Számítási modellek

2. előadás

Reguláris kifejezések

Motiváció

Mintaillesztés reguláris kifejezések alapján:

- Vírusbejegyzések pásztázása.
- Természetes nyelvfeldolgozás.
- Strukturális szövegkiemelés szövegszerkesztőkben.
- Programnyelvek specifikációja.
- Információ elérése digitális könyvtárakban.
- Genomika. (genom, mint karakterlánc)
- Szöveg szűrése (spam, malware).
- Adatbeviteli mezők (dátum, e-mail, URL, hitelkártya) érvényesítése.

Reguláris kifejezések

Definíció

Legyenek V és $V' = \{\varepsilon, \cdot, +, *, (,)\}$ diszjunkt ábécék. A V ábécé feletti **reguláris kifejezéseket** rekurzív módon a következőképpen definiáljuk:

1. $\emptyset$ és ε reguláris kifejezés V felett,
2. minden $a \in V$ reguláris kifejezés V felett,
3. Ha R reguláris kifejezés V felett, akkor R^* is reguláris kifejezés V felett,
4. Ha Q és R reguláris kifejezések V felett, akkor $(Q \cdot R)$ és $(Q + R)$ is reguláris kifejezések V felett

Reguláris kifejezések

Minden egyes reguláris kifejezés egy-egy reguláris nyelvet reprezentál, melyet az alábbi rekurzióval definiálunk.

Definíció

1. $\emptyset$ az $\emptyset$ nyelvet reprezentálja,
2. ε az $\{\varepsilon\}$ nyelvet reprezentálja,
3. a $(\in V)$ az $\{a\}$ nyelvet reprezentálja,
4. ha R az L nyelvet reprezentálja, akkor R^* az L^* nyelvet reprezentálja,
5. ha R_1 az L_1 és R_2 az L_2 nyelvet reprezentálja, akkor $(Q \cdot R)$ az $L_1 L_2$ nyelvet, míg $(Q + R)$ az $L_1 \cup L_2$ nyelvet reprezentálja.

Azaz a $*, \cdot$ és $+$ szimbólumok jelölik rendre a lezárt, a konkatenáció és az unió műveleteket.

Reguláris kifejezések

Legyen P, Q és R egy-egy reguláris kifejezés. Ekkor fennállnak a következő egyenlőségek:

1. $P + (Q + R) = (P + Q) + R$
2. $P \cdot (Q \cdot R) = (P \cdot Q) \cdot R$
3. $P + Q = Q + P$
4. $P \cdot (Q + R) = P \cdot Q + P \cdot R$
5. $(P + Q) \cdot R = P \cdot R + Q \cdot R$
6. $P^* = \varepsilon + P \cdot P^*$
7. $\varepsilon \cdot P = P \cdot \varepsilon = P$
8. $P^* = (\varepsilon + P)^*$

A műveletek precedenciasorrendje $*, \cdot, +$, ennek és az asszociatív szabályok figyelembevételével bizonyos zárójelpárok elhagyhatók.

A reguláris kifejezések kifejező ereje

Példa:

Az $(a+b)a^*$ és $aa^* + ba^*$ reguláris kifejezések által reprezentált nyelv ugyanaz. $\{aa^n \mid n \in \mathbb{N}\} \cup \{ba^n \mid n \in \mathbb{N}\}$.

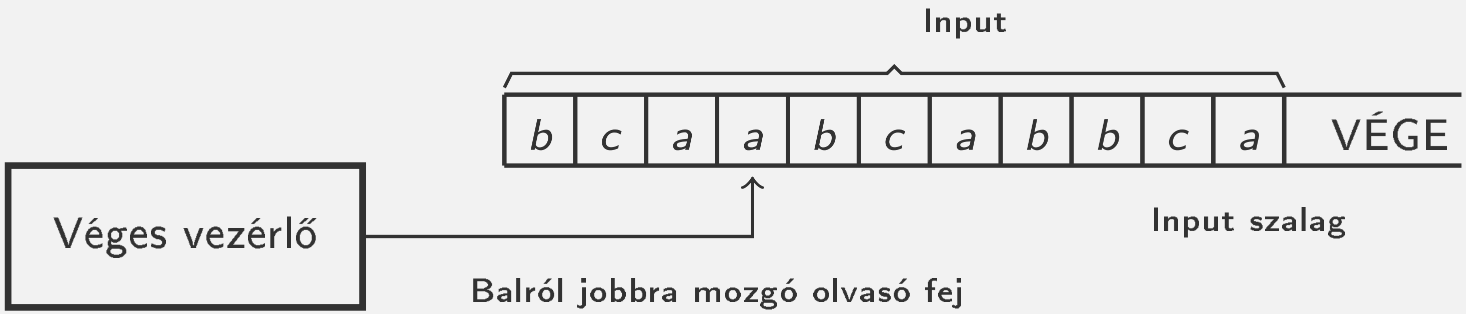
Másrészt az $a + ba^*$ által reprezentált nyelv $\{a, b, ba, ba^2, ba^3, \ldots\}$.

Tétel

- Minden reguláris kifejezés egy reguláris (3-típusú) nyelvet reprezentál.
- Mnden reguláris (3-típusú) nyelvhez megadható egy, ezen nyelvet reprezentáló reguláris kifejezés.

Megjegyzés: Mivel $\mathcal{L}_3$ nem csak a reguláris műveletekre (Kleene-lezárt, konkatenáció unió), hanem a metszet, különbség, komplementer műveletekre is zárt ($\mathcal{L}_2, \mathcal{L}_1, \mathcal{L}_0$ csak a reguláris műveletekre) ezért a gyakorlatban gyakran ezen műveletekkel kiterjesztett reguláris kifejezéseket használnak.

Véges automata



- Formális nyelvek azonosítása nemcsak generatív eszközökkel, hanem felismerő eszközökkel is lehetséges. Ilyenek az automaták, amelyek szavak feldolgozására és azonosítására alkalmasak.
- A grammatikák szintetizáló, míg az automaták analitikus megközelítést alkalmaznak.
- Az automata egy szó feldolgozása után kétféleképpen viselkedhet, vagy elfogadja (igen), vagy elutasítja (nem).

