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 szelektorhalmazukkal, faosztályok leírása nyelveken értelmezett rekurzióval

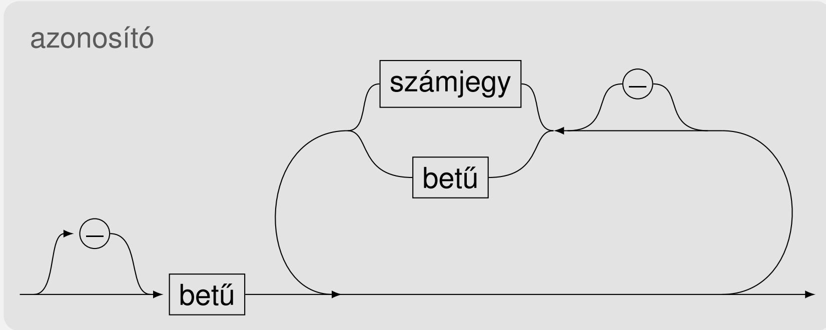
2008/09 I. félév

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)

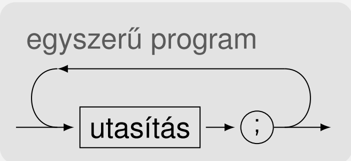


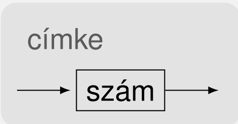
⟨azonosító⟩ ::={ |_}⟨betű⟩@{{ |_}{⟨betű⟩ | ⟨számjegy⟩}}

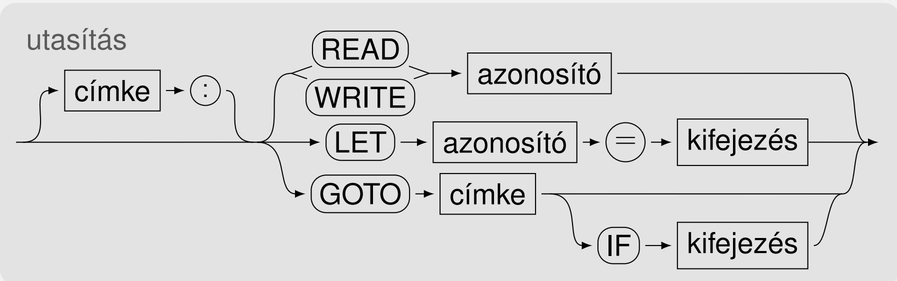
Házi feladatok megoldása

2. feladat

EP teljes átírása szintaxis gráfokkal (folytatás).



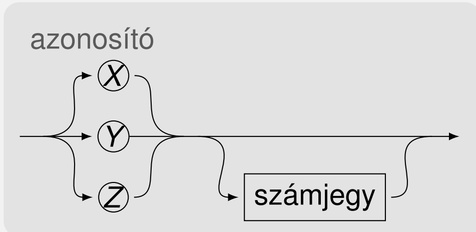


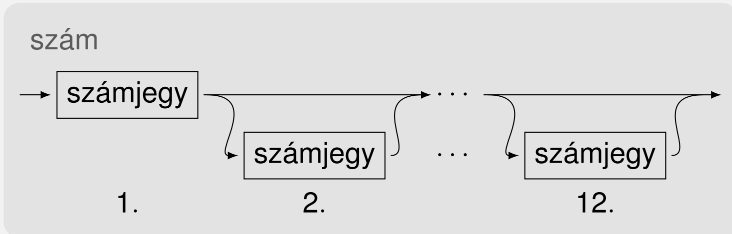


Házi feladatok megoldása

2. feladat

EP teljes átírása szintaxis gráfokkal (folytatás).

