Alkalmazott Matematikai Lapok **24** (2007), 187-207.

AZ ITERÁLT SHUFFLE MŰVELET ALKALMAZÁSA REGULÁRIS NYELVEKEN

IVÁN SZABOLCS

A párhuzamos folyamatok elméletében központi szerepet játszanak a *shuffle-szorzás* ($\sqcup\!\sqcup$) és iterált megfelelője, a *shuffle-iterált* ($^{*\sqcup\!\sqcup}$) műveletek. Nem ismert, hogy egy adott reguláris (vagy véges) nyelv shuffle-iteráltja mikor reguláris, illetve környezetfüggetlen; vizsgálatainkban bizonyos speciális reguláris nyelvekre megválaszoljuk e kérdéseket, továbbá adunk egy reguláris nyelvosztályt is, mely zárt mindkét műveletre.

1. Bevezetés

A *shuffle-szorzat* egy jól ismert kommutatív és asszociatív kétváltozós művelet nyelvek között. Ha L_1, L_2 nyelvek a Σ ábécé fölött, shuffle-szorzatuk, $L_1 \sqcup\!\sqcup L_2$ szintén Σ fölötti nyelv, amit a következőképp definiálunk:

$$L_1 \sqcup\!\sqcup L_2 = \{x_1y_1 \dots x_ny_n : x_1x_2 \dots x_n \in L_1,\ y_1y_2 \dots y_n \in L_2, x_i, y_i \in \Sigma^*\}\,.$$

Egy Σ fölötti L nyelv *shuffle-iteráltja*, $L^{*\sqcup\!\sqcup}$ szintén Σ fölötti nyelv; ezt a következőképp definiálhatjuk:

$$L^{*\sqcup\!\sqcup} = \{\lambda\} \cup L \cup (L \sqcup\!\sqcup L) \cup (L \sqcup\!\sqcup L \sqcup\!\sqcup L) \cup \dots.$$

Jól ismert tény, hogy a Chomsky-hierarchia $\mathcal{L}_1$ és $\mathcal{L}_0$ osztályai zártak mindkét műveletre, lásd pl. [1]. A reguláris nyelvek osztálya, $\mathcal{L}_3$ a shuffle-szorzásra zárt, ám nem zárt a shuffle-iterált képzésére; a környezetfüggetlen nyelvek osztálya, $\mathcal{L}_2$ pedig nem zárt egyik műveletre sem, lásd [1]. A *shuffle nyelvek* $\mathcal{SL}$ osztályát [1, 9] úgy definiáljuk, mint a legszűkebb olyan nyelvosztályt, mely tartalmazza a véges (vagy reguláris) nyelveket, és zárt mind a reguláris műveletekre, mind a shuffle-szorzásra és a shuffle-iterált képzésére. A fenti zártsági tulajdonságok alapján $\mathcal{SL} \subseteq \mathcal{L}_1$. [13]-ban igazolást nyert, hogy $\mathcal{SL} \subseteq \mathcal{P}$, ahol $\mathcal{P}$ a polinom időben determinisztikusan eldönthető nyelvek osztálya. Továbbá, [15] szerint léteznek $\mathcal{SL}$-ben $\mathcal{P}$-teljes problémák is. Egy $\mathcal{SL}$-en belüli végtelen hierarchiát adtak meg [10]-ben. Az iterált shuffle művelettel lehet $\mathcal{NP}$-teljes nyelveket is definiálni, lásd [16]. Reguláris nyelvek varietásainak (ld. [14]) a shuffle műveletre való lezártjáról lásd [3].

[2]-ben igazolták, hogy eldönthetetlen a következő kérdés: adott egy shuffle nyelv (egy „shuffle-kifejezéssel”, melyben a reguláris műveleti jelek, a $\sqcup\!\sqcup$ és a ${}^{*\sqcup\!\sqcup}$ szerepelnek), L reguláris nyelv-e? További eldönthetetlenségi eredmények találhatók az [5, 8, 11, 12] cikkekben.

Ugyanez a kérdés *kommutatív* shuffle-nyelvekre eldönthető, lásd [4]. Jelenleg nem ismert, hogy eldönthető-e a következő kérdés: adott egy reguláris nyelv, igaz-e, hogy shuffle-iteráltja reguláris (vagy hogy környezetfüggetlen). A probléma nyitott *véges* nyelvekre is. A [4]-beli eredmények szerint a kérdés eldönthető kommutatív reguláris nyelvekre; egy közvetlen bizonyítás található [7]-ben. Azonban az ezen munkákban megjelent módszerek nem adnak explicit felső korlátot a kérdés bonyolultságára.

Ebben a cikkben szükséges és elégséges feltételeket adunk arra nézve, hogy bizonyos reguláris nyelvek shuffle-iteráltja reguláris vagy környezetfüggetlen legyen.

2. Definíciók, jelölések

Jelen cikkben $\mathcal{N}$ jelöli a természetes számok $\{0, 1, \dots\}$ halmazát. Amennyiben $n \in \mathcal{N}$, $[n]$ jelenti a $\{0, 1, \dots, n\}$ halmazt.

A formális nyelvek elméletében szereplő standard jelöléseket (ld. [6]) és elnevezéseket fogjuk használni: *ábécének* nevezünk bármely véges halmazt, egy ábécé elemei a betűk. Az ábécék jelölésére a Σ, Γ betűket használjuk, a betűket az $a, b, c, \dots$ betűkkel jelöljük, szükség esetén indexelve.

Σ ábécé fölötti *szónak* nevezünk bármely, Σ elemeiből képzett véges sorozatot, azaz $a_1a_2\dots a_n$ alakú sorozatokat, ahol $n \in \mathcal{N}$ és minden $1 \le i \le n$-re $a_i \in \Sigma$. Ekkor a szó hossza, $|a_1\dots a_n| = n$; a benne levő a betűk számát (ahol $a \in \Sigma$) $|a_1\dots a_n|_a$ jelzi. Ha $n = 0$, a szó az üres szó, aminek jele a λ. A Σ fölötti összes szavak halmazát Σ^* jelöli. Szavakra általában az u, v, w, x, y jelöléseket használjuk. Amennyiben w egy szó, $i \in \mathcal{N}$ és a w_i jelölés még nem definiált, úgy w_i a w szó i. betűjét jelenti.

Ha adott egy Σ ábécé fölötti két szó,

$$a_1a_2\dots a_n \text{ és } b_1\dots b_k \ (a_1, \dots, a_n, b_1, \dots, b_k \in \Sigma),$$

akkor (ilyen sorrendű) *konkatenáltjuk* az $a_1\dots a_nb_1\dots b_k$ szó. Σ^* a konkatenáció műveletével monoidot alkot, melynek egységeleme λ, az üres szó.

Σ ábécé feletti szavak bármely (üres, véges vagy végtelen) halmazát *nyelvnek* nevezzük. Az egyelemű nyelveket azonosítjuk egyetlen elemükkel; ez nem okoz majd félreértést. Nyelveket általában K vagy L jelöl, szükség szerint indexelve.

Nyelvek között értelmezhetjük a halmazelméleti műveletek mellett a konkatenáció műveletét: ha K és L két nyelv, akkor legyen $KL = \{uv : u \in K, v \in L\}$. A konkatenáció segítségével definiálhatjuk nyelvek nemnegatív egész kitevőjű hatványait: ha L egy nyelv, legyen

– $L^0 = \{\lambda\}$;

– $L^{n+1} = L^n L$.

Egy nyelvnek definiáljuk továbbá *(Kleene-)iteráltját* is: ha L egy nyelv, legyen

$$L^* = \bigcup_{i=0}^{\infty} L^i.$$

Nyelveket bonyolultságuk szerint osztályokba sorolhatunk. Cikkünk a *reguláris* és a *környezetfüggetlen* nyelvek osztályával foglalkozik; ezen osztályokat definiálhatjuk például a következőképpen:

– Egy L nyelv reguláris, ha előáll a véges nyelvekből a konkatenáció, Kleene-iteráció és a halmazelméleti műveletek véges sokszori alkalmazásával.

– Egy L nyelv környezetfüggetlen, ha létezik olyan P veremautomata, melyre $L_\emptyset(P) = L$ (ld. lentebb).

Veremautomatának nevezünk egy a $P = (Q,\, \Sigma,\, \Gamma,\, \delta,\, q_0,\, Z_0,\, F)$ rendszert, ahol

– Q állapotok véges halmaza;

– Σ a véges inputábécé;

– Γ a véges veremábécé;

– δ egy $Q \times (\Sigma \cup \{\lambda\}) \times \Gamma$-ból $Q \times \Gamma^*$ véges részhalmazaiba képező függvény, az átmenetfüggvény;

– $q_0 \in Q$ a kezdőállapot, $Z_0 \in \Gamma$ a verem-kezdőszimbólum, $F \subseteq Q$ a végállapotok halmaza.

A fenti P veremautomata egy konfigurációja egy $Q \times \Sigma^* \times \Gamma^*$-beli elem.

Konfigurációk között értelmezzük a $\vdash$ *rákövetkezési* relációt: $(p, x, X) \vdash (q, y, Y)$ pontosan akkor, ha az alábbiak valamelyike teljesül:

– $x = ay$ valamely $a \in \Sigma$ betűre, $X = \gamma X'$, $Y = \gamma_1 \dots \gamma_n X'$ valamely $X' \in \Gamma^*$, $\gamma, \gamma_1, \dots, \gamma_n \in \Gamma$ veremszimbólumokra, és $\delta(p, a, \gamma) \ni (q, \gamma_1 \dots \gamma_n)$;

– $x = y$, $X = \gamma X'$, $Y = \gamma_1 \dots \gamma_n X'$ valamely $X' \in \Gamma^*$ szóra, $\gamma, \gamma_1, \dots, \gamma_n \in \Gamma$ veremszimbólumokra és $\delta(p, \lambda, \gamma) \ni (q, \gamma_1 \dots \gamma_n)$.

