# Reg. izrazi in avtomati, neregularnost

Uroš Čibej

# Pregled

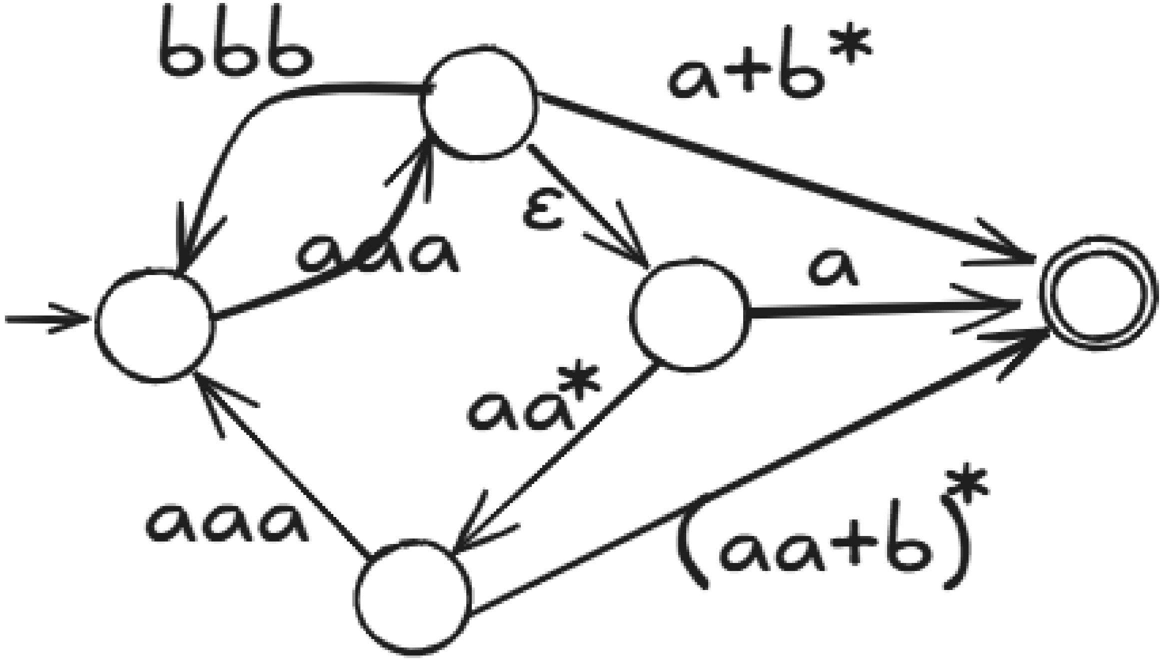
- KA $\rightarrow$ RI
- Lema o napihovanju
- Uporaba leme o napihovanju

# Literatura

- Sipser razdelka 1.3 in 1.4
- https://introtcs.org/public/lec_05_infinite.html (za radovedne, razdelek 6.5)

# Posplošeni NKA

Avtomati, ki imajo prehode preko reg. izrazov



# Posebna oblika (PNKA)

1. začetno stanje ima prehode do vseh stanj in nobenega prehoda do njega
2. obstaja zgolj eno končno stanje, ki ima prehode iz vseh stanj in nobenega prehoda iz njega
3. Vsa ostala stanja imajo prehode do vseh ostalih stanj (razen začetnega) in tudi do sebe.

# Vsak PNKA -> posebno obliko

1. dodamo dve novi stanji, eno začetno eno končno $(q_s, q_e)$
2. dodamo $\varepsilon$ prehod iz novega začetnega do starega začetnega in od starih končnih do novega končnega
3. Vsi ostali manjkajoči prehodi so preko $\emptyset$

# Formalna definicija PNKA (posebne oblike)

$$M = \langle Q, \Sigma, \delta, q_s, q_e \rangle$$

- $\delta : (Q \setminus \{q_e\}) \times (Q \setminus \{q_s\}) \longrightarrow RI$

