

# Nyelvek használata adatszerkezetek, képek leírására

## Formális nyelvek, 2. gyakorlat

**Célja:** A formális nyelvek elmélete alapfogalmainak gyakorlása, formális nyelvek néhány alkalmazási lehetőségének bemutatása (adatszerkezetek szintaktikus leírása, teknőc grafika képek reprezentálása, fák és nyelvek).

**Fogalmak:** A formális nyelvek elmélete alapfogalmainak gyakorlása, formális nyelvek néhány alkalmazási lehetőségének bemutatása (adatszerkezetek szintaktikus leírása, teknőc grafika képek reprezentálása, fák és nyelvek).

**Feladatok jellege:** A lista és a fa adatszerkezet leírása kétszintű nyelvtannal, a Koch-szigetek teknőc-grafikával való leírásának tanulmányozása, fák reprezentációja szelektorhalmazokkal, faosztályok leírása nyelveken értelmezett rekurzióval

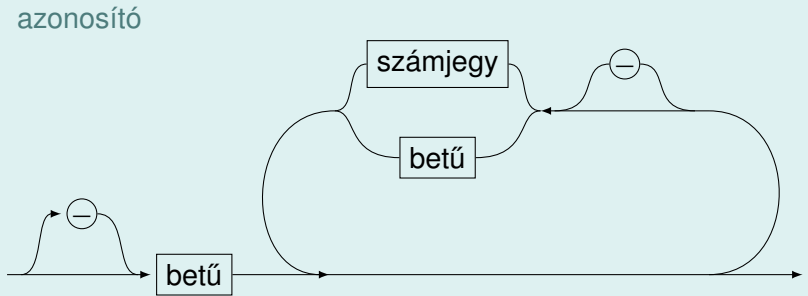
2008/09 I. félév

# Házi feladatok megoldása

## 1. feladat

Módosított azonosító: belsejében lehet \_ jel is. Kezdődhet, de nem végződhet vele, két aláhúzás nem lehet egymás mellett. Van nem "\_" karaktere és első nem "\_" karakter betű. Írjuk fel (E)BNF formulákkal!

**Megoldás:** (az első nem \_ szimbólum betű)



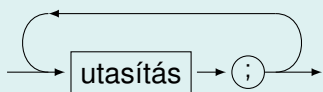
$(\text{azonosító}) ::= \{ \_ \} \{ \text{betű} \} @ \{ \{ \_ \} \{ \text{betű} \} \mid \{ \text{számjegy} \} \}$

# Házi feladatok megoldása

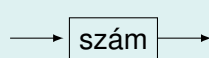
## 2. feladat

EP teljes átírása szintaxis gráfokkal (folytatás).

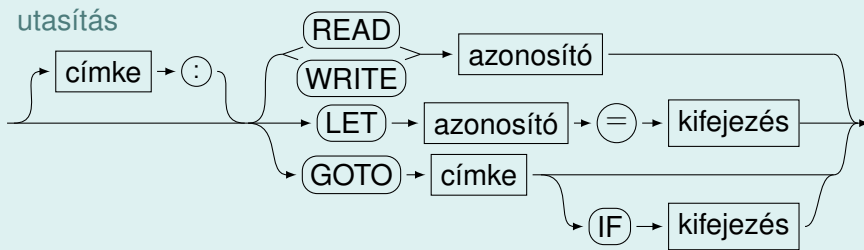
egyszerű program



címke



utasítás

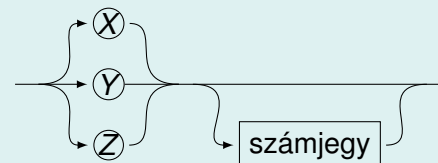


# Házi feladatok megoldása

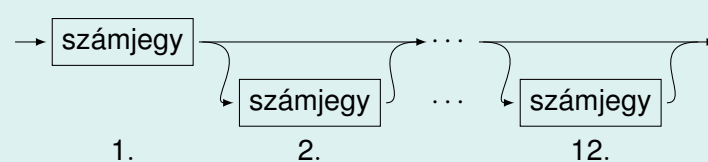
## 2. feladat

EP teljes átírása szintaxis gráfokkal (folytatás).