A fenti P veremautomata által üres veremmel felismert nyelv

$$L_\emptyset(P) = \{w \in \Sigma^* : (q_0, w, Z_0) \vdash^* (q, \lambda, \lambda) \text{ valamely } q \in Q\text{-ra}\};$$

itt * a reflexív-tranzitív lezárás jele.

Nyelvek között értelmezzük a *shuffle-szorzás* műveletét is: ha K, L nyelvek, akkor legyen

$$K \sqcup\!\sqcup L = \{x_1 y_1 x_2 y_2 \dots x_n y_n : x_1 x_2 \dots x_n \in K,\ y_1 y_2 \dots y_n \in L,\ x_i,\, y_i \in \Sigma^*\}.$$

A shuffle-szorzás a konkatenációhoz hasonlóan asszociatív, így ennek a műveletnek a segítségével is definiálhatjuk nyelvek nemnegatív egész kitevőjű hatványait: ha L egy nyelv, legyen

- $L^{0\sqcup\sqcup} = \{\lambda\}$;
- $L^{(n+1)\sqcup\sqcup} = L^{n\sqcup\sqcup} \sqcup\sqcup L$.

A Kleene-iteráláthoz hasonlóan kapjuk egy L nyelv *shuffle-iteráltját*: $L^{*\sqcup\sqcup} = \bigcup_{i=0}^{\infty} L^{i\sqcup\sqcup}$.

Használni fogjuk továbbá a *legfeljebb n. shuffle-iterált* jelölését: $L^{\leq n\sqcup\sqcup} = \bigcup_{i=0}^{n} L^{i\sqcup\sqcup}$.

A *részszó* fogalmát többféleképp szokták definiálni. Ebben a cikkben ha x és y két szó, akkor mondjuk, hogy x részszava y-nak, ha létezik olyan w szó, melyre $y \in x \sqcup\sqcup w$.

A shuffle és a shuffle-iterált műveletekre számos jól ismert összefüggés áll fenn, a teljesség igénye nélkül néhány:

- A $\sqcup\sqcup$ művelet kommutatív, asszociatív.
- Tetszőleges K, L_1, L_2 nyelvekre $K \sqcup\sqcup (L_1 \cup L_2) = (K \sqcup\sqcup L_1) \cup (K \sqcup\sqcup L_2)$.
- Tetszőleges K, L nyelvekre $(K \cup L)^{*\sqcup\sqcup} = K^{*\sqcup\sqcup} \sqcup\sqcup L^{*\sqcup\sqcup}$.
- A reguláris nyelvek osztálya zárt a shuffle-szorzásra, nem zárt viszont a shuffle-iterált képzésére.
- A környezetfüggetlen nyelvek osztálya sem a shuffle-szorzásra, sem a shuffle-iterált képzésére nem zárt.

Ha $\Sigma = \{a_1, \ldots, a_k\}$ egy ábécé, és w egy Σ fölötti szó, akkor w *Parikh-képe*, $\Phi(w)$ az az $\mathcal{N}^k$-beli vektor, melynek i. koordinátáján $|w|_{a_i}$ áll.

Nyelvekre is kiterjesztjük a Parikh-függvényt: ha $L \subseteq \Sigma^*$, legyen

$$\Phi(L) = \{\Phi(w) : w \in L\}.$$

3. Elemi eredmények

Első vizsgálataink a legegyszerűbb olyan reguláris nyelvekre irányulnak, melyekre a kérdés nemtriviális: az egyetlen szóból álló nyelvekre. Kezdjük egy egyszerű állítással:

3.1. Állítás. *Legyen Σ tetszőleges ábécé és u, v_1, v_2, ..., $v_n \in \Sigma^*$ úgy, hogy $u \in v_1 \sqcup\sqcup v_2 \sqcup\sqcup \ldots \sqcup\sqcup v_n$. Ekkor $\Phi(u) = \Phi(v_1) + \Phi(v_2) + \cdots + \Phi(v_n)$.*

Az állítás néhány egyszerű következménye:

3.1. Következmény. *Tetszőleges L nyelvre $\Phi(L^{*\sqcup\sqcup}) = \Phi(L^*)$ (ezt úgy is mondjuk, hogy $L^{*\sqcup\sqcup}$ és L^* betűekvivalens).*

3.2. KÖVETKEZMÉNY. *Tetszőleges* u *és* v *szavakra ha* $u \in v^{*\sqcup\sqcup}$, *akkor* $\Phi(u) = k\Phi(v)$ *valamely* $k \in \mathcal{N}$*-re.*

Most már be tudjuk bizonyítani első eredményünket:

3.1. TÉTEL. *Tetszőleges* w *szóra* $w^{*\sqcup\sqcup}$ *pontosan akkor reguláris, ha* $w \in a^*$ *valamely* a *betűre.*

Bizonyítás. Az elegendőség triviális, hiszen ha $w \in a^*$, akkor $w^{*\sqcup\sqcup} = w^*$, ami reguláris nyelv. Belátjuk, hogy ha nem létezik olyan a betű, melyre $w \in a^*$, akkor $w^{*\sqcup\sqcup}$ nem reguláris. Legyen tehát $w = a^n by$ valamely $a \neq b \in \Sigma$, $n > 0$, $y \in \Sigma^*$-ra. Tegyük fel, hogy $w^{*\sqcup\sqcup}$ reguláris. Ekkor a pumpáló lemma szerint kell létezzen olyan $N \in \mathcal{N}$ konstans, melyre tetszőleges $u \in w^{*\sqcup\sqcup}$ esetén, ha $|u| > N$, úgy u felírható $u = u_1u_2u_3$ alakban oly módon, hogy a következők fennállnak:

- $|u_1u_2| \leq N$;
- $u_2 \neq \lambda$;
- tetszőleges $i \in \mathcal{N}$ esetén $u_1u_2^iu_3 \in w^{*\sqcup\sqcup}$.

Vegyük az $u = a^{nN}b^Ny^N \in w^{*\sqcup\sqcup}$ szót! E szóra a feltételek fennállnak. Tekintve bármely, a fenti három kikötésnek eleget tevő u_1, u_2, u_3 szót, az $|u_1u_2| \leq N$ feltétel miatt (mivel u első N betűje csupa a) $u_2 = a^j$ valamely $j > 0$-ra. Ekkor viszont $u' = u_1u_2^2u_3$ nem lehet $w^{*\sqcup\sqcup}$-beli szó, hiszen $|u'|_b = N|u|_b$, de $|u'|_a > N|w|_a$. Ebből következően nem létezik olyan $k \in \mathcal{N}$ szám, melyre $\Phi(u') = k\Phi(w)$. Ellentmondásra jutottunk, így $w^{*\sqcup\sqcup}$ valóban nem reguláris. □

Folytatva a $w^{*\sqcup\sqcup}$ alakban előálló nyelvek vizsgálatát, a következő eredmény is adódik:

3.2. TÉTEL. *Tetszőleges* w *szóra, melyben az* **alternálások** *száma legalább* 2 *(azaz bármely* a, b *betűket is tekintünk,* $w \notin a^*b^*$ *fennáll),* $w^{*\sqcup\sqcup}$ *nem lehet környezetfüggetlen.*

Bizonyítás. Írjuk fel w-t $w = a_1^{n_1}a_2^{n_2}\dots a_k^{n_k}$ alakban, ahol minden $1 \leq i < k$ esetén $a_i \neq a_{i+1}$, és minden $1 \leq i \leq k$ esetén $n_i > 0$. A feltétel szerint tehát ekkor $k > 2$. Tekintsük az $L = w^{*\sqcup\sqcup} \cap a_1^*a_2^*\dots a_k^*$ nyelvet. Könnyen láthatóan ekkor $L = \left\{a_1^{n_1N}a_2^{n_2N}\dots a_k^{n_kN} : N \in \mathcal{N}\right\}$. Mivel a környezetfüggetlen nyelvek osztálya zárt a reguláris nyelvvel való metszésre, elegendő belátni, hogy L nem környezetfüggetlen. Ehhez a Bar-Hillel lemmát alkalmazzuk: amennyiben L környezetfüggetlen nyelv, úgy léteznek p, $q > 0$ egészek úgy, hogy tetszőleges $u \in L$-beli szó, melyre $|u| > p$, felírható $u = u_1u_2u_3u_4u_5$ alakban úgy, hogy a következők fennállnak:

- $|u_2u_3u_4| < q$;
- $u_2u_4 \neq \lambda$;
- minden $i > 0$ egész esetén $u_1u_2^iu_3u_4^iu_5 \in L$.