Véges automata

- A véges automata diszkrét időintervallumokban végrehajtott lépések sorozatán keresztül működik.
- A véges automata a kezdőállapotból vagy kezdeti állapotból indul, az inputszó az inputszalagon helyezkedik el, az olvasófej pedig az inputszó legbaloldalibb szimbólumán áll.
- Az automata, miután elolvasott egy szimbólumot, az olvasófejet egy pozícióval jobbra mozgatja, majd állapotot vált az állapot-átmenet függvény szerint.
- Ha az automata elolvasta az inputot, megáll (felismeri vagy elutasítja a szót).

Véges automata

Alkalmazási területek:

- Önműködő ajtó
- Kávéautomata
- Mintafelismerés
- Matematikai rejtvények (pl. átkelés a folyón)
- Mintafelismerés Markov-láncokban.
- Beszédfeldolgozás, optikai karakterfelismerés.
- Piaci részesedések eloszlásának előrejelzése.

Véges automata

Jelölés: ha X egy halmaz, $\mathcal{P}(X)$ jelöli X **hatványhalmazát**, azaz X részhalmazainak halmazát.

Definíció

A **véges automata** egy rendezett ötös, $A = \langle Q, T, \delta, Q_0, F \rangle$, ahol

- Q az állapotok egy véges, nemüres halmaza,
- T az inputszimbólumok véges ábécéje,
- $\delta : \; Q \times T \to \mathcal{P}(Q)$ ún. állapot-átmenet függvény,
- $Q_0 \subseteq Q$ a kezdőállapotok halmaza,
- $F \subseteq Q$ az elfogadó állapotok halmaza.

Definíció

Ha $\forall (q, a) \in Q \times T$ esetén $|\delta(q, a)| = 1$ és $|Q_0| = 1$, akkor **determinisztikus véges automatáról** beszélünk.

Ilyenkor δ egy $Q \times T \to Q$ (totális) függvénynek tekinthető, a kezdőállapot pedig $q_0 \in Q$.

Determinisztikus és nemdeteminisztikus véges automata

Determinisztikus véges automata: (VDA) A δ függvény egyértékű. Ez azt jelenti, hogy minden (q, a) párra, ahol $(q, a) \in Q \times T$ pontosan egy olyan s állapot létezik, amelyre $\delta(q, a) = s$ fennáll. Egyetlen $q_0 \in Q$ kezdőállapot van.

Nemdeterminisztikus véges automata: (VNDA) Többértékű állapot-átmenet függvény is megengedett, azaz δ definiálható úgy, mint egy $Q \times T \to \mathcal{P}(Q)$ leképezés. Ekkor nemdeterminisztikus véges automatáról beszélünk. Több kezdőállapot is megengedett (a kezdőállapotok $Q_0 \subseteq Q$ halmaza). Előfordulhat, hogy az $\delta(q, a) = \emptyset$ valamely (q, a)-ra, azaz elakad a gép. Több kezdőállapot lehet, $Q_0 \subseteq Q$ a kezdőállapotok halmaza.

Véges automata

Alternatív jelölés: Az állapot-átmeneteket $qa \rightarrow p$ alakú szabályok formájában is megadhatjuk $p \in \delta(q, a)$ esetén.

Jelöljük M_δ -val az $A = \langle Q, T, \delta, Q_0, F \rangle$ nemdeterminisztikus véges automata δ állapot-átmenet függvénye által az előbbi módon származó szabályok halmazát.

Ekkor, ha minden egyes (q, a) párra M_δ pontosan egy $qa \rightarrow p$ szabályt tartalmaz, akkor a véges automata determinisztikus, egyébként nemdeterminisztikus.

Véges automaták – egy- és többlépéses redukció

Definíció

Legyen $A = \langle Q, T, \delta, Q_0, F\rangle$ egy véges automata és legyenek $u, v \in QT^*$ szavak. Az A automata az u szót **egy lépésben** (közvetlenül) a v szóra **redukálja** (jelölés: $u \Rightarrow_A v$), ha van olyan $qa \rightarrow p \in M_\delta$ szabály (vagyis $p \in \delta(q, a)$) és olyan $w \in T^*$ szó, hogy $u = qaw$ és $v = pw$ teljesül.

Példa: Ha $\delta(q, a) = \{r, s\}$, akkor $qabbab \Rightarrow sbbab$. (VNDA)

Definíció

A $\Rightarrow_A^* \subseteq QT^* \times QT^*$ reláció a $\Rightarrow_A$ reláció reflexív, tranzitív lezártja. Ha $u \Rightarrow_A^* v$, akkor azt mondjuk hogy A az u szót több lépésben (közvetetten) a v szóra **redukálja**.

Példa: Ha $\delta(q, a) = \{r, s\}$ és $\delta(s, b) = \{q, r\}$ akkor $qabbab \Rightarrow sbbab \Rightarrow rbab$ és így $qabbab \Rightarrow^* rbab$. (VNDA)

Véges automaták által felismert nyelv

Definíció

Az $A = \langle Q, T, \delta, Q_0, F\rangle$ nemdeterminisztikus véges automata által **elfogadott/felismert nyelv**:
$L(A) = \{u \in T^* \mid q_0 u \Rightarrow_A^* p$ valamely $q_0 \in Q_0$-ra és $p \in F$-re$\}$

Megjegyzés: Determinisztikus esetben $Q_0 = \{q_0\}$ egyelemű, és minden $u \in T^*$-ra $q_0 u$ legfeljebb egyféleképp redukálható valamely $p \in F$-re.

Tehát a determinisztikus esetben az felismert nyelv definíciója így egyszerűsödik:

Definíció

Az $A = \langle Q, T, \delta, q_0, F\rangle$ determinisztikus véges automata által **felismert nyelv**: $L(A) = \{u \in T^* \mid q_0 u \Rightarrow_A^* p,$ ahol $p \in F\}$

Véges automata

Példa: $T = \{a, b, c\}$. Adjunk VDA-t mely a legfeljebb 5 hosszú szavakat fogadja el!