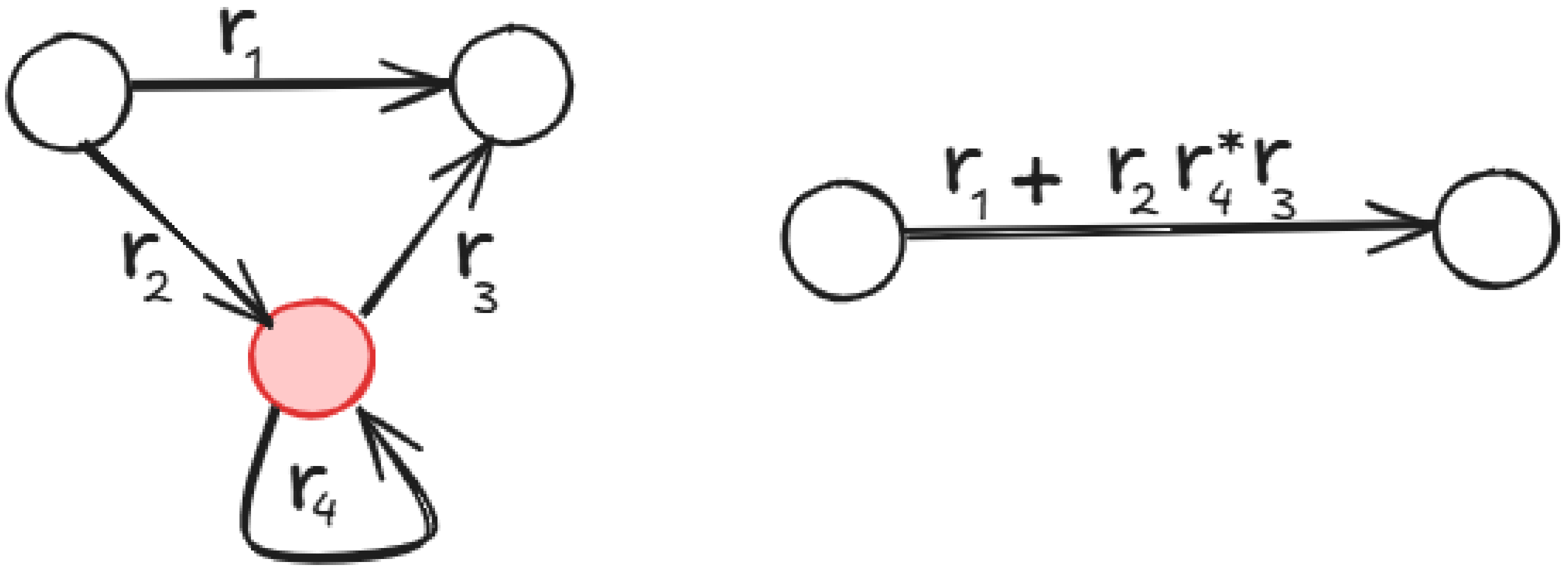
# Jezik PNKA

$$L(M) = \{w = w_1 w_2 \dots w_n \mid \exists q_s q_1 \dots q_e, w_i \in L(\delta(q_{i-1}, q_i))\}$$

$q_0 = q_s$

$q_n = q_e$

# Odstranjevanje stanj



# Algoritem $KA \rightarrow RI$

*convert*($G$):

- **if** $|Q| == 2$: **return** $\delta(q_s, q_e)$
- izberi poljubno vozlišče $q_x \notin \{q_s, q_e\}$
- zgradimo avtomat $G'$, ki nima več stanja $q_x$ in funkcijo $\delta'$
- $\delta'(q_i, q_j) = r_1 r_2^* r_3 + r_4$, kjer so:
  - $r_1 = \delta(q_i, q_x)$,$r_2 = \delta(q_x, q_x)$, $r_3 = \delta(q_x, q_j)$, $r_4 = \delta(q_i, q_j)$
- **return** *convert*($G'$)

# Dokaz pravilnosti

**Izrek.**

$$\forall G \in PNKA : L(G) = L(convert(G))$$

# Indukcija pa številu stanj $G$

**|Q|==2**:

obstaja samo prehod $q_s \to q_e$ in to je tudi edino možno zaporedje stanj iz začetnega v končno stanje. Jezik avtomata je torej enak $\delta(q_s, q_e)$ kar tudi *convert* vrne.