Tegyük fel tehát, hogy L környezetfüggetlen és a p, q egészek rendelkeznek a fenti tulajdonsággal. Vegyük az $u = a_1^{(p+q)n_1} a_2^{(p+q)n_2} \dots a_k^{(p+q)n_k} \in L$ szót. Erre a szóra $|u| > p$, így a lemma szerint felírható $u = u_1u_2u_3u_4u_5$ alakban úgy, hogy a fenti három feltétel fennáll. Tekintsük az $u' = u_1u_2^2u_3u_4^2u_5$ szót! u konstrukciója és az $|u_2u_3u_4| < q$ feltétel miatt az $u_2u_3u_4$ szó legfeljebb két szomszédos blokkot érinthet (itt egy u-beli blokkon azonos betűk maximális hosszú sorát értjük, azaz blokkok az $a_1^{(p+q)n_1}$, $a_2^{(p+q)n_2}$,…,$a_k^{(p+q)n_k}$ szavak). Mivel w-ben az alternálások száma legalább 2, így u-ban legalább 3 blokk kell legyen; tekintve tehát az $u' = u_1u_2^2u_3u_4^2u_5$ szót, azt kapjuk, hogy u'-ben u-hoz képest legalább egy (mondjuk az i.) blokkban több a betűk száma (mert $u_2u_4 \neq \lambda$), viszont legalább egy (mondjuk a j.) blokkban ugyanannyi. Ez L fenti formája szerint nem lehetséges (ugyanis az i. és a j. blokk hosszának aránya $n_i : n_j$ kellene maradjon). Ellentmondást kaptunk, így sem L, sem $w^{*\sqcup\!\sqcup}$ nem környezetfüggetlen. $\square$

A fenti föltétel nemcsak szükséges, hanem elégendő is. Ennek igazolását egy általánosabb tételben végezzük el, amihez előbb ismét szükséges néhány állítást belátni. Ahhoz, hogy az állításokat ki tudjuk mondani, bevezetjük az alábbi jelöléseket:

3.1. Definíció. Legyen Σ ábécé. Definiáljuk minden $i \in \mathcal{N}$-re a

$$h_i : \Sigma^* \to (\Sigma \times \mathcal{N})^*$$

homomorfizmust a következőképp: minden $a \in \Sigma$ esetén legyen $h_i(a) = (a, i)$.

A fenti homomorfizmusra akkor lesz szükség, ha valamely L nyelvre és $w \in L^{*\sqcup\!\sqcup}$ szóra „nyomon akarjuk követni", a w szó L-beli „összetevőit". Továbbá definiáljuk az alábbi homomorfizmust is:

3.2. Definíció. Legyen Σ ábécé. Ekkor a $d : (\Sigma \times \mathcal{N})^* \to \Sigma^*$ homomorfizmus legyen a következő: minden $a \in \Sigma$ és $i \in \mathcal{N}$ esetén $d((a, i)) = a$.

Ezt a homomorfizmust pedig akkor használjuk, ha az előbbi homomorfizmuscsaláddal kapott szavakból vissza akarjuk kapni Σ^*-beli megfelelőjüket.

Ahhoz pedig, hogy egy L nyelv és egy $w \in L^{*\sqcup\!\sqcup}$ szó esetében tudjunk beszélni a w szó „i. komponenséről" (ami így messze nem egyértelmű, de ha a h_i homomorfizmusokkal jelölt komponensekről beszélünk, akkor az lesz), definiáljuk az alábbi homomorfizmus-családot is:

3.3. Definíció. Legyen Σ ábécé. Tetszőleges $i \in \mathcal{N}$ esetén legyen

$$s_i : (\Sigma \times \mathcal{N})^* \to \Sigma^*$$

a következő homomorfizmus: $a \in \Sigma$ és $j \in \mathcal{N}$ esetén

- $s_i((a, j)) = a$, ha $i = j$;
- $s_i((a, j)) = \lambda$, egyébként.

Megjegyzés. A két utolsó definícióban s a *select*, d pedig a *delete* szavak kezdőbetűjéből származik.

Az imént definiált homomorfizmusok és az iterált shuffle között az alábbi összefüggés áll fenn:

3.2. ÁLLÍTÁS. *Legyen Σ ábécé, $L \subseteq \Sigma^*$ tetszőleges nyelv, $w \in (\Sigma \times \mathcal{N})^*$ pedig egy szó. Ha minden $i \in \mathcal{N}$-re $s_i(w) \in L \cup \{\lambda\}$, akkor $d(w) \in L^{*\sqcup\sqcup}$.*

Mivel w pontosan akkor van $L^{*\sqcup\sqcup}$-ban, ha szét lehet particionálni diszjunkt L-beli részszavakra, az állítás nyilvánvaló. Továbbá, az alábbi megfordítás is igaz:

3.3. ÁLLÍTÁS. *Legyen Σ ábécé, $L \subseteq \Sigma^*$ egy nyelv és $w \in L^{*\sqcup\sqcup}$ egy szó. Ekkor létezik olyan $w' \in (\Sigma \times \mathcal{N})^*$ szó és $n \in \mathcal{N}$ szám, melyre minden $i \in \mathcal{N}$ esetén $s_i(w') \in L$, ha $i < n$; és $s_i(w') = \lambda$, ha $i \geq n$.*

Ezen bevezető definíciók után elkezdhetjük vizsgálni a következő nyelvosztályt. A továbbiakban jelölje $F = \{a^{k_1}b^{n_1}, a^{k_2}b^{n_2}, \ldots, a^{k_f}b^{n_f}\}$ az a^+b^+ nyelv tetszőleges véges részhalmazát.

3.4. ÁLLÍTÁS. *Legyen $w \in \{a, b\}^*$. $w \in F^{*\sqcup\sqcup}$ pontosan akkor teljesül, ha létezik olyan $w' \in (\{a, b\} \times \mathcal{N})^*$ szó és $n \in \mathcal{N}$, melyre az alábbiak fennállnak:*

- $d(w') = w$;
- $s_i(w') \in F$, *ha* $i < n$;
- $s_i(w') = \lambda$, *ha* $i \geq n$;
- *tetszőleges* $i < j < n$ *indexek esetén* (b, i) *összes előfordulása megelőzi* (b, j) *összes előfordulását.*

Bizonyítás. Ha létezik ilyen w' szó, úgy már az első három feltételből következik, hogy $w \in F^{*\sqcup\sqcup}$. Az ellentétes irányú tartalmazás igazolásához legyen $w \in F^{*\sqcup\sqcup}$. Ekkor léteznek olyan $u_0, u_1, \ldots, u_{n-1} \in F$ szavak úgy, hogy $w \in u_0 \sqcup\sqcup u_1 \sqcup\sqcup \ldots \sqcup\sqcup u_{n-1}$ ($w = \lambda$ esetén $n = 0$ választása adja az üres szorzatot). Ekkor létezik (esetleg több) olyan w' szó, melyre az alábbiak fennállnak:

- $w' \in h_0(u_0) \sqcup\sqcup h_1(u_1) \sqcup\sqcup \ldots \sqcup\sqcup h_{n-1}(u_{n-1})$ (azaz $s_0(w') = u_0, s_1(w') = u_1, \ldots, s_{n-1}(w') = u_{n-1}$, és minden más i-re $s_i(w') = \lambda$);
- w'-ben az első $(b, 0)$ megelőzi az első $(b, 1)$-et, az az első $(b, 2)$-t és így tovább (ezt a shuffle-szorzat kommutativitása miatt tehetjük fel);
- $d(w') = w$.

Legyen egy ilyen $w' \in (\Sigma \times \mathcal{N})^*$ szó esetén $Alt(w') \subseteq \mathcal{N} \times \mathcal{N}$ az a halmaz, melyben egy (i, j) elempár pontosan akkor van benne, ha a következő feltételek fennállnak:

- $i < j \leq |w'|$ (azaz i és j pozíciók w'-ben);
- $w'_i = (b, i')$ és $w'_j = (b, j')$ úgy, hogy $i' > j'$.

Legyen w' a fenti tulajdonságú szavak közül egy olyan, melyre $|Alt(w')|$ minimális. Állítjuk, hogy ekkor $Alt(w') = \emptyset$.

Tegyük fel ugyanis, hogy $Alt(w') \neq \emptyset$. Vegyük ekkor azt az $(l, r) \in Alt(w')$ párt, melyre r minimális és ezen belül l maximális. Tekintsük azt a w'' szót, melyet w'-ből kapunk az l. és az r. pozíción álló betűk felcserélésével. Ekkor $|Alt(w'')|$ könnyen láthatóan kisebb, mint $|Alt(w')|$; ha belátjuk tehát, hogy w'' is teljesíti a feltételeket, igazoltuk, hogy $Alt(w') = \emptyset$ valóban fennáll.

Először belátjuk, hogy w''-re továbbra is fennáll az, hogy ha $i < j < n$, akkor w''-ben az első (b, i) megelőzi az első (b, j)-t.

Valóban, tegyük fel, hogy w'' nem ilyen: előzze meg az első (b, j) az első (b, i)-t az $i < j$ számokra. Ebből következik, hogy kell legyen olyan i index is, melyre az első $(b, i+1)$ megelőzi az első (b, i)-t. Ekkor a következő esetek lehetségesek:

- Az első (b, i)-t cseréltük az első $(b, i+1)$-gyel. Ez nem lehet, hiszen $i < i+1$, ez a pár nem volt $Alt(w')$-ben.
- Az első (b, i)-t a szóban hátrafelé cseréltük, így az első $(b, i+1)$ mögé került. Ekkor egy i-nél kisebb i'-re egy (b, i') betűvel kellett cseréljük, mely az első $(b, i+1)$ után állt w'-ben. Viszont a cserélendő pár választási stratégiája miatt ekkor semmiképp nem cserélhettük (b, i)-vel ($(b, i+1)$ is jó lenne a cserére, és hátrább áll w'-ben). Így ez sem lehetséges.
- Az első $(b, i+1)$-et cseréltük a szóban egy, az első (b, i)-t megelőző olyan betűvel, mely (b, i') alakú, ahol $i' > i+1$. Ekkor w'-ben az első (b, i')-nek meg kellene előznie az első $(b, i+1)$-et, ami szintén nem lehetséges.

