#B241/JAG/LEC

Teorie jazyků a automatů— 2. týden

Marie Demlová
http://math.feld.cvut.cz/en/people/demlova

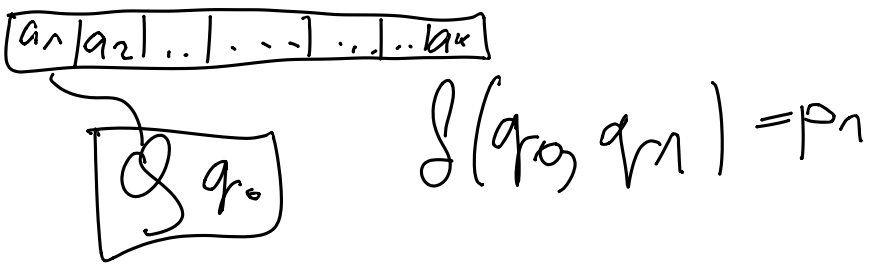
30. 9. 2024

Konečné automaty — DFA

Co bylo na minulé přednášce

DFA: $M = (Q, \Sigma, \delta, q_0, F)$ $\delta : Q \times \Sigma \to Q$

Ještě jiný pohled na DFA.



Konečné automaty — DFA

Pumping lemma pro regulární jazyky dává **nutnou** podmínku pro to, aby jazyk L byl regulární.

Pumping lemma.

Pro každý regulární jazyk L nad abecedou Σ existuje přirozené číslo n s touto vlastností: Každé slovo $u \in L$, které je delší než n, lze rozdělit na tři slova $u = xwy$ tak, že

1. $|xw| \leq n$,
2. $w \neq \varepsilon$
3. a pro každé přirozené číslo $i = 0, 1, \ldots$ platí $xw^i y \in L$.

$\exists n \in \mathbb{N}, \forall u \in L, |u| > n \quad \exists x,w,y \in \Sigma^*$, takž e

Logicky $\forall n \in \mathbb{N}, \exists u \in L, |u| > n \; \forall x,w,y \; u = xwy \longrightarrow$ takž e $1 \wedge 2 \Rightarrow 3$

Př. $L = \{0^n 1^n \mid n \geq 0\}$

Vezmeme lib. n, $n \in \mathbb{N}$, zvolíme $w = 0^n 1^n \in L$ $|w| = 2n > n$,

kdyby $0^n 1^n = xwy$ a $|xw| \leq n$

implikuje $xw = 0^k$ $k \leq n$ $\underbrace{000\ldots0}_{n}\underbrace{111\ldots1}_{n}$

$w \neq \varepsilon$ tj $w = 0^\ell$

$\ell \geq 1$

$\ell \leq k$

$xw^2y = 0^{n+\ell} 1^n \notin L$

$n + \ell > n$

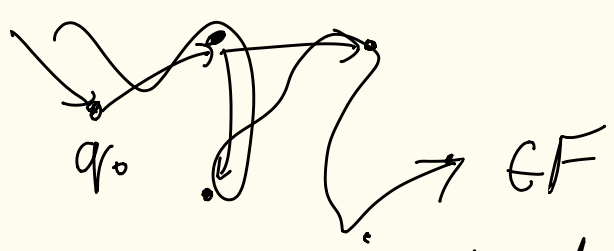
Důkaz pumping lemma

$L \in Reg$, tj existuje DFA $M = (Q, \Sigma, \delta, q_0, F)$, že $L = L(M)$

$n := |Q|$

stavový diagram

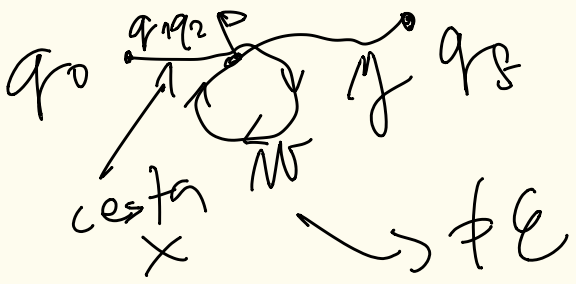
$w \in L(M)$ iff $\delta^*(q_0, w) \in F$



lib $w \in L(M)$

→ sled co musí cesta, nějaký uzel se musí opakovat

$|w| > n$ w



Konečné automaty — DFA

Použití Pumping lemmatu.