# Indukcija pa številu stanj $G$

**Indukcijski korak** predpostavimo, da izrek drži za $k-1$ stanj (avtomat $G'$) in pokažimo, da drži tudi za $G$. Pokazati moramo:

$$w \in L(G) \iff w \in L(G')$$

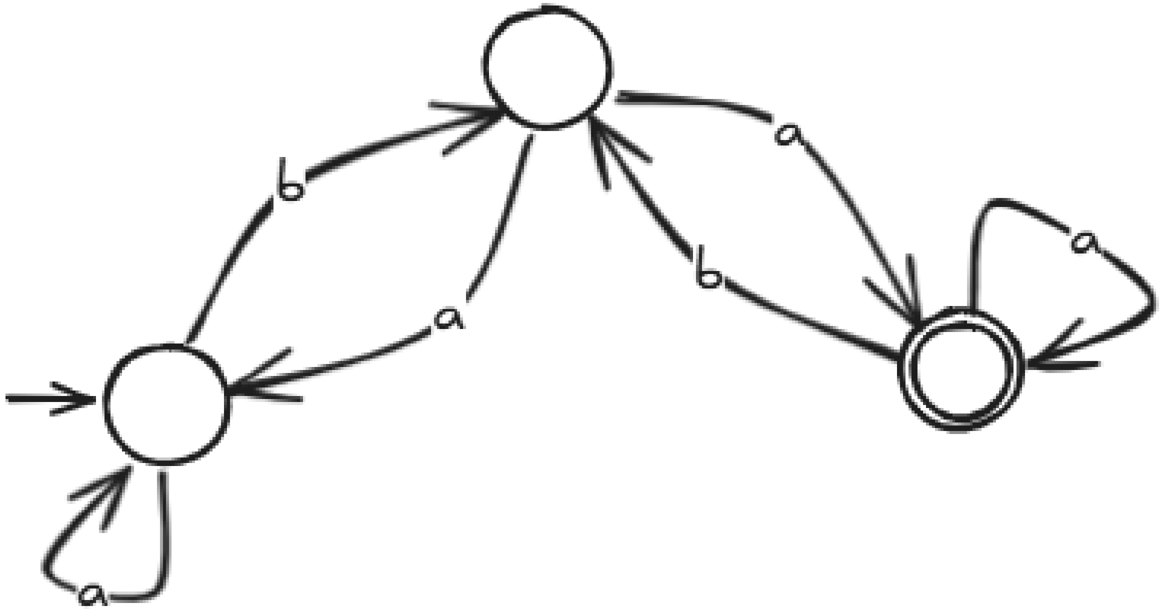
- $\implies$
  - $q_s q_1, \ldots, q_e$ je zaporedje, ki sprejema $w$ v avtomatu $G$ in $w_1 w_2 \ldots w_n$ je pripadajoče razbitje besede $w$.
    a. $q_x$ ni v tem zaporedju, potem isto zaporedje sprejema $w$ v $G'$
    b. $q_x$ je v tem zaporedju, potem odstranimo vse pojavitve $q_x$ iz zaporedja (pripadajoče besede staknemo) in dobimo veljavno zaporedje stanj.

