Formális nyelvek és automaták

A második zárthelyi dolgozat megoldásai

2006. május 20.

A csoport

1. Feladat: Tekintsük a következő grammatikát:

$G = \langle \{a, b\}, \{S, A, B, C\}, \mathcal{P}, S \rangle$

$\mathcal{P}: \quad S \to AB \,|\, CC$
$\quad\quad A \to CC$
$\quad\quad B \to BC \,|\, a$
$\quad\quad C \to a \,|\, b$

A CYK algoritmus segítségével döntse el, hogy a nyelv tartalmazza-e a *baaba* szót!

Megoldás:

$\{S\}$				
$\{S\}$	$\{B\}$			
$\{S\}$	$\{B\}$	$\{S,B\}$		
$\{S,A\}$	$\{S,A,B\}$	$\{S,A,B\}$	$\{S,A\}$	
$\{C\}$	$\{B,C\}$	$\{B,C\}$	$\{C\}$	$\{B,C\}$
b	a	a	b	a

Mivel $S \in H_{1,1}$ ("piramis" csúcsa), ezért $baaba \in L(G)$.

2. Feladat: Készítsen az alábbi nyelvhez véges determinisztikus automatát (VDA-t)!

$L = \{u \in \{a, b, c\}^* \,|\, ac$ nem részszava az u szónak és $\ell_a(u) + \ell_c(u)$ páros$\}$

Megoldás:

I. mo.

		a	b	c
$\leftrightarrows$	$q_{\neg a0}$	q_{a1}	$q_{\neg a0}$	$q_{\neg a1}$
	$q_{\neg a1}$	q_{a0}	$q_{\neg a1}$	$q_{\neg a0}$
$\leftarrow$	q_{a0}	q_{a1}	$q_{\neg a0}$	q_{zs}
	q_{a1}	q_{a0}	$q_{\neg a1}$	q_{zs}
	q_{zs}	q_{zs}	q_{zs}	q_{zs}

q_{a0}, q_{a1}: az utolsó olvasott betű a, páros ill. páratlan az eddig olvasottra az $\ell_a(u) + \ell_c(u)$ összeg.
$q_{\neg a0}$, $q_{\neg a1}$: az utolsó olvasott betű NEM a, páros ill. páratlan az eddig olvasottra az $\ell_a(u) + \ell_c(u)$ összeg.
q_{zs}: zsákutcaállapot

II. mo. Az L_u maradéknyelvek $L_\varepsilon, L_c, L_a, L_{aa}, L_{ac}$ valamelyikével egyeznek meg attól függően, hogy u tartalmazza-e az ac résszót, a-ra végződik-e, és $\ell_a(u) + \ell_c(u)$ páros-e vagy páratlan. A maradéknyelvek segítségével konstruált automata izomorf a fentivel.

3. Feladat: A tanult algoritmusok segítségével adjon véges determinisztikus automatát (VDA-t), mely az alábbi grammatika által generált nyelvet fogadja el!

$G = \langle \{a, b\}, \{S, A, B, C\}, \mathcal{P}, S \rangle$

$\mathcal{P}: \quad S \to bS \,|\, bC$
$\quad\quad A \to bA \,|\, \varepsilon$
$\quad\quad B \to aC \,|\, bS \,|\, bA \,|\, bC \,|\, \varepsilon$
$\quad\quad C \to aB \,|\, aC \,|\, bA$

Megoldás:

NDA

		a	b
$\rightarrow$	S		S, C
$\leftarrow$	A		A
$\leftarrow$	B	C	S, A, C
	C	B, C	A

VDA

		a	b
$\rightarrow$	$\{S\}$	$\{\}$	$\{S, C\}$
	$\{\}$	$\{\}$	$\{\}$
	$\{S, C\}$	$\{B, C\}$	$\{S, A, C\}$
$\leftarrow$	$\{B, C\}$	$\{B, C\}$	$\{S, A, C\}$
$\leftarrow$	$\{S, A, C\}$	$\{B, C\}$	$\{S, A, C\}$

4. Feladat: Bizonyítsa be, hogy a következő nyelv szigorú típusa 2-es!

$L = \{u \in \{a, b, c\}^* \mid \ell_a(u) = \ell_b(u)$ és b betű után közvetlen nem állhat c betű$\}$

A szigorú típus bizonyításaként adjon a nyelvhez környezetfüggetlen (2-es típusú) grammatikát, valamint lássa be, hogy 3-as típusú grammatika nem adható hozzá!