Megoldás: **I.** (Képlettel)

$$\langle \{q_0, \ldots, q_6\}, \{a, b, c\}, \delta, q_0, \{q_0, \ldots, q_5\} \rangle$$

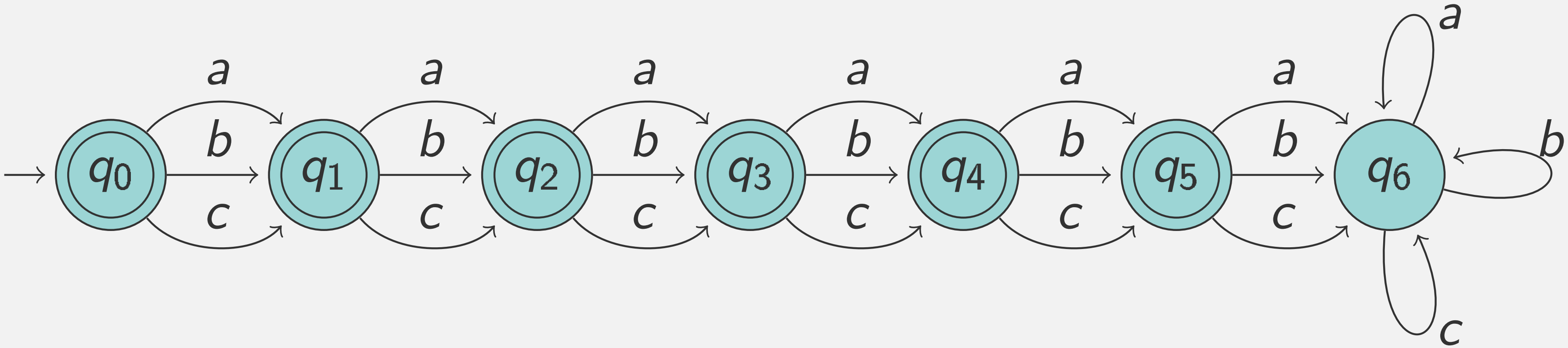
$$\delta(q_i, t) = q_{i+1}, \quad (i{=}0, \ldots, 5,\ \ t \in \{a, b, c\})$$

$$\delta(q_6, t) = q_6. \ (t \in \{a, b, c\})$$

II. (Táblázattal)

	a	b	c
$\leftrightarrows q_0$	q_1	q_1	q_1
$\leftarrow q_1$	q_2	q_2	q_2
$\leftarrow q_2$	q_3	q_3	q_3
$\leftarrow q_3$	q_4	q_4	q_4
$\leftarrow q_4$	q_5	q_5	q_5
$\leftarrow q_5$	q_6	q_6	q_6
q_6	q_6	q_6	q_6

III. (Átmenetdiagrammal)



Véges automaták determinizálása

Tétel

Minden $A = \langle Q, T, \delta, Q_0, F \rangle$ nemdeterminisztikus véges automatához megkonstruálható egy $A' = \langle Q', T, \delta', q_0', F' \rangle$ determinisztikus véges automata úgy, hogy $L(A) = L(A')$ teljesül.

A konstrukció:

$Q' := \mathcal{P}(Q), \quad q_0' := Q_0, \quad F' := \{q' \in Q' \mid q' \cap F \neq \emptyset\},$

$$\delta'(q', a) := \bigcup_{q \in q'} \delta(q, a).$$

Véges automaták determinizálása – Példa

	a	b
$\rightarrow q_0$	$\{\}$	$\{q_1, q_2\}$
$\leftrightarrows q_1$	$\{q_0\}$	$\{\}$
$\leftarrow q_2$	$\{q_1\}$	$\{q_2\}$

VNDA

	a	b
$\{\}$	$\{\}$	$\{\}$
$\{q_0\}$	$\{\}$	$\{q_1, q_2\}$
$\leftarrow \{q_1\}$	$\{q_0\}$	$\{\}$
$\leftarrow \{q_2\}$	$\{q_1\}$	$\{q_2\}$
$\leftrightarrows \{q_0, q_1\}$	$\{q_0\}$	$\{q_1, q_2\}$
$\leftarrow \{q_0, q_2\}$	$\{q_1\}$	$\{q_1, q_2\}$
$\leftarrow \{q_1, q_2\}$	$\{q_0, q_1\}$	$\{q_2\}$
$\leftarrow \{q_0, q_1, q_2\}$	$\{q_0, q_1\}$	$\{q_1, q_2\}$

VDA

Megjegyzés: Q' q_0'-ből elérhetetlen állapotai elhagyhatóak.

A véges automaták számítási ereje

Tétel

- Minden A nemdeterminisztikus véges automatához meg tudunk adni egy 3-típusú G grammatikát úgy, hogy $L(G) = L(A)$ teljesül.
- Minden 3-típusú G grammatikához meg tudunk adni egy A véges automatát úgy, hogy $L(A) = L(G)$ teljesül.

Következmény

- Minden reguláris kifejezéshez van olyan véges automata, amelyik a reguláris kifejezés által reprezentált nyelvet ismeri fel.
- A véges automaták által felismert nyelvek reprezentálhatóak reguláris kifejezéssel.

Vagyis a VDA-k, VNDA-k, reguláris kifejezések számítási ereje megegyezik a reguláris grammatikáéval, éppen az $\mathcal{L}_3$-beli nyelveket tudjuk megadni velük.

Véges automata – Myhill-Nerode tétel

Definíció

Legyen L egy T ábécé feletti nyelv. Az L nyelv által indukált E_L reláció alatt egy olyan bináris relációt értünk a T^*-on, amelyre teljesül, hogy bármely $u, v \in T^*$-ra uE_Lv akkor és csak akkor, ha nincs olyan $w \in T^*$, hogy az uw és vw szavak közül pontosan az egyik eleme L-nek.

E_L ekvivalenciareláció és jobb-invariáns. (Jobb-invariáns: ha uE_Lv, akkor uwE_Lvw is fennáll minden $w \in T^*$ szóra.)

Az E_L reláció indexén ekvivalenciaosztályainak számát értjük.

Tétel(Myhill-Nerode)

$L \subseteq T^*$ akkor és csak akkor ismerhető fel determinisztikus véges automatával, ha E_L véges indexű.

Minimális (állapotszámú) véges automata

Definíció

Az A determinisztikus véges automata minimális állapotszámú (minimális), ha nincs olyan A' determinisztikus véges automata, amely ugyanazt a nyelvet ismeri fel, mint A, de A' állapotainak száma kisebb, mint A állapotainak száma.

Tétel

Az L reguláris nyelvet elfogadó minimális (állapotszámú) determinisztikus véges automata az izomorfizmus erejéig egyértelmű.

