Megoldás a nehezebb/ érdekesebb feladatokra

dr. Bach Iván előadásai és
Csima Judit gyakorlatai
alapján készítette:
Szabó Péter
<pts@inf.bme.hu>

2000. október 1–7.

Feladatok

1. Adjunk CF nyelvtant az alábbi nyelvhez:**
$\{x \in \{\mathrm{a}, \mathrm{b}\}^* \mid x \text{ nem } ww \text{ alakú}\}$

2.* Mutassuk meg, hogy minden DVA, ami elfogadja a $(\mathrm{a}+\mathrm{b})^* a(\mathrm{a}+\mathrm{b})^{n-1}$ nyelvet, legalább 2^n állapotú, noha nemdeterminisztikus véges automata már van $n+1$ állapottal is.

3. Nyelvtan kell az $\{\mathrm{a}^{F_n} \mid n \geq 1,\ F_n \text{ az } n\text{-edik Fibonacci szám}\}$ nyelvre. (Nem baj, ha 0-s lesz a nyelvtan.)**

4.* Nyelvtan kell az $\{\mathrm{a}^p \mid p \text{ prím}\}$ nyelvre. (Itt se baj, ha 0-s lesz a nyelvtan, sőt már az is jó, ha nagyjából el tudod mesélni, hogy hogyan működjön, nem kell rögtön szabályokra lebontva leírni.)**

5.1. Reguláris-e $L_1 = \{xx^R w \mid x \in \{0,1\}^+, w \in \{0,1\}^*\}$?**

5.2. Reguláris-e $L_2 = \{xwx^R \mid x \in \{0,1\}^+, w \in \{0,1\}^*\}$?

6.1. Reguláris-e $L_1 = \{0^i 1^j \mid \mathrm{lnko}(i,j) \neq 1\}$?

6.2.* Reguláris-e $L_2 = \{0^{F_n} \mid n \geq 1,\ F_n \text{ az } n\text{-edik Fibonacci szám}\}$?

7. Tegyük fel, hogy az L_1 nyelv reguláris, az L_2 nyelv pedig nem reguláris. Igazoljuk, hogy ekkor $L_1 L_2$ sem reguláris!**

8. Legyen L egy reguláris nyelv a Σ ábécé felett. $\frac{1}{2}L := \{x \in \Sigma^* \mid \exists y \in \Sigma^* : |y| = |x|, xy \in L\}$. Azaz $\frac{1}{2}L$-ben az L-beli szavak első felei vannak. Reguláris-e az $\frac{1}{2}L$ nyelv?**

9.** Bizonyítsd, hogy ha $L \subseteq 0^*$ és L CF nyelv, akkor L reguláris!**

Megoldások

1. Adjunk CF nyelvtant az alábbi nyelvhez:**
$\{x \in \{\mathrm{a}, \mathrm{b}\}^* \mid x \text{ nem } ww \text{ alakú}\}$

$\mathrm{S} \to \mathrm{AB} \mid \mathrm{BA} \mid \mathrm{A} \mid \mathrm{B}, \quad \mathrm{A} \to \mathrm{XAX} \mid \mathrm{a}, \quad \mathrm{B} \to \mathrm{XBX} \mid \mathrm{b}, \quad \mathrm{X} \to \mathrm{a}, \mathrm{b}.$

Ez a CF nyelvtan a páratlan hosszú (A | B), az $(\mathrm{a}+\mathrm{b})^i a(\mathrm{a}+\mathrm{b})^{i+j} b(\mathrm{a}+\mathrm{b})^j$, és az $(\mathrm{a}+\mathrm{b})^i b(\mathrm{a}+\mathrm{b})^{i+j} a(\mathrm{a}+\mathrm{b})^j$ alakú szavakra illeszkedik, tehát pontosan a nem ww alakúakra.

2.* Mutassuk meg, hogy minden DVA, ami elfogadja a $(\mathrm{a}+\mathrm{b})^* a(\mathrm{a}+\mathrm{b})^{n-1}$ nyelvet, legalább 2^n állapotú, noha nemdeterminisztikus véges automata már van $n+1$ állapottal is.