$\Longleftarrow$

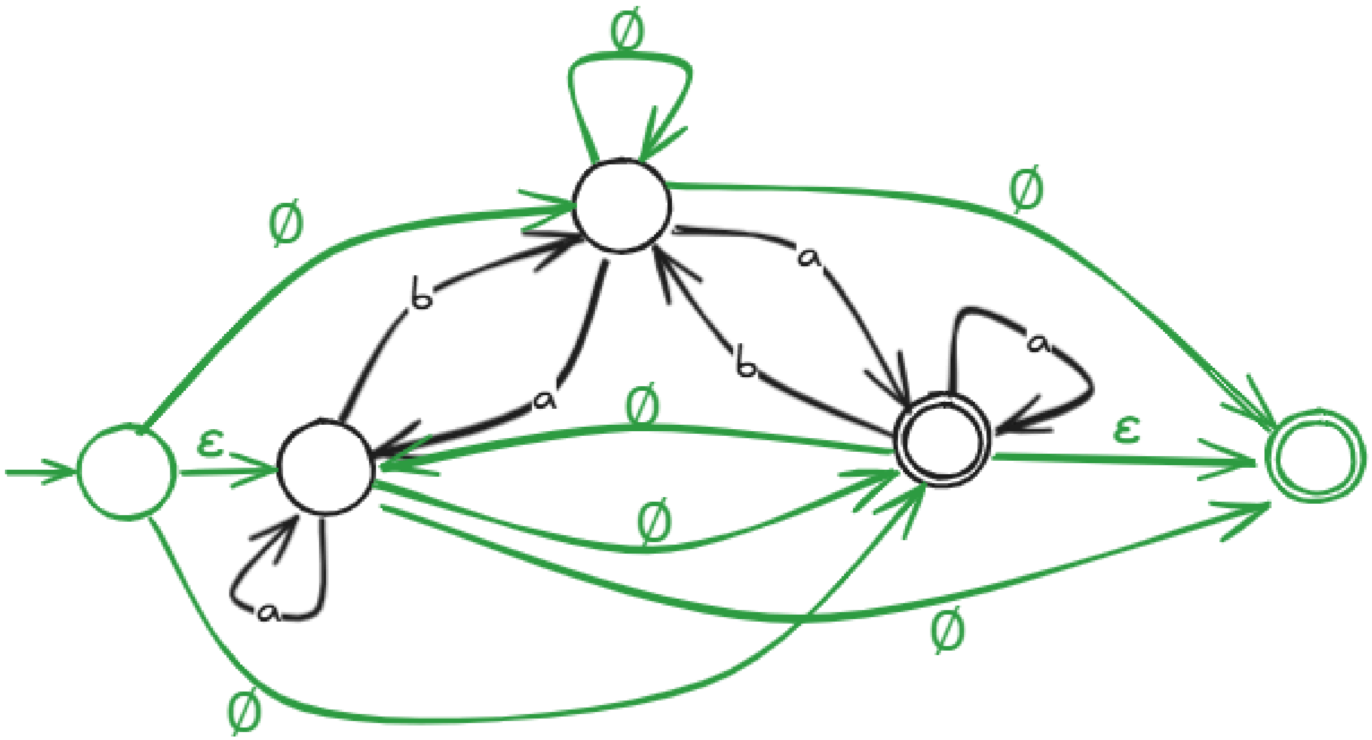
$w \in L(G')$

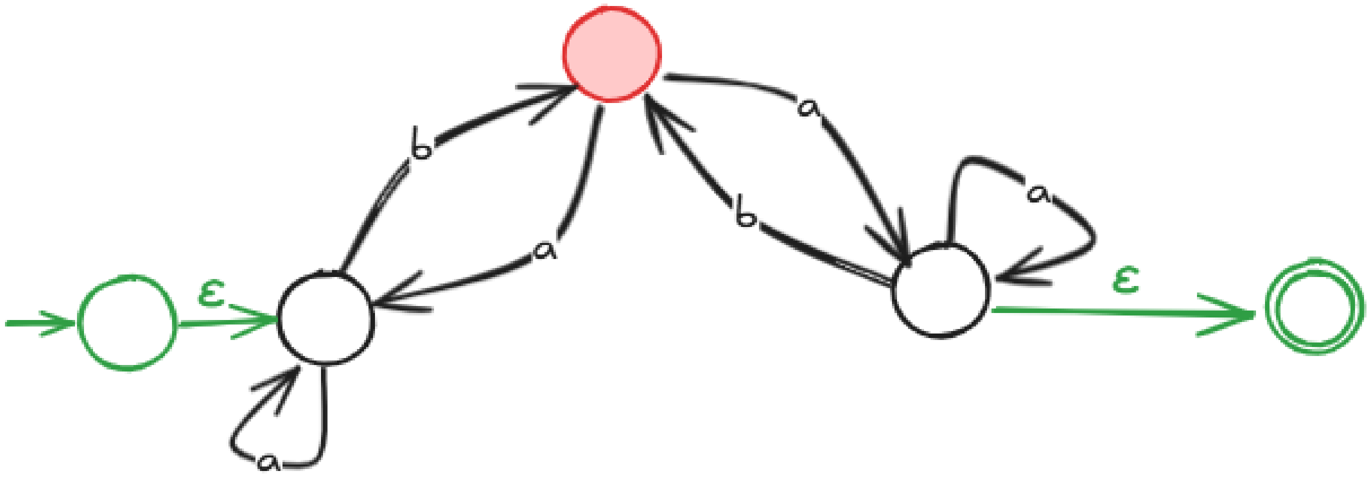
- vsak prehod v $G'$ $q_i q_{i+1}$ predstavlja dve stanji tudi v $G$
- regularni izraz $\delta'(q_i, q_{i+1}) = R_1 + R_2$:
  - če $w_i \in L(R_2)$ potem to predstavlja prehod $q_i q_{i+1}$ tudi v G
  - sicer $w_i$ ustrezno razbijemo s prehodi preko $q_x$