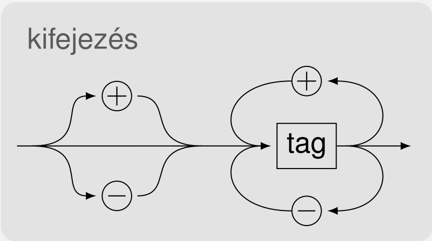


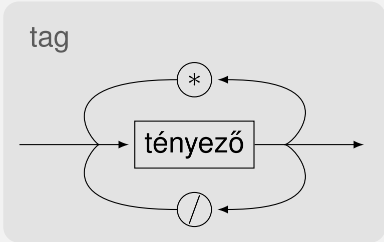


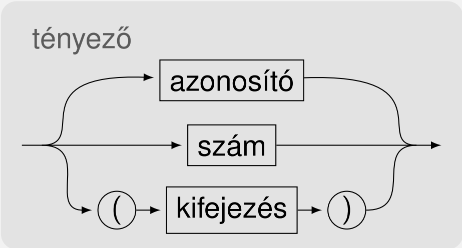
Házi feladatok megoldása

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:

⟨kifejezés⟩ ::= ⟨egyszerű kifejezés⟩{ |⟨relop⟩⟨egyszerű kifejezés⟩}
⟨relop⟩ ::= = | < | > | <> | <= | >= | IN
⟨egyszerű kifejezés⟩ ::={ | + | −}⟨tag⟩@{⟨addop⟩⟨tag⟩}
⟨addop⟩ ::= + | − | OR
⟨tag⟩ ::= ⟨tényező⟩@{⟨mpop⟩⟨tényező⟩}
⟨mpop⟩ ::= ∗ | / | DIV | MOD | AND
⟨tényező⟩ ::=⟨előjel nélküli konstans⟩|⟨változó⟩|
⟨függvénynév⟩{ |(⟨kifejezés⟩@{,⟨kifejezés⟩ })}|
(⟨kifejezés⟩) |NOT ⟨tényező⟩|[ ]|
[⟨kifejezés⟩{ |..⟨kifejezés⟩}@{,⟨kifejezés⟩{ |..⟨kifejezés⟩}}]
és így tovább . . .

Házi feladatok megoldása

4. feladat

Kifejezések fogalmának leírása W nyelvtannal.

Megoldás:

Metaszabályok:

⟨$\hat{X}$ kifejezés⟩ ::=⟨$\hat{X}$ tag⟩@{⟨$\hat{X}$ addop⟩⟨$\hat{X}$ tag⟩}
⟨$\hat{X}$ tag⟩ ::=⟨$\hat{X}$ tényező⟩@{⟨$\hat{X}$ mpop⟩⟨$\hat{X}$ tényező⟩}
⟨$\hat{X}$ tényező⟩ ::=⟨$\hat{X}$⟩|(⟨$\hat{X}$ kifejezés⟩)|⟨azonosító⟩

Hiperszabály:

$\hat{X}$:= egész | valós | Boole

⟨Boole addop⟩ ::= OR
⟨Boole mpop⟩ ::= AND
⟨egész addop⟩ ::= + | −
⟨egész mpop⟩ ::= ∗ | /

Szavak, nyelvek, konkatenáció

Ha H egy halmaz, jelölje H^* a H elemeiből képezhető összes véges sorozatok halmazát (beleértve az ε-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	**ábécé**
$t \in T$	a T ábécé egy **betűje**
$u \in T^*$	a T ábécé feletti **szó**
$L \in 2^{T^*}$ (vagyis $L \subseteq T^*$)	a T ábécé feletti **nyelv**
$\mathcal{L} \subseteq \bigcup_{T \subset U, \|T\| < \infty} 2^{T^*}$	**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 $\Leftrightarrow$ $\exists v' \in T^*\ u = v'v$

$\mathrm{pre}(u, \ell)$ ill. $\mathrm{suf}(u, \ell)$ az u ℓ hosszú prefixe ill. suffixe.
v valódi prefix ill. suffix, ha $v \neq \varepsilon, u$.

$\mathrm{Pre}(u)$=$\{v;\ v$ prefixe u-nak$\}$ $\qquad$ $\mathrm{Pre}(L)$=$\bigcup\limits_{u \in L} \mathrm{Pre}(u)$

$\mathrm{Suf}(u)$=$\{v;\ v$ suffixe u-nak$\}$ $\qquad$ $\mathrm{Suf}(L)$=$\bigcup\limits_{u \in L} \mathrm{Suf}(u)$

Azt mondjuk, hogy L zárt a prefix illetve suffix képzésre, ha $\mathrm{Pre}(L) = L$ illetve $\mathrm{Suf}(L) = L$ teljesül.