A nemdet. automata nyelvtana: $\mathrm{S} \to \mathrm{aS} \mid \mathrm{bS} \mid \mathrm{aA}_1$, $\mathrm{A}_i \to \mathrm{a}A_{i+1} \mid \mathrm{b}A_{i+1}$ $(i = 1 \ldots n)$, $\mathrm{A}_n \to \varepsilon$. Ez $n+1$ állapotú.

Tekintsünk egy DVA-t, ami a nyelvet fogadja el, de $< 2^n$ állapota van. A skatulyaelv szerint a 2^n db $(\mathrm{a}+\mathrm{b})^n$ alakú szó között van kettő (x és y), amely a DVA-t ugyanabba az állapotba viszi. x-nek és y-nak egyezzen meg az első i db betűje, de az $i+1$-edik már különbözzön, mégpedig úgy, hogy x-ben ez a betű a, y-ban pedig b legyen!

Ekkor az $x\mathrm{a}^i$ és $y\mathrm{a}^i$ szavak a DVA-t ugyanabba az állapotba viszik. Másrészt viszont $x\mathrm{a}^i$ nincs benne a nyelvben (hátulról az n-edik betűje a), noha $y\mathrm{a}^i$ benne van (hátulról az n-edik betűje b). Ez ellentmondás, tehát eredeti feltevésünk hamis volt, így bármely, e nyelvet felismerő DVA $\geq 2^n$ állapotú.

3. Nyelvtan kell az** $\{\mathrm{a}^{F_n} \mid n \geq 1,\ F_n \text{ az } n\text{-edik Fibonacci szám}\}$ **nyelvre. (Nem baj, ha 0-s lesz a nyelvtan.)**

S → a	S → aa	S → aaa	S → CABBD	BD → LE
BL → LB	AL → AAM	MB → BM	BME → LAE	BMA → LAA
AMA → ONAA	AON → OPB	PBA → APB	PBE → NBE	PBB → NBB
AN → NA	CON → CQQ	QQA → AQQ	QQB → BQQ	QQE → BBD
CA → aR	RA → aR	RB → aR	RD → a0	

Ez egy nemcsökkentő, tehát CS nyelvtan. F_n-t $\mathrm{CA}^{F_{n-2}-1}\mathrm{B}^{F_{n-1}-1}\mathrm{D}$ alakban ábrázoljuk. A rekurziós képlet szerint a B-ket az A-hoz adva először $\mathrm{CA}^{F_n-3}\mathrm{ONA}^{F_{n-1}-1}\mathrm{D}$-t, majd ebből mozgatással $\mathrm{CA}^{F_{n-1}-1}\mathrm{A}^{F_n-1}\mathrm{D}$-et vezetjük le, ami épp F_{n+1}-et adja. Illusztrációul bemutatjuk $F_6 = 8$ levezetését:

S ⇒ CABBD (2+3) ⇒ CABLE :-) ⇒ CALBE ⇒ CAAMBE ⇒ CAABME ⇒ CAALAE ⇒ CAAAMAE ⇒ CAAONAAE (5+3) ⇒ CAOPBAAE ⇒ CAOAPBAE ⇒ CAOAAPBE ⇒ CAOAANBE ⇒ CAOANABE ⇒ CAONAABE ⇒ COPBAABE ⇒ COAPBABE ⇒ COAAPBBE ⇒ COAANBBE ⇒ COANABBE ⇒ CONAABBE ⇒ CQQAABBE ⇒ CAQQABBE ⇒ CAAQQBBE ⇒ CAABQQBE ⇒ CAABBQQE ⇒ CAABBBBD ⇒ 0RABBBBD ⇒ 00RBBBBD $\Rightarrow^4$ 000000RD ⇒ 00000000 (8).

4.* Nyelvtan kell az $\{a^p \mid p \text{ prím}\}$ nyelvre. (Itt se baj, ha 0-s lesz a nyelvtan, sőt már az is jó, ha nagyjából el tudod mesélni, hogy hogyan működjön, nem kell rögtön szabályokra lebontva leírni.)**

CS nyelvtant adunk, de csak „nagyjából”. Először generálunk egy akármekkora számot, és – a nála kisebb számokkal vett osztási maradék alapján – ellenőrizzük, hogy prím-e. Csak akkor érhet véget a levezetés, ha a szám prím. Az n számot $KA^{d-1}B^{dk-d}C^{n-dk-1}V$ alakban ábrázoljuk, melynek hossza n. Az első d írásjel az aktuális osztót jelöli (kezdetben $d = 2$, végül pedig $d = n - 1$), az első dk írásjel pedig az osztó többszöröse, melyet n-ből már levontunk annak érdekében, hogy meghatározzuk n maradékát modulo d (kezdetben $k = 1$).