Fakt. Jazyk $L = \{0^m 1^m \,|\, m \geq 0\}$ není regulární.

Poznámka. Podmínka z pumping lemmatu je pouze nutná, není postačující.

Pumping lemma nelze použít, chceme-li dokázat, že jazyk je regulární.

Konečné automaty — DFA

Nerodova věta dává nutnou a postačující podmínku pro to, aby byl jazyk regulární.

Nerodova věta.

Jazyk L nad abecedou Σ je regulární právě tehdy, když existuje ekvivalence T na množině všech slov $\Sigma^\star$ taková, že

1. L je sjednocení některých tříd ekvivalence T.
2. Jestliže pro nějaké $u, v \in \Sigma^\star$ platí $u\ T\ v$, pak pro každé slovo $w \in \Sigma^\star$ platí také $uw\ T\ vw$.
3. T má pouze konečně mnoho tříd ekvivalence.

$L \in Reg$ iff ekvivalence T na Σ^* taková, že

1) L je sjednocení některých tříd ekvivalence T
2) iff $u T v$ pro $u, v \in \Sigma^*$, tak pro každé $w \in \Sigma^*$

$$uw \, T \, vw$$

3) T má jen konečně mnoho tříd

Př

$L = \{0^n 1^n\}$ není regulární

zvolme $0, 0^2, 0^3, 0^4, \ldots, 0^i \quad i > 0$

Protože T má konečně mnoho tříd, tak existuje $i \neq j$, že $0^i T 0^j$ Potom $w = 1^i$, pak (2)

$$0^i 1^i T 0^j 1^i$$

a $0^i 1^i \in L$, ale $0^j 1^i \notin L \quad (i \neq j)$

neplatí 1.

Důkaz

1) $L = Reg \quad M = (Q, \Sigma, \delta, q_0, F) \quad L = L(M)$

Definujeme T $u T v$ iff $\delta^*(q_0, u) = \delta^*(q_0, v)$

if uTv, tak pro každé $w \in \Sigma^*$ je $uw \, T \, vw$ T je ekvivalence

potom $\delta^*(q_0, uw) = \delta^*(q_0, vw)$

$$L(M) = \{u \mid \delta^*(q_0, x) \in F\} = \bigcup_{q_f \in F} \{u \mid \delta^*(q_0, u) = q_f\}$$

Tříd je nejvýše tolik jako $|Q|^{|Q|}$

D 2) máme T s vlastnostmi 1) 2) 3)

$Q := \{T[u] \mid u \in \Sigma^*\}$ $T[u]$ je třída T-ekvivalence u

$q_0 := T[\varepsilon]$ $\delta(T[u], a) := T[ua]$

$F := \{T[u] \mid T[u] \subseteq L\}$

$\delta^*(T[\varepsilon], a_1 a_2 \dots a_k) = T[a_1 a_2 \dots a_k]$ (q0)

$u \in L(M)$ iff $\delta^*(T[\varepsilon], u) \in F$

$T[\varepsilon u] \in F$ iff $T[u] \subseteq L$ iff $u \in L$

Konečné automaty — DFA

Použití Nerodovy věty k důkazu, že L není regulární.

Fakt. Jazyk $L = \{0^m 1^m \,|\, m \geq 0\}$ není regulární.

Konečné automaty — DFA

Použití Nerodovy věty k důkazu, že L je regulární.

Použijeme invarianty; tj. každému stavu přiřadíme popis všech slov u, pro které $\delta^\star(q_0, u) = q$ a to tak, že

1. ε splňuje invariant pro q_0,
2. každé u splňuje právě jeden invariant,
3. pro každé $q \in Q$ s $a \in \Sigma$ platí: Jestliže u odpovídá invariantu q, pak ua odpovídá invariantu $\delta(q, a)$,
4. $u \in L$ iff u splňuje invariant pro některý $q \in F$.

