1. príklad

Dokážte, že nasledujúce jazyky nie sú regulárne.

(a) $L_1 = \{0^n 1^n \mid n \in \mathbb{N}\}$

(b) $L_2 = \{w\#w \mid w \in \{a,b\}^*\};\ \Sigma = \{a,b,\#\}$

(c) $L_3 = \{ww \mid w \in \{0,1\}^*\}$

(d)* $L_4 = \{ww^R \mid w \in \{a,b\}^*\}$

(e)* $L_5 = \{a^{2^n} \mid n \in N\}$

(f) $L_6 = \{a^{n^2} \mid n \in N\}$

(g)* $L_7 = \{a^{n!} \mid n \geq 1, n \in N\}$

(h)* $L_8 = \{a^p \mid p \text{ je prvočíslo}\}$

(i) $L_9 = \{w \in \{0,1\}^* \mid 4|w|_0 = 2|w|_1\}$

(j)* $L_{10} = \{a^n b^m c^{mn} \mid n, m \in N\}$

(k)* $L_{11} = \{0^x 1^y \mid x, y \in N; x \; mod \; y = 0\}$

Riešenie príkladu 1a

Čo znamená, že je jazyk regulárny? Že vieme zostrojiť konečný stavový automat, ktorý bude tento jazyk rozpoznávať. Ak je jazyk neregulárny, takýto automat zostrojiť nevieme. Zamyslime sa, či budeme vedieť zostrojiť automat pre jazyk L_1. Čo by sme si museli pamätať?

Najprv by sme potrebovali mať stavy, v ktorých si pamätáme, koľko núl sme prečítali. Potom, z každého takéhoto stavu by sme išli do stavov, v ktorých by sme si počítali počet jednotiek. Je ale možné takýto automat zostrojiť pre $n \in \mathbb{N}$? Vieme povedať, aký by bol najväčší počet núl a jednotiek? Nekonečno. Náš automat má však iba konečný počet stavov, čiže tento jazyk nevie rozoznávať, a preto je neregulárny.

Dokážme to sporom, čiže dokazujeme opačné tvrdenie a to, že jazyk L_1 je regulárny. Potom, pre takýto jazyk vieme zostrojiť deterministický automat A, ktorý má konečný počet stavov, čiže $|Q| = m$. Pre každý deterministický automat platí léma 3.12, ktorá hovorí, že ak dve slová skončia v rovnakom stave a prireťazíme k nim obom rovnaké slovo, tak obe tieto nové slová buď budú z jazyka $L(A)$, alebo nebudú z $L(A)$.

Kedy dve slová určite skončia v rovnakom stave? Ak ich je viac, ako stavov. Vyberme teda takú množinu slov, ktorá bude aspoň o jedna väčšia, ako počet stavov m. Nech sú to slová tvaru 0^k, kde $1 \leq k \leq m+1$. Vyberme z nich dve rôzne, ktoré skončia v rovnakom stave. Nevieme povedať, ktoré presne to budú, preto budeme používať premenné. Keďže nebudú rovnaké, jedno bude dlhšie ako druhé.

Nech $\hat{\delta}(q_0, 0^i) = \hat{\delta}(q_0, 0^j)$, kde $1 \leq i < j \leq m + 1$.

Potom z Lémy 3.12 vyplýva, že $0^i z \in L(A) \Leftrightarrow 0^j z \in L(A)$, pre každé $z \in \Sigma^*$ (pretože ak sme pri čítaní 0^i a 0^j skončili v rovnakom stave, keď z tohto stavu začneme čítať z, vieme skončiť iba v jednom stave, čiže $\hat{\delta}(q_0, 0^i z) = \hat{\delta}(q_0, 0^j z)$. A ak obe skončia v rovnakom stave a jedno by bolo z jazyka a druhé nie, vedeli by sme o tomto stave povedať, či je akceptačný?).

Teraz si musíme zvoliť z tak, aby jedno slovo bolo z jazyka L_1 a druhé nie, tak získame spor s tvrdením, že jazyk je regulárny. Ako vyzerajú slová z jazyka? Majú niekoľko núl a po nich rovnaký počet jednotiek. Zvoľme preto $z = 1^i$. Potom:

- $0^i 1^i \in L(A)$
- $0^j 1^i \notin L(A)$, keďže $i < j$, čiže v tomto slove bude viac núl ako jednotiek.

Dostali sme sa do sporu s tvrdením, že jazyk L_1 je regulárny, platí preto pôvodné tvrdenie – jazyk L_1 je neregulárny.