A nyelvtan úgy működik, hogy d db C jelet B-re cserélünk (a szó elején levő K-ból és minden A-ból egy-egy cserejelet „indítunk el”, ami C-hez érve azt B-re cseréli). Ha mind a d-t le tudtuk cserélni, akkor újabb d-s cserének futunk neki. Ha csak $d - 1$-et tudunk cserélni (mert a d-edik jel V), akkor $d \mid n$, n összetett, a levezetés akadjon el. Ha ennél is kevesebbett tudtunk cserélni, akkor $d \not| \ n$; a B-ket visszaírjuk C-be, majd eggyel hosszabb d-vel próbálkozunk újra. Amint az A-k elérik V-t, már biztosak lehetünk benne, hogy a szám prím, tehát minden jelet 0-ra cserélünk.

A nyelvtan CS lesz, mivel a kezdeti (hossznövelő) generálás után már nem változik a mondatszerű forma hossza. A nyelvtan eleje: S → 00, S → KAG, G → W, G → CW, majd az „osztópróbáló” szabályok jönnek (hasonlóan az előző feladathoz), s végül AV → H0, AH → H0, KH → 00.

$n = 11$-re a főbb állomások: S $\Rightarrow^*$ KAG $\Rightarrow^*$ KACCCCCCCCG $\Rightarrow^*$ KAACCCCCCCV $\Rightarrow^*$ KAABBBCCCCV $\Rightarrow^*$ KAABBBBBBCV $\Rightarrow^*$ KAAACCCCCCV $\Rightarrow^*$ KAAABBBBCCV $\Rightarrow^*$ KAAAACCCCCV $\Rightarrow^*$ KAAAABBBBBV $\Rightarrow^*$ KAAAAACCCCV $\Rightarrow^*$ KAAAAAAAAAV $\Rightarrow^*$ KAAAAAAAAH0 $\Rightarrow^*$ 00000000000.

5.1. Reguláris-e $L_1 = \{xx^Rw \mid x \in \{0,1\}^+, w \in \{0,1\}^*\}$?**

*(Szerintem ez volt a **-os.)* Tegyük fel, hogy L_1 reguláris. Tekintsük a $(01)^n$ szót, és hosszabbítsuk meg egy tetszőleges olyan u szóval, hogy az eredmény benne legyen L_1-ben (pl. $u = (10)^n$ megfelel).

(1) $xx^Rw = (01)^n u$. $|x| \geq |2n|$, mert ellenkező esetben x utolsó és x^R első betűje (a két betű ugyanaz!) is $(01)^n$-ben követné egymást, ami ellentmond annak, hogy $(01)^n$-ben nincs betűismétlődés. $|x| \geq 2n$ miatt viszont $|u| > |w|$, tehát u előáll $u = vw$ alakban. Ezt (1)-be írva: $xx^Rw = (01)^n vw$, tehát $xx^R = (01)^n v$. $4n \leq 2|x| = 2n + |v|$, tehát $|u| \geq |v| \geq 2n$, tehát egy $(01)^n$ alakú szót $\geq 2n$ betűvel kell kiegészítenünk, hogy L_1 minimálautomatája elfogadja. Ez azt jelenti, hogy az az állapot, ahová az automata $(01)^n$ hatására kerül, $\geq 2n$ távolságra van a legközelebbi elfogadó állapottól, tehát az automata $\geq 2n + 1$ állapotú. n-et az automata állapotszámánál nagyobb-

nak választva ellentmondást kapunk, tehát az eredeti feltevésünk hamis volt. L_1 **nem reguláris.**

