Veremautomaták

Formális nyelvek, 12. gyakorlat

Célja: A környezet-független nyelvek használatával kapcsolatos alapfeladatok begyakorlása

Fogalmak: Szintaxis-fa, legbal és legjobb levezetés, nagy Bar-Hillel lemma, felülről-lefelé és alulról-felfelé elemzés, LL(k), LR(k) nyelvtanok, verem-automaták.

Feladatok jellege: Néhány szintaxis-fa egy konkrét 2. típusú nyelvtanban. Kicsit bonyolultabb nyelvtan esetében (adott szóhoz) a felülről-lefelé és az alulról-felfelé elemzés bemutatása. Konkrét nyelvtanra az LL, LR tulajdonság deteketálása, illetve a nem teljesülés kimutatása. Nagy Bar-Hillel lemma alkalmazása konkrét nyelvre. 1 verem építése a kifejezésekhez, kettő verem a dadogós nyelvhez.

2008/09 I. félév

Házi feladatok megoldása

1. feladat

Oldjuk meg a következő egyenletrendszert!
$a^*b^*X \cup (ba)^*bY \cup b^* = X$
$ba^*X \cup bY \cup a^*b^* = Y$

Megoldás:

$X = (a^*b^*)^*(b^* \cup (ba)^*bY)$.

A második egyenletbe helyettesítve:

$(ba^*(a^*b^*)^*(ba)^*b \cup b)Y \cup (ba^*(a^*b^*)^*b^* \cup a^*b^*) = Y$, amiből

$Y = (ba^*(a^*b^*)^*(ba)^*b \cup b)^*(ba^*(a^*b^*)^*b^* \cup a^*b^*)$.

Hasonlóan:

$X = (a^*b^* \cup (ba)^*b^2b^*a^*)^*((ba)^*bb^*a^*b^* \cup b^*)$.

Házi feladatok megoldása

2. feladat

Határozzuk meg reguláris kifejezéssel az alábbi véges determinisztikus automata által elfogadott nyelvet!

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

Megoldás:

$X = aY \cup bZ$
$Y = aV \cup \varepsilon$
$Z = aV$
$V = aX \cup bV \cup \varepsilon$

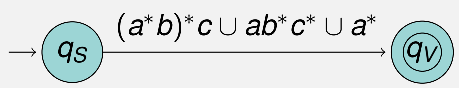
$V = b^*(aX \cup \varepsilon)$,
$Y = ab^*(aX \cup \varepsilon) \cup \varepsilon$,
$X = a(ab^*(aX \cup \varepsilon) \cup \varepsilon) \cup bab^*(aX \cup \varepsilon)$,
$X = (a \cup b)ab^*aX \cup (a \cup (a \cup b)ab^*)$,
$X = ((a \cup b)ab^*a)^*(a \cup (a \cup b)ab^*)$.

Házi feladatok megoldása

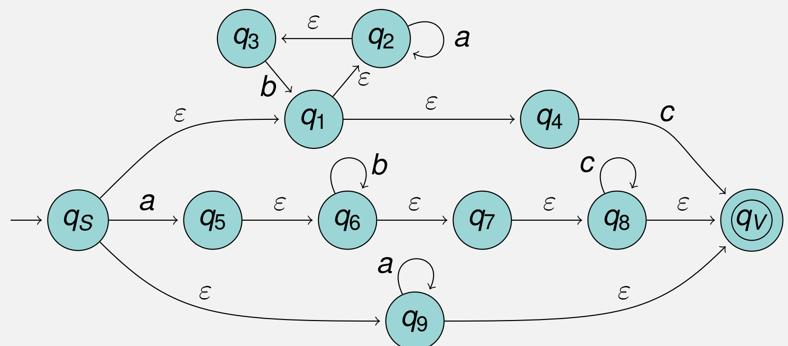
3. feladat

Készítsünk VDA-t a következő reguláris kifejezéshez!
$(a^*b)^*c \cup ab^*c^* \cup a^*$.