Megjegyzés: $\forall$ L reguláris nyelvhez Myhill-Nerode tétel alapján készíthető egy $A_{\mathrm{MN}} = \langle Q, T, \delta, q_0, F \rangle$ minimális automata.

Legyenek $w_0, \ldots, w_{n-1}$ E_L ekvivalenciaosztályainak 1-1 reprezentánsa, ahol $w_0 = \varepsilon$. $Q := \{q_i \mid 0 \leqslant i \leqslant n-1\}$, $\delta(q_i, t) := q_j$, akkor és csak akkor, ha $w_i t E_L w_j$. $F := \{q_i \mid w_i \in L\}$.

Kis Bar-Hillel lemma

A Myhill-Nerode tétel szükséges és elégséges feltételt ad egy nyelv $\mathcal{L}_3$-ba tartozására.

A Kis Bar-Hillel lemma szükséges feltételt ad egy nyelv $\mathcal{L}_3$-ba tartozására, így arra használható, hogy egy nyelvről bizonyítsuk, hogy nem reguláris: ha egy L nyelvre nem teljesül a Kis Bar-Hillel lemma feltétele, akkor $L \notin \mathcal{L}_3$.

Tétel (Kis Bar-Hillel lemma)

Minden $L \in \mathcal{L}_3$ nyelvhez van olyan $n \in \mathbb{N}$ konstans, hogy minden $w \in L$ szó esetén ha tekintjünk egy tetszőleges olyan $w = uw'v$ felbontását, ahol $|w'| \geqslant n$, akkor van w'-nek olyan y részszava ($w' = xyz$), hogy $0 < |y| \leqslant n$, és minden $i \geqslant 0$ esetén $uxy^izv \in L$.

Megjegyzés A tétel gyakran használt másik elnevezése: "pumpálási lemma"

Sztochasztikus automata

- A **sztochasztikus automata** a nemdeterminisztikus véges automata olyan általánosítása, ahol mind a kezdőállapotot, mind pedig egy adott (átmenet, betű) párra az új állapotot egy valószínűségi eloszlás alapján véletlenszerűen választjuk.
- Jelölje $s_1, \ldots, s_n$ a sztochasztikus automata állapotait. Ekkor x inputszimbólum hatására az automata s állapotból valamely s_i állapotba megy, $p_i(s, x)$ valószínűséggel, ahol minden s-re és x-re fennáll a következő:

$$\sum_{i=1}^{n} p_i(s, x) = 1, \quad p_i(s, x) \geqslant 0$$

- A kezdőállapot helyett definiáljuk a **kezdőállapotok eloszlását**, azaz minden állapot kezdőállapot valamilyen rögzített valószínűséggel.
- Az **elfogadott nyelv** $L(PA, S_1, \eta)$ függ a végállapotok S_1 halmazától és a $0 \leqslant \eta < 1$ valós számtól, az ún. **vágási ponttól**.

Sztochasztikus automata

- Az elfogadott nyelv $L(PA, S_1, \eta)$ az összes olyan szót tartalmazza, amely által az automata állapot-átmenetek sorozatán keresztül valamely S_1 -beli állapotba jut, ahol a valószínűség nagyobb, mint η.
- n-dimenziós **sztochasztikus mátrix** alatt egy $(p_{ij})_{1 \leqslant i,j \leqslant n}$ négyzetes mátrixot értünk, melyre (1) $p_{ij} \geqslant 0$ $(1 \leqslant i, j \leqslant n)$ (2) $\sum_{j=1}^{n} p_{ij} = 1$ $(1 \leqslant i \leqslant n)$.
- n-dimenziós **sztochasztikus sorvektornak** (oszlopvektornak) egy olyan n-dimenziós sorvektort (oszlopvektort) nevezünk, amelynek komponensei nemnegatívak és a komponensek összege 1.
- Ha a sztochasztikus sorvektornak csak egy komponense 1, akkor **koordinátavektorról** beszélünk.
- Az n-dimenziós E_n egységmátrix sztochasztikus mátrix.

Sztochasztikus automata

Definíció

A **véges sztochasztikus automata** egy V ábécé felett egy $PA = \langle S, s_0, M\rangle$ rendezett hármas, ahol

- $S = \{s_1, \ldots, s_n\}$ az állapotok egy véges, nemüres halmaza,
- s_0 egy n-dimenziós sztochasztikus sorvektor, a kezdeti állapotok eloszlása
- M egy leképezés, amely V -t leképezi az n-dimenziós sztochasztikus mátrixok halmazába.

Valamely $x \in V$ -re az $M(x)$ mátrix (i, j)-dik eleme $p_j(s_i, x)$, annak a valószínűsége, hogy az x szimbólum hatására PA az s_i állapotból az s_j állapotba lép.

Példa Tekintsük a következő sztochasztikus automatát:

$PA_1 = \langle\{s_1, s_2\}, (1, 0), M\rangle$ az $\{x, y\}$ ábécé felett, ahol

$$M(x) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad M(y) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

Sztochasztikus automata

Legyen $PA = \langle S, s_0, M \rangle$ egy V ábécé feletti véges sztochasztikus automata. Ekkor az M függvény V -ről a következőképpen terjeszthető ki V^*-ra:

- $\hat{M}(\varepsilon) := E_n$
- $\hat{M}(x_1 \cdots x_n) := M(x_1)M(x_2) \cdots M(x_n)$, ahol $k \geqslant 2, x_i \in V$.

$\hat{M}$ helyett a továbbiakban M-et írunk.

Valamely $w \in V^*$ szóra az $M(w)$ mátrix (i, j)-edik elemét $p_j(s_i, w)$ jelöli, amely annak a valószínűsége, hogy az automata a w inputszó feldolgozása után az s_i állapotból éppen az s_j állapotba jut.

Sztochasztikus automata

Legyen $PA = \langle S, s_0, M\rangle$ egy V ábécé feletti véges sztochasztikus automata és legyen $w \in V^*$. Az $s_0 M(w)$ sztochasztikus sorvektor a w **eredményeként kapott állapoteloszlás**, melyet $PA(w)$-vel jelölünk.

Észrevétel: $PA(\varepsilon) = s_0$.