5.2. Reguláris-e $L_2 = \{xwx^R \mid x \in \{0,1\}^+, w \in \{0,1\}^*\}$?

Vegyük észre, hogy L_2-be pontosan a ≥ 2 hosszú, azonos kezdő- és záróbetűvel rendelkező szavak tartoznak. (Valóban: az ilyen szavak L_2-beliek, ha x-et a kezdőbetűnek választjuk; és az L_2-beli szavak ilyenek.) Tehát L_2-t generálja a következő reguláris nyelvtan: $S \to aA|bB$, $A \to aA|bA|a$, $B \to aB|bB|b$. Tehát L_2 **reguláris.**

6.1. Reguláris-e $L_1 = \{0^i 1^j \mid \text{lnko}(i,j) \neq 1\}$?

Pumpálási lemmával bizonyítjuk, hogy L_1 nem reguláris. Ha ugyanis az volna, akkor vegyünk egy „elég nagy” p prímet, $i := p$, $j := p$, és pumpáljuk 0^{p-1}-et. Ez azt jelenti, hogy $\exists m < p : 0^{p+m}1^p \in L_1$, tehát $1 \neq \text{lnko}(p, p+m) = \text{lnko}(p,m) = 1$, ami ellentmondás. Tehát L_1 **nem reguláris.**

6.2.* Reguláris-e $L_2 = \{0^{F_n} \mid n \geq 1,\ F_n$ az n-edik Fibonacci szám$\}$?

Pumpálási lemmával bizonyítjuk, hogy L_2 nem reguláris. Ha ugyanis az volna, akkor vegyünk egy „elég nagy” i-t, és pumpáljuk 0^{F_i}-t. Ez azt jelenti, hogy $\exists m < F_i : \forall k\ 0^{F_i+mk} \in L_2$, tehát $F_i + mk$ Fibonacci-szám. F_n különbségsorozata F_{n-2}, és $F_{n-2} \to \infty$, tehát két szomszédos Fibonacci-szám különbsége előbb-utóbb $> m$, noha mi végtelen sok m különbségű Fibonacci-számpárt találtunk. Ez ellentmondás, tehát L_2 **nem reguláris.**

7. Tegyük fel, hogy az L_1 nyelv reguláris, az L_2 nyelv pedig nem reguláris. Igazoljuk, hogy ekkor L_1L_2 sem reguláris!**

Az állítás nem igaz, vagyis tudunk mutatni olyan L_1 reguláris és L_2 nem reguláris nyelvet, hogy L_1L_2 reguláris legyen. Pl. $L_1 := \{\mathrm{a}^n \mid n \geq 0 \text{ egész}\}$, $L_2 := \{\mathrm{a}^f \mid f \geq 1 \text{ Fibonacci-szám}\}$, $L_1L_2 = \{\mathrm{a}^n \mid n \geq 1 \text{ egész}\}$. L_1 és L_1L_2 közismerten reguláris, L_2-ről pedig 6.2-ben megmutattuk, hogy nem reguláris.

Megjegyzés: Ha a feladat szövege L_1L_2 helyett $L_1 \cup L_2$-et kérné, az előző bekezdés ellenpéldája itt is ellenpélda lenne, míg ha $L_1 \cap L_2$ nemregularitását kéne bizonyítani, $L_1 := \{\varepsilon\}$, $L_2 := \{\mathrm{a}^f \mid f \geq 1 \text{ Fibonacci-szám}\}$ lenne ellenpélda.

8. Legyen L egy reguláris nyelv a Σ ábécé felett. $\frac{1}{2}L := \{x \in \Sigma^* \mid \exists y \in \Sigma^* : |y| = |x|, xy \in L\}$. Azaz $\frac{1}{2}L$-ben az L-beli szavak első felei vannak. Reguláris-e az $\frac{1}{2}L$ nyelv?**

Előbb a 9. feladat megoldását érdemes elolvasni, mert merítünk az ott leírtakból. Azt állítjuk, hogy $\frac{1}{2}L$ **reguláris.**

Tegyük fel, hogy az ábécé egyetlen betűből áll. Készítsük el L-hez a 9. feladatban leírt M automatát, amely véges sok kivételtől eltekintve a szóhossz valamely m számmal vett osztási maradékából állapítja meg, hogy a szó L-beli-e. Az m-mel való osztási maradék azonban (véges sok kivételtől eltekintve) azt is meghatározza, hogy a szó $\frac{1}{2}L$-beli-e: pontosan akkor $\frac{1}{2}L$-beli, ha a dupla akkora szóhossz L-ben megengedett maradékot ad. Ez alapján $\frac{1}{2}L$-hez is készíthetünk egy $\frac{1}{2}M$ automatát, tehát $\frac{1}{2}L$ reguláris.