Megoldás:

$G = \langle\{a, b, c\}, \{S_0, S, A, B\}, \mathcal{P}, S_0\rangle$

$\mathcal{P}$: $S_0 \rightarrow cS_0 \mid S$

$S \rightarrow ASB \mid BSA \mid SS \mid \varepsilon$

$B \rightarrow b$

$A \rightarrow Ac \mid a$

Myhill-Nerode-dal: $L_{a^k} \subseteq \{u \in \{a, b, c\}^* \mid \ell_b(u) - \ell_a(u) = k\}$. De ezek ∞ sok páronként diszjunkt halmaz. Tehát L-nek ∞ sok különböző maradéknyelve van. Tehát a Myhill-Nerode tétel szerint $L \notin \mathcal{L}_3$-beli.

Kis Bar-Hillel-lel: Indirekt. Ekkor $a^n b^n \in L$ első n betűjéből álló szónak (azaz a^n-nek) egy részszava (ami persze $a^j, 0 < j \leq n$ alakú) beiterálható (amennyiben n elég nagy, azaz nagyobb mint a Kis BH lemma nyelvfüggő konstansa). De ekkor a beiterált szavak $a^{n+(i-1)j} b^n$ alakúak, de ezek $i \neq 1$ esetén nem L-beliek, ami ellentmondás.

5. Feladat: Adjon veremautomatát az alábbi nyelvhez!

$L = \{u \in \{aa, b\}^* \mid \ell_a(u) = 2 \cdot \ell_b(u)\}$

Megoldás: $\mathcal{V} = \langle\{q_0, q_1, q_F\}, \{a, b\}, \{+, -, \#\}, \delta, q_0, \#, \{q_F\}\rangle$

$\delta(q_1, a, \sigma) = (q_0, +\sigma) \quad \sigma \in \{+, \#\}$

$\delta(q_1, a, -) = (q_0, \varepsilon)$

$\delta(q_0, b, \sigma) = (q_0, -\sigma) \quad \sigma \in \{-, \#\}$

$\delta(q_0, b, +) = (q_0, \varepsilon)$

$\delta(q_0, a, \sigma) = (q_1, \sigma) \quad \sigma \in \{+, -, \#\}$

$\delta(q_0, \varepsilon, \#) = (q_F, \#)$

Magyarázat: Minden olvasott aa pár után berakunk a verembe egy '+'-t vagy kitörlünk egy '–'-t. Minden olvasott b után berakunk a verembe egy '–'-t vagy kitörlünk egy '+'-t. A veremben vagy csak '+'-ok vagy csak '–'-ok vannak a #-en kívül, attól függően, hogy eddig aa-ból vagy b-ből volt több. A q_1 állapotba akkor jutunk, ha páratlanadik a-t olvasunk, de ekkor csak a olvasása esetén tudunk továbblépni.

B csoport

1. Feladat: Tekintsük a következő grammatikát:

$G = \langle\{a, b\}, \{S, A, B, U, V, W, X\}, \mathcal{P}, S\rangle$

$\mathcal{P}$: $S \rightarrow AU \mid BV$

$U \rightarrow AX \mid BS \mid b$

$V \rightarrow AS \mid BW \mid a$

$W \rightarrow VV \mid BU$

$X \rightarrow UU \mid AV$

$A \rightarrow a$

$B \rightarrow b$

A CYK algoritmus segítségével döntse el, hogy a nyelv tartalmazza-e az *abbaaa* szót!

