1. príklad

Dokážte že nasledujúce jazyky nie sú regulárne:

(a) $L_1 = \{ww \mid w \in \{0,1\}^*\}$

(b) $L_2 = \{ww^R \mid w \in \{0,1\}^*\}$

(c)* $L_3 = \{a^{2^n} \mid n \in N\}$

(d) $L_4 = \{a^{n^2} \mid n \in N\}$

(e)* $L_5 = \{a^{n!} \mid n \geq 1, n \in N\}$

(f) $L_6 = \{w \in \{0,1\} \mid 4|w|_0 = 2|w|_1\}$

(g)* $L_7 = \{a^n b^m c^{mn} \mid n, m \in N\}$

(h) $L_8 = \{0^x 1^y \mid x, y \in N; x \; mod \; y = 0\}$

Riešenie príkladu 1(a)

Jazyk nie je regulárny. Dokazujeme sporom. T.j. dokazujeme tvrdenie, že jazyk $L_1 = \{ww \mid w \in \{a,b\}^*\}$ je regulárny. Potom musí existovať automat $A = \{Q, \Sigma, \delta, q_0, F\}$, pre ktorý platí že $L(A) = L_2$. Nech počet stav tohto (konečného) automatu je n.

Keď vezmeme slová $a, a^2, a^3, ..., a^{n+1}$, tak potom musia aspoň dve slová skončiť v rovnakom stave, čiže $\exists x, y \in \Sigma^*, \hat{\delta}(q_0, x) = \hat{\delta}(q_0, y)$.

Nech $x = a^i$ a $y = a^j$, pričom $1 < i < j \leq n+1$. Čiže z Lémy 3.12 $\hat{\delta}(q_0, a^i) = \hat{\delta}(q_0, a^j) \Longrightarrow \hat{\delta}(q_0, a^j z) = \hat{\delta}(q_0, a^i z)$, pre každé $z \in \Sigma^*$

Ako si zvoliť z? Možno vám intuitívne napadne, že veď stačí zvoliť a^i alebo a^j. To ale nie je správne. Prečo? Aké slová tvaru a^k budeme vedieť zapísať ako ww? Jednoducho slovo rozdelíme na polovice, teda $ww = a^{k/2}a^{k/2}$, a to vieme spraviť iba vtedy, ak k bude párne. Čo vieme povedať o dĺžke slov $a^i a^j$ a $a^j a^j$? Dĺžka prvého bude $i+j$, o čom nevieme povedať nič viac. Dĺžka druhého bude $i+i = 2i$, čiže párna. Vieme, že druhé slovo určite budeme môcť rozdeliť na dve rovnaké za sebou idúce časti (ww). Vieme s určitosťou povedať, že $i+j$ bude párne alebo nepárne? Nie, takže toto je nesprávne z.

Teraz by ste si mohli povedať, že veď tak upravíme slová, aby nám to vyšlo. Napríklad pri a^i a a^{i+1} by sme vedeli povedať čosi o párnosti dĺžky druhého slova. Ale aj toto je nesprávne. Prečo? No lebo nami vybraté slová nemôžu závisieť jedno od druhého.

Vráťme sa preto k pôvodným slovám a hľadajme iné z. Čo nám pomohlo v príklade 4b? Že sme vedeli ľahko určiť oddelenie slov pomocou mriežky #. Vieme tu použiť niečo iné, čo by slúžilo rovnako? Máme okrem a ešte niečo v abecede? No predsa b.

Zvoľme si $z = ba^ib$. Potom $\hat{\delta}(q_0, a^jba^ib) = \hat{\delta}(q_0, a^iba^ib)$.

- $a^jba^ib \notin L(A)$, pretože $a^jb \neq a^ib$
- $a^iba^ib \in L(A)$

Dospeli sme k sporu (buď akceptujeme obe slová, čiže aj slovo, ktoré nie je z jazyka, alebo neakceptujeme žiadne z týchto slov a teda ani slovo, ktoré patrí do jazyka), daný automat nerozoznáva jazyk L_1. Platí preto pôvodné tvrdenie. Jazyk L_1 nie je regulárny.

