Prvi kolokvij: rešitve in točkovanje praktičnega dela

Naloga

Naj bosta L_k in P_k jezika nad abecedo $\{1, 2, \ldots, k\}$ za nek $k \in \{1, \ldots, 9\}$. Definirajmo ju takole:

- L_k je množica besed $a_1 a_2 \ldots a_n$ ($n \geqslant 2$), pri katerih obstaja tak $i \in \{1, \ldots, n-1\}$, da je $a_i = a_{i+1}$, pri vseh ostalih indeksih j pa velja $a_j \neq a_{j+1}$.
- Jezik P_k je definiran enako, le da zahtevamo, da se indeks i nahaja v prvi polovici besede ($i \leqslant \lfloor n / 2 \rfloor$).

Primeri besed v jeziku L_3 so 33, 112, 3221, 13323, 31221 in 1312311. Vse naštete besede razen zadnjih dveh so tudi v jeziku P_3. Besede ε, 2, 111 in 3233211 niso niti v L_3 niti v P_3.

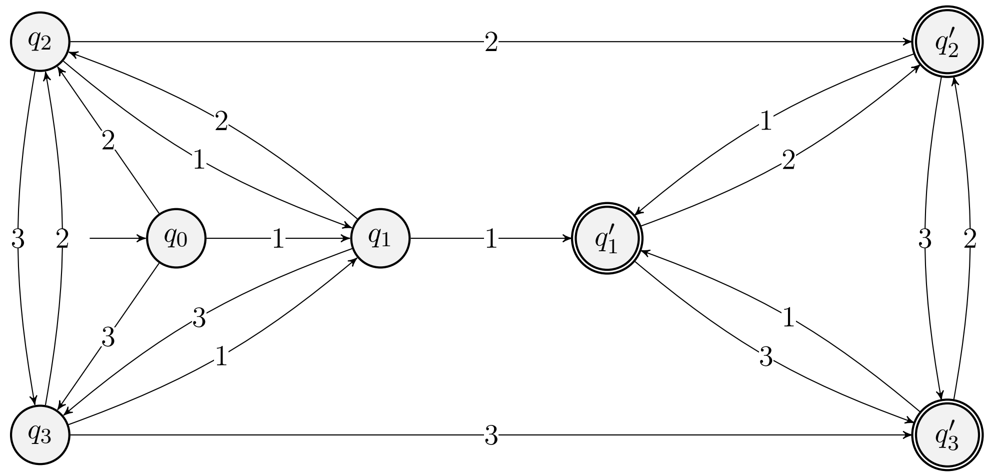
(a) Zgradite (ne nujno deterministični) končni avtomat za jezik L_3.

(b) Dokažite, da je jezik P_k regularen natanko tedaj, ko je $k = 1$.

(c) Zapišite kontekstno neodvisno gramatiko za jezik P_2.

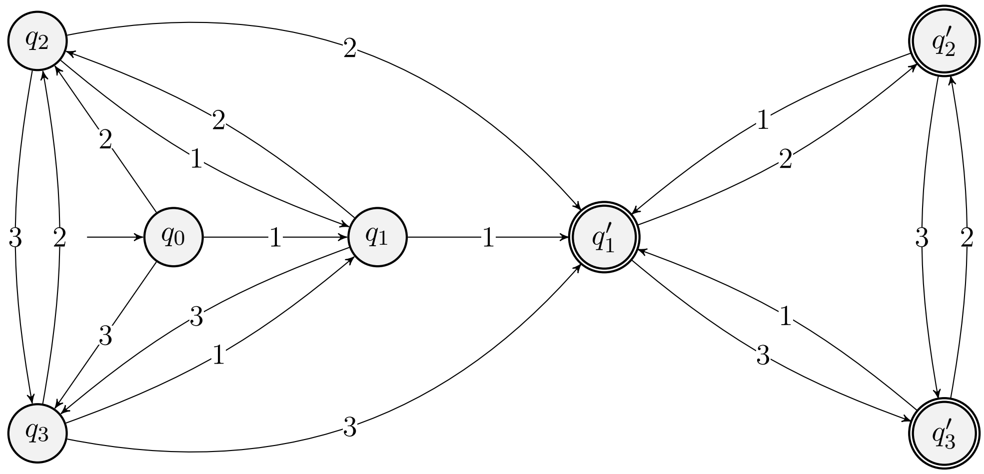
Rešitev naloge (a)

[6 točk] Dokler ne preberemo dveh enakih zaporednih znakov, krmarimo med stanji q_1, q_2 in q_3 (v stanje q_k preidemo, ko preberemo znak k). Takoj ko preberemo dva enaka zaporedna znaka k, preidemo v stanje q'_k, nato pa na enak način kot v prvem delu niza prehajamo med stanji q'_1, q'_2 in q'_3. Ker gradimo NKA, »slepega« stanj ne potrebujemo, seveda pa ni nič narobe, če ga kljub temu vključimo.



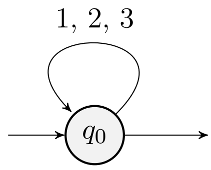
Pogoste napake:

- **[$\leqslant$ 3 točke]** Nekatere rešitve so izgledale (približno) takole:



  Ko pridemo do stanja q_1', »pozabimo« nazadnje prebrani simbol, zato lahko, denimo, sprejmemo besedo 1222.

- **[$\leqslant$ 2 točki]** Marsikdo je avtomat pričel risati takole:



  To je groba napaka (oziroma dokaz nerazumevanja delovanja nedeterminističnih končnih avtomatov): do trenutka, ko se avtomat »odloči«, da bo zapustil stanje q_0 in jo mahnil končnemu stanju naproti, je lahko prebral že zelo veliko parov enakih simbolov.