Mi legyen, ha az ábécében többféle betű is van? A betűket egyformának tekintve készítsük el $\frac{1}{2}M$-et az előző bekezdésben leírt módon. Vegyük még M-et, L minimálautomatáját. M állapotszáma legyen n.

$\frac{1}{2}M$ minden egyes állapotát cseréljük le n db részállapotra, melyek M állapotainak felelnek meg. Egy részállapot akkor legyen elfogadó, ha M-ben is elfogadó volt, *és* az ő $\frac{1}{2}M$-beli „gyűjtőállapota" is elfogadó. A nyilak a részállapotok között ugyanúgy fussanak, mint M-ben, de ne a gyűjtőállapoton belül, hanem mindig a következő gyűjtőállapot megfelelő részállapotába.

Ezzel elértük, hogy az új automata pont az $\frac{1}{2}L$-ben megengedett hosszúságú szavakat fogadja el, és – mivel a részállapotok pont jól alakulnak egymásba – minden elfogadott szó egy L-beli szó előtagja lesz. Tehát az új automata pontosan $\frac{1}{2}L$-t ismeri fel. Tehát $\frac{1}{2}L$ reguláris. Q.E.D.

9.** Bizonyítsd, hogy ha $L \subseteq 0^*$ és L CF nyelv, akkor L reguláris!**

Bevezetünk néhány elnevezést. *betű:* terminális szimbólum. *írásjel:* grammatikai szimbólum. *forma:* mondatszerű forma. S: a mondatszimbólum. *szó:* mondat. *kezdés:* S-ből induló, szabályszerű átalakítás. *befejezés:* szóra végződő, szabályszerű átalakítás. *levezetés:* kezdés és befejezés is egyben. egy írásjel *elérhető:* van levezetés, melyben szerepel. egy írásjel *befejezhető:* van olyan befejezés, ami vele kezdődik.

Írjuk át L CF nyelvtanát úgy, hogy továbbra is CF maradjon, de $J \to \varepsilon$ alakú szabály ne legyen benne, leszámítva esetleg $S \to \varepsilon$-t. (Ezt tanultuk, hogyan kell.)

(Ez a bekezdés lényegében a CNF-fé alakítást írja le.) Mivel az elérhetetlen írásjelek nincsenek hatással a levezetésekre, tehát a nyelvre sem, hagyjuk el őket! Az elhagyás után már minden J írásjel elérhető, tehát minden J-hez létezik egy levezetés, melyben ő szerepel. Ezt a levezetést alakítsuk át J befejezésévé, mégpedig úgy, hogy a levezetésben J-t és „leszármazottait" aláhúzzuk, majd csak az aláhúzott formákat tekintjük; ezek sorozata épp J egy befejezését adja. Azt kaptuk, hogy minden írásjel befejezhető. Emiatt azonban minden forma befejezhető, pl. úgy, hogy sorra befejezzük az írásjeleit.

A bizonyítandó állítást először azokra a nyelvtanokra látjuk be, melyekben S-ből eljuthatunk S-et is tartalmazó, ≥ 2 hosszú formához (f). Az ilyen nyelvtanokat nevezzük *hízlalhatónak.* f írásjeleit (egyetlen S-t kivéve) fejezzük be az előző bekezdésben leírt módon. Ekkor egy olyan f' formához jutunk, amely 1 db S-et, és $m \geq 1$ db 0-t tartalmaz. Az így kapott $S \Rightarrow \ldots \Rightarrow f'$ átalakítássorozatot nevezzük az L nyelv *hízlalásának*.

Vegyük L-nek egy tetszőleges (0^k alakú) szavát, és tekintsük ennek a szónak a levezetését. Ha a levezetést az első átalakítás helyett a hízlalással kezdjük, majd az eredeti levezetést a hízlalás eredményéül kapott f'-ben levő S-ből indulva hajtjuk végre, akkor egy 0^{k+m} hosszú szót kapunk. Ezzel beláttuk, hogy $(\forall k)\ 0^k \in L \quad \Longrightarrow \quad 0^{k+m} \in L$.