Tehát w''-ben továbbra is fennáll, hogy tetszőleges $i < j < n$ esetén az első (b, i) megelőzi az első (b, j)-t.

Most belátjuk azt is, hogy tetszőleges $i < n$-re az első (b, i) pozíciója ugyanaz w'-ben és w''-ben is. Tegyük fel ugyanis, hogy nem így van. Ekkor az alábbi esetek lehetségesek:

- Az első (b, i)-t cseréltük ki egy őt megelőző (b, j)-vel, ahol ezek szerint $j > i$. Ekkor az első (b, j)-nek értelemszerűen meg kellene előznie az első (b, i)-t, amiről az előbb megállapítottuk, hogy nem lehetséges.
- Egy (nem az első) (b, i)-t cseréltünk ki egy, az első (b, i)-t megelőző pozíción lévő betűvel (így a szóban előrébb került). Csakhogy ekkor az adott betűvel az első (b, i)-t is kicserélhettük volna; a választási stratégiából következően ekkor az első (b, i)-t kellett volna választanunk (vagy egy még előbb álló betűt). Ez tehát nem lehetséges.
- Az első (b, i)-t cseréltük ki egy, a szóban hátrébb álló betűvel, átugorva a második (esetleg több) (b, i)-t. Ez ismét nem lehet, hiszen a választási stratégia szerint az átugrott betűt kellett volna válasszuk (vagy egy még hátrébb álló betűt). Tehát ez sem lehetséges.

Összefoglalva: a w'-ből kapott w''-ben minden $0 \leq i < n$-re az első (b, i) pozíciója megmaradt. Így tetszőleges $0 \leq i < n$-re fennáll, hogy az utolsó (a, i) (amit nem cserélhettünk) még mindig megelőzi az első (b, i)-t w''-ben is; mivel a csere nem változtatja sem az (a, i), sem a (b, i) betűk számát, így $s_i(w'') = u_i$ továbbra is fennáll. Természetesen $d(w'') = d(w')$ is teljesül, hiszen két olyan betűt cseréltünk ki, melyek d melletti képe megyegyezik.

Tehát beláttuk, hogy ha egy, a feltételeket teljesítő w'-re $Alt(w')$ nemüres, akkor megadható olyan w'' szó is, mely szintén teljesíti a feltételeket, és melyre $|Alt(w'')| < |Alt(w')|$. Mivel az $Alt(w')$ halmaz véges minden w'-re, valóban kell létezzen olyan w', mely teljesíti a feltételeket, és melyre $Alt(w') = \emptyset$. □

Analóg gondolatmenettel belátható, hogy az (a, i)-k is rendezhetők ebben az értelemben; így végeredményben azt kapjuk, hogy $w \in F^{*\sqcup\sqcup}$ pontosan akkor teljesül egy $w \in \{a, b\}^*$ szóra, ha van olyan $w' \in (\{a, b\} \times \mathcal{N})^*$ szó és $n \in \mathcal{N}$ konstans, melyekre:

- $s_i(w') \in F$, ha $i < n$;
- $s_i(w') = \lambda$, ha $i \geq n$;
- tetszőleges $i < j < n$-re bármelyik (a, i) betű megelőzi az összes (a, j) betűt w'-ben;
- tetszőleges $i < j < n$-re bármelyik (b, i) betű megelőzi az összes (b, j) betűt w'-ben.

Jelöljön $w \in F^{*\sqcup\sqcup}$ esetén egy hozzá tartozó ilyen tulajdonságú $(\Sigma \times \mathcal{N})^*$-beli szót (mondjuk a lexikografikusan legkisebbet) $\hat{w}$.

Ez lehetővé teszi az $F^{*\sqcup\sqcup}$-beli szavak nemdeterminisztikus, egyfajta „mohó" módon történő felismerését egy P veremautomatával a következő módon: ha P a betűt olvas, azt beteszi a verembe. Amint elolvassa az első b betűt, nemdeterminisztikusan megsejti, hogy ez melyik $a^i b^j \in F$-beli szónak az első b betűje (vagyis: mivel lesz egyenlő $s_0(\hat{w})$, ahol w az input szó), a megfelelő számú a-t kiveszi a veremből, és a további $j - 1$ darab b-t (amik megfelelnek a többi $(b, 0)$-nak $\hat{w}$-ben) pedig a verem tartalmát nem változtatva olvassa be. Ezután ha találkozik még egy b-vel, ismét megsejti, hogy ez melyik F-beli szó első b-je (hogy mi $s_1(\hat{w})$), és így tovább. A beolvasás végén pontosan akkor ürül ki a verem, ha $w \in F^{*\sqcup\sqcup}$ valóban fennállt, és a $\hat{w}$-beli sorrendet is jól sejtettük meg.

A fenti veremautomata formális megadása a következő: $P = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, \emptyset)$, ahol

- $Q = \left[\max_{v \in F} |v|_a\right] \times \left[\max_{v \in F} |v|_b\right]$;
- $\Sigma = \{a, b\}$;
- $\Gamma = \{Z_0, a\}$;
- $q_0 = (0, 0)$;

- végül, δ a következő függvény:
  - $\delta((0,j),a,\gamma)=\{((0,j),a\gamma)\}$ minden $0\le j\le \max\limits_{v\in F}|v|_b$, $\gamma\in\Gamma$ esetén;
  - $\delta((0,j),b,\gamma)=\{(0,j-1),\gamma\}$ minden $0<j\le \max\limits_{v\in F}|v|_b$ $\gamma\in\Gamma$ esetén;
  - $\delta((0,0),b,\gamma)\ni((i,j-1),\gamma)$ minden $a^ib^j\in F$, $\gamma\in\Gamma$ esetén;
  - $\delta((i,j),\lambda,a)=\{((i-1,j),\lambda)\}$ minden $0<i\le \max\limits_{v\in F}|v|_a$, $0\le j\le$ $\le \max\limits_{v\in F}|v|_b$ esetén;
  - $\delta((0,0),\lambda,Z_0)=\{((0,0),\lambda)\}$.

Tehát az eddigiek szerint erre a P veremautomatára $L_\emptyset(P)=F^{*\sqcup\sqcup}$.

Eddig tehát beláttuk, hogy ha F véges részhalmaza a^+b^+-nak, akkor $F^{*\sqcup\sqcup}$ környezetfüggetlen. Ezt könnyen kiterjeszthetjük a^*b^*-ra is:

3.3. Tétel. *Ha F véges részhalmaza a^*b^*-nak, akkor $F^{*\sqcup\sqcup}$ környezetfüggetlen.*

Bizonyítás. F ekkor felírható $F=F_1\cup F_2\cup F_3$ alakban, ahol:

- $F_1\subseteq a^+b^+$;
- $F_2\subseteq a^*$;
- $F_3\subseteq b^*$.

Ekkor
$$F^{*\sqcup\sqcup}=(F_1\cup F_2\cup F_3)^{*\sqcup\sqcup}=F_1^{*\sqcup\sqcup}\sqcup\sqcup F_2^{*\sqcup\sqcup}\sqcup\sqcup F_3^{*\sqcup\sqcup}.$$
Azt már láttuk, hogy $F_1^{*\sqcup\sqcup}$ környezetfüggetlen; mivel $F_2^{*\sqcup\sqcup}=F_2^*$ és $F_3^{*\sqcup\sqcup}=F_3^*$ (ugyanis ábécéjük egyelemű), e két nyelv reguláris. Mivel pedig a környezetfüggetlen nyelvek osztálya zárt a reguláris nyelvvel való shuffle-szorzásra, így $F^{*\sqcup\sqcup}$ valóban környezetfüggetlen. $\square$

3.3. Következmény. *Tetszőleges $w\in a^*b^*$ szóra $w^{*\sqcup\sqcup}$ környezetfüggetlen.*

Maradva a^*b^* véges részhalmazainál, vizsgáljuk most azt, hogy milyen feltételek mellett lesz shufle-iteráltjuk reguláris! Erre a kérdésre az alábbi szükséges és elégséges feltételt adjuk:

3.4. Tétel. *Legyen $F\subseteq a^*b^*$ véges nyelv. $F^{*\sqcup\sqcup}$ pontosan akkor reguláris, ha az alábbi feltételek valamelyike fennáll:*

- *$a^+b^+\cap F=\emptyset$;*
- *$a^+\cap F$ és $b^+\cap F$ egyike sem üres.*

Bizonyítás. Amennyiben $a^+b^+\cap F=\emptyset$, természetesen $F^{*\sqcup\sqcup}$ reguláris: ekkor F felírható $F=F_1\cup F_2$ formában, ahol $F_1\subseteq a^*$ és $F_2\subseteq b^*$. Így $F^{*\sqcup\sqcup}=F_1^{*\sqcup\sqcup}\sqcup\sqcup F_2^{*\sqcup\sqcup}$, ami ismét két reguláris nyelv shuffle-szorzata, így reguláris.