Definíció

Legyen $PA = \langle S, s_0, M\rangle$ egy V ábécé feletti véges sztochasztikus automata, $0 \leqslant \eta < 1$ egy valós szám, és $\bar{s}_1$ egy n-dimenziós oszlopvektor, amelynek minden komponense vagy 0, vagy 1.

Az **$\bar{s}_1$ által η vágási ponttal elfogadott nyelvet** az $L(PA, \bar{s}_1, \eta) = \{w \in V^* \mid s_0 M(w)\bar{s}_1 > \eta\}$ halmaz definiálja.

$\bar{s}_1$-ra úgy gondolhatunk, hogy 1 koordinátái kijelölik a végállapotok halmazát. $L(PA, \bar{s}_1, \eta)$ ekkor azon szavak halmaza, amelyekre η-nál nagyobb valószínűséggel kerül PA végállapotba.

Sztochasztikus automata

Definíció

Egy L nyelvet **η-sztochasztikusnak** mondunk, ha valamely $PA = \langle S, s_0, M \rangle$ véges sztochasztikus automatára és $\bar{s}_1$ oszlopvektorra $L = L(PA, \bar{s}_1, \eta)$ teljesül.

Definíció

Egy L nyelvet **sztochasztikusnak** nevezünk, ha valamely $0 \leqslant \eta < 1$-re η-sztochasztikus.

Az alábbi tételt nem bizonyítjuk.

Tétel

Minden reguláris nyelv sztochasztikus, de nem minden sztochasztikus nyelv reguláris. A 0-sztochasztikus nyelvek regulárisak.

Sztochasztikus automata

Példa: Tekintsük az előző példában szereplő automatát.
Emlékeztetőül:
$V = \{x, y\}$, $PA_1 = \langle\{s_1, s_2\}, (1, 0), M\rangle$, ahol

$$M(x) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad M(y) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

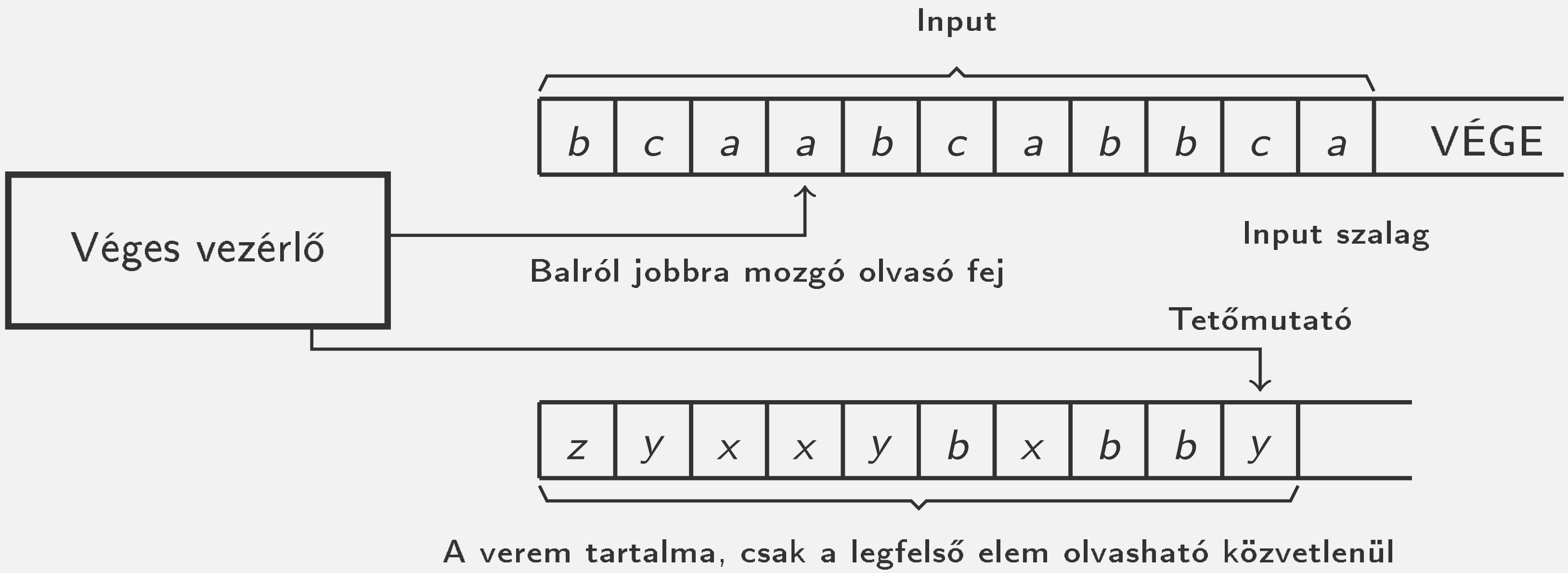
Ekkor
$PA_1(x^n) = (1, 0)M(x^n) = (1, 0)$, ha n páros,
$PA_1(x^n) = (0, 1)$, ha n páratlan, és
$PA_1(w) = (1/2, 1/2)$, ha w legalább egy y -t tartalmaz.

Tehát $\bar{s}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ esetén

$$L(PA, \bar{s}_1, \eta) = \begin{cases} V^* - (xx)^* & \text{ha } 0 \leqslant \eta < 1/2 \\ x(xx)^* & \text{ha } 1/2 \leqslant \eta < 1 \end{cases}$$

Tehát $V^* - (xx)^*$ 1/3-sztochasztikus, míg $x(xx)^*$ 2/3-sztohasztikus nyelv. Így mindkettő sztochasztikus nyelv.

Veremautomata



- A veremautomata a véges automata általánosítása potenciálisan végtelen veremmel és véges kontrollal.
- A verem esetében az új adat mindig a már meglévő veremtartalom tetejéhez adódik, kivétele fordított sorrendben történik.
- alapértelmezetten nemdeterminisztikus

Veremautomata

Jelölés: ha X egy halmaz, jelölje $\mathcal{P}_{\text{véges}}(X)$ az X véges részhalmazainak halmazát.