Soroljuk L szavait a hosszuk alapján m szerinti maradékosztályokba. Bizonyos maradékok végtelen sokszor fordulnak elő (K_∞), mások pedig csak véges sokszor (K_f). Mivel K_f-et csak véges sok szó képviseli, ezért egy bizonyos hossz felett már csak K_∞-beli szavak lehetnek. Másrészt azonban minden K_∞-beli maradék ≥ 1-szer előfordul (valóban: hisz' végtelen sokszor elfordul), és bármely L-beli szót m-mel hízlalhatunk (lásd előző bekezdés), tehát egy bizonyos hossz felett már minden K_∞-beli szó L-beli. Összefoglalva: egy bizonyos hossz felett pontosan a K_∞-beli szavak lesznek az L-beliek. Másként fogalmazva: véges sok kivételtől eltekintve a szóhossz m szerinti maradéka határozza meg, hogy a szó L-beli-e.

Ekkor viszont könnyedén készíthetünk L-hez reguláris nyelvtant. Csinálunk egy m hosszú irányított kört, melyen pontosan az K_∞ elemeinek megfelelő állapotok az elfogadóak. A véges sok kivételt pedig úgy valósítjuk meg, hogy a kezdőállapotot egy elegendően hosszú irányított úttal csatlakoztatjuk a körhöz. Az út elfogadó állapotait a véges sok kivétel figyelembevételével jelöljük ki.

Az állítást már csak a nem hízlalható nyelvtanokra kell belátnunk. Ezekben S csak a levezetések elején fordulhat elő. Az írásjelek számára (k) vonatkozó teljes indukciót almalmazunk. $k = 1$ esetén L csak véges sok szóból áll, tehát L reguláris. Ha feltesszük, hogy $k < n$-re már igaz, akkor $k = n$-re a következőképpen érvelünk: Ha a mondatszimbólumnak S helyett egy másikat ($J_2, \ldots, J_n$) jelölünk ki, akkor S elérhetetlen, tehát S-et elhagyhatjuk a nyelvtanból. Az így kapott $L_2, \ldots, L_n$ nyelvtanok írásjeleinek száma $< n$, tehát L_i vagy hízlalható (tehát reguláris), vagy az indukciós feltevés miatt reguláris. L viszont felépíthető az L_i-k konkatenációjából és úniójából az $S \to$-sel kezdődő levezetés szabályok szerint. Tehát L is reguláris. Q.E.D.

$$
\begin{aligned}
\mathsf{S} &\rightarrow \mathsf{a}\\
\mathsf{S} &\rightarrow \mathsf{aa}\\
\mathsf{S} &\rightarrow \mathsf{aaa}\\
\mathsf{S} &\rightarrow \mathsf{CABBD}\\
\mathsf{BD} &\rightarrow \mathsf{LE}\\
\mathsf{BL} &\rightarrow \mathsf{LB}\\
\mathsf{AL} &\rightarrow \mathsf{AAM}\\
\mathsf{MB} &\rightarrow \mathsf{BM}\\
\mathsf{BME} &\rightarrow \mathsf{LAE}\\
\mathsf{BMA} &\rightarrow \mathsf{LAA}\\
\mathsf{AMA} &\rightarrow \mathsf{ONAA}\\
\mathsf{AON} &\rightarrow \mathsf{OPB}\\
\mathsf{PBA} &\rightarrow \mathsf{APB}\\
\mathsf{PBE} &\rightarrow \mathsf{NBE}\\
\mathsf{PBB} &\rightarrow \mathsf{NBB}\\
\mathsf{AN} &\rightarrow \mathsf{NA}\\
\mathsf{CON} &\rightarrow \mathsf{CQQ}\\
\mathsf{QQA} &\rightarrow \mathsf{AQQ}\\
\mathsf{QQB} &\rightarrow \mathsf{BQQ}\\
\mathsf{QQE} &\rightarrow \mathsf{BBD}\\
\mathsf{CA} &\rightarrow \mathsf{aR}\\
\mathsf{RA} &\rightarrow \mathsf{aR}\\
\mathsf{RB} &\rightarrow \mathsf{aR}\\
\mathsf{RD} &\rightarrow \mathsf{a0}
\end{aligned}
$$