azonosító



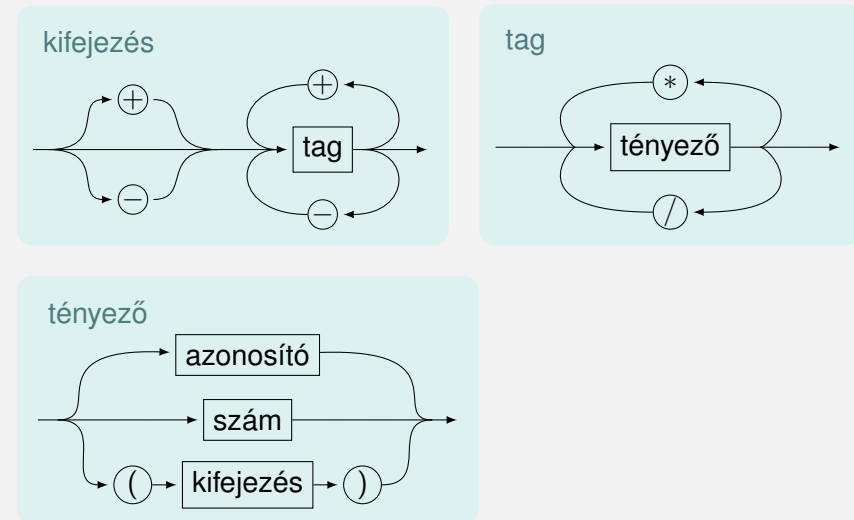
szám



## Házi feladatok megoldása

### 2. feladat

EP teljes átírása szintaxis gráfokkal (folytatás).



## Házi feladatok megoldása

### 3. feladat

A kiadott PASCAL szintaxis tanulmányozása. A kifejezések átírása (E)BNF-re.

#### Megoldás:

$\langle \text{kifejezés} \rangle ::= \langle \text{egyszerű kifejezés} \rangle \{ | \langle \text{relop} \rangle \langle \text{egyszerű kifejezés} \rangle \}$   
 $\langle \text{relop} \rangle ::= = | < | > | <> | <= | >= | \text{IN}$   
 $\langle \text{egyszerű kifejezés} \rangle ::= \{ | + | - \} \langle \text{tag} \rangle @ \{ \langle \text{addop} \rangle \langle \text{tag} \rangle \}$   
 $\langle \text{addop} \rangle ::= + | - | \text{OR}$   
 $\langle \text{tag} \rangle ::= \langle \text{tényező} \rangle @ \{ \langle \text{mpop} \rangle \langle \text{tényező} \rangle \}$   
 $\langle \text{mpop} \rangle ::= * | / | \text{DIV} | \text{MOD} | \text{AND}$   
 $\langle \text{tényező} \rangle ::= \langle \text{előjel nélküli konstans} \rangle | \langle \text{változó} \rangle |$   
 $\langle \text{függvénynév} \rangle \{ | ( \langle \text{kifejezés} \rangle @ \{ \langle \text{kifejezés} \rangle \} ) \}$   
 $\langle \text{kifejezés} \rangle | \text{NOT} \langle \text{tényező} \rangle [ [ ] ]$   
 $\{ \langle \text{kifejezés} \rangle \{ | .. \langle \text{kifejezés} \rangle \} @ \{ \langle \text{kifejezés} \rangle \{ | .. \langle \text{kifejezés} \rangle \} \}$   
 és így tovább ...

## Házi feladatok megoldása

### 4. feladat

Kifejezések fogalmának leírása  $W$  nyelvtannal.

#### Megoldás:

Metaszabályok:

$\langle \hat{X} \text{ kifejezés} \rangle ::= \langle \hat{X} \text{ tag} \rangle @ \{ \langle \hat{X} \text{ addop} \rangle \langle \hat{X} \text{ tag} \rangle \}$   
 $\langle \hat{X} \text{ tag} \rangle ::= \langle \hat{X} \text{ tényező} \rangle @ \{ \langle \hat{X} \text{ mpop} \rangle \langle \hat{X} \text{ tényező} \rangle \}$   
 $\langle \hat{X} \text{ tényező} \rangle ::= \langle \hat{X} \rangle | ( \langle \hat{X} \text{ kifejezés} \rangle ) | \langle \text{azonosító} \rangle$

Hiperszabály:

$\hat{X} := \text{egész} | \text{valós} | \text{Boole}$

$\langle \text{Boole addop} \rangle ::= \text{OR}$

$\langle \text{Boole mpop} \rangle ::= \text{AND}$

$\langle \text{egész addop} \rangle ::= + | -$

$\langle \text{egész mpop} \rangle ::= * | /$

## Szavak, nyelvek, konkatenáció

Ha  $H$  egy halmaz, jelölje  $H^*$  a  $H$  elemeiből képezhető összes véges sorozatok halmazát (beleértve az  $\epsilon$ -nal jelölt üres sorozatot is).