Megoldás:

$\{X,W\}$					
$\{V\}$	$\{V\}$				
$\{S\}$	$\{S\}$	$\{W,X\}$			
$\{U\}$	$\{U\}$	$\{V\}$	$\{U\}$		
$\{S\}$	$\{W,X\}$	$\{S\}$	$\{X,W\}$	$\{X,W\}$	
$\{A,V\}$	$\{B,U\}$	$\{B,U\}$	$\{A,V\}$	$\{A,V\}$	$\{A,V\}$
a	b	b	a	a	a

Mivel $S \notin H_{1,1}$ (”piramis” csúcsa), ezért $abbaaa \notin L(G)$.

2. Feladat: Adjon az $(aba)^* \cup a^*$ reguláris kifejezéssel definiált nyelvhez véges determinisztikus automatát (VDA-t)!

Megoldás:

I. mo. εNDA

		a	b	ε
$\rightarrow$	1			2, 5
	2	3		6
	3		4	
	4	2		
	5	5		6
$\leftarrow$	6			

NDA

		a	b
$\leftrightarrows$	1	3, 5	
$\leftarrow$	2	3	
	3		4
	4	2	
$\leftarrow$	5	5	
$\leftarrow$	6		

VDA

		a	b
$\leftrightarrows$	$\{1\}$	$\{3,5\}$	$\{\}$
$\leftarrow$	$\{3,5\}$	$\{5\}$	$\{4\}$
	$\{\}$	$\{\}$	$\{\}$
$\leftarrow$	$\{5\}$	$\{5\}$	$\{\}$
	$\{4\}$	$\{2\}$	$\{\}$
$\leftarrow$	$\{2\}$	$\{3\}$	$\{\}$
	$\{3\}$	$\{\}$	$\{4\}$

II. mo. A különböző maradéknyelvek:
$L_\varepsilon = L = (aba)^* \cup a^*$, $L_a = ba(aba)^* \cup a^*$, $L_b = \emptyset$, $L_{aa} = a^*$, $L_{ab} = a(aba)^*$, $L_{aba} = (aba)^*$, $L_{abaa} = ba(aba)^*$. Ezek alapján a VDA:

		a	b
$\leftrightarrows$	L_ε	L_a	L_b
$\leftarrow$	L_a	L_{aa}	L_{ab}
	L_b	L_b	L_b
$\leftarrow$	L_{aa}	L_{aa}	L_b
	L_{ab}	L_{aba}	L_b
$\leftarrow$	L_{aba}	L_{abaa}	L_b
	L_{abaa}	L_b	L_{ab}

Megjegyzés: A kapott VDA izomorf az első megoldásban kapott VDA-val.

3. Feladat: A tanult algoritmussal adjon az alábbi VDA-val ekvivalens minimális állapotszámú automatát!

$\mathcal{A} = \langle \{1,2,3,4,5,6,7,8,9\}, \{a,b\}, \delta, 1, \{2,3,9\} \rangle$

		a	b
$\rightarrow$	1	4	6
$\leftarrow$	2	3	5
$\leftarrow$	3	2	5
	4	9	2
	5	2	3
	6	8	7
	7	8	1
	8	9	3
$\leftarrow$	9	9	9

Megoldás:

$\overset{0}{\sim}$: $\{1,4,5,6,7,8\}$ $\{2,3,9\}$

$\overset{1}{\sim}$: $\{1,6,7\}$ $\{4,5,8\}$ $\{2,3\}$ $\{9\}$

$\overset{2}{\sim}$: $\{1,6,7\}$ $\{4,8\}$ $\{5\}$ $\{2,3\}$ $\{9\}$

$\overset{3}{\sim} = \overset{2}{\sim} = \sim$

		a	b
$\rightarrow$	$q_{1,6,7}$	$q_{4,8}$	$q_{1,6,7}$
$\leftarrow$	$q_{2,3}$	$q_{2,3}$	q_5
	q_5	$q_{2,3}$	$q_{2,3}$
	$q_{4,8}$	q_9	$q_{2,3}$
$\leftarrow$	q_9	q_9	q_9

