Regularni izrazi

- Regularni izrazi nad alfabetom Σ formiraju se na sledeći način:
  - $\emptyset$, ε i s, za svako $s \in \Sigma$, jesu regularni izrazi;
  - ako su α i β regularni izrazi, onda su to i $(\alpha \cup \beta)$ i $(\alpha\beta)$;
  - ako je α regularan izraz, onda je to i α^*.
- Svakom regularnom izrazu α nad Σ, na prirodan način pridružujemo jedan regularan jezik $L(\alpha) \subseteq \Sigma^*$:
  - $L(\emptyset) = \emptyset$, $L(\varepsilon) = \{\varepsilon\}$, $L(s) = \{s\}$, $s \in \Sigma$;
  - $L(\alpha \cup \beta) = L(\alpha) \cup L(\beta)$, $L(\alpha\beta) = L(\alpha)L(\beta)$,
  - $L(\alpha^*) = L(\alpha)^*$.

Теорема

Jezik $L \subseteq \Sigma^*$ je regularan akko postoji regularan izraz α nad Σ takav da je $L = L(\alpha)$.

Regularni izrazi

Lema

Neka su $K \subseteq \Sigma^+$ i $L \subseteq \Sigma^*$ regularni jezici. Tada jednačina $X = XK \cup L$ ima jedinstveno rešenje $X = LK^*$ koje je regularan jezik.

- Za automat $\mathbb{M} = (\{q_0, q_1, \dots, q_n\}, q_0, F, \Sigma, \delta)$ definišimo sledeće skupove:

  $L_i = L(\mathbb{M}_i)$, gde je $\mathbb{M}_i = (\{q_0, q_1, \dots, q_n\}, q_0, \{q_i\}, \Sigma, \delta)$, $0 \leqslant i \leqslant n$;
  $K_{ij} = \{s \in \Sigma \mid \delta(q_j, s) = q_i\}$, $0 \leqslant i, j \leqslant n$.

- Ovi jezici su povezani sledećim jednakostima:

$$(\star) \qquad \begin{cases} L_0 = L_0 K_{00} \cup L_1 K_{01} \cup \cdots \cup L_n K_{0n} \cup \{\varepsilon\}, \\ L_1 = L_0 K_{10} \cup L_1 K_{11} \cup \cdots \cup L_n K_{1n}, \\ \vdots \\ L_n = L_0 K_{n0} \cup L_1 K_{n1} \cup \cdots \cup L_n K_{nn}. \end{cases}$$

- Važi da je $L(\mathbb{M}) = \bigcup_{q_i \in F} L_i$.

Jezici koji nisu regularni

- Za svaki alfabet Σ postoji samo prebrojivo mnogo regularnih izraza, dok podskupova od Σ^* ima neprebrojivo mnogo, odakle direktno sledi da postoje jezici nad Σ koji nisu regularni

Lema naduvavanja za regularne jezike

Теорема

Za svaki regularan jezik L nad nekim alfabetom Σ postoji prirodan broj n takav da se svaka reč w iz L dužine bar n ($|w| \geqslant n$) može zapisati u obliku $w = xyz$, za neke $x, y, z \in \Sigma^*$ tako da važi:

- $y \neq \varepsilon$;
- $|xy| \leqslant n$;
- $xy^k z \in L$, za svako $k \geqslant 0$.

Lema naduvavanja za regularne jezike

Posledica

Neka je $\mathbb{M} = (Q, q_0, F, \Sigma, \delta)$ *konačni automat.*

1) $L(\mathbb{M}) \neq \emptyset$ *akko postoji* $w \in L(\mathbb{M})$ *takva da je* $|w| < |Q|$*;*
2) $L(\mathbb{M})$ *je beskonačan akko postoji* $w \in L(\mathbb{M})$*takva da je* $|Q| \leqslant |w| < 2|Q|$*.*

Ekvivalentne reči

Дефиниција

Reči x i y su ekvivalentne u odnosu na jezik L, u oznaci $x \approx_L y$, akko $\{z \mid xz \in L\} = \{z \mid yz \in L\}$

- Relacija $\approx_L$ može da se uvede i na sledeći način:
  - Ako postoji z tako da $xz \in L$ i $yz \notin L$ tada $x \not\approx_L y$
  - Ako postoji z tako da $xz \notin L$ i $yz \in L$ tada $x \not\approx_L y$
  - U suprotnom $x \approx_L y$

Дефиниција

Dve reči $x, y \in \Sigma^*$ su ekvivalentne u odnosu na automat $M = (Q, q_0, F, \Sigma, \delta)$, u oznaci $x \sim_M y$, akko $q_0 x \vdash_{\mathbb{M}}^* t$ i $q_0 y \vdash_{\mathbb{M}}^* t$.

Ekvivalentne reči

Теорема

Za proizvoljni (deterministički) konačni automat $M = (Q, q_0, F, \Sigma, \delta)$ i proizvoljne reči $x, y \in \Sigma^*$, ako je $x \sim_M y$ tada je $x \approx_{L(M)} y$.

Posledica

Svaki konačni automat M koji prepoznaje jezik L mora da ima najmanje onoliko stanja koliko ima klasa ekvivalencije za relaciju $\approx_L$.

Posledica

Ako je jezik L regularan, tada relacija $\approx_L$ ima konačno mnogo klasa ekvivalencije.

Minimalni automat

Теорема (Myhill-Nerode teorema)

Neka je $L \subseteq \Sigma^*$ regularan jezik. Tada postoji minimalni konačni automat M koji ima broj stanja jednak broju klasa ekvivalencije relacije $\approx_L$.

Ekvivalentna stanja

Дефиниција

Za svako $i \geqslant 0$ definišemo relaciju ekvivalencije $\equiv_i$ na skupu stanja Q automata M tako da je $p \equiv_i q$ akko za svaku reč $z \in \Sigma^*$ i $|z| \leqslant i$, $\hat{\delta}(p, z) \in F \Leftrightarrow \hat{\delta}(q, z) \in F$.

Lema

Neka je $L = L(M)$. Neka su $p, q \in Q$ i neka je svako stanje automata M dostižno iz početnog stanja. Tada je $p \equiv_i q$ za svako $i \geqslant 0$ akko klase ekvivalencije relacije $\sim_M$ koje odgovaraju stanjima p i q odgovaraju istim klasama ekvivalencije za relaciju $\approx_L$.

Minimizacija automata

Algoritam na osnovu kog odredjujemo klase ekvivalencije za svako $\equiv_i$:

I korak Klase ekvivalencije za $\equiv_0$ su skupovi F i $Q \setminus F$.

II korak Za date klase ekvivalencije za $\equiv_i$ odredjuju se klase ekvivalencije za $\equiv_{i+1}$:
stanja koja su istoj klasi ekvivalencije za $\equiv_{i+1}$ moraju pripadati istoj klasi ekvivalencije i za $\equiv_i$. Dodatno $p \equiv_{i+1} q$ akko $p \equiv_i q$ i za svako $a \in \Sigma$, $\delta(p, a) \equiv_i \delta(q, a)$.

III korak Ako su klase ekvivalencije za $\equiv_{j+1}$ i $\equiv_j$ jednake, sa deljem klasa se staje sa zaključkom da nema novih klasa ni za naredne vrednosti i. Relacija $\equiv_j$ se zove fiksna tačka.