

Rozoznávanie jazykov konečnými monoidmi

Ročníkový projekt

Michaela Tokárová

1.12.2024

Obsah

1	Definície	5
2	Rozoznávanie jazykov monoidmi	7

Kapitola 1

Definície

Na úvod definujeme základné pojmy používané v tomto texte.

Definícia 1.0.1. Abeceda je ľubovoľná konečná neprázdna množina. Prvky abecedy nazývame symboly alebo písmená.

Definícia 1.0.2. Slovo nad abecedou Σ je konečná postupnosť symbolov množiny Σ . Prázdnu postupnosť (prázdne slovo) označujeme ε .

Definícia 1.0.3. Jazyk nad abecedou Σ je ľubovoľná množina slov nad abecedou Σ .

Definícia 1.0.4. Pologrupa je dvojica $(M, *)$ kde M je množina a $*$ je asociatívna binárna operácia na množine M .

Definícia 1.0.5. Neutrálny prvok pologrupy $(M, *)$ je prvok $e \in M$ taký, že pre všetky $x \in M$ platí

$$e * x = x = x * e$$

Neutrálne prvky budeme označovať ako e alebo 1 .

Definícia 1.0.6. Monoid je pologrupa obsahujúca neutrálny prvok. (Monoid budeme zapisovať aj ako trojicu $(M, *, e)$).

Príklad 1.0.7. Medzi monoidy patrí napríklad $(\mathbb{N}, +, 0)$ alebo $(\Sigma^*, \cdot, \varepsilon)$ kde $\cdot$ je operácia zretiazenia slov.

Príklad 1.0.8. Konečným monoidom je napr. $(\mathbb{Z}_n, +, 0)$ alebo tiež $(F, \circ, \text{Id}_S)$ kde F je množina funkcií $f : S \rightarrow S$, $\circ$ je operácia skladania funkcií a Id_S je identické zobrazenie na konečnej množine S .

Definícia 1.0.9. Konečný monoid je monoid $(M, *, e)$ taký, že množina M je konečná.

Definícia 1.0.10. Akcia Pod akciou pologrupy $(S, \cdot)$ rozumieme zobrazenie $\delta : Q \times S \rightarrow Q$ také, že pre $\forall x, y \in S$ a pre $\forall q \in Q$ platí $\delta(\delta(q, x), y) = \delta(q, xy)$. Ak S je monoid $(S, \cdot, 1)$, musí navyše pre $\forall q \in Q$ platiť $\delta(q, 1) = q$.

Definícia 1.0.11. Deterministický konečný automat je päťica $A = (S, Q, \delta, q_0, F)$, kde S je pologrupa, Q je konečná množina stavov, $\delta : Q \times S \rightarrow Q$ je akcia na S , $q_0 \in Q$ je počiatočný stav a $F \subseteq Q$ je množina akceptačných stavov. **Jazyk** akceptovaný automatom A je množina $L(A) = \{x \in S \mid \delta(q_0, x) \in F\}$.

Definícia 1.0.12. Konfigurácia deterministického konečného automatu je prvok (q, w) kde q je stav automatu a w je nespracovaná časť vstupného slova.

Definícia 1.0.13. Prechodový monoid Nech A je konečný automat $A = (S, Q, \delta, q_0, F)$. Označme $M_A = (\{\delta_x \mid x \in S\}, \circ, \delta_\varepsilon)$ kde δ_x je funkcia z Q do Q definovaná ako $\delta_x(q) = \delta(x, q)$. Potom M_A nazveme prechodovým monoidom, pretože pozostáva z funkcií, ktoré sú definované ako prechodové funkcie na slovách z S .

Definícia 1.0.14. Homomorfizmus h definujeme ako také zobrazenie z A do B , že pre $\forall x, y \in A$ platí $h(x \cdot y) = h(x) \cdot h(y)$. Pod obrazom jazyka L pri zobrazení homomorfizmom h rozumieme $h(L) = \{h(w) \mid w \in L\}$. **Homomorfizmus monoidov:** Pod homomorfizmom z monoidu $(M, \cdot, e)$ do monoidu $(N, *, 1)$ rozumieme zobrazenie $h : M \rightarrow N$ s nasledovnými vlastnosťami:

- i. pre všetky $x, y \in M$ platí $h(x \cdot y) = h(x) * h(y)$
- ii. $h(e) = 1$.

Definícia 1.0.15. Antihomomorfizmus je zobrazenie z A do B , pre ktoré platí, že pre $\forall x, y \in A : h(x \cdot y) = h(y) \cdot h(x)$. **Antihomomorfizmus monoidov** $(M, \cdot, e)$ a $(N, *, 1)$ je zobrazenie $h : M \rightarrow N$ spĺňajúce podmienky:

- i. pre všetky $x, y \in M$ platí $h(x \cdot y) = h(y) * h(x)$
- ii. $h(e) = 1$.

