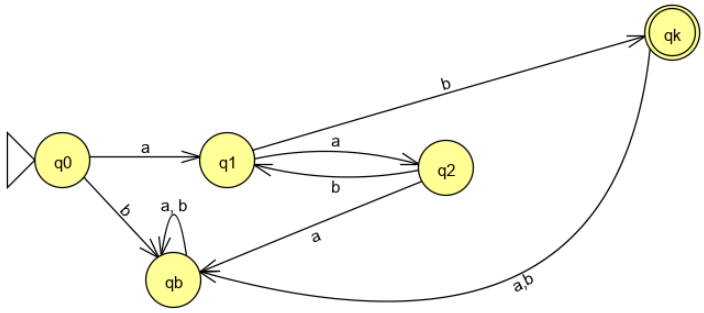
Dôkazy neexistencie 2 - pumpovacia léma

Máme automat A zadaný nasledujúcim diagramom.



$\delta(q0,a) = q1$; $\delta(q1,a) = q1$; $\delta(q2,a) = qb$; $\delta(qb,a) = qb$; $\delta(qk,a) = qb$
$\delta(q0,b) = qb$; $\delta(q1,b) = qk$; $\delta(q2,b) = q1$; $\delta(qb,b) = qb$; $\delta(qk,b) = qb$

Úloha 1

Kam sa dostane automat A **zo stavu *q_1*** po prečítaní slova:

- ab ?
- abab ?
- ababababab ?
- $(ab)^{21}$?

Zovšeobecnite po prečítaní akých slov sa automat dostane zo stavu q_1 do stavu q_1.

Úloha 2

Automat začal v počiatočnom stave a momentálne sa nachádza v stave *q1*. Aké slovo doteraz prečítal? Uveďte príklady:

- ....
- ....
- ....

Vieme to nejako zovšeobecniť?
Vieme povedať, koľko kópií slova *ab* automat prečítal, ak sa nachádza v stave *q1*?

Úloha 3

Pridajme (prireťazme) k všetkým slovám z Úlohy 2 slovo *b*. Čo vieme povedať o slovách:

- aabb
- aababb
- aababababb
- ab

Akceptuje ich automat alebo neakceptuje? Napíšte ďalšie príklady slov, ktoré automat akceptuje. Ako vo všeobecnosti vyzerajú slová, ktoré automat akceptuje?

Úloha 4

Prireťazme k všetkým slovám z Úlohy 2 slovo *a*. Čo vieme povedať o slovách:

- aaba
- aababa
- aabababa
- aa

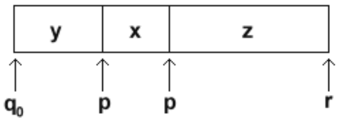
Akceptuje ich automat alebo neakceptuje? Napíšte ďalšie príklady slov, ktoré automat neakceptuje. Ako vo všeobecnosti vyzerajú slová, ktoré automat neakceptuje?

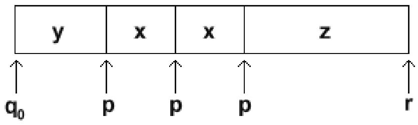
Vidíme, že v stave q_1 nedokážeme rozlíšiť, či automat prečítal jedno slovo *ab*, desať, sto či nula slov *ab*. Vieme ale, že keď ku všetkým slovám, po dočítaní ktorých automat skončí v stave q_1, prireťazíme **rovnaký symbol**, tak takto vzniknuté slová automat **buď všetky akceptuje alebo všetky zamietne**.

Skúsime to trochu zovšeobecniť.

Ak sa automat zo stavu *p* po prečítaní slova *x* dostane znova do stavu *p*, potom sa zo stavu *p* dostane do stavu *p* aj po prečítaní *xx*, *xxx*, *xxxx* ..., po prečítaní hocikoľkých (aj nula) kópií slova *x*. Formálne **ak $\hat{\delta}(p, x) = p$, potom $\hat{\delta}(p, x^i) = p$ pre $i \in$ N**. Automat si nedokáže zapamätať, koľkokrát prečítal slovo *x*, a teda nedokáže rozlišovať medzi slovami s rôznym počtom opakovaní slova *x*.

