Regularni jeziki in končni avtomati

Rok Kaufman, Gimnazija Vič, Ljubljana
Filip Kozarski, Gimnazija Bežigrad, Ljubljana
Mentor: Gašper Zadnik, UL FMF

Leta 1936 je angleški matematik Alan Turing predstavil idejo o abstraktnem modelu stroja, ki bi bil sposoben s pomočjo poljubno dolgega spominskega traku in nekaj osnovnimi operacijami izvajati zahtevnejše algoritme. Enajst let kasneje je svojo zamisel razširil z idejo o t.i. splošnem Turingovem stroju, ki bi bil sposoben oponašati delovanje vseh možnih Turingovih strojev.

Poseben primeri Turingovih strojev so med drugim tudi končni avtomati (Finite state automata). Delijo se na deterministične in nedeterministične; slednjih ne bomo obravnavali. Kljub preprostosti so deterministični končni avtomati (DKA) zelo aplikabilni. So srce iskalnih algoritmov, npr. iskalnikov nizov v urejevalnikih besedil, ali celo medmrežnih brskalnikov, kot je Google [2].

1 Definicije in osnovne lastnosti

Definicija 1. *Deterministični končni avtomat (DKA)* definira peterica

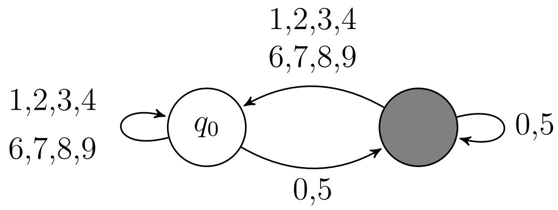
$$M = (Q, \Sigma, \delta, q_0, F),$$

kjer

- Q predstavlja končno množico *možnih stanj*,
- Σ predstavlja končno množico znakov, ki jo imenujemo *abeceda*,
- $\delta : Q \times \Sigma \to Q$ predstavlja *prehodno funkcijo*,
- $q_0 \in Q$ predstavlja *začetno stanje*,
- $F \subseteq Q$ predstavlja množico *sprejemljivih stanj*.

Vsakemu DKA lahko priredimo tudi tranzicijski graf, da si lažje predstavljamo potek njegovega delovanja. Oglejmo si tak graf na preprostem primeru ugotavljanja deljivosti števila (zapisanega v desetiškem sestavu) s pet (slika 1).

Abecedo v tem primeru sestavljajo števke desetiškega sestava. Avtomat ima dve stanji. V sprejemljivo stanje pridemo, če je prebrana števka 5 ali 0, sicer se preusmeri v drugo stanje (ki je hkrati tudi začetno). Oznake iz definicije 1 in tranzicijski graf povežemo:

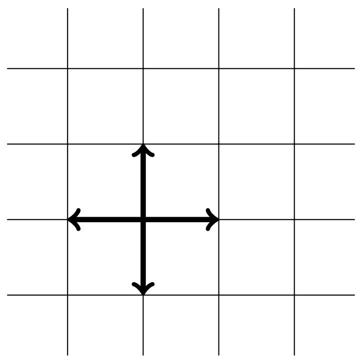


Slika 1: Tranzicijski graf, ki prikazuje DKA, ki sprejema večkratnike števila pet.

- množico Q možnih stanj predstavljata oba krogca s slike 1;
- *abeceda* Σ vsebuje vse števke od 0 do 9;
- *tranzicijsko funkcijo* δ predstavljajo puščice na sliki 1;
- *začetno stanje* q_0 je bel krogec na isti sliki;
- *končno stanje* je krogec, obarvan s sivo barvo.

2 Problemi

- Imamo neskončno veliko kvadratno mrežo. Iz enega vozlišča se lahko po črtah premaknemo le v eno od sosednjih štirih vozlišč. Naš način potovanja po mreži lahko opišemo s preprostim jezikom, ki uporablja štiri znake; vsak pomeni premik v eno izmed štirih smeri. Nekoliko manjši jezik so vsi sprehodi po mreži, ki se začnejo in končajo v isti tocki. Ali obstaja DKA, ki prepozna slednji jezik?



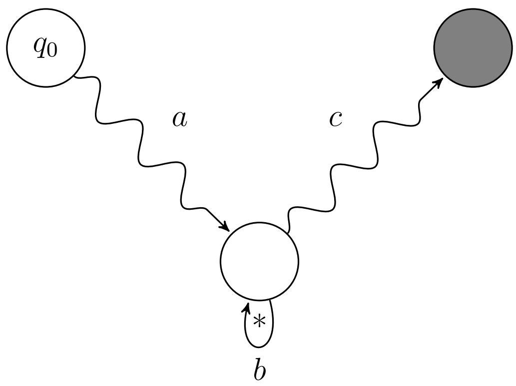
Slika 2: Mreža, s katero se v tem primeru ukvarjamo, in možne poteze.

Denimo, da obstaja DKA z n stanji, ki prepozna ta jezik. Gibanje po mreži ni omejeno, zato obstaja sprehod dolžine $(n+1)$, ki gre po samih različnih točkah mreže (npr. $(n+1)$ pomikov v levo). Po Dirichletovem načelu zato obstaja vsaj eno stanje v avtomatu, ki je med tem sprehodom obiskano (vsaj) dvakrat. Med prvim in drugim obiskom tega stanja smo naredili $m > 0$ korakov levo po mreži, prav tako pa bi jih lahko naredili $2m$ in avtomat tega ne bi zabeležil (DKA ne bi opazili razlike med $(n+1)$ in $(n+m+1)$ pomiki v levo). Tu pridemo do protislovja, saj se pri dveh različnih številih potrebnih korakov za povratek v začetno stanje nahajamo v istem stanju avtomata.