Megoldás:



$\Downarrow$ (lebontással)



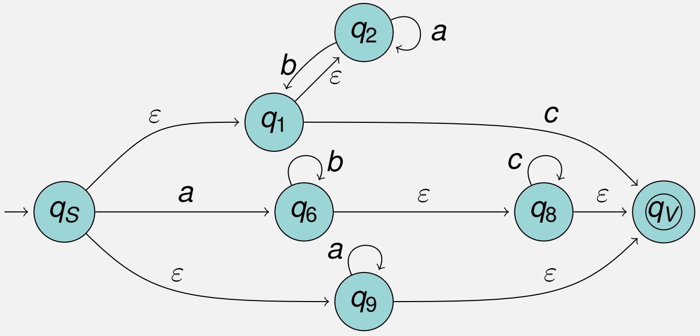
Házi feladatok megoldása

3. feladat

Készítsünk VDA-t a következő reguláris kifejezéshez!
$(a^*b)^*c \cup ab^*c^* \cup a^*$.

Megoldás:

$\Downarrow$ (egyszerűsítve)

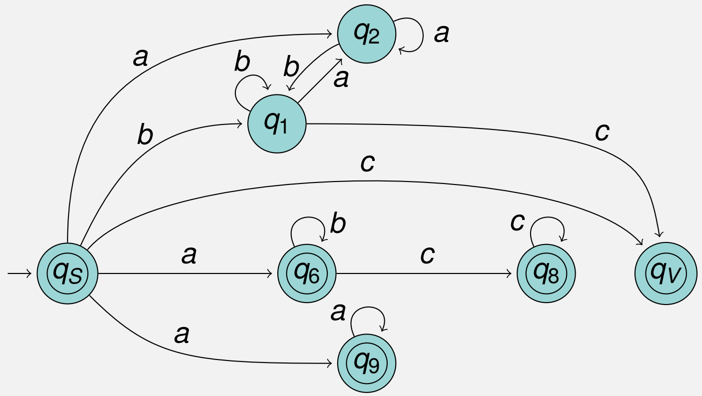


Házi feladatok megoldása

3. feladat

Készítsünk VDA-t a következő reguláris kifejezéshez!
$(a^*b)^*c \cup ab^*c^* \cup a^*$.

Megoldás: $\Downarrow$ ($\varepsilon NDA \rightarrow NDA$)



Házi feladatok megoldása

3. feladat

Készítsünk VDA-t a következő reguláris kifejezéshez!
$(a^*b)^*c \cup ab^*c^* \cup a^*$.

Megoldás: VDA

		a	b	c
$\leftrightarrows$	$\{q_S\}$	$\{q_2, q_6, q_9\}$	$\{q_1\}$	$\{q_V\}$
$\leftarrow$	$\{q_2, q_6, q_9\}$	$\{q_2, q_9\}$	$\{q_1, q_6\}$	$\{q_8\}$
	$\{q_1\}$	$\{q_2\}$	$\{q_1\}$	$\{q_V\}$
$\leftarrow$	$\{q_V\}$	$\{\}$	$\{\}$	$\{\}$
$\leftarrow$	$\{q_2, q_9\}$	$\{q_2, q_9\}$	$\{q_1\}$	$\{\}$
$\leftarrow$	$\{q_1, q_6\}$	$\{q_2\}$	$\{q_1, q_6\}$	$\{q_8, q_V\}$
$\leftarrow$	$\{q_8\}$	$\{\}$	$\{\}$	$\{q_8\}$
	$\{q_2\}$	$\{q_2\}$	$\{q_1\}$	$\{\}$
	$\{\}$	$\{\}$	$\{\}$	$\{\}$
$\leftarrow$	$\{q_8, q_V\}$	$\{\}$	$\{\}$	$\{q_8\}$

Veremautomaták

Veremautomata (1-verem) alatt a következő 7-est értjük:
$\mathcal{V} = \langle Q, T, \Sigma, \delta, q_0, \sigma_0, F \rangle$, ahol

Q	az állapotok (véges) halmaza
T	egy ábécé, a bemenő ábécé
Σ	a verem ábécéje
δ	állapotátmeneti függvény, $\delta : Q \times (T \cup \{\varepsilon\}) \times \Sigma \rightarrow 2^{Q \times \Sigma^*}$
$q_0 \in Q$	kezdőállapot
$\sigma_0 \in \Sigma$	a verem kezdőszimbóluma
$F \subseteq Q$	a végállapotok halmaza.