Definíció

A **veremautomata** egy $A = \langle Z, Q, T, \delta, z_0, q_0, F \rangle$, rendezett hetes, ahol

- Z a veremszimbólumok véges halmaza (veremábécé),
- Q az állapotok véges halmaza,
- T az inputszimbólumok véges halmaza (inputábécé),
- $\delta : Z \times Q \times (T \cup \{\varepsilon\}) \to \mathcal{P}_{\text{véges}}(Z^* \times Q)$, az ún. átmeneti függvény,
- $z_0 \in Z$ a kezdeti (kezdő) veremszimbólum,
- $q_0 \in Q$ a kezdeti állapot (kezdőállapot),
- $F \subseteq Q$ az elfogadó állapotok vagy végállapotok halmaza.

Veremautomata

- A verem tetején lévő szimbólum, az aktuális állapot és az inputszimbólum együttesen határozzák meg a következő átmenetet.
- Minden lépésben mindenképpen kiveszünk egyetlen egy elemet a verem tetejéről és beteszünk helyette néhányat. $(0, 1, 2, \ldots$ darabot)
- Ha $\delta(z, q, \varepsilon)$ nem üres, akkor ún. **ε-átmenet** (ε-lépés, ε-mozgás) hajtható végre, ami lehetővé teszi, hogy a veremautomata anélkül változtassa meg az állapotát, hogy valamilyen szimbólumot olvasson az inputszalagról.
- ε-mozgásra lehetőség van már az első inputszimbólum elolvasása előtt is illetve még az utolsó inputszimbólum elolvasása után is.

Veremautomata konfigurációi

Definíció

A veremautomata **konfigurációja** alatt egy zqw alakú szót értünk, ahol $z \in Z^*$ a verem aktuális tartalma és $q \in Q$ az aktuális állapot és $w \in T^*$ az input még feldolgozatlan része.

z első betűje van a verem alján, míg utolsó betűje a verem tetején.

Az input olvasófeje w első betűjén áll.

Így a q baloldalán lévő szimbólum van a verem tetején, míg a jobboldalán lévő szimbólum az input következő feldolgozandó betűje.

Definíció

Az $A = \langle Z, Q, T, \delta, z_0, q_0, F \rangle$ veremautomata $w \in T^*$ bemenethez tartozó **kezdőkonfigurációja** $z_0 q_0 w$.

Alapvető veremműveletek megvalósítása

Legyen $t \in T \cup \{\varepsilon\}$, $q, r \in Q$ és $z \in Z$

- $(\varepsilon, r) \in \delta(z, q, t)$: a z elemet kivehetjük a veremből (POP művelet)
- $(z, r) \in \delta(z, q, t)$: a verem tartalma változatlan maradhat
- $(z', r) \in \delta(z, q, t)$: z-t lecserélhetjük z'-re a verem tetején $(z' \in Z)$
- $(zz', r) \in \delta(z, q, t)$: z'-t a verem tetejére (z-re rá) tehetjük $(z' \in Z)$ (PUSH művelet)
- Egyéb lehetőségek, például $(zz'z'', r) \in \delta(z, q, t)$: $z'z''$-t a verem tetejére tehetjük, z'' lesz a tetején $(z', z'' \in Z)$.
- Általánosan $(w, r) \in \delta(z, q, t)$, ahol $w \in Z^*$ tetszőleges Z feletti szó. A w szó kerül z helyére és w utolsó betűje lesz a verem tetején.

Veremautomata – egylépéses redukció

Definíció

Az A veremautomata az $\alpha \in Z^*QT^*$ konfigurációt a $\beta \in Z^*QT^*$ konfigurációra **redukálja egy lépésben**, amelyet $\alpha \Rightarrow_A \beta$-val jelölünk, ha létezik olyan $z \in Z, q, p \in Q, a \in T \cup \{\varepsilon\}$, $r, u \in Z^*$ és $w \in T^*$, hogy $(u, p) \in \delta(z, q, a)$ és $\alpha = rzqaw$ és $\beta = rupw$ teljesül.

Példák:

- ha A-ban $\delta(c, q_1, a) = \{(dd, q_2), (\varepsilon, q_4)\}$ és $z_0cddcq_1ababba$ egy konfiguráció, akkor $z_0cddcq_1ababba \Rightarrow_A z_0cddddq_2babba$ és $z_0cddcq_1ababba \Rightarrow_A z_0cddq_4babba$ is teljesül,
- ha A-ban $\delta(c, q_3, \varepsilon) = \{(dd, q_2)\}$ és $z_0cddcq_3ababba$ egy konfiguráció, akkor $z_0cddcq_3ababba \Rightarrow_A z_0cddddq_2ababba$
- ha A-ban $\delta(c, q_5, \varepsilon) = \emptyset$ és $\delta(c, q_5, a) = \emptyset$, akkor nem létezik olyan C konfiguráció, melyre $z_0ccq_5aab \Rightarrow_A C$

Veremautomata – többlépéses redukció

Definíció

Az A veremautomata az $\alpha \in Z^*QT^*$ konfigurációt a $\beta \in Z^*QT^*$ konfigurációra **redukálja**, amelyet $\alpha \Rightarrow_A^* \beta$-val jelölünk, ha vagy $\alpha = \beta$, vagy létezik olyan $\alpha_1, \ldots, \alpha_n$ szavakból álló véges sorozat, ahol $\alpha = \alpha_1, \beta = \alpha_n$ és $\alpha_i \Rightarrow_A \alpha_{i+1}$, $1 \leqslant i \leqslant n-1$.

Tehát $\Rightarrow_A^* \subseteq Z^*QT^* \times Z^*QT^*$ a $\Rightarrow_A$ reláció reflexív, tranzitív lezártja.

Példa:

Ha $\delta(d, q_6, b) = \{(\varepsilon, q_5)\}$ és $\delta(d, q_5, \varepsilon) = \{(dd, q_2), (\varepsilon, q_4)\}$ akkor

$\#cddq_6bab \Rightarrow_A \#cdq_5ab \Rightarrow_A \#cddq_2ab$ és

$\#cddq_6bab \Rightarrow_A \#cdq_5ab \Rightarrow_A \#cq_4ab$.

Tehát $\#cddq_6bab \Rightarrow_A^* \#cddq_2ab$ és $\#cddq_6bab \Rightarrow_A^* \#cq_4ab$.