- Ponovno naj abecedo tvorijo števke v desetiškem sestavu. Ali obstaja končni avtomat, ki prepoznava jezik praštevil?

Tak končni avtomat ne obstaja, kar bomo dokazali s protislovjem.

Dokaz. Na razpolago imamo končno število možnih stanj. Izberimo praštevilo p, ki ima več števk, kot je stanj v DKA. Zopet po Dirichletovem načelu obstaja neko stanje DKA, ki je med branjem števila p obiskano vsaj dvakrat, med dvema obiskoma tega stanja se torej zgodi neka zanka b. Označimo del števila p pred pojavitvijo zanke b z a, del števila p, ki b sledi, pa s c (število beremo od leve proti desni, torej je p oblike $\overline{abc}$, seveda pa je tudi $\overline{ac}$ praštevilo).



Slika 3: Shema zanke (*), ki v končnem avtomatu za prepoznavanje praštevil mora obstajati. Opozorilo: b je niz števk, ki ga dobimo, ko obhodimo zanko (*).

Torej DKA kot praštevilo prepozna vsako število oblike $\overline{a\underbrace{bb\ldots b}_{n}c}$. Naj bo m dolžina niza b v decimalnem zapisu in l dolžina niza c. Imejmo še poljubno praštevilo q, ki je različno od 2 in 5, torej je tuje vsem potencam števila 10. Brez škode za splošnost je dolžina števila b deljiva s $(q-1)$ (sicer namesto b v zgoraj opisani družini praštevil zamenjamo z $\underbrace{\overline{bb\ldots b}}_{q-1}$).

Razlika med prašteviloma, ki imata n in $(n+1)$ b-jev, je:

$$\overline{a\underbrace{bb\dots b}_{n+1}c} - \overline{a\underbrace{bb\dots b}_{n}c} =$$

$$10^{m(n+1)+l}a + 10^{mn+l}b + \dots 10^l b + c - (10^{mn+l}a + 10^{(m(n-1)+l)}b + \dots + 10^l b + c) =$$

$$a \cdot 10^{m(n+1)+l} - a \cdot 10^{mn+l} + b \cdot 10^{mn+l} =$$

$$= 10^{mn+l}\underbrace{(a \cdot 10^m - a + b)}_{z} =$$

$$= 10^{mn+l}z$$

Naj bo p praštevilo oblike $\overline{ac}$. Iz konstrukcije avtomata vemo, da je

$$p + z(10^{m+l} + 10^{2m+l} + \dots + 10^{km+l}) =$$

$$p + z \cdot 10^l(10^m + 10^{2m} + \dots + 10^{km})$$

praštevilo za vsak $k \geq 0$. Po malem Fermatovem izreku je $10^{(q-1)} \equiv 1(\text{mod } q)$. Ker $(q-1)$ deli m, je tudi $10^{sm} \equiv 1(\text{mod } q)$ za vsak $s \in \mathbb{N}$. Ker je vsak člen v oklepaju $(10^m + 10^{2m} + \dots + 10^{km})$ kongruenten 1 po modulu q, je vsota v oklepaju kongruentna k, ki je poljubno naravno število. Če je z tuj q, potem da $z \cdot 10^l(10^m + 10^{2m} + \dots + 10^{km})$ tudi poljuben ostanek pri deljenju s q (tj. za vsak r obstaja $k \in \mathbb{N}$, da je $z \cdot 10^l(10^m + 10^{2m} + \dots + 10^{km}) \equiv r(\text{mod } q)$). Torej je tudi celotna vrednost izraza

$$p + z \cdot 10^l(10^m + 10^{2m} + \dots + 10^{km})$$

po modulu q lahko kongruentna kateremukoli številu. Torej obstaja tudi k (celo neskončno mnogo takšnih), pri katerem je

$$r := p + z \cdot 10^l(10^m + 10^{2m} + \dots + 10^{km}) \equiv 0(\text{mod } q)$$

V tem primeru bi bil r deljiv s q, DKA pa bi ga sprejel kot praštevilo. Torej z ne more biti tuj q. Ker pa je bilo q poljubno praštevilo (razen 2 in 5), nas to pripelje v protislovje, saj bi moral biti z deljiv z vsemi praštevili (razen morda 2 in 5). □

- Kakšen pa je končni avtomat, ki prepozna potence števila dve, ostala števila pa zavrne?

Za enostavno rešitev število najprej pretvotimo v dvojiški zapis. Tu so sprejemljiva števila, ki imajo enico le na prvem mestu, za kar potrebujemo le preprost avtomat. (Ko imamo prvo enico pridemo v sprejemljivo stanje. Ničla nas ohranja v tem položaju, če pa se pojavi še ena enica, pademo ven in ne moremo več nazaj v sprejemljivo stanje.)

Kako pa bi izgledal avtomat, ki bi presodil, ali je število oblike 2^k, v primeru, da ohranimo desetiški sestav? Je dvojiški sestav v tem primeru neprimerljivo boljši od desetiškega oz. DKA v desetiškem sestavu sploh obstaja? Predstavili smo dovolj primerov, da ta problem lahko razišče bralec sam.

Literatura

[1] Peter Linz: *An Introduction to Formal Languages and Automata.* Jones and Bartlett publishers, Sudbury, Massachusetts, 2001.

[2] `http://www.google.si`

[3] `http://en.wikipedia.org`