Poznámka 1.0.16. Môžeme si všimnúť, že ak pre automat $A = (S, Q, \delta, q_0, F)$ definujeme zobrazenie $h : S \rightarrow M_A$, kde M_A je prechodový monoid, ako $h(x) = \delta_x$ a pod operáciou $\circ$ budeme rozumieť také skladanie zobrazení, že $(g \circ f)(x) = g(f(x))$, potom zobrazenie h bude antihomomorfizmom. Z definície akcie totiž platí, že $\delta_{xy} = \delta_y \circ \delta_x$. Z tohto dôvodu budeme často predpokladať, že sa zobrazenia skladajú v opačnom poradí, teda že $(g \cdot f)(x) = f(g(x))$. Pri takomto skladaní bude δ_{xy} rovné $\delta_x \cdot \delta_y$ a zobrazenie h teda bude homomorfizmom. Namiesto zápisu $h(x)(q)$ resp. $\delta_x(q)$ potom môžeme používať stručnejší zápis $q \cdot x$.

Príklad 1.0.17. Majme monoid $A = (\Sigma^*, \cdot, \varepsilon)$, monoid $B = (\mathbb{N}, +, 0)$ a homomorfizmus $f : \Sigma^* \rightarrow \mathbb{N}$ definovaný ako $f(w) = |w|$. Potom f je homomorfizmom monoidov A a B .

Kapitola 2

Rozoznávane jazykov monoidmi

Definícia 2.0.1. Jazyk rozoznávaný homomorfizmom: Hovoríme, že jazyk $L \subseteq \Sigma^*$ je rozoznávaný homomorfizmom $h : \Sigma^* \rightarrow M$, ak existuje $X \subseteq M$ také, že $L = h^{-1}(X)$

Definícia 2.0.2. Jazyk rozoznávaný monoidom: Jazyk $L \subseteq \Sigma^*$ je rozoznávaný monoidom $(M, \cdot, e)$, ak existuje homomorfizmus $h : \Sigma^* \rightarrow M$ rozoznávajúci jazyk L .

Veta 2.0.3. *Jazyk L je regulárny $\Leftrightarrow$ je rozoznávaný nejakým konečným monoidom*

Dôkaz 2.0.4. $\boxed{\Leftarrow}$ Nech L je jazyk rozoznávaný monoidom $(M, \cdot, e)$ cez homomorfizmus h a množinu X . Chceme dokázať, že potom jazyk L musí byť regulárny. Vieme, že jazyk L je regulárny $\Leftrightarrow$ existuje konečný automat A , pre ktorý platí $L(A) = L$. Ukážeme, že takýto automat naozaj existuje. Nech $A_M = (\Sigma^*, M, \delta, e, X)$ kde prechodová funkcia δ je definovaná ako $\delta(m, a) = m \cdot h(a)$, kde m je stav a a je znak vstupnej abecedy Σ . Potom pre každé $a_1, a_2, \dots, a_n \in \Sigma$ platí $\delta(m, a_1 a_2 \dots a_n) = m \cdot h(a_1) \cdot h(a_2) \cdot \dots \cdot h(a_n)$. Keďže na začiatku čítania slova je stav (m) automatu A_M počiatočný stav e , ktorý je zároveň neutrálnym prvkom vzhľadom na operáciu $\cdot$, automat každé vstupné slovo $w = a_1 \dots a_n$ dočíta v stave $e \cdot h(a_1 \dots a_n) = h(a_1 \dots a_n) = h(w)$. Vieme, že slovo w patrí do jazyka automatu práve vtedy, keď existuje akceptačný výpočet na slove w , teda ak sa automat po dočítaní slova w nachádza v nejakom zo svojich akceptačných stavov. Množina akceptačných stavov automatu A_M zodpovedá množine X , pre ktorú platí $L = h^{-1}(X)$ (teda $L = \{x \in \Sigma^* \mid h(x) \in X\}$). Jazyk automatu A_M je teda rovný množine slov w nad abecedou Σ , pre ktoré platí $h(w) \in X$, čiže $L(A_M) = \{w \in \Sigma^* \mid h(w) \in X\}$. Zrejme platí rovnosť $L(A_M) = L$, takže sme naozaj zostrojili konečný automat akceptujúci jazyk L , a jazyk L je naozaj regulárny.

$\boxed{\Rightarrow}$ Predpokladajme teraz, že L je regulárny jazyk; dokážeme, že existuje konečný monoid, ktorý ho rozoznáva. Ak L je regulárny jazyk, potom existuje deterministický konečný automat $A = (\Sigma^*, Q, \delta, q_0, F)$ taký, že $L(A) = L$. Nech $M_A = (M, \cdot, Id_K)$ je prechodový monoid, $M = \{\delta_x \mid x \in \Sigma^*\}$ a nech h je homomorfizmus zo Σ^* do M definovaný ako $h(x) = \delta_x$ (pozri poznámku). Homomorfizmus h teda každému vstupnému slovu priradí funkciu, ktorá pre dané vstupné slovo a stav, v ktorom sa automat nachádza na začiatku čítania slova, vráti stav, v ktorom bude automat, keď slovo dočíta. Nech ďalej X je podmnožina M , pre ktorú platí $X = \{\delta_x \mid \delta_x(q_0) \in F\}$. X je teda množina práve tých funkcií od slova x , že automat A dočíta x v nejakom zo svojich akceptačných stavov. Potom $h^{-1}(X)$ je zrejme $= L(A)$, pretože $h^{-1}(X)$ zodpovedá množine slov, na ktorých výpočet automatu A skončí v stave $q \in F$. Jazyk L je teda skutočne rozoznávaný konečným monoidom. $\square$