Amennyiben $a^+ \cap F$ és $b^+ \cap F$ egyike sem üres, úgy létezik $i, j > 0$ úgy, hogy $a^i \in F$ és $b^j \in F$. Írjuk fel F-et $F_1 \cup F_2 \cup F_3$ alakban, ahol $F_1 \subseteq a^+b^+$, $F_2 \subseteq a^*$ és $F_3 \subseteq b^*$. Tekintsünk egy $a^l b^r \in F_1$ szót. Mivel

$$(a^l b^r)^{(ij)\sqcup\sqcup} \subseteq (a^l)^{(ij)\sqcup\sqcup} \sqcup\sqcup (b^r)^{(ij)\sqcup\sqcup} \subseteq (a^i)^* \sqcup\sqcup (b^j)^*,$$

így

$$(a^l b^r)^{*\sqcup\sqcup} \subseteq (a^l b^r)^{(\leq ij)\sqcup\sqcup} \sqcup\sqcup F_2^{*\sqcup\sqcup} \sqcup\sqcup F_3^{*\sqcup\sqcup}$$

fennáll. Ebből következően$F_1^{*\sqcup\sqcup} \subseteq (F_1)^{(\leq |F_1| ij)\sqcup\sqcup} \sqcup\sqcup F_2^{*\sqcup\sqcup} \sqcup\sqcup F_3^{*\sqcup\sqcup}$. Ezt visszahelyettesítve kapjuk, hogy

$$F^{*\sqcup\sqcup} = F_1^{*\sqcup\sqcup} \sqcup\sqcup F_2^{*\sqcup\sqcup} \sqcup\sqcup F_3^{*\sqcup\sqcup} \subseteq (F_1^{(\leq |F_1| ij)\sqcup\sqcup} \sqcup\sqcup F_2^{*\sqcup\sqcup} \sqcup\sqcup F_3^{*\sqcup\sqcup}) \sqcup\sqcup F_2^{*\sqcup\sqcup} \sqcup\sqcup F_3^{*\sqcup\sqcup} =$$
$$=((F_1)^{(\leq |F_1| ij)\sqcup\sqcup} \sqcup\sqcup F_2^{*\sqcup\sqcup} \sqcup\sqcup F_3^{*\sqcup\sqcup}) \subseteq F^{*\sqcup\sqcup}.$$

Így a teljes $F^{*\sqcup\sqcup}$ előáll $((F_1)^{(\leq |F_1| ij)\sqcup\sqcup} \sqcup\sqcup F_2^{*\sqcup\sqcup} \sqcup\sqcup F_3^{*\sqcup\sqcup})$ alakban – az első tényező reguláris nyelv véges shuffle-szorzatainak véges uniója, ami reguláris; a másik két tényező pedig egybetűs ábécé feletti nyelvek shuffle-iteráltja, ami reguláris. Mivel a reguláris nyelvek zártak a shuffle-szorzásra és a véges unió képzésére, kapjuk, hogy $F^{*\sqcup\sqcup}$ valóban reguláris.

Az elegendőséget tehát beláttuk; most igazoljuk a feltételek szükségességét. Tegyük fel tehát, hogy $a^+b^+ \cap F \neq \emptyset$ és (mondjuk) $a^+ \cap F = \emptyset$. Ekkor tehát létezik az $\epsilon := \min\limits_{u \in F} \frac{|u|_b}{|u|_a}$ hányados, mely pozitív (itt a 0 nevezőjű törtek értéke, beleértve a $\frac{0}{0}$-t is, ∞): mivel van a^+b^+-beli szó, a hányados véges és mert minden a-t tartalmazó szóban van b is, így a minimum nem 0.

Ebből következik, hogy – mivel $F^{*\sqcup\sqcup}$ betűekvivalens F^*-gal – ugyanez az ϵ a minimuma az $F^{*\sqcup\sqcup}$-beli u szavak esetében is az $\frac{|u|_b}{|u|_a}$ hányadosnak.

Ismét alkalmazzuk a pumpáló lemmát $L = F^{*\sqcup\sqcup} \cap a^*b^*$-ra (ami végtelen nyelv, hiszen $a^+b^+ \cap F \neq \emptyset$). Tegyük fel, hogy L reguláris. Legyen $N \in \mathcal{N}$ a pumpáló lemma szerint ekkor létező olyan konstans, melyre tetszőleges $u \in L$-re ha $|u| > N$, akkor u felírható $u = u_1u_2u_3$ alakban úgy, hogy $|u_1u_2| < N$, $u_2 \neq \lambda$, és minden $i \in \mathcal{N}$-re $u_1u_2^iu_3 \in L$.

Legyen $a^l b^r \in F \cap (a^+b^+)$! Ekkor $u = a^{lN}b^{rN}$ eleget tesz a feltételeknek; viszont tetszőleges $u = u_1u_2u_3$ felírást tekintve, melyre $|u_1u_2| < N$, az u_2 szó a^+-beli kell legyen. Ekkor viszont tekintve az $u^{(i)} := u_1u_2^iu_3$ szavakat, azt kapjuk, hogy $\lim\limits_{i \to \infty} |u^{(i)}|_a = \infty$, miközben $|u^{(i)}|_b$ konstans. Így tehát létezik olyan $u^{(i)}$, melyre $\frac{|u^{(i)}|_b}{|u^{(i)}|_a} < \epsilon$, és így ez az $u^{(i)}$ nem lehet L-beli. Ellentmondást kaptunk tehát, amiből következik, hogy sem L, sem $F^{*\sqcup\sqcup}$ nem reguláris. $\square$

Megjegyzés. A másik eset (amikor is $F \cap a^+b^+ \neq \emptyset$ és $F \cap b^+ = \emptyset$) bizonyítása analóg, azzal a különbséggel, hogy nemcsak a hányadosokat kell megfordítani, hanem a pumpáló lemma azon változatát kell alkalmazni, amikor is az $|u_1u_2| < N$ feltétel helyett az $|u_2u_3| < N$ feltétel szerepel.

4. Kommutatív nyelvek

Mivel tetszőleges kommutatív nyelvekre igaz, hogy shuffle-szorzatuk a legszűkebb kommutatív nyelv, mely tartalmazza konkatenáltjukat, továbbá tetszőleges kommutatív nyelv shuffle-iteráltja a legszűkebb kommutatív nyelv, mely tartalmazza a nyelv Kleene-iteráltját, erős kapcsolat áll fenn a kommutatív nyelvek és a shuffle műveletek között. Ezért e fejezetben kommutatív reguláris nyelvek shuffle-iteráltját vizsgáljuk. Annak eldöntésére, hogy egy adott L kommutatív nyelv shuffle-iteráltja pontosan mikor reguláris, már ismert algoritmus, lásd [4, 7]. Azonban bizonyos szűkebb nyelvosztályokra sikerült könnyen ellenőrizhető feltételeket adni; továbbá vannak környezetfüggetlenséget biztosító feltételeink is.

A legegyszerűbb kommutatív nyelvek ($\emptyset$ és $\{\lambda\}$ után) azok, melyek valamely egyszavas nyelv kommutatív lezártjaként allnak elő.

Az alábbi állítás igaz kommutatív nyelvekre:

4.1. Állítás. *Legyen Σ ábécé, $L \subseteq \Sigma^*$ kommutatív nyelv és $x \in \Sigma^*$ szó. Az $x \in L^{*\sqcup\sqcup}$ tartalmazás pontosan akkor áll fenn, ha léteznek $u_1, \ldots, u_n \in L$ szavak úgy, hogy $\Phi(x) = \sum_{i=1}^{n} \Phi(u_i)$.*

Bizonyítás. Ha $w \in L^{*\sqcup\sqcup}$, akkor $w \in u_1 \sqcup\sqcup \ldots \sqcup\sqcup u_n$ valamely $u_1, \ldots, u_n \in L$ szavakra. Ekkor természetesen $\Phi(x) = \sum_{i=1}^{n} \Phi(u_i)$ is fennáll. A másik irány bizonyítását n szerinti indukcióval végezzük el:

- Ha $n = 0$, azaz $\Phi(x) = \underline{0}$, úgy $x = \lambda$ és ekkor $x \in L^{*\sqcup\sqcup}$ bármely L-re igaz.
- Ha $n = 1$, azaz $\Phi(x) = \Phi(u_1)$ valamely $u_1 \in L$-re, akkor L kommutativitása miatt igaz az állítás.
- Tegyük fel, hogy az állítás igaz minden olyan szóra, melynek Parikh-képe felírható legfeljebb n darab L-beli szó Parikh-képének összegeként. Legyen $x \in \Sigma^*$-ra $\Phi(x) = \sum_{i=1}^{n+1} \Phi(u_i)$ valamely $u_1, \ldots, u_{n+1} \in L$ szavakra. Ekkor x felbontható két diszjunkt részszóra, x_1-re és x_2-re úgy, hogy $\Phi(x_1) = \sum_{i=1}^{n} \Phi(u_i)$ és $\Phi(x_2) = \Phi(u_{n+1})$. Az indukciós feltevés szerint ekkor $x_1 \in L^{*\sqcup\sqcup}$ és $x_2 \in L^{*\sqcup\sqcup}$ is fennáll. Mivel így x felbomlik két $L^{*\sqcup\sqcup}$-beli részszóra, így $x \in L^{*\sqcup\sqcup} \sqcup\sqcup L^{*\sqcup\sqcup} = L^{*\sqcup\sqcup}$ is fennáll. □

Speciálisan ha $L = \Phi^{-1}\Phi(w)$ valamely w-re, ezt az állítást a következő formában mondhatjuk ki:

4.2. Állítás. *Legyen Σ ábécé és $x, w \in \Sigma^*$. Az $x \in \left(\Phi^{-1}\Phi(w)\right)^{*\sqcup\sqcup}$ tartalmazás pontosan akkor áll fenn, ha $\Phi(x) = k\Phi(w)$ valamely $k \in \mathcal{N}$-re.*

Ezt felhasználva a következő eredmény született:

4.1. TÉTEL. *Legyen w tetszőleges szó.* $(\Phi^{-1}\Phi(w))^{*\sqcup\sqcup}$ *pontosan akkor reguláris, ha* $|Alph(w)| \leq 1$.

