Harmadik típusú nyelvek és véges automaták

Formális nyelvek, 10. gyakorlat

Célja: Az automata-nyelvekre vonatkozó tételek elmélyítése, gyakorlati alkalmazása

Fogalmak: reguláris kifejezés, Kleene-tétel, általánosított reguláris kifejezés, direkt szorzat automata, maradéknyelvek és tulajdonságaik, MYHILL-NERODE tétel, kis Bar-Hillel lemma

Feladatok jellege: Néhány reguláris és általánosított reguláris kifejezés felírása. Egy egyszerű automata Kleene-nyelveinek elkészítése. Automatakészítés a szimmetrikus differenciához valamely konkrét, nagyon egyszerű VDA-k esetén. Konkrét nyelvek maradéknyelvei halmazának felírása, ezekből MYHILL-NERODE alapján következtetések levonása. Konkrét nyelvről kis Bar-Hillel lemmával kimutatni, hogy nem reguláris.

2008/09 I. félév

Házi feladatok megoldása

1. feladat

Melyik nyelvet fogadja el a következő automata?

	a	b	c
$\rightarrow q_0$	q_1	q_2	q_3
q_1	q_4	q_2	q_4
q_2	q_0	q_4	q_3
$\leftarrow q_3$	q_4	q_3	q_4
q_4	q_4	q_4	q_4

Megoldás:

$(aba \cup ba)^*(abc \cup bc \cup c)b^*$

Házi feladatok megoldása

2. feladat

Készítsünk KMP automatát a következő mintához!
babbcabc $(T = \{a, b, c\})$

Megoldás:

	a	b	c
$\rightarrow q_\varepsilon$	q_ε	q_b	q_ε
q_b	q_{ba}	q_b	q_ε
q_{ba}	q_ε	q_{bab}	q_ε
q_{bab}	q_{ba}	q_{babb}	q_ε
q_{babb}	q_{ba}	q_b	q_{babbc}
q_{babbc}	q_{babbca}	q_b	q_ε
q_{babbca}	q_ε	$q_{babbcab}$	q_ε
$q_{babbcab}$	q_{ba}	q_b	q_F
$\leftarrow q_F$	q_F	q_F	q_F

VDA-hoz 3NF nyelvtan készítése

1. Feladat: Adjunk 3NF nyelvtant, mely ugyanezt a nyelvet generálja, amit az automata elfogad!

	a	b
$\rightarrow q_0$	q_0	q_2
q_1	q_1	q_4
q_2	q_1	q_3
$\leftarrow q_3$	q_0	q_4
$\leftarrow q_4$	q_2	q_3

Megoldás:

$G = \langle\{a, b\}, \{q_0, q_1, q_2, q_3, q_4\}, \mathcal{P}, q_0\rangle$

$q_0 \rightarrow aq_0 \mid bq_2$

$q_1 \rightarrow aq_1 \mid bq_4$

$q_2 \rightarrow aq_1 \mid bq_3$

$q_3 \rightarrow aq_0 \mid bq_4$

$q_4 \rightarrow aq_2 \mid bq_3$

$q_3 \rightarrow \varepsilon$

$q_4 \rightarrow \varepsilon$

VDA konstruálása 3.típusú nyelvtanhoz

1. lépés: Normálformára hozzuk a nyelvtant.

2. lépés: Az automata elkészítése

Legyen $\mathcal{G} = \langle T, N, \mathcal{P}, S \rangle$ 3NF nyelvtan.

Készítünk egy $\mathcal{A}$ VDA-t, melyre $L(\mathcal{A}) = L(\mathcal{G})$.

$\mathcal{A} = \langle 2^N, T, \delta, \{S\}, F \rangle$, ahol

$\delta(\{A_1, \ldots, A_k\}, t) = \{B \in N \mid \exists 1 \leq i \leq k, A_i \to tB \in \mathcal{P}\}$,

$F = \{X \in 2^N \mid X \cap \{A \in N \mid A \to \varepsilon \in \mathcal{P}\} \neq \emptyset\}$.

A gyakorlatban csak azon állapotokra határozzuk meg az átmeneteket, melyek a kezdőállapotból elérhető.