Ak sa tento automat z počiatočného stavu q_0 dostane po prečítaní nejakého slova *y* do stavu *p* a z neho po prečítaní nejakého slova *z* do stavu *r*, potom sa takýto automat vie z počiatočného stavu q_0 dostať do stavu *r* aj po prečítaní slova *yxz*, *yxxz*, *yxxxz*, ...





všeobecne po prečítaní yx^iz pre ľubovoľné $i \in$ N i, t.j. **$\hat{\delta}(q_0, y) = p \wedge \hat{\delta}(p, z) = r \Rightarrow \hat{\delta}(p, yx^iz) = r$, pre $i \in$ N**.

Vidíme, že všetky slová yx^iz patria do tej istej Kl triedy Kl[*r*], pre $r \in Q$. Ak stav *r* je:

- akceptujúci, tak automat akceptuje všetky slová z množiny $\{yx^iz \mid i \in N\}$,
- neakceptujúci, tak automat neakceptuje žiadne slovo z množiny $\{yx^iz \mid i \in \mathrm{N}\}$.

Úloha 5

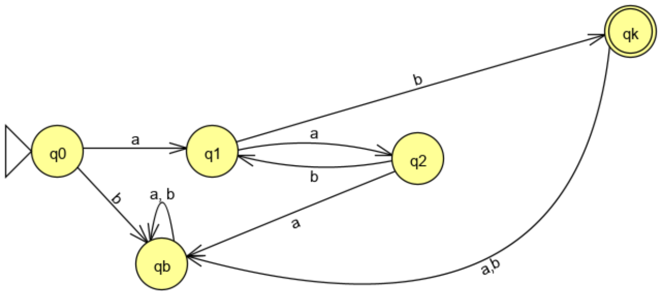
Zapíšte nasledujúce slová ako zreťazenie troch slov yx^iz pre nejaké $i \in$ N, t.j. nájdite *y*, *x*, *z* a *i*.

a) 11101001001001010
b) cbabaacbaacbaacbaacbaacbaacbaacccb
c) 000000000111
d) abc
e) abaaaaaaaaaaaaaac

Dalo sa to pri každom slove?
Dalo sa to pri každom slove jednoznačne?

Úloha 6

Vráťme sa ešte k automatu z úvodu.



- Aké je najdlhšie slovo, v ktorom nezopakujete dva stavy?
  - Aké je najdlhšie také slovo, ktoré automat akceptuje?
  - Aké je najdlhšie také slovo, ktoré automat neakceptuje?
- Pri slove akej dĺžky vieme s určitosťou povedať, že sa počas výpočtu zopakovali dva stavy
  - v tomto automate?
  - vo všeobecnosti?

Léma 3.24 (Pumpovacia léma)
Pre každý regulárny jazyk L existuje konštanta $n_0 \in$N taká, že každé slovo $w \in \Sigma^*$ dĺžky aspoň n_0 môžeme vyjadriť ako (t.j. existujú x, y, $z \in \Sigma^*$ také, že) $w = yxz$, kde

1. $|yx| \leq n_0$,
2. $|x| \geq 1$ a
3. buď $\{ yx^kz \mid k \in N) \subseteq L$ (buď všetky "napumpované" slová sú z L)
   alebo $\{ yx^kz \mid k \in N) \cap L = \varnothing$ (žiadne napumpované slovo nie je z L)

Úloha 7
Aká je konštanta n_0 pre jazyk, ktorý akceptuje náš automat z úvodu?

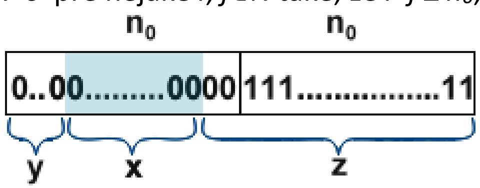
Použitie pumpovacej lémy

Pr. 1
Dokážte, že jazyk $L = \{0^n1^n \mid n \in N\}$, $\Sigma=\{0,1\}$ je neregulárny.
Dokážeme to sporom. Nech L je regulárny. $\Rightarrow^{\text{Pump. léma}}$ existuje konštanta $n_0 \in$N taká, že každé slovo $w \in \Sigma^*$ dĺžky aspoň n_0 môžeme vyjadriť ako $w = yxz$, pričom:
1) $|yx| \leq n_0$, 2) $|x| \geq 1$, 3) $(\forall k \in N: yx^kz \in L) \vee (\forall k \in N: yx^kz \notin L)$

Vezmime slovo z jazyka L $w = 0^{n0}1^{n0}$. Dĺžka tohto slova je $2n_0 \geq n_0$.