Rögzítsünk egy  $U$  megszámlálhatóan végtelen számosságú halmazt, melyet **univerzális ábécének** nevezünk és feltesszük, hogy tartalmaz minden számunkra szükséges karaktert.

$T \subset U$ , ha  $|T|$  véges

$t \in T$

$u \in T^*$

$L \in 2^{T^*}$  (vagyis  $L \subseteq T^*$ )

$\mathcal{L} \subseteq \bigcup_{T \subset U, |T| < \infty} 2^{T^*}$

**ábécé**

a  $T$  ábécé egy **betűje**

a  $T$  ábécé feletti **szó**

a  $T$  ábécé feletti **nyelv**

**nyelvcsalád** (nyelvek egy összessége)

Ha  $u = a_1 \cdots a_n$  ( $a_i \in T, 1 \leq i \leq n$ ) és  $v = b_1 \cdots b_m$

( $b_i \in T, 1 \leq i \leq m$ ), akkor az  $uv = a_1 \cdots a_n b_1 \cdots b_m$  szót az  $u$  és  $v$  szavak **konkatenációjának** nevezzük. Általában  $uv \neq vu$ .

Ha  $a \in T$  és  $L \subseteq T^*$ , jelölje  $aL = \{au \mid u \in L\}$ .

## Prefixek és suffixek

Legyen rögzítve egy  $T$  ábécé.

### Prefix és suffix

$v \in T^*$  **prefixe**  $u \in T^*$ -nak  $\Leftrightarrow \exists v' \in T^* u = v v'$

$v \in T^*$  **suffixe**  $u \in T^*$ -nak  $\Leftrightarrow \exists v' \in T^* u = v' v$

$\text{pre}(u, \ell)$  ill.  $\text{suf}(u, \ell)$  az  $u$   $\ell$  hosszú prefixe ill. suffixe.

$v$  valódi prefix ill. suffix, ha  $v \neq \varepsilon, u$ .

$\text{Pre}(u) = \{v; v \text{ prefixe } u\text{-nak}\}$        $\text{Pre}(L) = \bigcup_{u \in L} \text{Pre}(u)$

$\text{Suf}(u) = \{v; v \text{ suffixe } u\text{-nak}\}$        $\text{Suf}(L) = \bigcup_{u \in L} \text{Suf}(u)$

Azt mondjuk, hogy  $L$  zárt a prefix illetve suffix képzésre, ha

$\text{Pre}(L) = L$  illetve  $\text{Suf}(L) = L$  teljesül.

**Például:**  $\text{suf}(\text{abbaba}, 4) = \text{baba}$ ,

$\text{Pre}(\text{abbaa}) = \{\varepsilon, a, ab, abb, abba, abbaa\}$ ,

$\text{Suf}(\{ab, abb, bb\}) = \{\varepsilon, b, ab, bb, abb\}$

## $r$ -áris fák

**$r$ -áris fa:** Olyan gyökeres fa, ahol minden csúcsnak legfeljebb  $r$  gyereke van, a gyerekekhez vezető élek a  $0, 1, \dots, r-1$  számok valamelyikével vannak címkézve. Egy csúcs gyerekeihez induló élek címkéje különböző.

**Elemi szelektor:** a  $\{0, 1, \dots, r-1\}$  halmaz egy eleme.

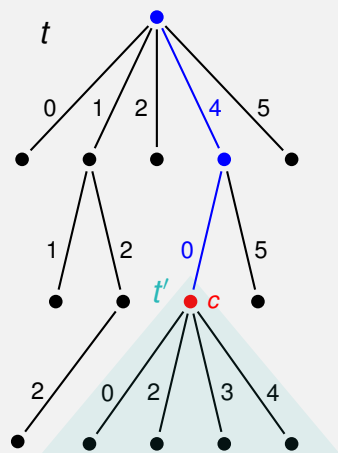
**Szelektor:** a  $\{0, 1, \dots, r-1\}^*$  halmaz egy eleme, azaz elemi szelektorok egy véges sorozata.