VDA konstruálása 3.típusú nyelvtanhoz
Példa

2. Feladat: Készítsünk VDA-t, mely ugyanazt a nyelvet fogadja el, melyet a nyelvtan generál!

$S \to abA \mid bS$
$A \to bbA \mid B \mid a \mid b$
$B \to S \mid abB$

Megoldás:

1.a.) lépés: Láncmentesítés

$S \to abA \mid bS$
$A \to bbA \mid abB \mid abA \mid bS \mid a \mid b$
$B \to abA \mid abB \mid bS$

1.b.) lépés: Hosszredukció

$S \to aK_1 \mid bS$
$A \to bK_1 \mid aK_2 \mid aK_1 \mid bS \mid a \mid b$
$B \to aK_1 \mid aK_2 \mid bS$
$K_1 \to bA$ $\quad K_2 \to bB$

1.c.) lépés: "$A \to a$" alakú szabályok eliminálása

$S \to aK_1 \mid bS$
$A \to bK_1 \mid aK_2 \mid aK_1 \mid bS \mid aF \mid bF$
$B \to aK_1 \mid aK_2 \mid bS$
$K_1 \to bA$
$K_2 \to bB$
$F \to \varepsilon$

VDA konstruálása 3.típusú nyelvtanhoz
Példa

$S \to aK_1 \mid bS$
$A \to bK_1 \mid aK_2 \mid aK_1 \mid bS \mid aF \mid bF$
$B \to aK_1 \mid aK_2 \mid bS$
$K_1 \to bA$
$K_2 \to bB$
$F \to \varepsilon$

		a	b
$\to$	$\{S\}$	$\{K_1\}$	$\{S\}$
	$\{K_1\}$	$\{\}$	$\{A\}$
	$\{\}$	$\{\}$	$\{\}$
	$\{A\}$	$\{K_1, K_2, F\}$	$\{K_1, S, F\}$
$\leftarrow$	$\{K_1, K_2, F\}$	$\{\}$	$\{A, B\}$
$\leftarrow$	$\{K_1, S, F\}$	$\{K_1\}$	$\{A, S\}$
	$\{A, B\}$	$\{K_1, K_2, F\}$	$\{K_1, S, F\}$
	$\{A, S\}$	$\{K_1, K_2, F\}$	$\{K_1, S, F\}$

A VDA-k éppen az $\mathcal{L}_3$-beli nyelveket ismerik fel.

Minimális automata készítése

Cél: Adott automatához minél kevesebb állapotú, az eredeti automata által felismert nyelvet elfogadó automatát megadni.

Minimális automata: minimális állapotszámú ilyen automata.

1) Összefüggővé alakítás:
$H_0 = \{q_0\}$.
$H_{i+1} = H_i \cup \{q \in Q \mid \exists q' \in H_i, t \in T, q = \delta(q', t)\}$.
Tartalmazásra nézve monoton növő, felülről korlátos halmazsorozat, így stabilizálódik, azaz $\exists i$, hogy $H_i = H_{i+1} =: H$.
Hagyjuk el a $Q \setminus H$-beli állapotokat és ezekre vonatkozó állapotátmeneteket.

2) Redukció:
Ha a q és q' állapotokból ugyanazon szavak hatására kerül az automata végállapotba, akkor az elfogadás szempontjából mindegy, hogy a működés során a q vagy a q' állapotba kerül az automata. Célunk az ilyen állapotok összevonása.

Redukció
Ekvivalens állapotok

Az $q \in Q$ állapotra vonatkozó maradék nyelv:
$L(\mathcal{A}, q) := \{v \mid \delta(q, v) \in F\}$.

Legyenek $q, q' \in Q$ állapotok. q és q' ekvivalensek, ha
$L(\mathcal{A}, q) = L(\mathcal{A}, q')$. Jelölése: $q \sim q'$.

$\sim$ ekvivalencia- és jobbkongruenciareláció és analóg módon kiterjeszthető két automata állapotai közötti relációvá.

$\mathcal{A}, \mathcal{A}'$ automaták (q_0 és q_0' kezdőállapotokkal) ekvivalensek, (jelölésben $\mathcal{A} \sim \mathcal{A}'$) ha $q_0 \sim q_0'$.