Riešenie príkladu 1(b)

Čo znamená zápis ww^R? Že k slovu w prireťazíme jeho reverz, čiže to isté slovo odzadu. Ak by sme toto slovo tvorili iba z núl, alebo iba z jednotiek, platilo by to isté čo v predchádzajúcom príklade – takýto počet (núl alebo jednotiek) by musel byť párny. Opäť tak použime oddeľovač a tentokrát skúsme použiť Pumpovaciu lému, čiže Lému 3.14.

Dokazujeme sporom tvrdenie, že jazyk L_2 je regulárny. Pre každý regulárny jazyk platí Léma 3.14, ktorá hovorí, že všetky slová w dlhšie nanajvýš rovné konštante n_0[1] vieme rozdeliť tak, že $w = yxz$, kde $|yx| \leq n_0$, $|x| > 1$. Potom buď všetky slová yx^kz pre $k \in N$ sú z jazyka, alebo žiadne takéto slovo do jazyka nepatrí.

Zvoľme si slovo $1^{n_0}001^{n_0}$. Toto slovo patrí do jazyka L_2 ($w = 1^{n_0}0; w^R = 01^{n_0}$), je dlhšie ako n_0 a preto ho vieme rozdeliť podľa Lémy 3.14 tak, aby platili vyššie napísané podmienky.

Vieme, že dĺžka yx má byť menšia nanajvýš rovná n_0, v našom slove to znamená, že podslová y a x sa budú nachádzať v prefixe 1^{n_0}. Zvoľme $y = 1^i; i \in N$; $x = 1^j$, kde $j \geq 1$.

Nevieme, či $i + j = n_0$ alebo $i + j < n_0$ (zamyslite sa prečo), preto z zapíšme takto: $z = 1^{n_0-i-j}001^{n_0}$.

$w = yxz = 1^i1^j1^{n_0-i-j}001^{n_0}$

Pumpujeme poslovo x, to znamená, že ho môžeme odobrať (napumpujeme ho nulakrát), alebo ľubovoľne veľakrát do slova pridať. To vieme zapísať ako $yx^kz, k \in N$, čiže pre naše slovo: $1^i1^{kj}1^{n_0-i-j}001^{n_0}$.

[1] n_0 je rovné počtu stavov DKA, ktorý tento jazyk rozoznáva

Ak zvolíme za $k = 0$, ako bude vyzerať toto slovo? $1^i 1^{0j} 1^{n_0-i-j} 001^{n_0} = 1^i 1^{n_0-i-j} 001^{n_0} = 1^{n_0-j} 001^{n_0}$. Patrí toto slovo do jazyka L_2? Nie, pretože $(1^{n_0-j}0)^R \neq 01^{n_0}$, keďže $j \geq 1$.

Čím sme prišli do sporu a platí pôvodné tvrdenie, jazyk L_2 je neregulárny.

Čo ak by sme zvolili slovo $1^m 001^m$, pre nejaké m také, že $n_0/2 \leq m < n_0$? Toto slovo by bolo dlhšie ako n_0, ale ako by sa rozdelilo na podslová? Treba si uvedomiť, že my vieme, že existuje také rozdelenie v prefixe dĺžky n_0, ale nevieme povedať[2]*, kde konkrétne končí y a kde začína x. V tomto slove by sa tak mohlo stať, že x by obsahovalo aj 0, alebo 00, alebo 001^i, pre nejaké i, alebo niečím z toho začínalo a y by tiež mohlo obsahovať 0. Aby ste dokázali, že jazyk je neregulárny, museli by ste ukázať, že pre* VŠETKY *tieto rozdelenia prídete do sporu. Preto je jednoduchšie použiť na tento dôkaz slovo, v ktorom to rozdelenie určite nastane v jednej časti slova.*

Riešenie príkladu 1(d)

Idea dôkazu pomocou Lémy 3.12.

- Predpokladajme, že jazyk L_4 je regulárny. Potom vieme zostrojiť automat, ktorý ho rozpoznáva, a má k stavov. Ak budeme mať (aspoň) $k+1$ slov, aspoň dve slová musia skončiť v rovnakom stave.

- Majme slová tvaru a^{2m+1}, kde $m \in N$. Vidíme, že počet takýchto slov je nekonečný, preto ich určite bude viac ako k. Nech $1 \leq i < j \leq k+1$ a $\hat{\delta}(q_0, a^{2i+1}) = \hat{\delta}(q_0, a^{2j+1})$

- Potom musí platiť aj $\hat{\delta}(q_0, a^{2i+1}x) = \hat{\delta}(q_0, a^{2j+1}x)$ pre akékoľvek $x \in \Sigma^*$ (z Lémy 3.12).