Zamyslite sa, či aj jazyk $L = \{0^n 1^n \mid 0 < n < 2021\}$ je neregulárny, alebo preň vieme zostrojiť automat, ktorý ho bude rozpoznávať.

Riešenie príkladu 1b

Dokazujeme sporom. T.j. dokazujeme tvrdenie, že jazyk $L_2 = \{w\#w \mid w \in \{a, b\}^*\};\ \Sigma = \{a, b, \#\}$ 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[1] musia aspoň dve slová skončiť v rovnakom stave, čiže $\hat{\delta}(q_0, x) = \hat{\delta}(q_0, y)$.

Nech $x = a^i$ a $y = a^j$, pričom $1 \leq i < j \leq n + 1$. (Sú to teda nejaké dve rôzne slová z postupnosti vyššie[2].) Čiže $\hat{\delta}(q_0, a^i) = \hat{\delta}(q_0, a^j)$. A z Lémy

[1] z Pigeonhole principle

[2] Častou chybou je, že vyberiete nejaké konkrétne slová. To spraviť nemôžeme, lebo

3.12 vieme, že potom platí $a^j z \in L(A) \Leftrightarrow a^i z \in L(A)$ (pretože $\hat{\delta}(q_0, a^j z) = \hat{\delta}(q_0, a^i z)$).

Zvoľme si $z = \#a^i$. Takže $\hat{\delta}(q_0, a^j\#a^i) = \hat{\delta}(q_0, a^i\#a^i)$.

- $a^j\#a^i \notin L(A)$, pretože $i < j$
- $a^i\#a^i \in L(A)$

Dostali sme sa do sporu. Preto platí pôvodné tvrdenie a jazyk L_2 nie je regulárny.

Mohli by ste sa pýtať, prečo sme zvolili a^i, a^j a nikde v slove sa nenachádza b. Je na nás, akú množinu slov si zvolíme, nemusíme použiť všetky znaky abecedy, ale na konci potrebujeme dostať spor – čiže ak k našim slovám prireťazíme nejaké iné slovo, jedno bude z pôvodného jazyka a druhé nie. Čo okrem $\#a^i$ by sme mohli k slovám ešte prireťaziť, aby sme sa dostali do sporu? Napríklad $b^3\#a^i b^3$, $ba\#a^i ba$... Čiže to, čo dáme pred mriežku, zopakujeme aj za ňou.

Čo by sa stalo, ak by sme prireťazili $\#a^j$? Dostali by sme spor? Áno, keďže slovo $a^j\#a^j \in L(A)$, ale $a^i\#a^j \notin L(A)$.

Riešenie príkladu 1c

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

nevieme, ktoré z nich skončia v rovnakom stave. Preto ich musíme zapísať takto všeobecnejšie a dávať pritom pozor, aby sme zaručili, že tie slová budú rôzne a z množiny väčšej ako je počet stavov automatu.

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 1f

Idea dôkazu pomocou Lémy 3.12.

- Predpokladajme, že jazyk L_6 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_6

- 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 $a^{j^2+2i+1} \notin L_6$.
- 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_6. A teda jazyk L_6 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_6 nie je regulárny.

Riešenie príkladu 1i

Dokazujeme sporom. T.j. dokazujeme tvrdenie, že jazyk $L_9 = \{w \in \{0,1\}^* \mid 4|w|_0 = 2|w|_1\}$ je regulárny. Potom musí existovať automat $A = \{Q, \Sigma, \delta, q_0, F\}$, pre ktorý platí že $L(A) = L_9$. Nech počet stav tohto (konečného) automatu je n.

Teraz sa zamyslime, aké slová patria do jazyka L_6. Veľmi častou chybou je predpoklad, že $w = 0^4 1^2 \in L_9$. 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_9 nie je regulárny.

2. príklad

Zamyslite sa nad nasledujúcimi otázkami:

(a) Existuje neregulárny jazyk nad jednopísmenkovou abecedou?

(b) Existuje nad jednopísmenkovou abecedou regulárny jazyk?

[3] Bude správne aj $0^i 1^{2i}$?

(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?

Hinty k 2. príkladu

(a) Pozrite si príklad 1. 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 1 – vidíte tam nejaké jeho podmnožiny?