Az $\mathcal{A}$ automata $\mathcal{A}/_\sim$ faktorautomatájája ekvivalens az eredetivel, redukált (nincsenek különböző ekvivalens állapotai), továbbá izomorfia erejéig az egyetlen összefüggő, redukált $\mathcal{A}$-val ekvivalens automata.

Állapotok ekvivalenciájának algoritmikus eldöntése

$q \sim q' \;\Leftrightarrow\; L(\mathcal{A}, q) = L(\mathcal{A}, q') \;\Leftrightarrow\; \forall u \in T^* : \big(\delta(q, u) \in F \;\Leftrightarrow\; \delta(q', u) \in F\big)$.

Azt mondjuk, hogy $q \overset{i}{\sim} q'$ (q i-ekvivalens q'-vel), ha minden $u \in T^{\leq i}$ esetén $\big(\delta(q, u) \in F \;\Leftrightarrow\; \delta(q', u) \in F\big)$ $(i \geq 0)$.

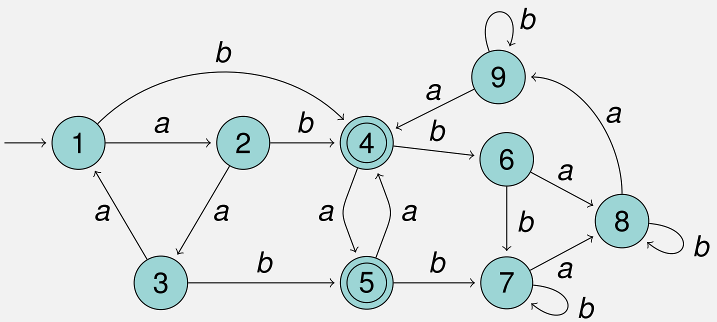
- $\overset{i}{\sim}$ ekvivalenciareláció,
- $q \overset{0}{\sim} q'$, ha ($q \in F \Leftrightarrow q' \in F$),
- minden $q, q' \in Q$-ra
  $q \overset{i+1}{\sim} q' \;\Leftrightarrow\; q \overset{i}{\sim} q' \wedge (\forall t \in T : \delta(q, t) \overset{i}{\sim} \delta(q', t))$,
- $\overset{0}{\sim} \prec \overset{1}{\sim} \prec \overset{2}{\sim} \prec \cdots \prec \sim$, továbbá $q \sim q' \;\Leftrightarrow\; \forall i \geq 0 : q \overset{i}{\sim} q'$.
  ($\varrho_1 \prec \varrho_2$, ha minden $q, q' \in Q$ esetén $q \varrho_2 q' \Rightarrow q \varrho_1 q'$.)

$i_0 := \min\{i \mid \overset{i}{\sim} = \overset{i+1}{\sim}\}$.
Ekkor $i_0 \leq |Q| - 1$, és így $\sim = \overset{|Q|-1}{\sim}$.

Redukálás

5. Feladat:

Redukáljuk a következő automatát!



Redukálás

Megoldás:

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

	a	b
4	{4,5}	{1,2,3,6,7,8,9}
5	{4,5}	{1,2,3,6,7,8,9}

	a	b
1	{1,2,3,6,7,8,9}	{4,5}
2	{1,2,3,6,7,8,9}	{4,5}
3	{1,2,3,6,7,8,9}	{4,5}
6	{1,2,3,6,7,8,9}	{1,2,3,6,7,8,9}
7	{1,2,3,6,7,8,9}	{1,2,3,6,7,8,9}
8	{1,2,3,6,7,8,9}	{1,2,3,6,7,8,9}
9	{4,5}	{1,2,3,6,7,8,9}