- Zvoľme si $x = a^{j^2}$. Musí teda platiť: $a^{2i+1}a^{j^2} \in L(A) \iff a^{2j+1}a^{j^2} \in L(A)$

- Upravme $a^{2j+1}a^{j^2} = a^{j^2+2j+1} = a^{(j+1)^2}$, čo určite patrí do jazyka L

- Upravme $a^{2i+1}a^{j^2} = a^{j^2+2i+1}$. Vidíme, že toto slovo má viac a ako a^{j^2}, a z podmienky, že i je menšie ako j, vieme aj, že to bude menej ako v $a^{(j+1)^2}$.

- Medzi a^{j^2} a $a^{(j+1)^2}$ sa nenachádza žiadna iná druhá mocnina prirodzeného čísla, preto určite $a^{j^2+2i+1} \notin L_3$.

[2] tak ako pri Léme 3.12 nevieme povedať, ktoré konkrétne dve slová skončia v rovnakom stave

- Dospeli sme tak do sporu s tvrdením, že automat rozoznáva daný jazyk. V rovnakom stave skončilo slovo z jazyka, aj slovo, ktoré do jazyka nepatrí, takže automat nerozoznáva správne jazyk L_3. A teda jazyk L_4 nie je regulárny.

Idea dôkazu pomocou (pumpovacej) Lémy 3.14

- Pri pumpovacej léme opäť predpokladáme, že jazyk je regulárny a tak preňho existuje také n_0, že všetky slová $w, |w| \geq n_0$ vieme vhodne rozdeliť a napumpovať tak, že všetky napumpované slová budú patriť do jazyka, alebo žiadne napumpované slovo nebude patriť do jazyka.
- Zvoľme si slová z jazyka $w = a^k$, kde $k = n^2$ (k je druhá mocnina nejakého čísla) a nech $n > n_0$.
- Rozdeľme slovo $a^k = a^m a^l a^{k-m-l}$, s tým, že $m + l \leq n_0,\ l \geq 1$.
- Z Lémy 3.14 teda bude platiť: $y = a^m$, $x = a^l$ a $z = a^{k-m-l}$. Slovo x môžeme napumpovať ľubovoľne veľakrát a všetky napumpovania by mali patriť do jazyka L_6, keďže slovo a^k doňho patrí.
- Napumpujme slovo x napr. dvakrát. Potom dostávame slovo: $a^m a^{2l} a^{k-m-l} = a^{m+2l+k-m-l} = a^{l+k} = a^{l+n^2}$.
- Slovo a^{l+k} môže byť druhou mocninou, iba ak by $l = 2n + 1$, teda by platilo $a^{l+k} = a^{l+n^2} = a^{2n+1+n^2} = a^{(n+1)^2}$.
- Z počiatočných podmienok vieme, že: $m + l \leq n_0 < n$, čiže $l < n$ a tak sa nemôže rovnať $2n + 1$.
- Keďže medzi a^{j^2} a $a^{(j+1)^2}$ sa nenachádza žiadna iná druhá mocnina prirodzeného čísla, dospeli sme k sporu. Naše napumpované slovo nepatrí do jazyka a teda jazyk L_4 nie je regulárny.

Riešenie príkladu 1(f)

Dokazujeme sporom. T.j. dokazujeme tvrdenie, že jazyk $L_6 = \{w \in \{0,1\}^* \mid 4|w|_0 = 2|w|_1\}$ je regulárny. Potom musí existovať automat $A = \{Q, \Sigma, \delta, q_0, F\}$, pre ktorý platí že $L(A) = L_6$. Nech počet stav tohto (konečného) automatu je n.

Teraz sa zamyslime, aké slová patria do jazyka L_6. Veľmi častou chybou je predpoklad, že $w = 0^4 1^2 \in L_7$. Ľahko ukážeme, že to nie je pravda: $4|w|_0 = 4|0^4 1^2|_0 = 16$ a $2|w|_1 = 2|0^4 1^2|_1 = 4$ a $4 \neq 16$.