Příklad Pro jazyk L, který se skládá ze všech binárních slov, které končí 010, zkonstruujte DFA, který ho přijímá.

jsou po dvou disjunktní

$\{0,1\}^*$ ε

končí 1
11
01

končí 0 < končí 010 / nebo končí 010

1 Věta o invariantech.

Mějme konečný automat $M = (Q, \Sigma, \delta, q_0, F)$. Označme L jazyk přijímaný automatem M. Pro každý stav $q \in Q$ mějme formuli (které budeme říkat *invariant*) $I_q(u)$, s proměnnou u. Předpokládejme, že množina všech invariantů splňuje tyto tři předpoklady:

I1: Platí $I_{q_0}(\varepsilon)$, tj. prázdné slovo ε splňuje invariant počátečního stavu q_0.

I2: Pro každý stav $q \in Q$, každé písmeno $a \in \Sigma$ a každé slovo $u \in \Sigma^\star$ platí implikace

$$I_q(u) \;\Rightarrow\; I_{\delta(q,a)}(ua)$$

tj. když jakékoli slovo u splňovalo invariant I_q, pak slovo ua splňuje invariant $I_{\delta(q,a)}$ následujícího stavu $\delta(q, a)$.

I3: Pro každé slovo $u \in \Sigma^\star$ existuje právě jeden stav $q \in Q$ takový, že $I_q(u)$, tj. takový, že slovo u splňuje invariant I_q.

Jsou-li tytp předpoklady splněny, pak:

T1: Platí

$$\text{pro všechna } u \in \Sigma^\star \quad (\; \delta^\star(q_0, u) = q \iff I_q(u) \;)\;, \tag{1}$$

jinak řečeno, zpracováním libovolného slova u se automat převede do přesně toho stavu q, jehož invariant slovo u splňuje.

T2: Platí

$$u \in L \iff \exists\, q_f \in F \;\; (I_{q_f}(u))\;, \tag{2}$$

tj. libovolné slovo u je přijímáno automatem právě tehdy, když splňuje invariant I_{q_f} některého koncového stavu $q_f \in F$.

T3: Platí

$$L \;=\; \bigcup_{q \in F} \{u \mid I_q(u)\}\;, \tag{3}$$

tj. jazyk L je sjednocením množin slov u, která splňují invariant $I_q(u)$ některého koncového stavu $q \in F$.

Důkaz. Nejprve indukcí podle délky slova u dokážeme implikci „⇒“, tedy implikaci

$$\delta^\star(q_0, u) = q \;\Rightarrow\; I_q(u) \quad . \tag{4}$$

Pro prázdné slovo $u = \varepsilon$ tvrzení (4) vyplývá z předpokladu I1.

Předpokládejme nyní, že tvrzení (4) platí pro všechna slova délky n, a dokážeme, že potom tvrzení (4) platí také pro slova délky $n + 1$. Vezměme libovolné slovo u délky $n + 1$, označme x jeho poslední písmeno a označme v počáteční úsek slova takový, že $u = vx$. Dále označme $p = \delta^\star(q_0, v)$ a také $q = \delta(p, x) = \delta^\star(q_0, u)$.

Délka slova v je n, proto podle indukčního předpokladu pro slovo v platí implikace (4), tedy slovo v splňuje invariant I_p, tj. platí $I_p(v)$.

Podle předpokladu I2 platí implikace $I_p(v) \Rightarrow I_{\delta(p,x)}(vx)$. Slovo $u = vx$ tedy splňuje invariant $I_{\delta(p,x)} = I_q$. Tím je dokázán druhý krok indukce. Implikace (4) je tedy dokázána pro slova u libovolné délky.