A veremautomata egy ütemben kiolvassa a központi egység állapotát, az input szó aktuális szimbólumát és a verem tetőelemét, ennek függvényében új állapotba kerül, a verem tetőelemét felülírja egy vagy több jellel (azaz egy szóval), az input szó következő betűjére áll az olvasófej (kivéve ε-mozgás) és a tetőmutató az új tetőelemre áll.

Veremautomaták
Elfogadás, determinisztikus veremautomata

$\Delta(q,u,\alpha)$ álljon azon állapot-veremtartalom párokból, melyeket az $u \in (T \cup \{\varepsilon\})^*$ inputsorozat végigolvasása után kaphatunk, ha kezdetben a verem tartalma α és az állapot q, azaz rekurzívan:
1. $\Delta(q,t,\sigma\beta) = \{(q',\tau\beta) \mid (q',\tau) \in \delta(q,t,\sigma)\}$
2. $\Delta(q,vt,\alpha) = \{(q'',\tau) \mid \tau = \beta\gamma, (q',\sigma\gamma) \in \Delta(q,v,\alpha), (q'',\beta) \in \delta(q',t,\sigma)\}$.

$(\tau,\alpha,\beta,\gamma \in \Sigma^*, \sigma \in \Sigma, u,v \in T^*, t \in T, q,q',q'' \in Q)$

Tehát a verem tartalmát egy $\tau \in \Sigma^*$ szó reprezentálja, a verem tetőmutatója a szó első betűjére mutat.

$\mathcal{V}$ végállapottal elfogad egy u szót, ha $\{q \in Q \mid \exists \beta \in \Sigma^*, (q,\beta) \in \Delta(q_0,u,\sigma_0)\} \cap F \neq \emptyset$.

$\mathcal{V}$ üres veremmel elfogad egy u szót, ha $\exists\, q \in Q$, hogy $(q,\varepsilon) \in \Delta(q_0,u,\sigma_0)$.

Veremautomaták
Példa

1. Feladat

Készítsünk végállapottal elfogadó veremautomatát a következő L nyelvhez! $L = \{u \in \{a,b,c\}^* \mid u = wcw^{-1}, w \in \{a,b\}^+\}$

Megoldás

$\mathcal{V} = \langle \{q_0,q_1,q_2,q_3\}, \{a,b,c\}, \{\#,a,b\}, \delta, q_0, \#, \{q_3\} \rangle$.

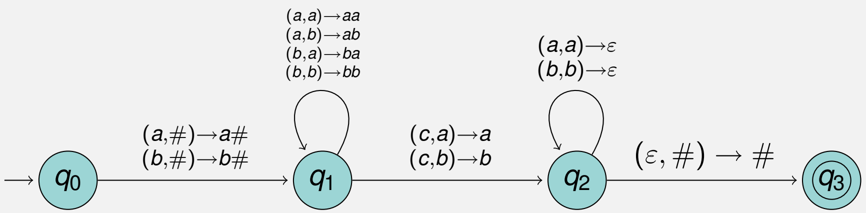
$\delta(q_0,t,\#) = (q_1,t\#)$	$\forall t \in \{a,b\}$
$\delta(q_1,t_1,t_2) = (q_1,t_1t_2)$	$\forall t_1,t_2 \in \{a,b\}$
$\delta(q_1,c,t) = (q_2,t)$	$\forall t \in \{a,b\}$
$\delta(q_2,t,t) = (q_2,\varepsilon)$	$\forall t \in \{a,b\}$
$\delta(q_2,\varepsilon,\#) = (q_3,\#)$	

Determinisztikus veremautomata: olyan veremautomata, melyre
- $\forall q \in Q, \sigma \in \Sigma, t \in T \cup \{\varepsilon\} : |\delta(q,t,\sigma)| \leq 1$.
- $\forall q \in Q, \sigma \in \Sigma : |\delta(q,\varepsilon,\sigma)| \neq 0 \implies \forall t \in T : |\delta(q,t,\sigma)| = 0$.

A fenti veremautomata determinisztikus.