4. Feladat: Bizonyítsa be, hogy a következő nyelv szigorú típusa 2-es!
$L = \{u \in \{a,b\}^* \mid u = u^{-1}$ (palindróma) és $\ell_a(u)$ osztható 4-gyel$\}$
A szigorú típus bizonyításaként adjon a nyelvhez környezetfüggetlen (2-es típusú) grammatikát, valamint lássa be, hogy 3-as típusú grammatika nem adható hozzá!

Megoldás:
$G = \langle\{a,b\},\{S_0,S_1,S_2,S_3\},\mathcal{P},S_0\rangle$
$\mathcal{P}$: $S_0 \rightarrow aS_2a \mid bS_0b \mid b \mid \varepsilon$
$S_2 \rightarrow aS_0a \mid bS_2b$

Myhill-Nerode-dal: A palindróma tulajdonság miatt $ba^{2k} \in L_{a^{2\ell}} \Leftrightarrow k = \ell$. Így ezek a maradéknyelvek páronként különböznek. Mivel ∞ sok különböző maradéknyelv van, ezért $L \notin \mathcal{L}_3$.

Kis Bar-Hillel-lel: Indirekt. Ekkor $a^{2n}ba^{2n} \in L$ első n betűjéből álló szónak (azaz a^n-nek) egy részszava (ami persze $a^j, 0 < j \leq n$ alakú) beiterálható (ha n elég nagy, azaz nagyobb mint a Kis BH lemma nyelvfüggő konstansa). De ekkor a beiterált szavak $a^{2n+(i-1)j}ba^{2n}$ alakúak. Ezek $i \neq 1$ esetén nem L-beliek, ami ellentmondás.

5. Feladat: Adjon veremautomatát az alábbi nyelvhez!
$L = \{u \in \{x,y,z\}^* \mid$ az u szóban nincsen xy részszó és $\ell_x(u) = \ell_y(u)\}$

Megoldás: $\mathcal{V} = \langle\{q_0,q_x,q_y,q_z,q_F\},\{x,y,z\},\{+,-,\#\},\delta,q_0,\#,\{q_F\}\rangle$

$\delta(q_0,x,\#) = (q_x,+\#)$
$\delta(q_0,y,\#) = (q_y,-\#)$
$\delta(q_0,z,\#) = (q_z,\#)$
$\delta(q_t,x,\sigma) = (q_x,+\sigma)$ $\quad \sigma \in \{+,\#\}, t \in \{x,y,z\}$
$\delta(q_t,x,-) = (q_x,\varepsilon)$ $\quad t \in \{x,y,z\}$
$\delta(q_t,y,\sigma) = (q_y,-\sigma)$ $\quad \sigma \in \{-,\#\}, t \in \{y,z\}$
$\delta(q_t,y,+) = (q_y,\varepsilon)$ $\quad t \in \{y,z\}$
$\delta(q_t,z,\sigma) = (q_z,\sigma)$ $\quad \sigma \in \{+,-,\#\}, t \in \{x,y,z\}$
$\delta(q_t,\varepsilon,\#) = (q_F,\#)$

Magyarázat: Minden olvasott x után berakunk a verembe egy '+'-t vagy kitörlünk egy '–'-t. Minden olvasott y után berakunk a verembe egy '–'-t vagy kitörlünk egy '+'-t. A veremben vagy csak '+'-ok vagy csak '–'-ok vannak a #-en kívül, attól függően, hogy eddig aa-ból vagy b-ből volt több. Figyelünk arra, hogy ne lehessen q_x-ből ne lehessen y olvasására továbblépni (q_t : t volt az utolsó olvasott betű, $t \in \{x,y,z\}$).