

Analiza lexicala. Alfabet. Limbaj.

Curs 2

ANALIZA LEXICALA

- › **Date de intrare:** codul sursa al programului
- › **lesire:** sirul atomilor lexicali
 - **FIP** (Forma Interna a Programului) cu doua coloane: cod atom si pozitie atom in TS
 - **TS** (Tabela de Simboluri)
- › Sau **Eroare lexicala**



Atomi lexicali

- › Clase:
 - Separatori
 - Operatori
 - Cuvinte rezervate
 - Identificatori
 - Constante

- Erori lexicale

Determinate pe baza specificarii limbajului de programare

Algoritmul de analiza lexicala

Cattimp (mai exista caractere necitite in codul sursa) **executa**
detecteaza atom

Daca atom e cuvant-cheie sau operator sau separator **atunci**
adauga in FIP codul corespunzator atomului

altfel

Daca e identificator sau constanta **atunci**

Daca nu exista deja in TS **atunci**

adauga atom in TS

Sf. Daca

adauga in FIP codul atomului si pozitia in TS

altfel raporteaza eroare

SfDaca

SfDaca

Sfcattimp

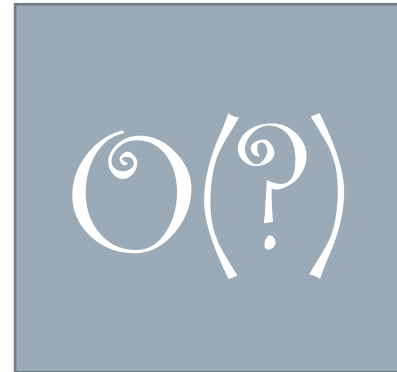
Cautarile in TS trebuie sa se
desfasoare rapid

Structuri de date???

Codul in FIP pentru identificatori – 0, constante - 1

Tabela de simboluri

- › 2 variante:
 - Unica pentru identificatori si constante
 - 2 tabele separate: una pentru identificatori, una pentru constante
- › Organizare:
 1. Tabela – identificatorii si constantele adaugate in ordinea aparitiei in codul sursa
 2. Tabela sortata lexicografic
 3. Arbore binar de cautare
 4. Tabela de dispersie



Alfabet, secventa, subsecventa

› Definitie:

Fie Σ o multime de elemente **finita si nevida**. O astfel de multime Σ o numim **alfabet**, iar elementele ei se numesc **simboluri**.

› Definitie:

O succesiune de simboluri ale alfabetului Σ o numim **secventa** peste alfabetul Σ , iar o *multime de elemente consecutive* dintr-o secventa o numim **subsecventa**.

$$w = a_1 a_2 \dots a_n, a_i \in \Sigma, i = 1, n \text{ secventa peste } \Sigma$$

Concatenare, prefix, sufix, lungimea unei secvente

Daca $x=a_1a_2...a_n$ si $y=b_1b_2...b_m$ atunci $z=xy$, adica $z=a_1a_2...a_nb_1b_2...b_m$ reprezinta **concatenarea** secventelor x si y .

Daca x, y, z sunt secvente peste alfabetul Σ , atunci x este un **prefix** al secventei xy , iar y un **sufix** al secventei xy , iar y in secventa xyz este o **subsecventa**.

Lungimea unei secvente w peste un alfabet Σ , notata $|w|$, este numarul simbolurilor din care este alcatuita. Secventa de lungime zero o numim **secventa vida** si o notam cu ϵ .

Σ^*

› Σ^* - multimea **tuturor secventelor** peste alfabetul Σ

Exemplu: $\Sigma=\{a\}$, $\Sigma^*=?$

Numim limbaj peste alfabetul Σ , o submultime L a multimii Σ^* , deci $L \subseteq \Sigma^*$.

$\forall w_1, w_2 \in \Sigma^*$ avem $w_1 w_2 \in \Sigma^*$
 $\forall w_1, w_2, w_3 \in \Sigma^*$ avem: $(w_1 w_2) w_3 = w_1 (w_2 w_3)$
 $\forall w \in \Sigma^*$, $\epsilon w = w \epsilon = w$

MULTIMI SPECIALE

$$\Sigma^n = \{w \mid w - \text{secventa peste } \Sigma, \\ |w| = n\}$$

$$\Sigma^* = \{w \mid w - \text{secventa peste } \Sigma, \\ 0 \leq |w| \}$$

$$\Sigma^+ = \{w \mid w - \text{secventa peste } \Sigma, \\ 0 < |w| \}$$

$$\Sigma^* = \Sigma^+ \cup \{\epsilon\}$$

Operatii pe limbaje

Daca L_1 este un limbaj peste alfabetul Σ_1 si L_2 este un limbaj peste alfabetul Σ_2 atunci prin operatiile:

$$L_1 \cup L_2 - \textit{reuniunea}$$

$$L_1 \cap L_2 - \textit{intersectia}$$

$$L_1 - L_2 - \textit{diferenta}$$

$$L_1 L_2 = \{uv \mid u \in L_1, v \in L_2\} - \textit{concatenarea}$$

$$L1 / L2 = \{w \in \Sigma^* \mid \exists y \in L2 : wy \in L1\} - \textit{catul la dreapta}$$

$$L1 \setminus L2 = \{w \in \Sigma^* \mid \exists y \in L2 : yw \in L1\} - \textit{catul la stanga}$$

obtinem de asemenea limbaje peste alfabete alese corespunzator.

Exercitii

- › Fie limbajele $L_1=\{a\}$ si $L_2=\{cc,d\}$.
- › Care este limbajul $L_1 \cup L_2$?
- › Care este limbajul $L_1 \cap L_2$?
- › Care este limbajul $L_1 L_2$?
- › Care este limbajul L_1^* ?

Operatii pe limbaje

$\bar{L} = \{x \in \Sigma^* \mid x \notin L\}$ *complementara*

$L^* = \bigcup_{n \geq 0} L^n$, unde $L^n = LL^{n-1}$ $L^0 = \{\varepsilon\}$ *inchiderea reflexiva si tranzitiva*

$L^+ = \bigcup_{n \geq 1} L^n$ sau $L^+ = LL^*$ sau $L^* = L^+ \cup \{\varepsilon\}$ *inchiderea tranzitiva*

Exemple de limbaje

- › Limbajul numerelor naturale in reprezentare zecimala
- › $\Sigma=?$ $L=?$
- › Limbajul identificatorilor
- › $\Sigma=?$ $L=?$
- › Limbajul numerelor reale pozitive in reprezentare binara
- › $\Sigma=?$ $L=?$