Veremautomata – felismert nyelv

Definíció

Az A veremautomata által **elfogadó állapottal (végállapottal) elfogadott nyelv**

$$L(A) = \{w \in T^* \mid z_0 q_0 w \Rightarrow_A^* up, \text{ ahol } u \in Z^*, p \in F\}.$$

Determinisztikus veremautomata

Definíció

Az $A = \langle Z, Q, T, \delta, z_0, q_0, F\rangle$ veremautomatát **determinisztikusnak** nevezzük, ha minden $(z, q, a) \in Z \times Q \times T$ esetén $|\delta(z, q, a)| + |\delta(z, q, \varepsilon)| = 1$.

Tehát minden $q \in Q$ és $z \in Z$ esetén

- vagy $\delta(z, q, a)$ pontosan egy elemet tartalmaz minden $a \in T$ inputszimbólumra és $\delta(z, q, \varepsilon) = \emptyset$,
- vagy $\delta(z, q, \varepsilon)$ pontosan egy elemet tartalmaz és $\delta(z, q, a) = \emptyset$ minden $a \in T$ inputszimbólumra.

Észrevétel: Ha minden $(z, q, a) \in Z \times Q \times T$ esetén $|\delta(z, q, a)| + |\delta(z, q, \varepsilon)| \leqslant 1$ akkor a veremautomata a felismert nyelv módosulása nélkül kiegészíthető determinisztikus veremautomatává. Így tágabb értelemben az ezt a feltételt teljesítő veremautomatákat is tekinthetjük determinisztikus veremautomatának.

Veremautomaták alternatív reprezentációi

- **Átírási szabályokkal:**
  A δ leképezést szabályok formájában is megadhatjuk. Az így nyert szabályhalmazt M_δ -val jelöljük. Tehát ezzel az alternatív jelöléssel:

  $$zqa \to up :\in M_\delta \iff (u,p) \in \delta(z,q,a),$$
  $$zq \to up :\in M_\delta \iff (u,p) \in \delta(z,q,\varepsilon).$$
  $$(p,q \in Q, a \in T, z \in Z, u \in Z^*)$$

- **Átmenetdiagrammal:**

  $p, q \in Q, a \in T \cup \{\varepsilon\}, z \in Z, u \in Z^*$ esetén:

  $$q \xrightarrow{a;\, z \to u} p \qquad \iff (u,p) \in \delta(z,q,a)$$

  A végállapotokat duplán karikázzuk. A kezdőállapotot $\to$ jelöli.

Veremautomata – példa

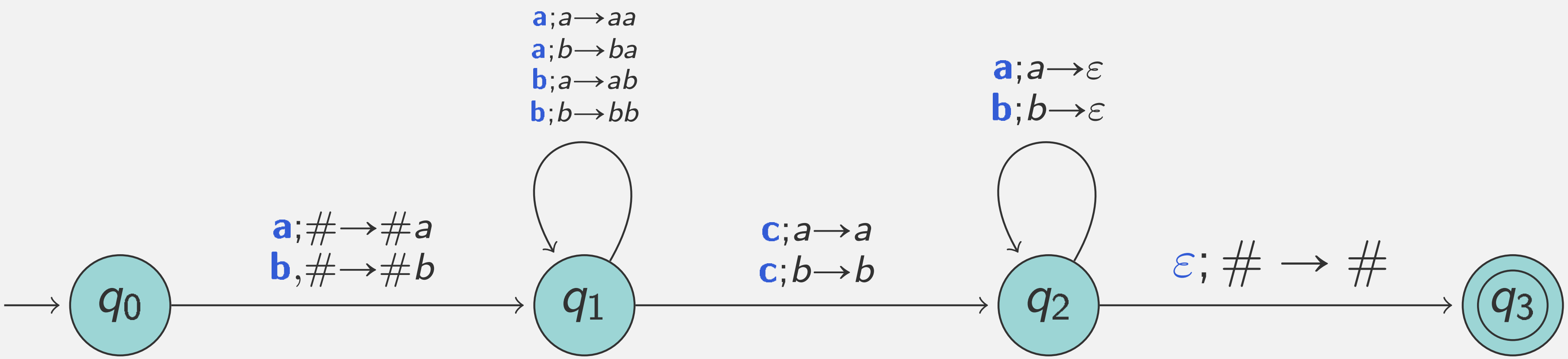
1. Példa: Legyen $L_1 = \{wcw^{-1} \mid w \in \{a,b\}^+\}$. Készítsünk egy A veremautomatát, melyre $L(A) = L_1$.

Megoldás:

$A = \langle \{q_0, q_1, q_2, q_3\}, \{\mathbf{a,b,c}\}, \{\#, a, b\}, \delta, q_0, \#, \{q_3\} \rangle$, ahol:

$$
\begin{array}{rll}
(\#t, q_1) \in \delta(\#, q_0, \mathbf{t}) & \forall t \in \{a,b\} \\
(zt, q_1) \in \delta(z, q_1, \mathbf{t}) & \forall z,t \in \{a,b\} \\
(z, q_2) \in \delta(z, q_1, \mathbf{c}) & \forall z \in \{a,b\} \\
(\varepsilon, q_2) \in \delta(t, q_2, \mathbf{t}) & \forall t \in \{a,b\} \\
(\#, q_3) \in \delta(\#, q_2, \varepsilon) &
\end{array}
$$

DETERMINISZTIKUS



Veremautomata – példa

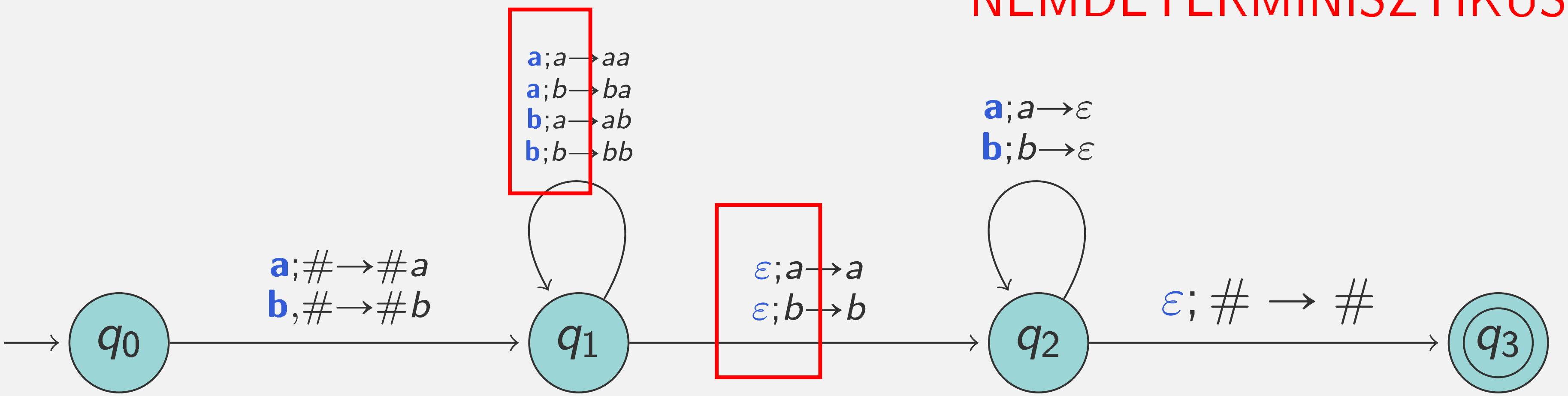
2. Példa: Legyen $L_2 = \{ww^{-1} \mid w \in \{a,b\}^+\}$. Készítsünk egy A veremautomatát, melyre $L(A) = L_2$.