Uvažujme o rozloženiach slova $w = 0^{n0}1^{n0}$ na yxz.

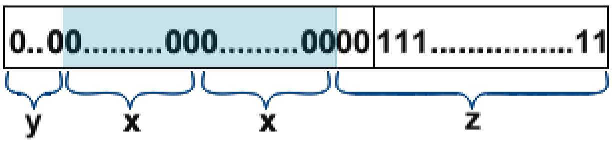
Podľa 1. $|yx| \leq n_0 \Rightarrow$ yx musí byť prefixom 0^{n0}, t.j. yx bude pozostávať len z núl, čo môžeme zapísať ako $y=0^i$ a $x=0^j$ pre nejaké $i, j \in$N také, že $i+j \leq n_0$,



Keďže $y=0^i$, $x=0^j$ a $yxz = 0^{n0}1^{n0}$, potom slovo z musia tvoriť zostávajúce nuly a všetky jednotky, presnejšie $z=0^{n0-i-j}1^{n0}$,

Podľa 2. $|x| \geq 1$ a keďže $x=0^j$, tak $j \neq 0$
Podľa 3. buď $\forall k \in N$ $yx^kz \in L$ (všetky napumpované slová sú z jazyka) alebo $\forall k \in N$ $yx^kz \notin L$ (žiadne napumpované slovo nie je z jazyka).
Kedže vieme, že slovo $w = yx^1z = \mathbf{0^{n0}1^{n0} \in L}$, tak platí 1. časť, t.j. $\forall k \in N$ $yx^kz \in L$.

Napumpujeme slovo x dvakrát: $\boldsymbol{yx^2z = 0^i0^{2j}0^{n0-i-j}1^{n0} = 0^{n0+j}1^{n0}}$.

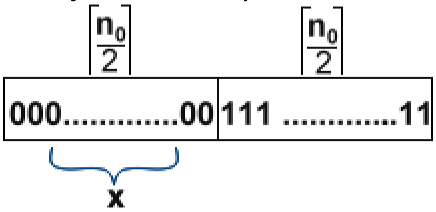


V slove $\mathbf{0^{n0+j}1^{n0}}$ nie je rovnaký počet núl a jednotiek (lebo $j \neq 0 \Rightarrow n_0 + j \neq n_0$), čo znamená, že slovo $\mathbf{0^{n0+j}1^{n0} \notin L}$ – dostali sme sa k sporu, a preto **L nie je regulárny**.

Poznámka k Pr. 1

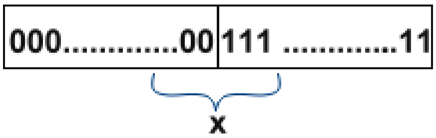
Slovo w v dôkaze z Pr. 1 by sme mohli zvoli aj takto: $w = 0^{\lceil \frac{n_0}{2} \rceil}1^{\lceil \frac{n_0}{2} \rceil}$. Dĺžka tohto slova je $\geq n_0$. Ak by sme chceli toto slovo w rozložiť na yxz spĺňajúce podmienky 1), 2) a 3), museli by sme uvažovať až o troch podprípadoch:

- slovo x je len zo samých núl



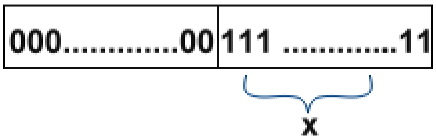
t.j. $y = 0^i \quad x = 0^j \quad z = 0^{\left\lceil \frac{n_0}{2} \right\rceil - i - j} 1^{\left\lceil \frac{n_0}{2} \right\rceil}$, pre $i, j \in$ N, $j \neq 0$

- slovo x je z núl a jednotiek



t.j. $y = 0^{\left\lceil \frac{n_0}{2} \right\rceil - j} \quad x = 0^j 1^m \quad z = 1^{\left\lceil \frac{n_0}{2} \right\rceil - m}$, pre $j, m \in$ N, $j \neq 0$, $m \neq 0$

- slovo x je len zo samých jednotiek



t.j. $y = 0^{\left\lceil \frac{n_0}{2} \right\rceil} 1^i \quad x = 1^m \quad z = 1^{\left\lceil \frac{n_0}{2} \right\rceil - i - m}$, pre $i, m \in$ N, $m \neq 0$

Pre každý z týchto podprípadov by sme mali vhodne napumpovať slovo x tak, aby sme dospeli k sporu. Keďže slovo w sme zvolili z jazyka L, mali by sme slovo x napumpovať tak, aby napumpované slovo nepatrilo jazyku.