Správne tam budú patriť napríklad slová $0^2 1^4$, alebo $0^4 1^8$. Všeobecne zapísané $0^{2i} 1^{4i}$. (Toto, samozrejme, nie sú všetky slová z jazyka, výraz vieme zjednodušiť [3], nuly a jednotky vieme vymeniť, ale aj dať cez seba. Pri dokazovaní (ne)regulárnosti však stačí vybrať nejakú podmnožinu jazyka, ktorá je aspoň o jedna väčšia ako počet stavov automatu. Lebo ak pre nejakú podmnožinu jazyka L ukážete, že sa nedá zostrojiť automat, ktorý by rozpoznával jazyk L (pozor, jazyk, nie tú danú podmnožinu), tak ho zjavne nezostrojíme ani pre zvyšné slová.)

Ak si vyberieme slová $0^2, 0^4, ..., 0^{2n}, 0^{2(n+1)}$ bude ich viac ako stavov, preto nejaké dve slová budú musieť skončiť v rovnakom stave $\hat{\delta}(q_0, x) = \hat{\delta}(q_0, y)$.

Nech $x = 0^{2i}, y = 0^{2j}$, kde $1 < i < j \leq n$. Potom z Lémy 3.12 platí pre ľubovoľné z že $\hat{\delta}(q_0, 0^{2i} z) = \hat{\delta}(q_0, 0^{2j} z)$.

Zvoľme $z = 1^{4i}$. Čiže $\hat{\delta}(q_0, 0^{2i} 1^{4i}) = \hat{\delta}(q_0, 0^{2j} 1^{4i})$.

- $0^{2i} 1^{4i} \in L(A)$, lebo $4 \,.\, 2i = 8i = 2 \,.\, 4i$
- $0^{2j} 1^{4i} \notin L(A)$, lebo $4 \,.\, 2j = 8j$ a $2 \,.\, 4i = 8i$. Keďže $i < j$, potom $8j > 8i$.

Dospeli sme k sporu, platí preto pôvodné tvrdenie. Jazyk L_6 nie je regulárny.

Riešenie príkladu 1(h)

Keď sa pozrieme na jazyk L_8, vidíme rozdelenie, ktoré vieme využiť pri dôkaze neregulárnosti pomocou Lémy 3.12. Vieme ale použiť aj Lému 3.14?

Kedy bude určite platiť, že počet núl je deliteľný počtom jednotiek? Pozrime sa na slová typu $0^k 1^k$, čiže také, ktoré majú rovnaký počet núl aj jednotiek. Tam o deliteľnosti nepochybujeme, tak to použime. Zvoľme si preto slovo: $0^{n_0+1} 1^{n_0+1}$.

Predpokladáme, že jazyk je regulárny, preto z lémy 3.14 vyplýva, že existuje konštanta n_0 a každé slovo dlhšie ako toto číslo vieme rozdeliť na tri časti (z ktorých prvé dve budú dokopy nanajvýš dĺžky n_0), tú strednú (s aspoň jedným znakom) napumpovať a buď všetky takto vytvorené slová budú z jazyka, alebo žiadne z takto vytvorených slov nebude patriť do jazyka.

Slovo $0^{n_0+1} 1^{n_0+1}$ patrí do jazyka, preto aj všetky napumpovania budú musieť byť z jazyka. Rozdeľme ho na časti. Z Lémy 3.14 vyplýva, že prvé dve časti slova majú dokopy dĺžku nanajvýš n_0, čiže vieme jednoznačne

[3] Bude správne aj $0^i 1^{2i}$?

povedať, že rozdelenie nastane v časti, kde sú 0. To vieme zapísať ako $0^i 0^j 0^{n_0+1-i-j} 1^{n_0+1}$, kde $i, j \in N, j \geq 1;\ i + j \leq n_0$.

Budeme pumpovať strednú časť. Pomôže nám, ak to napumpujeme dvakrát, trikrát alebo viackrát? Čo budeme vedieť povedať o počte núl a jednotiek? V takomto prípade bude väčší počet núl ako jednotiek a vtedy nebudeme vedieť povedať nič o tom, či je počet núl deliteľný počtom jednotiek. Čo ak ale použijeme napumovanie nulakrát? Čiže vymažeme strednú časť. Tá obsahuje aspoň jednu nulu, takže núl bude menej ako jednotiek. Čo to znamená, ak delíme menšie číslo väčším? Že zvyšok po delení bude to menšie číslo, a teda nie 0.