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

	a	b
4	{4,5}	{6,7,8}
5	{4,5}	{6,7,8}

	a	b
1	{1,2,3}	{4,5}
2	{1,2,3}	{4,5}
3	{1,2,3}	{4,5}

	a	b
6	{6,7,8}	{6,7,8}
7	{6,7,8}	{6,7,8}
8	{6,7,8}	{9}

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

Redukálás

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

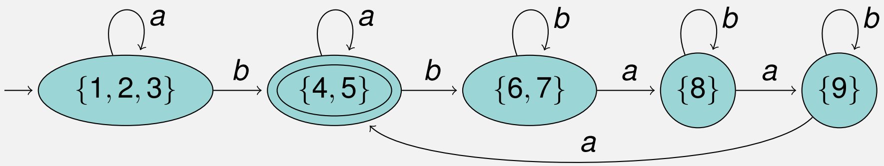
	a	b
4	{4,5}	{6,7}
5	{4,5}	{6,7}

	a	b
1	{1,2,3}	{4,5}
2	{1,2,3}	{4,5}
3	{1,2,3}	{4,5}

	a	b
6	{8}	{6,7}
7	{8}	{6,7}

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

A redukált automata:



Az átmenetek és elfogadó állapotok meghatározásához tetszőleges reprezentánst tekinthetünk. Az eredeti kezdőállapotot tartalmazó ekvivalenciaosztály lesz az új kezdőállapot.

$\mathcal{L}_3$-beli-e egy nyelv?

Egy L nyelv $p \in T^*$-ra vonatkozó maradéknyelve $L_p := \{v \mid pv \in L\}$.

Myhill-Nerode tétel

$L \in \mathcal{L}_3$ akkor és csak akkor, ha $\left|\{L_p\}_{p \in T^*}\right| < \infty$, ahol $T = T(L)$ az L nyelv ábécéje.

Kis Bar-Hillel lemma

Minden $L \in \mathcal{L}_3$ nyelvhez van olyan $n = n(L) \in \mathbb{N}$ nyelvfüggő konstans, hogy minden $u \in L$, $\ell(u) \geq n$ szó esetén van u-nak olyan $u = xyz$ felbontása ($x, y, z \in T(L)^*$), melyre

- $\ell(xy) \leq n$,
- $\ell(y) > 0$,
- minden $i \geq 0$ esetén $xy^iz \in L$.

Myhill-Nerode tétel

Maradéknyelvek

Határozzuk meg a maradéknyelveit az alábbi nyelveknek!

1. $L = \{a, abb, bb, b\}$
2. $L = \{0,1\}^*00 \cup \{0\}$
3. HE (helyes zárójelezések nyelve)

Megoldások:

1.feladat

$L_\varepsilon = L, L_a = \{\varepsilon, bb\}, L_b = \{\varepsilon, b\}, L_{ab} = \{b\}, L_{bb} = \{\varepsilon\}, L_{abb} = \{\varepsilon\}$.
$L_u = \emptyset \quad \forall u \notin \mathrm{Pre}(L)$.

A maradéknyelvek halmaza tehát:
$\{\emptyset, \{\varepsilon\}, \{b\}, \{\varepsilon, b\}, \{\varepsilon, bb\}, \{a, abb, bb, b\}\}$

Myhill-Nerode tétel

Maradéknyelvek

2.feladat

$$L_u = \begin{cases} L \setminus \{0\} & 1 \in \mathrm{Suf}(u) \\ L & u = \varepsilon \vee 0 \in \mathrm{Suf}(u) \wedge 00 \notin \mathrm{Suf}(u) \wedge u \neq 0 \\ L \cup \{\varepsilon\} & u = 0 \vee 00 \in \mathrm{Suf}(u) \end{cases}$$

3.feladat Legyen $k \in \mathbb{N}$-re:

$$\mathrm{HE}_k^P := \{u \in \{(,)\}^* \mid \ell_((u) - \ell_)(u) = k \wedge \ell_((v) \geq \ell_)(v), \forall v \in \mathrm{Pre}(u)\}$$

$$\mathrm{HE}_k^S := \{u \in \{(,)\}^* \mid \ell_)(u) - \ell_((u) = k \wedge \ell_)(v) \geq \ell_((v), \forall v \in \mathrm{Suf}(u)\}$$

Ekkor $$L_u = \begin{cases} \mathrm{HE}_k^S & u \in \mathrm{HE}_k^P \quad (k \in \mathbb{N}) \\ \emptyset & u \notin \bigcup_{k=0}^{\infty} \mathrm{HE}_k^P \end{cases}$$

Myhill-Nerode tétele alapján mivel az első két nyelvnek véges sok (6 illetve 3) maradéknyelve van, míg a harmadiknak végtelen, ezért az első két nyelv $\mathcal{L}_3$-beli, a helyes zárójelezések nyelve viszont nem.