Bizonyítás. Amennyiben $|Alph(w)| \leq 1$, úgy $w \in a^*$ valamely a betűre; erre

$$(\Phi^{-1}\Phi(w))^{*\sqcup\sqcup} = \{w\}^{*\sqcup\sqcup},$$

amire már láttuk, hogy reguláris nyelv.

Legyen $w \in a_1^{n_1} a_2^{n_2} \ldots a_k^{n_k}$ valamely $a_1, \ldots, a_k$ betűkre, ahol $a_i \neq a_{i+1}$ semmilyen $1 \leq i < k$ esetén, $k > 1$; és az n_i kitevők pozitívak, minden $1 \leq i \leq k$ esetén. Tegyük fel, hogy $L = \left(\Phi^{-1}\Phi(w)\right)^{*\sqcup\sqcup}$ reguláris nyelv, és legyen N a pumpáló lemma állításában szereplő, L-re jellemző konstans! Tekintsük az

$$u = a_1^{Nn_1} a_2^{Nn_2} \ldots a_k^{Nn_k} \in L$$

szót. Mivel $|u| > N$, a pumpáló lemma szerint u felírható $u = u_1 u_2 u_3$ alakban úgy, hogy $|u_1 u_2| < N$, $u_2 \neq \lambda$, és minden $i > 0$ esetén $u_1 u_2^i u_3 \in L$ is fennáll. Vegyük észre, hogy itt u_2 csak a_1 betűket tartalmazhat. Tekintve az $u_1 u_2^2 u_3$ szót, azt látjuk, hogy u-hoz képest csak a_1 betűből tartalmaz többet, így nem lehet, hogy $\Phi(u)$ és $\Phi(u_1 u_2^2 u_3)$ egyszerre legyen $k\Phi(w)$ alakú (hiszen w tartalmaz $a_2 \neq a_1$ betűt is). Tehát $u_1 u_2^2 u_3 \notin L$; ezzel ellentmondást kaptunk, és így beláttuk, hogy L tényleg nem reguláris nyelv. □

Ugyanezen nyelvekre a környezetfüggetlenséggel kapcsolatosan egy hasonló eredmény áll fenn:

4.2. TÉTEL. *Legyen w tetszőleges szó.* $(\Phi^{-1}\Phi(w))^{*\sqcup\sqcup}$ *pontosan akkor környezetfüggetlen, ha* $Alph(w) \leq 2$.

Bizonyítás. Legyen $Alph(w) = \{a_1, \ldots, a_k\}$, ahol $k > 2$. Vegyük először is az

$$L_1 = \left(\Phi^{-1}\Phi(w)\right)^{*\sqcup\sqcup} \cap a_1^* a_2^* \ldots a_k^*$$

nyelvet. Erre a nyelvre

$$L_1 = \left\{ a_1^{n|w|_{a_1}} a_2^{n|w|_{a_2}} \ldots a_k^{n|w|_{a_k}} : n \in \mathcal{N} \right\}.$$

Alkalmazzuk most L_1-re azt a

$$h : \{a_1, \ldots, a_k\}^* \to \{a_1, a_2, a_3\}^*$$

homomorfizmust, melyre

$$h(a_1) = a_1, \quad h(a_2) = a_2, \quad h(a_3) = a_3,$$

és minden $i > 3$ esetben $h(a_i) = \lambda$. Ezután $h(L_1)$-re alkalmazzuk azon h_2 homomorfizmus inverzét, melyre $h_2(a_i) = a_i^{|w|_{a_i}}$, minden $1 \leq i \leq 3$ esetben. Ekkor

$$h_2^{-1} h(L_1) = \{a_1^n a_2^n a_3^n : n \in \mathcal{N}\},$$

ami ismerten nem környezetfüggetlen nyelv. Mivel pedig a reguláris nyelvvel való metszés, a homomorfizmus alkalmazása és az inverz homomorfizmus képzése mind olyan műveletek, melyekre a környezetfüggetlen nyelvek osztálya zárt, így az eredeti $\left(\Phi^{-1}\Phi(w)\right)^{*\sqcup\!\sqcup}$ nyelv sem környezetfüggetlen.

Az elegendőség igazolásához az általánosság megszorítása nélkül feltehetjük, hogy $Alph(w) = \{a, b\}$ (ha $Alph(w) < 2$, úgy a nyelv reguláris, tehát környezetfüggetlen). Legyen $|w|_a = n$, $|w|_b = m$. Ekkor

$$L = \left(\Phi^{-1}\Phi(w)\right)^{*\sqcup\!\sqcup} = \{x : m|x|_a = n|x|_b\}\,.$$

Ezt a nyelvet pedig fel lehet ismerni egy olyan veremautomatával, mely vermét mint egészértékű számlálót használja úgy, hogy a beolvasásakor m-mel növeli a számláló értékét, b beolvasásakor pedig n-nel csökkenti; ha a szónak vége és a számláló értéke 0, felismeri, egyébként elveti azt. A következő veremautomata így működik: $P = (Q, \{a, b\}, \Gamma, \delta, q_0, Z_0, \emptyset)$, ahol:

- $Q = \{-n, -n + 1, \ldots, -1, 0, 1, \ldots, m\}$;
- $q_0 = 0$;
- $\Gamma = \{Z_0, a, b\}$;
- a δ átmenetfüggvény pedig a következő:
  - $\delta(0, a, \gamma) = \{(m, \gamma)\}$ minden $\gamma \in \Gamma$ esetén;
  - $\delta(0, b, \gamma) = \{(-n, \gamma)\}$ minden $\gamma \in \Gamma$ esetén;
  - $\delta(i, \lambda, b) = \{(i - 1, \lambda)\}$ minden $i > 0$ esetén;
  - $\delta(i, \lambda, \gamma) = \{(i - 1, a\gamma)\}$ minden $i > 0$, $\gamma \in \{a, Z_0\}$ esetén;
  - $\delta(i, \lambda, a) = \{(i + 1, \lambda)\}$ minden $i < 0$ esetén;
  - $\delta(i, \lambda, \gamma) = \{(i + 1, b\gamma)\}$ minden $i < 0$, $\gamma \in \{b, Z_0\}$ esetén;
  - $\delta(0, \lambda, Z_0) = \{(0, \lambda)\}$.

A verem a következőképpen reprezentálja a kívánt számlálót: ha a verem tartalma $a^n Z_0$, akkor értéke n; ha $b^n Z_0$, akkor értéke $-n$. Látható, hogy P csak a 0 állapotban olvashat be betűt; az is látható, hogy ha P egy $i < 0$ állapotba kerül, akkor i-vel csökkentve a számláló értékét kerül vissza a 0 állapotba, ha pedig egy $i > 0$ állapotba kerül, akkor i-vel növeli a számlálót, és úgy kerül vissza a 0 állapotba. Továbbá mivel a betű olvasásakor eszerint m-mel növeli, b olvasásakor pedig n-nel csökkenti a számláló értékét, valóban a kívánt módon kezeli a számlálóját. Amennyiben felismeri az x szót (ez pontosan akkor következik be, ha a szó elolvasása után kiveszi a veremből az egyetlen ott lévő szimbólumot, Z_0-t), úgy a számláló értéke 0 kell legyen x végigolvasása után. Ez valóban megfelel annak, hogy $m|x|_a = n|x|_b$.

Tehát $L_\emptyset(P) = L$, azaz L tényleg környezetfüggetlen. $\square$

A kommutatív nyelvek osztályán belüli osztályokat alkotnak a lokálisan $(1, m)$-küszöbtesztelhető (vagy *kommutatív csillagmentes*) nyelvek; a továbbiakban ezekkel foglalkozunk. Innen úgy tekintjük, hogy $\Sigma = \{a_1, \ldots, a_k\}$ adott.

4.1. Definíció. Legyen $m \in \mathcal{N}$. Definiáljuk a $\Phi_m : \Sigma^* \to \mathcal{N}^k$ függvényt a következőképpen: tetszőleges w szóra $\Phi_m(w)[i]$ (azaz a $\Phi_m(w)$ vektor i. koordinátája) legyen $\min\{m, \Phi(w)[i]\}$.

4.2. Definíció. Egy $L \subseteq \Sigma^*$ kommutatív nyelv lokálisan $(1, m)$-küszöbtesztelhető, ha megadhatók olyan $v_1, \ldots, v_n \in \mathcal{N}^k$ vektorok, melyekre tetszőleges $w \in \Sigma^*$ esetén $w \in L$ pontosan akkor áll fenn, ha valamely $1 \leq i \leq n$ esetén $\Phi_m(w) = v_i$.

4.3. Definíció. A lokálisan $(1, 1)$-küszöbtesztelhető nyelveket lokálisan 1-tesztelhető nyelveknek is nevezzük.

A lokálisan 1-tesztelhető nyelvekre igaz az alábbi állítás:

4.3. Állítás. *Egy $L \subseteq \Sigma^*$ nyelv pontosan akkor lokálisan 1-tesztelhető, ha megadhatóak $\Sigma_1, \Sigma_2, \ldots, \Sigma_n \subseteq \Sigma$ ábécék úgy, hogy tetszőleges $w \in \Sigma^*$ szó pontosan akkor van benne L-ben, ha valamely $1 \leq i \leq n$-re $Alph(w) = \Sigma_i$.*

Jelölje $\Sigma_1 \subseteq \Sigma$ esetén L_{Σ_1} azt a nyelvet, melyben pontosan azok a szavak vannak, amiknek ábécéje Σ_1. Azaz:

$$L_{\Sigma_1} = \bigcap_{a \in \Sigma_1} \Sigma_1^* a \Sigma_1^*;$$

láthatóan minden L_{Σ_1} alakú nyelv reguláris.

