Nyelvek generálása nyelvtanokkal

Formális nyelvek, 4. gyakorlat

Célja: A formális nyelvek leírására használatos eszközök bemutatása példákon keresztül, különös tekintettel a nyelvtanokra.

Fogalmak: Listázás, felsorolás, parciális és teljes eldöntés, logikai formula, reguláris kifejezés, Turing-gép, formális nyelvtan, levezetés (közvetlen, közvetett), generált nyelv.

Feladatok jellege: Néhány véges nyelv konkrét megadása, T^* -ot felsoroló algoritmus, parciális és totális eldöntő algoritmus valamilyen számhalmazt reprezentáló nyelvre, konkrét logikai formula és reguláris kifejezés, $L = \{uu, u \text{ eleme } T^*\}$-ra Turing-gép, egy egyszerű nyelvtanban levezetés, közvetett levezetés, elfogadott nyelv.

2008/09 I. félév

Házi feladatok megoldása

1. feladat

$L_1 = \{a^n b^n \mid n \geq 0\}$, $L_2 = \{ab\}$. $L_2^* \overset{?}{\subseteq} L_1^*$.

Megoldás:

Igen, hiszen $L_2 \subseteq L_1$ és egy múlt órai feladat miatt ekkor $L_2^* \subseteq L_1^*$.

Házi feladatok megoldása

2. feladat

$L^* = L^* L^*$

Megoldás:

"$\subseteq$"

Világos, hiszen $\varepsilon \in L^*$ és $\{\varepsilon\}L^* = L^*$.

"$\supseteq$"

Legyen $w \in L^* L^*$, ekkor $w = uv,\ u \in L^*,\ v \in L^*$.

Definíció szerint $u = u_1 u_2 \dots u_k$ és $v = v_1 v_2 \dots v_\ell$, valamely k, ℓ nemnegatív egész számokra, ahol $u_1, \dots, u_k, v_1, \dots, v_\ell \in L$.

De ez azt jelenti, hogy u és v konkatenációja L^*-ban van.

Házi feladatok megoldása

3. feladat

$(L^*)^* = L^*$.

Megoldás:

"$\supseteq$"

$L^* = (L^*)^1 \subseteq \bigcup_{i=0}^{\infty} (L^*)^i = (L^*)^*$.

"$\subseteq$"

Legyen $u \in (L^*)^*$. Ekkor $u = u_1 u_2 \dots u_k$ valamely k nemnegatív egész számra, ahol $u_1, \dots, u_k \in L^*$.

$\forall 1 \leq i \leq k$-ra $u_i = u_{i1} u_{i2} \dots u_{im_i}$ valamely m_i nemnegatív egész számokra, ahol $u_{i1}, \dots, u_{im_i} \in L$.

u összesen $\sum_{i=1}^{k} m_i$ darab L-beli szó konkatenációja, azaz L^*-beli.

Más megoldás: $(L^*)^i = L^*$ minden $i \geq 1$-re, az előző házi feladat miatt, tehát $(L^*)^* = \{\varepsilon\} \cup L^* \cup (L^*)^2 \cup \dots = L^*$.

Házi feladatok megoldása

4. feladat

$(L_1 \cup L_2)^* = (L_1^* L_2^*)^*$.

Megoldás:

"$\subseteq$"

$$\left.\begin{array}{l} L_1 \subseteq L_1^* \subseteq L_1^* L_2^* \\ L_2 \subseteq L_2^* \subseteq L_1^* L_2^* \end{array}\right\} \Rightarrow L_1 \cup L_2 \subseteq L_1^* L_2^*.$$

Tehát: $(L_1 \cup L_2)^* \subseteq (L_1^* L_2^*)^*$.

"$\supseteq$"

$(L_1^* L_2^*)^* \subseteq ((L_1 \cup L_2)^*(L_1 \cup L_2)^*)^* = ((L_1 \cup L_2)^*)^* = (L_1 \cup L_2)^*$.

(A 2. és 3. házi feladat eredményét felhasználva.)

Házi feladatok megoldása

5. feladat

$h(L^{-1}) \stackrel{?}{=} h(L)^{-1}$

Megoldás:

Nem igaz, legyen $L = a^*b$ és a homomorfizmus legyen $a \to a$, $b \to ba$.