Automata készítése maradéknyelvek segítségével

A maradéknyelvek segítségével készíthető 3. típusú nyelvekhez VDA:

$\mathcal{A} = \langle \{L_p\}_{p \in T^*}, T, \delta, L_\varepsilon, \{L_p \mid \varepsilon \in L_p\} \rangle$, ahol $\delta(L_p, t) = L_{pt}$.

1.feladat

	a	b
$\rightarrow \{a, abb, bb, b\}$	$\{\varepsilon, bb\}$	$\{\varepsilon, b\}$
$\leftarrow \{\varepsilon, bb\}$	$\emptyset$	$\{b\}$
$\leftarrow \{\varepsilon, b\}$	$\emptyset$	$\{\varepsilon\}$
$\emptyset$	$\emptyset$	$\emptyset$
$\{b\}$	$\emptyset$	$\{\varepsilon\}$
$\leftarrow \{\varepsilon\}$	$\emptyset$	$\emptyset$

2.feladat

	0	1
$\rightarrow L$	$L \cup \{\varepsilon\}$	$L \setminus \{0\}$
$\leftarrow L \cup \{\varepsilon\}$	$L \cup \{\varepsilon\}$	$L \setminus \{0\}$
$L \setminus \{0\}$	L	$L \setminus \{0\}$

Megjegyzés: az így kapott VDA minimális automata is lesz.

Kis Bar-Hillel Lemma

Legyen L_{Kif} a csak az a változót tartalmazó helyes kifejezések nyelve. (Kifejezések: az 1. gyakorlaton tanultak szerint.)

Feladat: $L_{\text{Kif}} \overset{?}{\in} \mathcal{L}_3$

Megoldás:

A Kis Bar-Hillel lemma segítségével bebizonyítjuk, hogy $L_{\text{Kif}} \notin \mathcal{L}_3$.

Elegendő belátni, hogy L_{Kif} nem rendelkezik a Kis Bar-Hillel lemmában leírt, minden $\mathcal{L}_3$-beli nyelvre igaz tulajdonsággal, azaz:

Minden $n \in \mathbb{N}$-re van olyan $u_n \in L_{\text{Kif}}$, $\ell(u_n) \geq n$ szó, hogy minden $u_n = xyz$, $\ell(xy) \leq n$, $\ell(y) > 0$ felbontás esetén létezik olyan $i \in \mathbb{N}$, melyre $xy^iz \notin L_{\text{Kif}}$.

Kis Bar-Hillel Lemma

Tehát minden n-re válasszunk egy u_n kifejezést. Legyen mondjuk $u_n = ({}^n a)^n$. Ekkor nyílván $u_n \in L_{\text{Kif}}$.

Azt kell belátni, hogy u_n n hosszúságú prefixének semelyik nemüres részszava sem iterálható be. De egy ilyen részszó mindenképp $y = ({}^d$ alakú, ahol $d > 0$. Ha y beiterálható lenne akkor a Kis Bar-Hillel lemma szerint $({}^{n-d-k}\{({}^d\}^i({}^k a)^n = ({}^{n+(i-1)d} a)^n$ (a képletben "{" és "}" metazárójelek!) is L_{Kif} eleme lenne minden i természetes számra. De ez $i \neq 1$ esetén nem igaz, mert nem ugyanannyi "(" és ")" szerepel benne.

Házi feladat

1. Készítsünk VDA-t a következő nyelvtanhoz!
   $S \rightarrow acA \mid bB \mid \varepsilon$
   $A \rightarrow B \mid b \mid C$
   $B \rightarrow S \mid abB \mid a$
   $C \rightarrow acbC \mid B$
2. Készítsünk a következő automatával ekvivalens *minimális állapotszámú* véges determinisztikus automatát!

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

3. Határozzuk meg a palindromák nyelvének ($L = \{u \in T^* \mid u = u^{-1}\}$) maradéknyelveit! ($T$ tetszőleges.)
4. Bizonyítsuk be, hogy a palindromák nyelve nem $\mathcal{L}_3$-beli a Myhill-Nerode tétel illetve a Kis Bar-Hillel lemma segítségével! ($|T| \geq 2$)