gramatici de tip 1 (gramatici dependente de context - GDC)**

$\alpha \rightarrow \beta$, $|\alpha| \leq |\beta|$, cu $\alpha \in (N \cup \Sigma)^* N (N \cup \Sigma)^*$
 $\beta \in (N \cup \Sigma)^*$

sau de forma

$S \rightarrow e$, caz în care S nu apare în partea dreaptă al nici unei alte productii.

- **gramatici de tip 2 (gramatici independente de context – GIC)**

$A \rightarrow \alpha$ cu $A \in N$, $\alpha \in (N \cup \Sigma)^*$

- **gramatici de tip 3 (gramatici regulate dreapta - GR)**

$A \rightarrow \alpha B$ cu $A \in N$, $B \in (N \cup \{e\})$, $\alpha \in \Sigma^*$

Clasificarea limbajelor

Ierarhia Chomsky