Veremautomaták
Példa

Az 1. Feladat megoldása átmenetdiagrammal:



Tehát a veremből *egy* σ betűt kivenni a $\delta(q,t,\sigma) = (q',\varepsilon)$, a verem tartalmát változatlanul hagyni a $\delta(q,t,\sigma) = (q',\sigma)$, egy σ' betűt betenni a $\delta(q,t,\sigma) = (q',\sigma'\sigma)$, a tetőelemet felülírni egy tetszőleges $\tau \in \Sigma^*$ szóval $\delta(q,t,\sigma) = (q',\tau)$ szabály megadásával lehet.

Megjegyzés: Nemdeterminisztikus veremautomata esetén $\delta(q,t,\sigma) = (q',\tau)$, valójában azt jelenti, hogy $(q',\tau) \in \delta(q,t,\sigma)$, e helyett $\delta(q,t,\sigma)$-t többértékűen adjuk meg.

Veremautomaták
Példa

2. Feladat Készítsünk végállapottal elfogadó veremautomatát a következő L nyelvhez! $L = \{u \in \{a,b\}^* \mid u = ww^{-1}, w \in \{a,b\}^+\}$

Megoldás

$\mathcal{V} = \langle \{q_0,q_1,q_2,q_3\}, \{a,b\}, \{\#,a,b\}, \delta, q_0, \#, \{q_3\} \rangle$.

$\delta(q_0,t,\#) = (q_1,t\#)$	$\forall t \in \{a,b\}$
$\delta(q_1,t_1,t_2) = (q_1,t_1t_2)$	$\forall t_1,t_2 \in \{a,b\}$
$\delta(q_1,t,t) = (q_2,\varepsilon)$	$\forall t \in \{a,b\}$
$\delta(q_2,t,t) = (q_2,\varepsilon)$	$\forall t \in \{a,b\}$
$\delta(q_2,\varepsilon,\#) = (q_3,\#)$	

Itt tehát $\delta(q_1,a,a)$-nak és $\delta(q_1,b,b)$-nek két értéke van, a veremautomata nemdeterminisztikus.

Veremautomaták

Példa

3. Feladat

Készítsünk üres veremmel elfogadó veremautomatát a csak az a változót tartalmazó helyes kifejezések nyelvéhez!

Megoldás

$\mathcal{V} = \langle \{q_0, q_1\}, \{a, +, -, *, /, (,)\}, \{\#, (\}, \delta, q_0, \#, \ \rangle.$

$$\begin{array}{ll}
\delta(q_0, a, \sigma) = (q_1, \sigma) & \forall \sigma \in \{\#, (\} \\
\delta(q_0, (, \sigma) = (q_0, (\sigma) & \forall \sigma \in \{\#, (\} \\
\delta(q_1, t, \sigma) = (q_0, \sigma) & \forall \sigma \in \{\#, (\}, t \in \{+, -, *, /\} \\
\delta(q_1,), () = (q_1, \varepsilon) & \\
\delta(q_1, \varepsilon, \#) = (q_1, \varepsilon) &
\end{array}$$

$\delta(q_1, \varepsilon, \#)$ és $\delta(q_1, -, \#)$ sem üres, tehát a második feltétel nem teljesül, azaz a veremautomata nemdeterminisztikus.

Veremautomaták

Példa

4. Feladat

Készítsünk üres veremmel elfogadó veremautomatát az $L = \{u \in \{a, b\}^* \mid \ell_a(u) = \ell_b(u)\}$ nyelvhez!

Megoldás

$\mathcal{V} = \langle \{q_0\}, \{a, b\}, \{\#, +, -\}, \delta, q_0, \#, \ \rangle.$

$$\begin{array}{ll}
\delta(q_0, a, \sigma) = (q_0, +\sigma) & \forall \sigma \in \{\#, +\} \\
\delta(q_0, a, -) = (q_0, \varepsilon) & \\
\delta(q_0, b, \sigma) = (q_0, -\sigma) & \forall \sigma \in \{\#, -\} \\
\delta(q_0, b, +) = (q_0, \varepsilon) & \\
\delta(q_0, \varepsilon, \#) = (q_0, \varepsilon) &
\end{array}$$

$\delta(q_0, \varepsilon, \#)$ és $\delta(q_0, a, \#)$ sem üres, tehát a második feltétel nem teljesül, azaz a veremautomata nemdeterminisztikus.