Igaz továbbá, hogy tetszőleges $\Sigma_1 \subseteq \Sigma$-ra $L_{\Sigma_1} ⧢ L_{\Sigma_1} \subseteq L_{\Sigma_1}$, hiszen ha két olyan szót fésülünk össze, melyeknek ábécéje Σ_1, egy olyan szót kapunk, melynek ábécéje ismét Σ_1.

Ebből következően $L_{\Sigma_1}^{*⧢} = \{\lambda\} \cup L_{\Sigma_1}$, tetszőleges $\Sigma_1 \subseteq \Sigma$ esetén.

Az eddigiekből kapjuk az alábbi egyszerű eredményt:

4.4. Állítás. *Tetszőleges L lokálisan 1-tesztelhető nyelv shuffle-iteráltja reguláris.*

Bizonyítás. Legyen

$$L = L_{\Sigma_1} \cup L_{\Sigma_2} \cup \cdots \cup L_{\Sigma_n}.$$

Ekkor

$$\begin{aligned} L^{*⧢} &= L_{\Sigma_1}^{*⧢} ⧢ L_{\Sigma_2}^{*⧢} ⧢ \ldots ⧢ L_{\Sigma_n}^{*⧢} = \\ &= (L_{\Sigma_1} \cup \{\lambda\}) ⧢ (L_{\Sigma_2} \cup \{\lambda\}) ⧢ \ldots ⧢ (L_{\Sigma_n} \cup \{\lambda\}), \end{aligned}$$

azaz előáll reguláris nyelvekből a véges unió képzésével és a shuffle-szorzással; mivel a reguláris nyelvek osztálya zárt ezen műveletekre, így $L^{*⧢}$ reguláris. $\square$

Ennél azonban erősebb összefüggést is kimondhatunk, amihez az alábbi állítás szükséges:

4.5. Állítás. *Ha* $\Sigma_1, \ldots, \Sigma_n \subseteq \Sigma$, *akkor* $L = L_{\Sigma_1} \sqcup\sqcup \ldots \sqcup\sqcup L_{\Sigma_n}$ *lokálisan* $(1, n)$-*küszöbtesztelhető nyelv.*

Bizonyítás. Legyen minden $1 \leq i \leq n$ esetén v_i a Σ_i halmaz karakerisztikus vektora (azaz $v_i[j] = 1$, ha $a_j \in \Sigma_i$, egyébként $v_i[j] = 0$). Legyen $v = \sum_{i=1}^{n} v_i$. Állítjuk, hogy L pontosan azon w szavakat tartalmazza, melyekre minden $1 \leq i \leq k$ esetén:

- vagy $v[i] = 0$ és $|w|_{a_i} = 0$;
- vagy $v[i] > 0$ és $|w|_{a_i} \geq v[i]$.

Ennek bizonyítását teljes indukcióval végezzük n szerint:

- Ha $n = 0$, akkor L csak λ-t tartalmazza, és ez minden i-re megfelel a $v[i] = 0$ és $|\lambda|_{a_i} = 0$ eseteknek.
- Ha $n = 1$, akkor egyetlen $\Sigma_1 \subseteq \Sigma$-ról van szó; ekkor v minden koordinátáján a 0 vagy 1 érték szerepel, és épp az L_{Σ_1} definíciójában szereplő feltételeket kapjuk.
- Tegyük föl, hogy az állítás igaz n-ig. Legyen $\Sigma_1, \ldots, \Sigma_{n+1} \subseteq \Sigma$, és tegyük fel, hogy $w \in L_{\Sigma_1} \sqcup\sqcup \ldots \sqcup\sqcup L_{\Sigma_{n+1}}$. Ekkor w két diszjunkt részszóra, w_1-re és w_2-re bontható úgy, hogy $w_1 \in L_{\Sigma_1} \sqcup\sqcup \ldots \sqcup\sqcup L_{\Sigma_n}$ és $w_2 \in L_{\Sigma_{n+1}}$. Vizsgáljuk meg minden $1 \leq i \leq k$ koordináta esetén $v[i]$ és $|w|_{a_i}$ viszonyát! Az alábbi esetek lehetségesek (ismét v_j jelöli Σ_j karakterisztikus vektorát):
- Ha $\sum_{j=1}^{n} v_j[i] = 0$ és $v_{n+1}[i] = 0$, akkor $|w_1|_{a_i} = 0$ (indukciós feltevés szerint) és $|w_2|_{a_i} = 0$ ($L_{\Sigma_{n+1}}$ definíciója szerint); ekkor $v[i] = 0$ és $|w|_{a_i} = 0$, az állítás igaz.
- Ha $\sum_{j=1}^{n} v_j[i] = 0$ és $v_{n+1}[i] = m > 0$, akkor $|w_1|_{a_i} = 0$ (indukciós feltevés szerint) és $|w_2|_{a_i} \geq m$ ($L_{\Sigma_{n+1}}$ definíciója szerint); ekkor $v[i] = m > 0$ és $|w|_{a_i} \geq m$, az állítás igaz.
- Ha $\sum_{j=1}^{n} v_j[i] = m > 0$ és $v_{n+1}[i] = 0$, akkor $|w_1|_{a_i} \geq m$ (indukciós feltevés szerint) és $|w_2|_{a_i} = 0$ ($L_{\Sigma_{n+1}}$ definíciója szerint); ekkor $v[i] = m > 0$ és $|w|_{a_i} \geq m$, az állítás igaz.
- Ha $\sum_{j=1}^{n} v_j[i] = m_1 > 0$ és $v_{n+1}[i] = m_2 > 0$, akkor $|w_1|_{a_i} \geq m_1$ (indukciós feltevés szerint) és $|w_2|_{a_i} \geq m_2$ ($L_{\Sigma_{n+1}}$ definíciója szerint); ekkor $v[i] = m_1 + m_2 > 0$ és $|w|_{a_i} \geq m_1 + m_2$, az állítás igaz.

Azaz beláttuk, hogy ha $w \in L$, akkor valóban állnak rá a feltételek.

Tegyük most fel, hogy w Parikh-képe megfelelő, azaz

$$\Phi(w) \geq v = \sum_{j=1}^{n+1} v_j,$$

és mindahányszor $v[i] = 0$, $\Phi(w) = 0$ is fennáll. Bontsuk Σ_{n+1}-et két diszjunkt részhalmaz uniójára, $S_1 \cup S_2$-re a következőképp: legyen

$$S_1 = \left(\bigcup_{i=1}^{n} \Sigma_i \right) \cap \Sigma_{n+1}$$

és $S_2 = \Sigma_{n+1} \backslash S_1$. Bontsuk most w-t két diszjunkt részszóra, w_1-re és w_2-re úgy, hogy w_2 minden S_1-beli betűből pontosan 1-et tartalmazzon, és tartalmazza az összes S_2-beli betűt! Könnyen látható, hogy $w_2 \in L_{\Sigma_{n+1}}$ fennáll; tekintve a $v' = \sum_{j=1}^{n} v_j$ vektort, igaz, hogy $\Phi(w_1)[i] \geq v'[i]$ minden $1 \leq i \leq k$-ra, és ha $v'[i] = 0$, akkor $\Phi(w_1)[i] = 0$, ugyanis:

- ha $a_i \notin \Sigma_{n+1}$, akkor $v[i] = v'[i]$ és w_2-be sem került egyetlen a_i sem a konstrukció miatt, így az egyenlőség fennáll;
- ha $a_i \in \Sigma$ és $a_i \notin \bigcup_{j=1}^{n} \Sigma_j$, akkor az összes a_i betű w_2-be kerül; ekkor $v'[i] = 0$ és $|w_1|_{a_i} = 0$;
- egyébként w_2-be pontosan egy a_i kerül, azaz $|w_1|_{a_i} = |w|_{a_i} - 1$; mivel $v_{n+1}[i] = 1$ ekkor, így $v'[i] = v[i] - 1$ is fennáll; és mivel $a_i \in \bigcup_{j=1}^{n} \Sigma_j$, így $v'[i] > 0$ is fennáll; összegezve $v'[i] > 0$ és $|w_1|_{a_i} \geq v'[i]$.

Mindezekből következően w_1-re az indukciós feltevés miatt valóban mondhatjuk, hogy $w_1 \in L_{\Sigma_1} ⧢ \ldots ⧢ L_{\Sigma_n}$, továbbá hogy $w_2 \in L_{\Sigma_{n+1}}$, így valóban, $w \in L$. □

Az állítás következménye pedig:

4.1. Következmény. *Ha* $\Sigma_1, \ldots, \Sigma_n \subseteq \Sigma$, *akkor* $L_{\Sigma_1} ⧢ L_{\Sigma_2} ⧢ \ldots ⧢ L_{\Sigma_n}$ *egy lokálisan* $(1, n)$*-küszöbtesztelhető nyelv.*

Mivel lokálisan $(1, n)$-küszöbtesztelhető nyelvek uniója is lokálisan $(1, n)$-küszöbtesztelhető, és felhasználva, hogy a Σ ábécének legfeljebb $2^{|\Sigma|}$ különböző részhalmaza lehet, kapjuk az alábbi, a regularitásnál erősebb állítást:

4.6. Állítás. *Tetszőleges* $L \subseteq \{a_1, \ldots, a_k\}^*$ *lokálisan* 1*-tesztelhető nyelvre igaz, hogy* $L^{*⧢}$ *egy lokálisan* $(1, 2^k)$*-küszöbtesztelhető nyelv.*

Megjegyzés. A korlát nem éles; valójában ha $k > 1$, bebizonyítható, hogy $L^{*⧢}$ egy lokálisan $(1, k-1)$-küszöbtesztelhető nyelv (ez már egy éles korlát).