Úloha 8

V jednotlivých podprípadoch z predchádzajúcej poznámky vytvorte slovo yx^kz pre rozumné k, a zdôvodnite, prečo toto slovo nepatrí jazyku L.

Pr. 2

Dokážte, že jazyk L = {vv | v∈{a,b}*}, Σ = {a,b} nie je regulárny.

Dokážeme to sporom. Nech L je regulárny. $\Rightarrow^{\text{Pump. léma}}$ existuje konštanta $n_0 \in$ N taká, že každé slovo $w \in \Sigma^*$ dĺžky aspoň n_0 môžeme vyjadriť ako $w = yxz$, pričom:

1) $|yx| \leq n_0$, 2) $|x| \geq 1$, 3) $(\forall k \in \text{N}: yx^kz \in \text{L}) \vee (\forall k \in \text{N}: yx^kz \notin \text{L})$

Vezmime $\boldsymbol{w = a^{n_0}ba^{n_0}b}$. $|w| = 2 + 2n_0 \geq n_0$.

Keďže $|yx| \leq n_0$ a $|x| \geq 1 \Rightarrow$ yx bude pozostávať len z á-čok, t.j. $y = a^i$, $x = a^j$, $z = a^{n_0 - i - j}ba^{n_0}b$, pre nejaké $i, j \in$ N, $j \neq 0$, $i + j \leq n_0$.

Slovo $\boldsymbol{w = yxz = a^{n_0}ba^{n_0}b \in \text{L}}$, a teda podľa 3) $\forall k \in$ N $yx^kz \in$ L.

Napumpujme slovo x nulakrát. $\boldsymbol{yx^0z = a^{n_0 - j}ba^{n_0}b \notin \text{L}}$ (lebo $j \neq 0 \Rightarrow n_0 - j \neq n_0$ //slovo sa nedá rozdeliť na dve rovnaké podslová

Dospeli sme k sporu $\Rightarrow$ jazyk L nie je regulárny.

Pr. 3

Dokážte, že jazyk L = {a^p| p je prvočíslo} nie je regulárny

Dokážeme to sporom. Nech L je regulárny. $\Rightarrow^{\text{Pump. léma}}$ existuje konštanta $n_0 \in$ N taká, že každé slovo $w \in \Sigma^*$ dĺžky aspoň n_0 môžeme vyjadriť ako $w = yxz$, pričom:

1) $|yx| \leq n_0$, 2) $|x| \geq 1$, 3) $(\forall k \in \text{N}: yx^kz \in \text{L}) \vee (\forall k \in \text{N}: yx^kz \notin \text{L})$

Nech p je prvočíslo, pre ktoré platí $p \geq n_0+2$ (musí také existovať, keďže prvočísiel je nekonečne veľa).

Vezmime slovo $\boldsymbol{w = a^p}$. $|w| = p \geq n_0$

Nech $|x| = m$. Potom $|yz| = p - m$.

Uvažujme o slove $yx^{p-m}z$.

$|yx^{p-m}z| = |yz| + |x^{p-m}| = |yz| + (p - m)*|x| = p - m + (p - m)*m = (m+1)*(p - m)$. Pozrime sa na čísla v tomto súčine:

- $m+1$:
  podľa 1) $|x| \geq 1$ a keďže $|x|=m$, tak $m \geq 1$
  ak $m \geq 1 \Rightarrow m+1 \geq 2 \Rightarrow m+1 \neq 1$.
- $p-m$:
  $m = |x| \leq |yx| \leq n_0 \Rightarrow m \leq n_0$
  $m \leq n_0 < n_0+2 \leq p \Rightarrow$ *m je minimálne o 2 menšie ako p* $\Rightarrow p - m \neq 1$.

Keďže $m+1 \neq 1$ aj $p - m \neq 1 \Rightarrow (m+1)*(p - m)$ je **zložené číslo** $\Rightarrow \boldsymbol{yx^{p-m}z \notin L}$. Ale $\boldsymbol{yxz \in L}$, takže sme dostali spor $\Rightarrow$ L nie je regulárny.

Použitá literatúra

Hromkovič Juraj: Teoretical Computer Science