Obrácenou implikaci dokážeme s využitím předpokladu I3. Předpokládejme, že slovo u splňuje invariant I_q, nějakého stavu q, tj. platí $I_q(u)$. Označme $p = \delta(q_0, u)$. Podle již dokázaného tvrzení (**??**) pak slovo u splňuje také invariant I_p. Podle předpokladu I3 pak musí platit $p = q$, tedy $\delta(q_0, u) = q$. Tím máme dokázáno tvrzení (1).

Z tvrzení (1) přímo vyplývá tvrzení (2) a odtud pak tvrzení (3).

2 Příklad.

		0	1
→	0	0	1
	1	0	2
←	2	0	2

Hypotézu, jaká slova automat přijímá, lze „vykoukat“ z diagramu. Hypotéza zní, že automat přijímá všechna slova končící 11.

2.1 Důkaz s využitím invariantů.

Nejprve je třeba invarianty vymyslet a napsat.

$I_0(u)$: $u = \varepsilon$ nebo u končí nulou.
$I_1(u)$: $u = 1$ nebo u končí 01.
$I_2(u)$: u končí 11.

Pak je třeba ověřit předpoklady I1, I2 a I3 věty o invariantech.

Předpoklad I1 je snadný, prázdné slovo splňuje invariant stavu 0.

Předpoklad I2 se ověřuje tak, že pro každou hranu stavového diagramu (nebo každou položku přechodové tabulky) ověřit příslušnou implikaci z věty o invariantech. Pro stručnost to ukážeme na hraně $1 \to 2$, která odpovídá vstupnímu písmenu 1. Jestliže nějaké slovo u splňovalo invariant I_1, tj. buď platilo $u = 1$ nebo u končilo 01, a automat přečetl písmeno 1, pak slovo $u1$ končí dvěma jedničkami a tedy splňuje invariant výsledného stavu 2.

Ověření předpokladu I3 v našem případě znamená provést rozbor, jak může slovo u končit. Buď končí 11, nebo končí 01, nebo končí 0, nebo je 1, nebo je prázdné.

RADA: Použijete-li metodu invariantů v písemce, nebo v domácím úkolu, stačí napsat správně formulované invarianty. Ověřování předpokladů I1 až I3 vypisovat nemusíte, ale samozřejmě to OVĚŘENÍ MUSÍTE UDĚLAT, jinak riskujete, že v důkazu budete mít chybu. Nejčastěji se chybuje v ověření předpokladu I3.

Redukce DFA

Ekvivalentní DFA

DFA M_1 a M_2 jsou ekvivalentní, jestliže $L(M_1) = L(M_2)$.

Dosažitelné stavy

Stav $q \in Q$ je dosažitelný, jestliže existuje slovo $u \in \Sigma^\star$ takové, že $\delta^\star(q_0, u) = q$.
Jinými slovy, jestliže ve stavovém diagramu existuje neorientovaný sled z q_0 do q ohodnocený u.

Postup nalezení dosažitelných stavů

$Q_0 := \{q_0\}$ BFS
`repeat`
$$Q_{i+1} := Q_i \cup \{\delta(q, a) \mid q \in Q_i, a \in \Sigma\}$$
`until` $Q_{i+1} = Q_i$.
`return` $Q' = Q_i$

Redukce DFA

Ekvivalence stavů $\sim$

Stavy $p, q \in Q$ jsou ekvivalentní, jestliže pro každé slovo $u \in \Sigma^\star$ platí

$$\delta^\star(p, u) \in F \quad \text{právě tehdy, když} \quad \delta^\star(q, u) \in F.$$

Fakt, že dva stavy p a q jsou ekvivalentní, zapisujeme $p \sim q$.

Redukovaný automat. DFA M je redukovaný, jestliže nemá nedosažitelné stavy a žádné jeho dva různé stavy nejsou ekvivalentní.

Jinými slovy, jestliže ekvivalence $\sim$ je identická ekvivalence na množině stavů Q. $M = (Q, \Sigma, \delta, q_0, F)$ na množině Q definujeme $\sim$ takto:
$q \sim p$ iff $\forall w \in \Sigma^* \left(\delta^*(q, w) \in F \text{ iff } \delta^*(p, w) \in F \right)$