Egy  $r$ -áris fában egy **élhez tartozó elemi szelektor:** az él címkéje.

Egy  $r$ -áris fában egy **csúcshoz tartozó szelektor:** A gyökérből a csúcsba vezető élsorozat éleihez tartozó elemi szelektorok sorozata.

## $r$ -áris fák szelektorai

Példa



$t$  6-áris fa

Mi a  $c$  csúcshoz tartozó  $\omega$  szelektor?

$\omega = 40$

A  $t$  fa szelektorainak halmazát jelölje  $\text{Sel}(t)$ .  $\text{Sel}(t)$  a  $\{0, 1, \dots, r-1\}$  ábécé feletti prefixképzésre zárt nyelv.

$\text{Sel}(t) = \{\varepsilon, 0, 1, 2, 4, 5, 11, 12, 40, 45, 122, 400, 402, 403, 404\}$

$\text{Sel}(t') = \{\omega' \subseteq \{0, 1, \dots, 5\}^* \mid 40\omega' \in \text{Sel}(t)\} = \{\varepsilon, 0, 2, 3, 4\}$

## $r$ -áris fák megadása szelektorainak halmazával

Legyen  $L \subseteq \{0, 1, \dots, r-1\}^*$  prefixképzésre zárt véges nyelv. Ekkor egyértelműen konstruálható egy olyan  $r$ -áris  $t$  fa, melyre  $\text{Sel}(t) = L$ .  $t$  rekurzívan építhető fel. Ha  $L \neq \emptyset$ , akkor legyen  $t_0$  az egyetlen csúcsból álló fa. Ha már felépítettünk egy  $t_k$   $k$  magasságú fát, melyre  $\text{Sel}(t_k)$  éppen  $L$  legfeljebb  $k$  hosszú szavait tartalmazza, akkor  $t_{k+1}$  legyen a következő. Ha  $u = u'a$  egy  $(k+1)$  hosszú szó (azaz  $u$  utolsó betűje  $a \in T$ ), akkor  $u' \in \text{Pre}(u)$ , és így  $u' \in L$  miatt van  $t_k$ -nak egy levele, melyhez tartozó szelektor  $u'$ . Legyen ennek egy gyermekéhez vezető él címkéje  $a$ , így ennek a gyermeknek a szelektora éppen  $u$  lesz. Ha  $k$  elég nagy,  $\text{Sel}(t_k) = L$ .

Tehát létezik egy kölcsönösen egyértelmű megfeleltetés az  $r$ -áris fák és  $\{0, 1, \dots, r-1\}^*$  prefixképzésre zárt véges részhalmazai között:

$$t \leftrightarrow \text{Sel}(t) \subseteq \{0, 1, \dots, r-1\}^*.$$

$\mathcal{L}_{\text{Bin}} = \{L; L \subseteq \{0, 1\}^* \wedge L \text{ zárt a prefix képzésre}\}$

$\mathcal{L}_{\text{Bin}}$  (bináris fák nyelvcsaládja) rekurzív definíciója:

- $\emptyset \in \mathcal{L}_{\text{Bin}}$
- $L_1, L_2 \in \mathcal{L}_{\text{Bin}}$ , akkor  $\{\varepsilon\} \cup 0L_1 \cup 1L_2 \in \mathcal{L}_{\text{Bin}}$

Ugyanis  $L = \emptyset$  megfelel az üres fának. Tegyük fel, hogy minden  $n$ -nél kisebb magasságú fát már definiáltunk. Ha egy  $t$   $n$  magasságú fa baloldali  $t_1$  illetve jobboldali  $t_2$  részfájának szelektorait az  $L_1$  illetve  $L_2$  nyelv írja le, akkor  $\{\varepsilon\} \cup 0L_1 \cup 1L_2$  írja le a  $t$  fa szelektorait.

**Teknőcgrafika:** Toll a papír felett, valamilyen irányban áll.

- $\langle F, d \rangle$  az adott pontból az adott irányba  $d$  hosszú vonalat húz véghelyzet: ahol a vonal végetér, irány változatlan
- $\langle f, d \rangle$  mint előbb, de nem húz vonalat
- $\langle +, \delta \rangle$   $\delta$  szöggel balra fordul (óramutató járásával ellentétesen)
- $\langle -, \delta \rangle$   $\delta$  szöggel jobbra fordul

Ha  $d$  és  $\delta$  rögzített, nem írjuk ki. Pl.  $d = 1, \delta = 90^\circ$ .