Megoldás:

$A = \langle\{q_0, q_1, q_2, q_3\}, \{\mathbf{a},\mathbf{b}\}, \{\#, a, b\}, \delta, q_0, \#, \{q_3\}\rangle$, ahol:

$$
\begin{aligned}
(\#t, q_1) &\in \delta(\#, q_0, \mathbf{t}) && \forall t \in \{a,b\}\\
(zt, q_1) &\in \delta(z, q_1, \mathbf{t}) && \forall z, t \in \{a,b\}\\
(z, q_2) &\in \delta(z, q_1, \varepsilon) && \forall z \in \{a,b\}\\
(\varepsilon, q_2) &\in \delta(t, q_2, \mathbf{t}) && \forall t \in \{a,b\}\\
(\#, q_3) &\in \delta(\#, q_2, \varepsilon)
\end{aligned}
$$

NEMDETERMINISZTIKUS



Üres veremmel elfogadott nyelv

Definíció

Az A veremautomata által **üres veremmel elfogadott nyelv** $N(A) = \{w \in T^* \mid z_0 q_0 w \Rightarrow_A^* p, \text{ ahol } p \in Q\}$.

Megjegyzés:
Vegyük észre, hogy ha a verem üres, akkor az automata működése blokkolódik, mivel nincs átmenet definiálva üres verem esetére. (lásd δ definíciója).

Így a verem az input teljes feldolgozása után, az utolsó átmenettel kell üressé váljon.

Szintén a blokkolás elkerülése végett definiáltuk úgy a kezdőkonfigurációt, hogy a veremábécé egy eleme (z_0) már eleve a veremben van.

Megjegyzés: Vegyük észre, hogy az elfogadó állapotok halmaza irreleváns $N(A)$ szempontjából.

Üres veremmel elfogadott nyelv

Példa: Az alábbi $A = \langle \{\$, a\}\{q_0, q_1\}, \{a, b\}, \delta, \$, q_0, \{\ \}\rangle$ veremautomata esetén $N(A) = \{a^n b^n \mid n \geqslant 1\}$, azaz ezt nyelvet ismeri fel üres veremmel.

M_δ :

$$
\begin{aligned}
\$q_0 a &\to \$aq_0 \\
aq_0 a &\to aaq_0 \\
aq_0 b &\to q_1 \\
aq_1 b &\to q_1 \\
\$q_1 \quad &\to q_1 .
\end{aligned}
$$

A determinisztikus, $a^2 b^3$-re:

$$\$q_0 aabbb \Rightarrow \$aq_0 abbb \Rightarrow \$aaq_0 bbb \Rightarrow \$aq_1 bb \Rightarrow \$q_1 b \Rightarrow q_1 b.$$

A elutasítja $aabbb$-t, mivel hiába lett üres a verem, még volt hátra az inputból.

A veremautomaták számítási ereje

Tétel

Bármely L nyelvre ekivalensek a következő állítások

- L környezetfüggetlen, azaz környezetfüggetlen (2-es típusú) grammatikával generálható
- L (nemdeterminisztikus) veremautomatával végállapottal felismerhető
- L (nemdeterminisztikus) veremautomatával üres veremmel felismerhető

A determinisztikus veremautomaták számítási ereje kisebb.

Tétel

Minden reguláris (3-as típusú) nyelv felismerhető determinisztikus veremautomatával, de létezik olyan (2-es típusú) környezetfüggetlen nyelv, ami nem ismerhető fel determinisztikus veremautomatával.

Ilyen nyelv például $L_2 = \{ww^{-1} \mid w \in \{a, b\}^+\}$.

Nagy Bar-Hillel lemma

Egy szükséges feltétel egy nyelv környezetfüggetlenségére (és így veremautomatával való felismerhetőségére is).

Tétel (Nagy Bar-Hillel lemma)

Minden L környezetfüggetlen nyelvhez meg tudunk adni két, p és q természetes számot úgy, hogy minden olyan szó L-ben, amely hosszabb, mint p felírható $uxwyz$ alakban, ahol $|xwy| \leqslant q, xy \neq \varepsilon$, továbbá, ekkor minden ux^iwy^iz, $i \geqslant 0$ alakú szó is benne van az L nyelvben $(u, x, w, y, z \in T^*)$.

Példák: ($|u|_t$: a t betűk száma u-ban)

$\in \mathcal{L}_3$	$\in \mathcal{L}_2 - \mathcal{L}_3$	$\in \mathcal{L}_1 - \mathcal{L}_2$
$\{u \mid abbab \subseteq u\}$	$\{u \in \{a,b\}^* \mid u = u^{-1}\}$	$\{uu \mid u \in \{a,b\}^*\}$
$\{u \mid abbab \nsubseteq u\}$	$\{a^nb^n \mid n \in \mathbb{N}\}$	$\{a^nb^nc^n \mid n \in \mathbb{N}\}$
7-tel osztható számok	$\{u \in \{a,b\}^* \mid \lvert u\rvert_a = \lvert u\rvert_b\}$	$\{a^{n^2} \mid n \in \mathbb{N}\}$
$((a+bb)^* + ab)^*$	helyes ()-k nyelve	$\{a^{2^n} \mid n \in \mathbb{N}\}$