Veremautomaták

Összefoglaló

A determinisztikus veremautomaták által elfogadott nyelvek osztálya valódi részhalmaza a veremautomaták által elfogadott nyelvek osztályának.

A végállapottal és az üres veremmel elfogadó veremautomaták által elfogadható nyelvek osztálya megegyezik (bármely veremautomatához készíthető egy másik típusú vele ekvivalens veremautomata).

A veremautomaták által elfogadott nyelvek osztálya megegyezik a 2-es típusú nyelvtanok által generált nyelvek osztályával ($\mathcal{L}_2$-vel).

2-vermek: Ezek már minden $\mathcal{L}_0$-beli nyelvet el tudnak fogadni, azaz az 1-vermekhez képest már két osztálynyi az ugrás.

Nagy Bar-Hillel lemma

Szükséges feltétel egy nyelv 2. típusba tartozására

Nagy Bar-Hillel-lemma

Minden $L \in \mathcal{L}_2$ esetén léteznek $p, q > 0$ nyelvfüggő egész konstansok ($p = p(L), q = q(L)$), amelyekre ha $u \in L$, és $\ell(u) > p$, akkor u-nak létezik $u = xyzvw$ felbontása, ahol $\ell(yv) > 0$, $\ell(yzv) \leq q$ és minden $i \geq 0$ egészre $xy^izv^iw \in L$.

Kevésbé formálisan a lényeget a következõképpen fejezhetjük ki: L minden elég hosszú szavában van két, egymáshoz közel lévõ, nem triviális, párhuzamosan beiterálható részszó.

Nagy Bar-Hillel lemma

5. Feladat:
$L = \{a^n b^n a^n \mid n \in \mathbb{N}\} \overset{?}{\in} \mathcal{L}_2$

Megoldás:

Nem. Indirekt, tegyük fel, hogy $L \in \mathcal{L}_2$. Ekkor a Nagy Bar-Hillel lemma szerint léteznek a nyelvfüggő p és q konstansok. Legyen $M = \max\{p, q\}$. Tekintsük az $u = a^M b^M a^M$ szót.

Mivel $\ell(u) > M \geq p$, ezért a Nagy Bar-Hillel lemma szerint létezik az u-nak $u = xyzvw$ felbontása, ahol $\ell(yv) > 0$, $\ell(yzv) \leq q \leq M = \ell(a^M) = \ell(b^M)$ és $\{xy^i zv^i w \mid i \geq 0\} \subseteq L$. Tehát vagy x, vagy w tartalmazza a^M-t részszóként. Tegyük fel, hogy x az (a másik eset teljesen analóg).

Vizsgáljuk meg, milyen szavakat kapunk y és v (a kettő közül legalább az egyik nem az üres szó) párhuzamos beiterálása során. A kapott szavak $\{xy^i zv^i w \mid i \geq 0\}$.

Nagy Bar-Hillel lemma

5. Feladat:
$L = \{a^n b^n a^n \mid n \in \mathbb{N}\} \overset{?}{\in} \mathcal{L}_2$

Megoldás:(folytatás)

Ha y vagy v valamelyike $b^k a^l$ alakú, ahol $k, l > 0$, akkor a beiterálás során olyan szavakat kapnánk amelyek felváltva a-ból és b-ből álló blokkokat tartalmaznak. Ha $i \geq 2$, akkor ezek a szavak nem lesznek L-beliek.

Ha viszont y és v a^k vagy b^k alakú (legalább az egyik kitevő pozitív), akkor az iterációval olyan szavakat kapunk, melyek $a^M b^{M_1} a^{M_2}$ alakúak. Így viszont $i \geq 2$-re $\max\{M_1, M_2\} > M$, azaz a kapott szó ez esetben sem L-beli, tehát a kezdeti, indirekt feltevésünk volt hamis.