$L^{-1} = ba^*$,

$h(L^{-1}) = ba^+$,

$h(L) = a^*ba$,

$h(L)^{-1} = aba^*$.

Házi feladatok megoldása

6. feladat

Adjunk reguláris kifejezést a legfeljebb 3 darab a*-t tartalmazó* $\{a, b\}^*$*-beli szavakra!*

Megoldás:

$b^* \cup b^*ab^* \cup b^*ab^*ab^* \cup b^*ab^*ab^*ab^*$.

Nyelvtanok

$G = \langle T, N, \mathcal{P}, S\rangle$

T a terminális jelek, N a nyelvtani jelek halmaza (minkettő véges halmaz és diszjunktak), $S \in N$ kezdőszimbólum, $\mathcal{P}$ véges szabályhalmaz.
$P \in \mathcal{P}$: $p \to q$, ahol $p, q \in (T \cup N)^*$, p tartalmaz nyelvtani jelet.

α mondatforma, ha $\alpha \in (T \cup N)^*$. u terminális szó, ha $u \in T^*$.

Az α mondatformából közvetlenül levezethető a β mondatforma, ha léteznek γ_1, γ_2 mondatformák és $p \to q \in \mathcal{P}$, hogy $\alpha = \gamma_1 p \gamma_2$ és $\beta = \gamma_1 q \gamma_2$. Jelölése: $\alpha \underset{G}{\to} \beta$.

Az α mondatformából közvetetten levezethető a β mondatforma, ha létezik $k \in \mathbb{N}$ és $\gamma_0, \gamma_1, \ldots, \gamma_k$ mondatformák, hogy $\alpha = \gamma_0, \beta = \gamma_k$ és minden $i \in [0, k-1]$ esetén $\gamma_i \underset{G}{\to} \gamma_{i+1}$. Jelölése: $\alpha \overset{*}{\underset{G}{\to}} \beta$, ($\alpha \overset{k}{\underset{G}{\to}} \beta$).

A G nyelvtan által generált nyelv: $L(G) = \{u \in T^* \mid S \overset{*}{\underset{G}{\to}} u\}$.

Példák nyelvtanokra

1. Feladat: $L(G) = ?$

a. $S \to aaS \mid a$,
$\{a^{2n+1} \mid n \in \mathbb{N}\}$

b. $S \to aSb \mid \varepsilon$,
$\{a^n b^n \mid n \in \mathbb{N}\}$

c. $S \to aA \mid Aa,\ A \to AAA \mid a \mid b$,
$\{v \in \{a,b\}^* \mid v = au \text{ vagy } v = ua \text{ és } 2 \not| \ell(u)\}$

d. $S \to ASB \mid \varepsilon,\ AB \to BA,\ BA \to AB,\ A \to a,\ B \to b$,
$\{u \in \{a,b\}^* \mid \ell_a(u) = \ell_b(u)\}$

e. $S \to ASA \mid BSB \mid \varepsilon,\ AB \to BA,\ BA \to AB,\ A \to a,\ B \to b$.
$\{u \in \{a,b\}^* \mid 2|\ell_a(u) \text{ és } 2|\ell_b(u)\}$

Példák nyelvtanokra

2. Feladat: Készítsünk nyelvtant az alábbi nyelvekhez!

a. $\{u \in \{0,1\}^* \mid u$ 1-gyel kezdődik és 00-ra végződik$\}$,
$S \to 1A,\ A \to 0A \mid 1A \mid 00$

b. $\{u \in \{0,1\}^* \mid u$ legalább 3 hosszúságú$\}$,
$S \to 0A_1 \mid 1A_1,\ A_1 \to 0A_2 \mid 1A_2,\ A_2 \to 0A_3 \mid 1A_3,\ A_3 \to 0A_3 \mid 1A_3 \mid \varepsilon$

c. $(\varepsilon \cup 1)10^*$,
$S \to 1A \mid 11A,\ A \to 0A \mid \varepsilon$

d. $\{a^n b^m c^n \mid n \geq 0, m \geq 2\}$,
$S \to aSc \mid bbA,\ A \to bA \mid \varepsilon$

e. $\{u \in \{a,b\}^* \mid u$ páros sok a-t és páratlan sok b-t tartalmaz$\}$.
$S \to ASA \mid BSB \mid B,\ AB \to BA,\ BA \to AB,\ A \to a,\ B \to b$

Számrendszereken alapuló nyelvek generálása 3-as típusú nyelvtannal

3. típusú nyelvtan: Csak "$A \to u$" és "$A \to uB$" alakú szabályok ($u \in T^*$)

L: 3-mal osztható decimális egészek, (nem állhat 0 az elején).