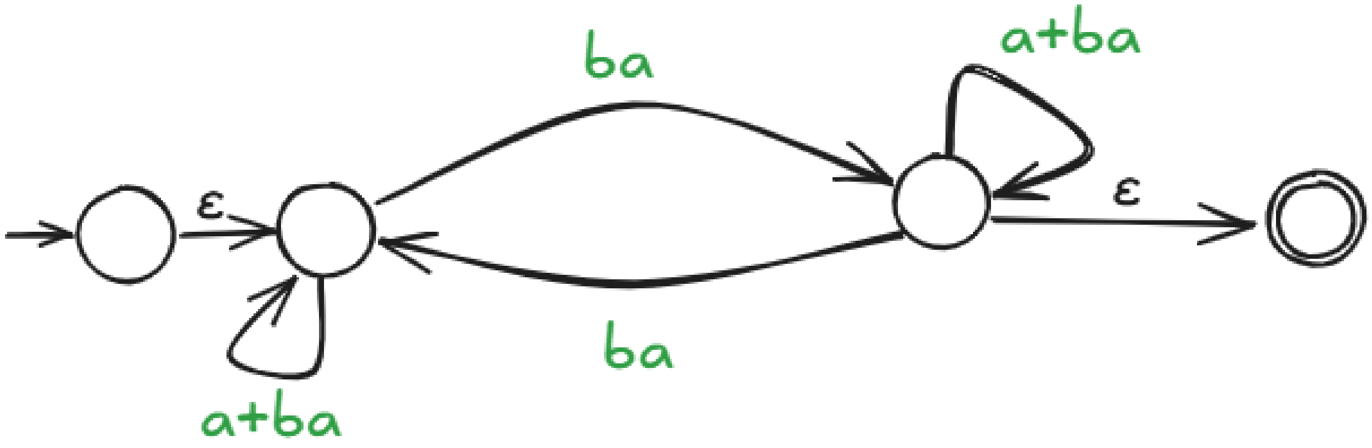
# Primer

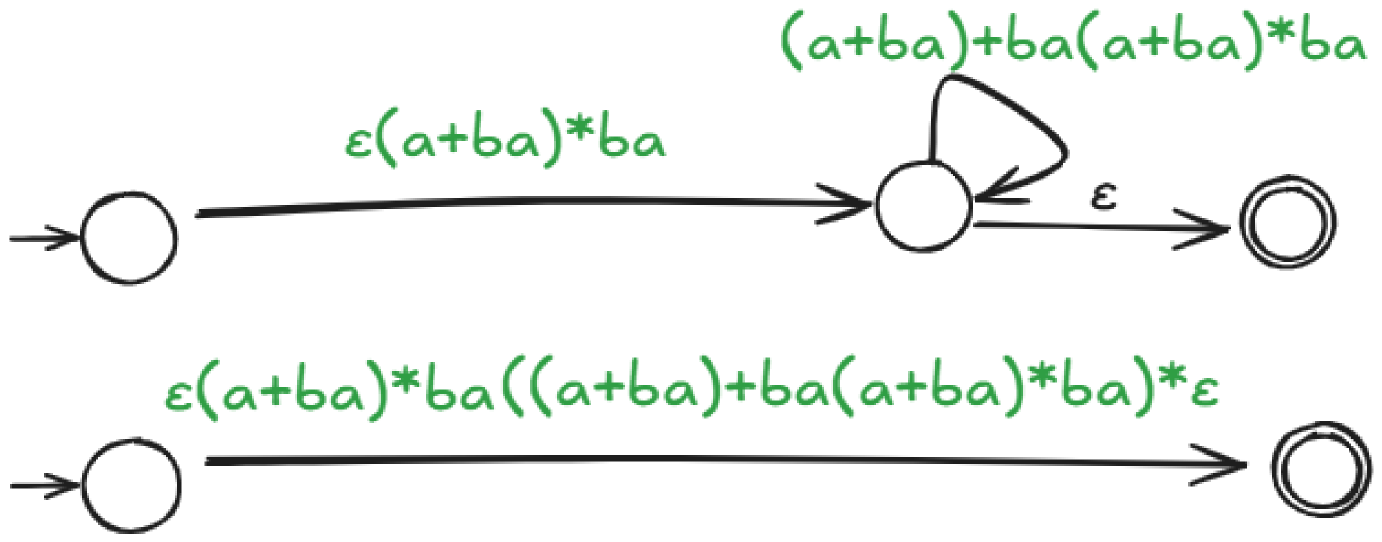


# Primer: pretvorba v posebno obliko




b
a
a
b
a
ε
a
ε


ba
a+ba
ε
ε
ba
a+ba

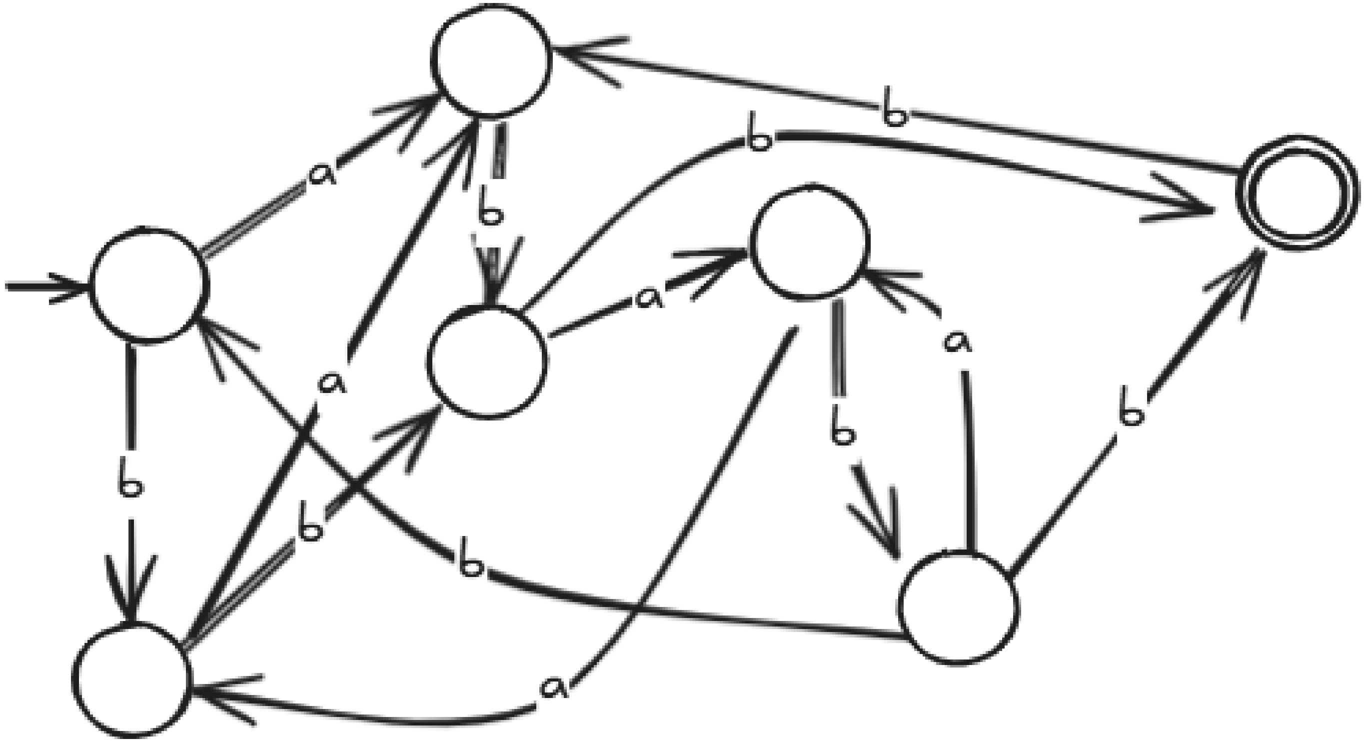

(a+ba)+ba(a+ba)*ba
ε(a+ba)*ba
ε
ε(a+ba)*ba((a+ba)+ba(a+ba)*ba)*ε

# Neregularnost

- česa avtomati niso sposobni?
- kako to dokazati?
- Ideja dokaza:
    i. identificirati lastnost, ki velja za vse regularne jezike
    ii. pokazati, da za nek jezik ta lastnost ne velja

# Izziv

Poiščite najdaljšo besedo $w \in L(M)$, pri kateri se stanje ne ponovi



# Lema o napihovanju

**Lema:** Za vsak regularni jezik $L$, obstaja konstanta $n$, da vsako besedo $w \in L$, ki je daljša od $n$ ($|w| \geq n$) lahko razbijemo na 3 komponente:

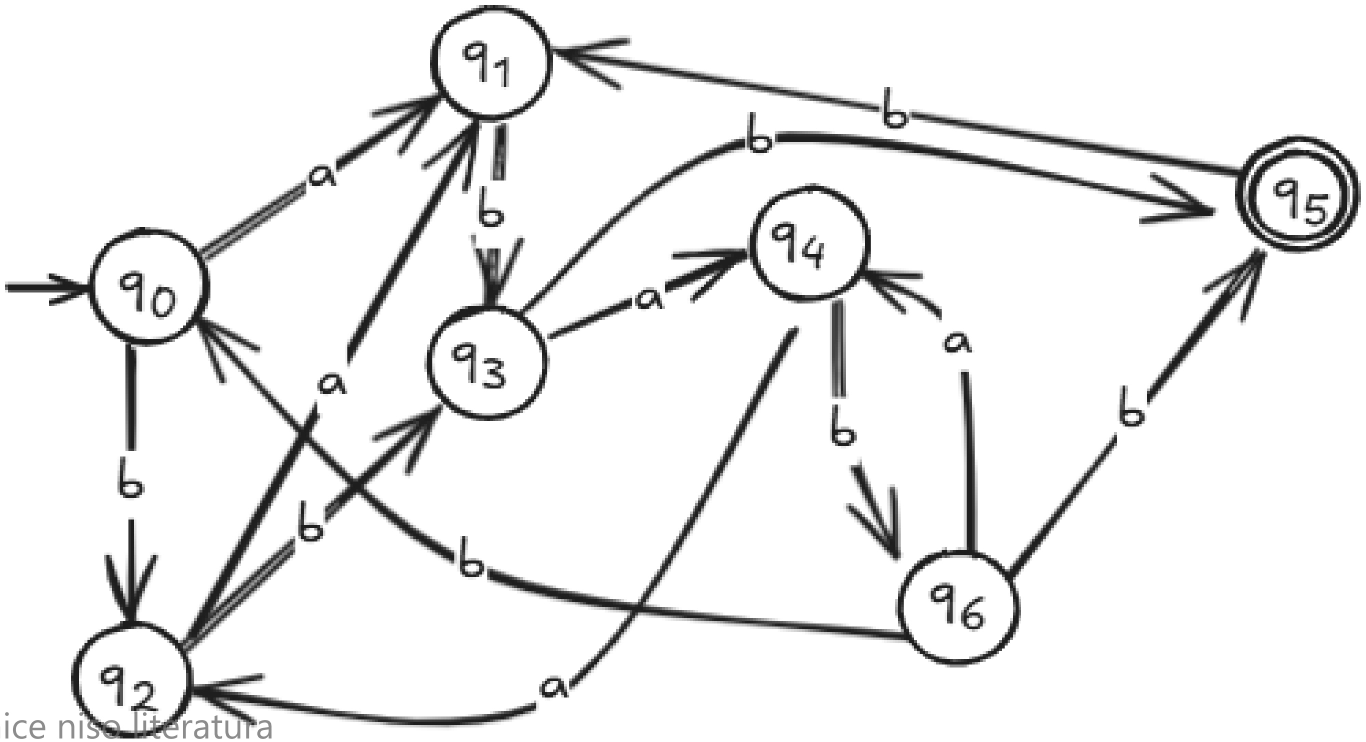
$$w = xyz$$

za katere velja:

1. $|y| > 0$
2. $|xy| \leq n$
3. $\forall i \geq 0 : xy^i z \in L$

# Preizkusimo

$w_1 = abbbbabb$
$w_2 = bababbabb$
$w_3 = abbbbbbbbbb$



# Dokaz - opis splošnega stanja

- Naj bo $M$ DKA$, $L(M) = L$ in $n = |Q|$.
- $w = a_1 a_2 \ldots a_k$, pri $k \geq n$ in $w \in L$.
- zaporedje stanj pri sprejemanju $r_1 r_2 \ldots r_k r_{k+1}$.
- $\exists j, l : j < l \leq n + 1$, $q_j = q_l$

## Dokaz - zaključki

- $x = a_1 \dots a_{j-1}, y = a_j \dots a_{l-1}, z = a_l \dots a_k$
- vemo $r_{k+1} \in F$ in besedo $xy^iz$ sprejmemo s ponavljanjem zaporedja stanj $q_j \dots q_l$
- ker $j < l \leq n + 1$ vemo $|y| > 0$ in $|xy| \leq n$

# Uporaba

Kako dokažemo, da za jezik $L$ LON ne velja:

1. Izberemo besedo $w \in L, |w| \geq n$
2. $\forall w = xyz \; |y| > 0, |xy| \leq n$
3. $\exists i : xy^i z \notin L$

# Interpretacija tega dokazovanja

Poljuben avtomat, ki bi se pretvarjal, da zna sprejeti $L$

1. $n$ predstavlja $|Q|$
2. izbrana beseda $w$ se v tem avtomatu "zacikla"
   - $w = xyz$
   - ker moramo biti pripravljeni na vse avtomate, pregledamo vse možne delitve
3. najdemo število ciklov $i$, pri katerem ta avtomat besedo sprejme, pa je ne bi smel

# Primer

$$L = \{a^k b^k | k \geq 0\}$$

Jezik ni regularen

# Dokaz

1. izberemo dovolj dolgo besedo, ki pripada jeziku ($n$ je konstanta iz leme)

$$w = a^n b^n$$

2. parametrično zapišimo vse možne delitve

$$w = xyz, |xy| \leq n, |y| > 0$$

   - $x = a^l$
   - $y = a^m$
   - $z = a^{n-l-m} b^n$

3. pri vseh delitvah pri $i = 2$ dobimo besedo

$$a^{n+m} b^n \notin L$$

# Primer

$$L = \{a^i b^j \mid i > j\}$$