Kezdőirány: vízszintes, kezdőpont: origó.

#### Homomorfizmus

$h: T_1^* \rightarrow T_2^*$  homomorfizmus, ha  $h(t_1 \cdots t_n) = h(t_1) \cdots h(t_n)$  minden  $t_1 \cdots t_n \in T_1^*$  szóra.

(A konkatenációtartás miatt tehát elég a betűk képét megadni.)

Álljon az  $L_{\text{Koch}}$  nyelv a következő szavakból:

$\omega_0 = F-F-F-F$  négyzet

Legyen a  $h: (F, f, +, -)^* \rightarrow (F, f, +, -)^*$  homomorfizmus a következő:

$h(F) = F-F+F+FF-F-F+F$ ,  $h(f) = f$ ,  $h(+)$   $= +$ ,  $h(-)$   $= -$ .

$L_{\text{Koch}}$  további szavai (Koch szigetek):  $\omega_1 = h(\omega_0)$ ,  $\omega_2 = h^2(\omega_0)$ , ...

Pl.  $\omega_1 = h(F)h(-)h(F)h(-)h(F)h(-)h(F) = h(F)-h(F)-h(F)-h(F) = F-F+F+FF-F-F+F-F-F+F+FF-F-F+F-F-F+F+FF-F-F+F$ .

$L_1 \cap L_2$  **metszet**

$L_1 \cup L_2$  **unió**

$L_1 \setminus L_2$  **különbség**

Ezeket, mint halmazműveleteket értelmezzük.

$L_1 L_2 = \{uv \mid u \in L_1, v \in L_2\}$  a két nyelv **konkatenációja**.

$L^0 = \{\varepsilon\}$ ,  $L^1 = L$ ,  $L^k = L^{k-1}L$ .

$L^* = \bigcup_{i=0}^{\infty} L^i$ : az  $L$  nyelv **lezártja**.

$L^+ = \bigcup_{i=1}^{\infty} L^i$ .

Egy  $u = t_1 \cdots t_k$ ,  $t_i \in T$ , ( $1 \leq i \leq k$ ) szó **megfordítása**  $u^{-1} = t_k \cdots t_1$ .

Egy  $L$  nyelv **megfordítása**  $L^{-1} = \{u^{-1} \mid u \in L\}$ .

## Műveletek nyelvek között

### Példák

Példa:  $T = \{a, b\}$

$L_1 = \{a^n b^n \mid n \geq 0\}$

$L_2 = \{a^{2n+1} b \mid n \geq 0\}$

$L_3 = \{a^{2^n} \mid n \geq 0\}$

1.  $L_1 \cup L_2 = ? \{x \mid (x = a^n b^n \wedge n \geq 0) \vee (x = a^{2^{n+1}} b \wedge n \geq 1)\}$ ,
2.  $L_1 \cap L_2 = ? \{ab\}$ ,
3.  $L_1 L_2 = ? \{a^n b^n a^{2^{k+1}} b \mid n \geq 0 \wedge k \geq 0\}$ ,
4.  $L_2 \cap L_3 = ? \emptyset$ ,
5.  $L_1 \cap L_3 = ? \{\varepsilon\}$ ,
6.  $L_2^* = ? \{a^{2^{k_1+1}} b \dots a^{2^{k_n+1}} b \mid n, k_1, \dots, k_n \geq 0\}$ .

## Házi feladat

1. Írjunk programot, mely kirajzolja  $n = 4$ -ig

a. a Koch szigeteket

b. az  $L = \{\omega_0, \omega_1, \dots\}$  nyelv szavait, ahol:

$\omega_0 = F + F + F + F$ ,  $\omega_{i+1} = h(\omega_i)$ , ahol

$h(F) = F + f - FF + F + FF + Ff + FF - f + FF - F - FF - Ff - FFF$

$h(f) = fffff$ ,  $h(+)=+$ ,  $h(-)=-$ .

2.  $(L_1 L_2)^{-1} = L_2^{-1} L_1^{-1}$ .
3.  $x$  palindrom  $\Leftrightarrow x^k$  palindrom ( $k \geq 1$ ).  
( $x$  palindrom, ha  $x = x^{-1}$ )