Rešitev naloge (b) — smer $k = 1 \implies P_k$ je regularen

[2 točki] Pri $k = 1$ je $P_k = \{11\}$. Ta jezik je končen in zato regularen.

Pogoste napake:

- **[0 točk]** Ni vas bilo tako malo, ki ste trdili, da je jezik P_1 regularen zato, ker besede 11 ni mogoče napihovati. Napačen sklep: argument »nenapihljivosti« uporabljamo ravno za dokazovanje, da jezik *ni* regularen.

Rešitev naloge (b) — smer P_k je regularen $\implies k = 1$

[6 točk] Trditev, da jezik P_k za $k \geqslant 2$ ni regularen, lahko dokažemo z lemo o napihovanju. Pri kateremkoli $k \geqslant 2$ lahko vzamemo besedo $z = (12)^n 11 (21)^n$. Beseda z je veljavna izbira, saj je za vsak $n \in \mathbb{N}$ dolga najmanj n in pripada jeziku P_k (za poljuben $k \geqslant 2$). Sedaj obravnavamo vse možne podnize v. Ta mora biti neprazen in se mora nahajati med prvimi n znaki besede z (torej znotraj podniza $(12)^{n/2}$):

- $v = (12)^p$ $(p \geqslant 1)$: Pri napihovanju z eksponentom 2 (ali več) bomo podniz 11 prestavili na desno polovico besede, zato ta ne bo več pripadala jeziku P_k.

- $v = (21)^p$ ($p \geqslant 1$): Ta primer je enakovreden prejšnjemu.
- $v = 1(21)^p$ ($p \geqslant 0$): Pri napihovanju z eksponentom 2 bo v besedi poleg obstoječega podniza 11 nastal še en podniz 11. Beseda tako izpade iz jezika.
- $v = 2(12)^p$ ($p \geqslant 0$): Pri napihovanju z eksponentom 2 bo v besedi poleg obstoječega podniza 11 nastal še podniz 22. Beseda tako izpade iz jezika.

Ker lahko pri vseh možnih položajih in dolžinah podniza v najdemo »napihovalni« eksponent, ki besedo izloči iz jezika, lahko zaključimo, da jezik P_k ni regularen.

Pogoste napake:

- **[$\leqslant$ 4 točke]** Primer $v = (12)^p$ ($p \geqslant 1$) ste obravnavali, na ostale pa pozabili.
- **[$\leqslant$ 2 točki]** Obravnavali ste samo eno konkretno izbiro za niz v (npr. $v = 12$). To je še hujša napaka: kdor je grešil v skladu s prejšnjo točko, je obravnaval vsaj celotno *družino* možnosti, četudi še vedno ni upošteval *vseh* možnosti.
- **[0 točk]** »Jezik P_2 ni regularen, ker si končni avtomat ne more zapomniti položaja sredine besede.« To ni dokaz!
- Nekateri ste dokazali, da jezik P_2 ni regularen, in potem sklepali, da to velja tudi za jezike $P_3, \ldots, P_9$, saj je $P_2 \subseteq P_3 \subseteq \ldots \subseteq P_9$. Napačen sklep! Jezik P_2 je tudi podmnožica jezika Σ^*, ta pa je, kot vemo, regularen.

Rešitev naloge (c)

[6 točk] Tukaj je več možnih pristopov, eden od njih pa gre takole. Najprej zgradimo gramatiko, ki tvori besede sode dolžine, v katerih se podvojitev (niz 11 ali 22) nahaja točno na sredini:

$$\begin{aligned} S &\to A \mid B \\ A &\to 1B1 \mid 11 \\ B &\to 2A2 \mid 22 \end{aligned}$$

Upoštevajmo še možnost, da lahko nizu $1 \ldots 1$ dodamo poljubno dolgo zaporedje izmenjujočih se dvojk in enic, nizu $2 \ldots 2$ pa poljubno dolgo zaporedje izmenjujočih se enic in dvojk:

$$\begin{aligned} S &\to AC \mid BD \\ A &\to 1B1 \mid 11 \\ B &\to 2A2 \mid 22 \\ C &\to 2D \mid \varepsilon \\ D &\to 1C \mid \varepsilon \end{aligned}$$

Pogoste napake:

- **[$\leqslant$ 4 točke]** Z morebitno izjemo končnega števila besed (tipično gre za posebne primere, ki jih ločeno obravnavate) tvori vaša gramatika izključno besede sode dolžine ali izključno besede lihe dolžine.
- **[$\leqslant$ 3 točke]** Z morebitno izjemo končnega števila besed tvori vaša gramatika samo besede, pri katerih podvojitev nastopa na sredini (npr. 12122121 ali 1211212).

- **[$\leqslant$ 2 točki]** Z morebitno izjemo končnega števila besed tvori vaša gramatika samo besede sode dolžine ali samo besede lihe dolžine, pri katerih podvojitev nastopa na sredini.
- **[$\leqslant$ 1 točka]** Z morebitno izjemo končnega števila besed tvori vaša gramatika besede, pri katerih lahko podvojitev nastopa kjerkoli, ali pa besede, pri katerih podvojitev nastopa na fiksni razdalji od začetka ali konca besede (npr. na začetku ali koncu besede). Ta poenostavitev je hujši prekršek od vseh prejšnjih, saj je jezik takšne gramatike *regularen*. Kot vemo, za razpoznavanje takšnega jezika sploh ne potrebujemo kontekstno neodvisne gramatike, ampak zadošča končni avtomat.