$G = \langle\{0,1,2,3,4,5,6,7,8,9\}, \{S, S_0, S_1, S_2\}, \mathcal{P}, S\rangle$

$$\mathcal{P}: \begin{array}{rcl} S & \to & 1S_1|2S_2|3S_0|\ldots|9S_0 \\ S_0 & \to & 0S_0|1S_1|2S_2|3S_0|\ldots|9S_0 \\ S_1 & \to & 0S_1|1S_2|2S_0|3S_1|\ldots|9S_1 \\ S_2 & \to & 0S_2|1S_0|2S_1|3S_2|\ldots|9S_2 \\ S_0 & \to & \varepsilon. \end{array}$$

Ez a nyelvtan épp L-et generálja, azaz $L(G) = L$.

Például: $S \to 7S_1 \to 71S_2 \to 718S_1 \to 7182S_0 \to 7182$.

Megjegyzés: ha az 1 illetve 2 maradékot adó számokat szeretnénk generálni, akkor egyszerűen $S_0 \to \varepsilon$ helyett $S_1 \to \varepsilon$-t illetve $S_2 \to \varepsilon$-t írhatunk.

Adott részszavakat nem tartalmazó nyelv generálása 3-as típusú nyelvtannal

$L = \{u \in \{a,b,c\}^*; a^3 \not\subseteq u\}$

$G = \langle\{a,b,c\}, \{S_\varepsilon, S_a, S_{\neg a}, S_{aa}\}, \mathcal{P}, S_\varepsilon\rangle$

$$\mathcal{P}: \begin{array}{rcl} S_\varepsilon & \to & aS_a|bS_{\neg a}|cS_{\neg a}|\varepsilon \\ S_a & \to & aS_{aa}|bS_{\neg a}|cS_{\neg a}|\varepsilon \\ S_{aa} & \to & bS_{\neg a}|cS_{\neg a}|\varepsilon \\ S_{\neg a} & \to & aS_a|bS_{\neg a}|cS_{\neg a}|\varepsilon. \end{array}$$

Ez a nyelvtan épp L-et generálja, azaz $L(G) = L$.

Például: $S_\varepsilon \to bS_{\neg a} \to baS_a \to baaS_{aa} \to baacS_{\neg a} \to baac$.

Négyzetszámhosszúságú szavakat generáló nyelvtan

Adjunk nyelvtant, mely a következő nyelvet generálja! $T = \{a\}$
$L = \{a^{n^2} \mid n \geq 0\}$.

Megoldás:

Ötlet: U^nV^n-ből V^nU^n-t csinálni $UV \rightarrow VU$ jellegű cserékkel n^2 lépés.

$S' \rightarrow \varepsilon \mid a \mid LSR$
$S \rightarrow XSY \mid YX$
$XY \rightarrow YaX$
$LY \rightarrow aL$
$XR \rightarrow Ra$
$L \rightarrow a$
$R \rightarrow a$
$aY \rightarrow Ya$
$Xa \rightarrow aX$

Házi feladat

1. Adjunk a következő nyelvet generáló 3. típusú nyelvtant! Azon M-áris számrendszerbeli számok, melyek d-vel osztva k maradékot adnak. (Nem állhat az elején 0.)
2. Adjunk a következő nyelvet generáló 3. típusú nyelvtant! $L = \{u \in \{a, b, c\}^* \mid ab, bc, ca \not\subseteq u\}$.
3. Adjunk olyan nyelvtant, mely az alábbi nyelvet generálja! $L = \{a^{2^n} \mid n \geq 0\}$.