Nagy Bar-Hillel lemma

6. Feladat:
$L = \{a^{n^2} \mid n \in \mathbb{N}\} \overset{?}{\in} \mathcal{L}_2$

Megoldás: Nem. Indirekt, tegyük fel, hogy $L \in \mathcal{L}_2$. Ekkor a Nagy Bar-Hillel lemma szerint léteznek a nyelvfüggő p és q konstansok. Legyen $M = \max\{p, q\}$. Tekintsük az $u = a^{M^2}$ szót.

Mivel $\ell(u) > M \geq p$, ezért a Nagy Bar-Hillel lemma szerint létezik az u-nak $u = xyzvw$ felbontása, ahol $K := \ell(yv) > 0$, $\ell(yzv) \leq q \leq M$. $xy^i zv^i w = a^{M^2 + (i-1)K}$. Mivel egy növekvő számtani sorozatban biztosan van nem négyzetszám, ezért a Nagy Bar-Hillel lemma feltétele nem teljesül, tehát $L \notin \mathcal{L}_2$.

2. típusú nyelvtanok feletti szintaxisfák

A szintaxisfa definíciója

Legyen $G = \langle T, N, \mathcal{P}, S \rangle$ tetszőleges 2-es típusú nyelvtan. A t nemüres fát G feletti szintaxisfának nevezzük, ha:

1) Pontjai $T \cup N \cup \{\varepsilon\}$ elemeivel vannak címkézve.
2) Belső pontjai N elemeivel vannak címkézve.
3) Ha egy belső pont címkéje X, a közvetlen leszármazottjainak címkéi pedig balról jobbra olvasva $X_1, X_2, \ldots, X_k$, akkor $X \to X_1 X_2 \ldots X_k \in \mathcal{P}$.
4) Az ε-nal címkézett pontoknak nincs testvére.

Szintaxisfákkal a levezetések szerkezetét ábrázoljuk. Jelölje egy adott t szintaxisfa leveleinek balról jobbra való összeolvasását $\mathrm{front}(t)$, a fa gyökerét pedig $\mathrm{gy}(t)$.

2. *típusú nyelvtanok feletti szintaxisfák*

Legbal és legjobb levezetések

Legbal illetve legjobb levezetés: Ha a levezetés folyamán valamely pozición (a mondatforma i. betûjén) helyettesítés történik, akkor a korábbi $(1, ..., i-1)$ illetve későbbi $(i+1, i+2, \ldots)$ pozíciókat a levezetés már nem érinti.

Legbal (legjobb) mondatforma: Valamely $L(G)$-beli szó legbal (legjobb) levezetése során elõforduló mondatforma.

Elemzés: u egy szintaxisfájának elkészítése, azaz amelyre $\mathrm{gy}(t) = S$, és $\mathrm{front}(t) = u$.

2. *típusú nyelvtanok feletti szintaxisfák*

Egyértelmű nyelv(tan)

A szóprobléma eldöntésének a szintaxisfa konstrukcióján alapuló módszere jól használható a programnyelvek elemzéséhez, ugyanis az eljárás így az elemzendõ szó szerkezetét is megadja.

Egyértelmű nyelvtan: minden $u \in L(G)$-nek pontosan egy szintaxisfája létezik.

Egyértelmű nyelv: Létezik 2. típusú egyértelmű nyelvtan, ami generálja.

Lényegesen nem egyértelmű nyelv: Ha nem létezik 2. típusú egyértelmű nyelvtan, ami generálja.

2. *típusú nyelvtanok feletti szintaxisfák*

7. Feladat:
$G = \langle\{a, b, c\}, \{S\}, \{S \to SaS \mid SbS \mid c\}, S\rangle$

a) Mutassunk példát egy legalább 7 hosszúságú szó legjobb levezetésére!

b) Egyértelmű-e a nyelvtan?

c) Egyeértelmű-e az $L(G)$ nyelv?

Megoldás:

a) $S \to SaS \to SaSbS \to SaSbSbS \to SaSbSbc \to SaSbcbc$
$\to Sacbcbc \to cacbcbc$

b) Nem, például $cacbc$. ($S \to SaS \to SaSbS$ és $S \to SbS \to SaSbS$ levezetéskezdetekhez más szintaxisfa tartozik.)

c) Igen, például:
$G = \langle\{a, b, c\}, \{S, A\}, \{S \to A \mid AaS,\ A \to c \mid cbA\}, S\rangle$.