Napumpovanie nulakrát znamená: $0^i 0^{0j} 0^{n_0+1-i-j} 1^{n_0+1} = 0^{n_0+1-j} 1^{n_0+1}$, a $(n_0 + 1 - j)\ mod\ (n_0 + 1) = n_0 + 1 - j$. Mohlo by byť $j = n_0 + 1$ (ak delíme 0, zvyšok po delení je vždy 0)? Pozrime sa na podmienky, ktoré sme si stanovili. Ak $i + j \leq n_0$, toto nemôže nastať[4]. Preto toto slovo nepatrí do jazyka L_8, čím sme sa dostali do sporu. Tento jazyk je neregulárny.

2. príklad

Zamyslite sa nad nasledujúcimi otázkami:

(a) Existuje neregulárny jazyk nad jednopísmenkovou abecedou?

(b) Existuje nad jednopísmenkovou abecedou regulárny jazyk?

(c) Ak je jazyk L regulárny, je aj komplement jazyka L^C regulárny?

(d) Ak je jazyk L neregulárny, aký je komplement jazyka L^C?

(e) Je prienik, rozdiel alebo zjednotenie dvoch regulárnych jazykov regulárny jazyk?

(f) Je prienik, rozdiel alebo zjednotenie dvoch neregulárnych jazykov vždy neregulárny jazyk?

(g) Najviac koľko stavov bude mať automat, ktorý vznikne simuláciou automatov A a B?

(h) Vieme ohraničiť aj minimálny počet stavov automatu, ktorý vznikne simuláciou automatov A a B?

(i) Sú všetky podmnožiny regulárneho jazyka regulárne jazyky?

[4] Zamyslite sa, čo by sa stalo, ak by sme zvolili slovo $0^{n_0} 1^{n_0}$. Mohlo by sa stať, že by takéto slovo po napumpovaní 0-krát bolo z jazyka L_8? Čo by sme museli zmeniť v predpise z jazyka, aby takto vytvorené (napumpované) slovo doňho nepatrilo?

Hinty k 2. príkladu

(a) Pozrite si príklad 2. Sú tam nejaké neregulárne jazyky nad jednopísmenkovou abecedou?

(b) Je konečný jazyk regulárny? Viete vytvoriť napríklad automat pre slová s párnou dĺžkou? A viete to spraviť aj na jednopísmenkovej abecede?

(c) Ak máme zostrojený deterministický konečný automat pre jazyk L, ako ho vieme ľahko upraviť, aby rozpoznával komplement jazyka? Čo by sme museli spraviť s jeho akceptačnými stavmi?

(d) To isté, čo v predchádzajúcom – ak nevieme vytvoriť automat pre jazyk L, budeme ho vedieť vytvoriť pre L^C?

(e) Ako vieme spraviť prienik, rozdiel alebo zjednotenie dvoch regulárnych jazykov (=automatov, ktoré ich rozpoznávajú)? Keď máme automat rozpoznávajúci nejaký jazyk L, je tento jazyk regulárny?

(f) Vie byť prienikom, rozdielom či zjednotením dvoch jazykov napríklad: prázdna množina, jazyk všetkých slov nad abecedou, konečná množina slov...?

(g) Ako vytvárame simuláciu (modulárnu konštrukciu) dvoch automatov? Koľko stavov zostrojujeme?

(h) Vieme povedať, či bude mať výsledný automat menej stavov ako automat A alebo ako automat B? Čo ak vznikne simuláciou prázdny jazyk alebo jazyk všetkých slov nad abecedou? Koľko stavov majú tieto automaty? Vieme teda povedať niečo o minimálnom počte stavov automatu, ktorý vznikol simuláciou?

(i) Je jazyk všetkých slov nad abecedou $\{a\}$ regulárny? Viete k nemu zostrojiť automat, ktorý ho rozpozná? No a teraz si pozrite príklad 2 – vidíte tam nejaké jeho podmnožiny?

3. príklad

Pomocou modulárnej konštrukcie (simulácie) zostrojte automat pre jazyk:

$$L = \{w \in \{a, b\}^* \mid \text{ počet } a \text{ je párny a zároveň neobsahuje podslovo } ba\}.$$

Vieme tento automat zostrojiť použitím modulárnej konštrukcie pre rozdiel dvoch automatov? Aký bude ten druhý jazyk? Dokážte správnosť oboch automaty (nezabudnite ukázať, že každé slovo patrí do KL triedy stavu, v ktorom skončí).