A lokálisan 1-tesztelhető nyelvek után most egy olyan kommutatív reguláris nyelvosztályt vizsgálunk, mely abból a szempontból „érdekesebb", mint az eddigiek, hogy zárt a shuffle-iterált képzésére.

4.4. Definíció. Legyen $\Sigma = \{a_1, \ldots, a_k\}$ és $N > 0$. Definiáljuk az $N - \mathcal{MOD}$ nyelvosztályt a következőképp: egy L nyelv pontosan akkor legyen benne $N - \mathcal{MOD}$-ban, ha létezik $\underline{c} \in \mathcal{N}^k$ vektor úgy, hogy

$$L = \{w \in \Sigma^* : \Phi(w) = \underline{c} + \lambda_1 N e_1 + \lambda_2 N e_2 + \cdots + \lambda_k N e_k\}$$

valamely $\lambda_1, \ldots, \lambda_k > 0$-ra. Itt e_i az i-edik egységvektort jelenti. Ekkor azt is mondjuk, hogy $\underline{c}$ az L nyelv *bázisa*.

4.5. Definíció. Legyen $N \in \mathcal{N}$ esetén $N - \mathcal{UMOD}$ az $N - \mathcal{MOD}$ osztály véges unióképzésre vett lezártja.

Ezen nyelvosztályokkal kapcsolatban kimondunk néhány állítást.

4.7. Állítás. *Legyen $N \in \mathcal{N}$ és K, L egy-egy $N - \mathcal{MOD}$ nyelv; K bázisa legyen $\underline{c}$, L bázisa pedig $\underline{d}$. Ekkor $K \sqcup\!\sqcup L$ is $N - \mathcal{MOD}$ nyelv, bázisa $\underline{c} + \underline{d}$.*

Bizonyítás. Legyen $w \in u \sqcup\!\sqcup v$ az $u \in K$, $v \in L$ szavakra. Ekkor

$$\Phi(u) = \underline{c} + NI\nu \text{ és } \Phi(v) = \underline{d} + NI\mu$$

valamely $\nu, \mu \in \mathcal{N}^k$ vektorokra (itt I a k-dimenziós egységmátrix). Tehát

$$\Phi(w) = \underline{c} + NI\nu + \underline{d} + NI\mu = \underline{c} + \underline{d} + NI(\nu + \mu),$$

és ekkor valóban fennáll, hogy w benne van az adott bázisú $N - \mathcal{MOD}$ nyelvben. Ha pedig w benne van e nyelvben, úgy $\Phi(w) = \underline{c} + \underline{d} + NI(\mu)$ valamely $\mu \in \mathcal{N}^k$-ra; ekkor szétbontható két diszjunkt részszóra, w_1-re és w_2-re úgy, hogy $\Phi(w_1) = \underline{c}$ (és így $w_1 \in K$), és $\Phi(w_2) = \underline{d} + NI\mu$ (azaz $w_2 \in L$). □

Ennek az egyszerű állításnak a következménye:

4.2. Következmény. *Ha L egy $N - \mathcal{MOD}$ nyelv a $\underline{c}$ bázissal, akkor*

$$L^{(N+1)\sqcup\!\sqcup} \subseteq L.$$

Bizonyítás. Az előző állítás szerint ekkor $L^{(N+1)\sqcup\!\sqcup}$ is egy $N - \mathcal{MOD}$ nyelv, az $(N+1)\underline{c}$ bázissal. Csakhogy ha $\Phi(w) = (N+1)\underline{c} + NI\mu$ valamely $\mu \in \mathcal{N}^k$-ra, akkor $N\underline{c}$-t kiemelve $\Phi(w) = \underline{c} + NI(\mu + \underline{c})$, azaz $w \in L$ is fennáll.

Ebből az eredményből következik, hogy ha $L \in N - \mathcal{MOD}$, akkor

$$L^{*\sqcup\!\sqcup} = \bigcup_{i=0}^{N} L^{i\sqcup\!\sqcup}, \text{ vagyis } L^{*\sqcup\!\sqcup} \in N - \mathcal{UMOD}.$$

Legyen most $L \in N - \mathcal{UMOD}$ tetszőleges! Ekkor L felírható

$$L = \bigcup_{i=0}^{n} L_i$$

alakban valamely $n \geq 0$-ra és $L_1, \ldots, L_n \in N - \mathcal{MOD}$ nyelvekre. Felírva és kifejtve L shuffle-iteráltját, kapjuk, hogy

$$L^{*\sqcup\sqcup} = \Big(\bigcup_{i=0}^{n} L_i\Big)^{*\sqcup\sqcup} = \coprod_{i=1}^{n} L_i^{*\sqcup\sqcup}.$$

□

Az előzőek szerint az $L_i^{*\sqcup\sqcup}$ nyelvek $N - \mathcal{UMOD}$-beliek, azaz felírhatóak

$$L_i = \bigcup_{j=0}^{n_i} L_{i,j}$$

alakban. Erre a disztributivitást alkalmazva

$$\coprod_{i=1}^{n} L_i^{*\sqcup\sqcup} = \coprod_{i=1}^{n} \bigcup_{j=0}^{n_i} L_{i,j} = \bigcup_{\substack{1 \leq i_j \leq n_j \\ \text{minden } 1 \leq j \leq n\text{-re}}} \coprod_{j=1}^{n} L_{j,i_j}.$$

Az unióban $N - \mathcal{MOD}$-beli nyelvek véges shuffle-szorzatai állnak, amik szintén $N - \mathcal{MOD}$-beli nyelvek; így $L^{*\sqcup\sqcup}$ végeredményben $N - \mathcal{MOD}$-beli nyelvek véges uniójaként áll elő, azaz $L^{*\sqcup\sqcup} \in N - \mathcal{UMOD}$.

Így igazoltuk a következőt:

4.8. Állítás. *Tetszőleges $N \in \mathcal{N}$-re $N - \mathcal{UMOD}$ zárt a shuffle-iterált és a shuffle-szorzás műveletekre.*

Hivatkozások

[1] T. Araki, Y. Tsujino, N. Tokura: *Descriptive power of flow expressions and event expressions.* Systems, Computers, Controls **11** (1980) 76–83

[2] T. Araki, N. Tokura: *Decision problems for regular expressions with shuffle and shuffle closure operators.* Systems, Computers, Controls **12** (1981) 1069–1073

[3] Z. Ésik, I. Simon: *Modeling literal morphisms by shuffle.* Semigroup Forum **56** (1998) 225–227

[4] P. Gohon: *An algorithm to decide whether a rational subset of $\mathcal{N}^k$ is recognizable.* Theoretical Computer Science **41** (1985) 51–59

[5] Y. Hirshfeld: *Undecidability of language equivalence for generalized regular expressions.* Fundamenta Informaticae **26** (1996) 95–102

[6] J. E. Hopcroft, J. D. Ullman: *Introduction to automata theory, languages and computation.* Addison-Wesley (1979)

[7] B. Imreh, M. Ito, M. Katsura: *On shuffle closures of commutative regular languages.* Proceedings of DMTS'96, Combinatorics, Complexity, and Logic. Springer-Verlag (1997)

[8] K. Iwama: *The universe problem for unrestricted flow languages.* Acta Informatica **19** (1983) 85–96

[9] M. Jantzen: *The power of synchronizing operations on strings.* Theoretical Computer Science **14** (1981) 127–154

[10] : J. Jędrzejowicz: *Infinite hierarchy of expressions containing shuffle closure operator.* Information Processing Letters **28** (1988) 33–37

[11] J. Jędrzejowicz: *An undecidable problem for shuffle languages.* Developments in Language Theory II (1995) 112–118

[12] J. Jędrzejowicz: *Undecidability results for shuffle languages.* Journal of Automata, Languages and Combinatorics **1** (1996) 147–159

[13] J. Jędrzejowicz, A Szepietowski: *Shuffle languages are in P.* Theoretical Computer Science **250** (2001) 31–53

[14] J. E. Pin: *Varieties of formal languages.* Plenum (1986)

[15] T. Shoudai: *A P-complete language describable with iterated shuffle.* Information Processing Letters **41** (1992) 233–238

[16] M. K. Warmuth, D. Haussler: *On the complexity of iterated shuffle.* Journal of Computer and System Scinces **28** (1984) 345–358

(Beérkezett: 2005. december 13.)

IVÁN SZABOLCS
SZEGEDI TUDOMÁNYEGYETEM
INFORMATIKAI TANSZÉKCSOPORT, SZÁMÍTÁSTUDOMÁNY ALAPJAI TANSZÉK
6701 SZEGED, PF. 652.
szabivan@inf.u-szeged.hu

ON THE ITERATED SHUFFLE OF SOME REGULAR LANGUAGES

Szabolcs Iván

In the paper the iterated shuffle operation is investigated on some regular languages. We give necessary and sufficient conditions for regularity and context-freeness of languages that are the iterated shuffle of some singleton language, or a commutative closure of some singleton language. We prove that for any finite language $L \subseteq a^+b^+$ it holds that the iterated shuffle of L is context-free, and characterize the cases when such a language is even regular. Also the iterated shuffle of any 1-testable language is proven to be locally threshold testable. Finally, a family of subclasses of the commutative regular languages is identified such that each member of this family is closed under both the shuffle product and the iterated shuffle operations.