Poznámky

Dôkaz neregulárnosti pomocou Lémy 3.12

- Dokazujeme sporom, takže predpokladáme, že jazyk L je regulárny a teda existuje DKA A, ktorý ho rozoznáva $L = L(A)$.
- Ak máme konečný automat s n stavmi, tak ak budeme mať čo i len o jedno slovo viac, ako je stavov $(n+1)$, musia výpočty na aspoň dvoch slovách skončiť v rovnakom stave (Pigeonhole principle).
- Zvolíme si nejakú množinu slov, ktorých je dostatočne veľa (aspoň $n+1$ pre ľubovoľné n), takže napr. 0^n, a^{2k}... Treba vysvetliť/ukázať, prečo je ich určite viac ako stavov automatu. Z nich vyberiete dve slová. Pozor, nie ľubovoľné! Treba si uvedomiť, že vy neviete, ktoré konkrétne dve slová skončia v rovnakom stave, ale viete, že ak ste vytvorili aspoň $n + 1$ slov, tak to mohlo byť ľubovolné i-te a j-te. Preto *nevhodné slová* sú také, kde používate konštanty a taktiež také, kde jedno závisí od druhého napr. $1^i, 1^{2i}$ alebo a^5, a^{5+k}.
- Z Lémy 3.12 vieme, že ak výpočet na slovách u a v skončí v rovnakom stave q, tak aj výpočty na ux a vx skončia v jednom stave p, čiže slovo ux patrí do jazyka L práve vtedy keď aj slovo vx.
- Slovo x je preto vhodné vybrať tak, aby jedno slovo z dvojice ux, vx patrilo do jazyka a druhé nie. Stav p by tak mal jedno slovo akceptovať a druhé nie, to DKA robiť nevie. Takže buď ich obe akceptuje alebo nie, čiže nerozoznáva daný jazyk, čím sme sa dostali do sporu s tvrdením, že jazyk je regulárny.
- Odporúčam prečítať si komentáre k úlohe 5.2 z roku 2017/18

Dôkaz neregulárnosti pomocou pumpovacej lémy (Léma 3.14)

- Dokazujeme sporom, takže predpokladáme, že jazyk L je regulárny a teda existuje DKA A, ktorý ho rozoznáva $L = L(A)$.
- Ak máme konečný automat s n_0 stavmi, tak každé slovo $w, w \geq n_0$ vieme rozdeliť tak, že $w = yxz$, kde $|yx| \leq n_0, |x| > 1$.
- Potom platí, že buď všetky slová yx^kz patria do L alebo žiadne takéto slovo do jazyka nepatrí. Čiže ak strednú časť slova (podslovo x) napumupujeme (znásobíme) ľubovoľne-veľakrát, výsledné slovo bude vždy z jazyka alebo z neho nebude nikdy.
- Máte dve možnosti (1) buď začnete so slovom, ktoré do jazyka L patrí a potom ukážete také napumpovanie, ktoré pokazí podmienky jazyka a výsledné slovo po napumpovaní do jazyka patriť nebude. Alebo (2) začnete so slovom, ktoré do jazyka L nepatrí a ukážete také napumpovanie, ktoré splní podmienky jazyka a takéto slovo po napumpovaní do jazyka bude patriť.
- Opäť je dôležité si uvedomiť, že my nevieme, ako dlhé sú jednotlivé časti slova - okrem podmienok pre x a yx vyššie. Odporúčam preto nájsť slovo, v ktorom budete vedieť, v akej časti prichádza k pumpovaniu. Napríklad pre slovo $w_1 = a^kb^k, k > n_0$ bude stredná časť (x, ktorú pumpujeme) v časti a^k. Pre slovo $w_2 = a^kb^k, 2k > n_0$ to určiť nevieme. Môžeme pumpovať podslovo x ktoré sa bude skladať z a^i alebo a^ib^j alebo b^j. Pri takto zvolenom slove by ste preto museli ukázať, že nech x bude hociktorá z týchto možností, pumpovanie pokazí podmienku z tretej odrážky.
- Slovo je preto vhodné navrhnúť tak, aby ste jednoznačne vedeli určiť, kde sa bude nachádzať x (samozrejme, nie konkrétne slovo).

... LET'S ASSUME THERE EXISTS SOME FUNCTION F(a,b,c...) WHICH PRODUCES THE CORRECT ANSWER—
HANG ON.
THIS IS GOING TO BE ONE OF THOSE WEIRD, DARK-MAGIC PROOFS, ISN'T IT? I CAN TELL.
WHAT? NO, NO, IT'S A PERFECTLY SENSIBLE CHAIN OF REASONING.
ALL RIGHT...
NOW, LET'S ASSUME THE CORRECT ANSWER WILL EVENTUALLY BE WRITTEN ON THIS BOARD AT THE COORDINATES (x, y). IF WE—
I KNEW IT!