Például: $\mathrm{suf}(abbaba, 4) = baba$,
$\mathrm{Pre}(abbaa) = \{\varepsilon, a, ab, abb, abba, abbaa\}$,
$\mathrm{Suf}(\{ab, abb, bb\}) = \{\varepsilon, b, ab, bb, abb\}$

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.

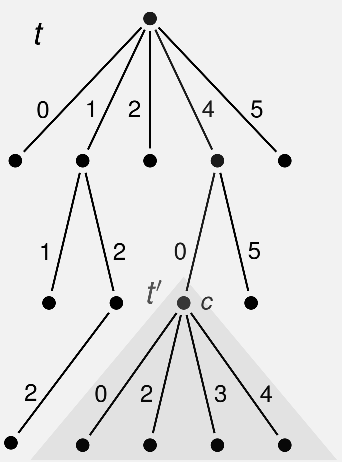
Szelektor: a $\{0, 1, \dots, r-1\}^*$ halmaz egy eleme, azaz elemi szelektorok egy véges sorozata.

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 ω szelektor?
ω =40

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

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

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

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 $\mathrm{Sel}(t) = L$. 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 $\mathrm{Sel}(t_k)$ éppen L legfeljebb k hosszú szavait tartalmazza, akkor t_{k+1} legyen a következõ. Ha $u = u'a$ egy $(k+1)$ hosszú szó (azaz u utolsó betűje $a \in T$), akkor $u' \in \mathrm{Pre}(u)$, 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, $\mathrm{Sel}(t_k) = L$.

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

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

Bináris fák nyelvcsaládja

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

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

1. $\emptyset \in \mathcal{L}_{\text{Bin}}$

2. $L_1, L_2 \in \mathcal{L}_{\text{Bin}}$, 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
Szabályok

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
$\langle +, \delta\rangle$	δ szöggel balra fordul (óramutató járásával ellentétesen)
$\langle -, \delta\rangle$	δ szöggel jobbra fordul

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

Kezdőirány: vízszintes, kezdőpont: origó.

Teknőcgrafika
Képek, mint a Koch nyelv szavai

Homomorfizmus

$h : T_1^* \to T_2^*$ 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 betûk képét megadni.)

Álljon az L_{Koch} nyelv a következő szavakból:

$\omega_0 = F{-}F{-}F{-}F$ négyzet

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

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

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

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

Műveletek nyelvek között
Unió, metszet, konkatenáció, ...

$L_1 \cap L_2$ **metszet**

$L_1 \cup L_2$ **unió**

$L_1 \setminus L_2$ **különbség**

Ezeket, mint halmazműveleteket értelmezzük.

$L_1L_2 = \{uv \mid u \in L_1, v \in L_2\}$ 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 $u = t_1 \cdots t_k,\ t_i \in T,\ (1 \le i \le k)$ 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: $T = \{a, b\}$
$L_1 = \{a^n b^n \mid n \geq 0\}$
$L_2 = \{a^{2n+1} b \mid n \geq 0\}$
$L_3 = \{a^{2n} \mid n \geq 0\}$

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

Házi feladat

1. Írjunk programot, mely kirajzolja $n = 4$-ig
  a. a Koch szigeteket
  b. az $L = \{\omega_0, \omega_1 \ldots\}$ nyelv szavait, ahol:
  $\omega_0 = F{+}F{+}F{+}F$, $\omega_{i+1} = h(\omega_i)$, ahol
  $h(F) = F{+}f{-}FF{+}F{+}FF{+}Ff{+}FF{-}f{+}FF{-}F{-}FF{-}Ff{-}FFF$
  $h(f) = ffffff$, $h(+) = +$, $h(-) = -$.

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 $x = x^{-1}$)