Automaty a gramatiky

Roman Barták, KTIML

bartak@ktiml.mff.cuni.cz
http://ktiml.mff.cuni.cz/~bartak

Organizační záležitosti

Rozsah: 3/2 Z, Zk 12 lekcí
úterý 9.00 - 11.15

Cvičení: Mgr. Caha (1),
Dr. Majerech (2),
Dr. Plátek (2)

Konzultace: pracovna č. 51 (3. patro)
Čt 10.30 - 11.30

Zkouška: písemná a ústní část

Zdroje a literatura

M. Chytil: *Automaty a gramatiky*, SNTL Praha, 1984
V. Koubek: *Automaty a gramatiky*, elektronický text, 1996
R. Barták: *Automaty a gramatiky: on-line*, 2001
http://ktiml.mff.cuni.cz/~bartak/automaty

M. Chytil: *Teorie automatů a formálních jazyků*, skripta
M. Chytil: *Sbírka řešených příkladů z teorie automatů a formálních jazyků*, skripta
M. Demlová, V. Koubek: *Algebraická teorie automatů*, SNTL Praha, 1990

J.E. Hopcroft, J.D. Ullman: *Introduction to Automata Theory, Languages and Computation*, Addison-Wesley, 1979

O čem bude přednáška?

studium konečného popisu nekonečných objektů

dvě větve: automaty a gramatiky

konečné automaty	regulární gramatiky
zásobníkové automaty	bezkontextové gramatiky
lineárně omezené automaty	kontextové gramatiky
Turingovy stroje	gramatiky typu 0

Pohled do historie

Počátky ve druhé čtvrtině 20. století
první formalizace pojmu algoritmus (1936)
Turing, Kleene, Post, Church, Markov

Polovina 20. století
neuronové sítě (1943)
konečné automaty (Kleene 1956 neuronové sítě ≈ KA)

60. léta 20. století
gramatiky (Chomsky)
zásobníkové automaty
formální teorie konečných automatů

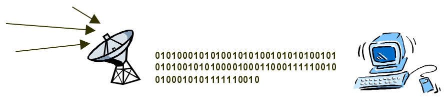
Praktické využití

- ✓ návrh a popis hardware
  - integrované obvody, stroje, automaty
- ✓ realizace pomocí software
  - formální popis → program
- ✓ překladače
- ✓ zpracování přirozeného jazyka
- ✓ aplikace v biologii
  - simulace růstu
  - celulární automaty
  - sebe-reprodukce automatů

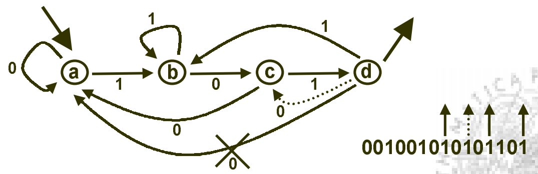
Úvod do konečných automatů

Projekt SETI (Search Extra-Terrestrial Intelligence)

analýza signálů - hledání vzorku



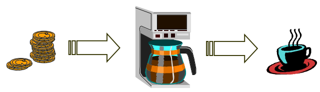
Hledání vzorku „101“



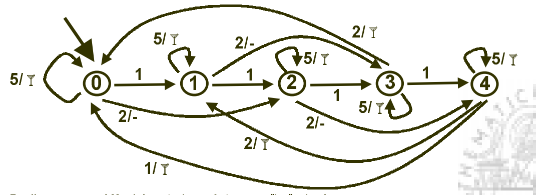
Úvod do konečných automatů 2

Stroj na kávu

stroj signalizuje vydání kávy po vhození potřebného obnosu



Vstupem stroje jsou mince 1,2,5 Kč, káva stojí 5 Kč



Realizace pomocí Mealyho stroje s výstupem při přechodu.

Formalizace konečného automatu

Konečným automatem nazýváme pětici

$$A = (Q,X,\delta,q_0,F),$$

kde:

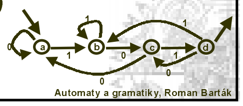
Q - konečná neprázdná množina stavů (*stavový prostor*)

X - konečná neprázdná množina symbolů (*vstupní abeceda*)

δ - zobrazení $Q \times X \to Q$ (*přechodová funkce*)

$q_0 \in Q$ (*počáteční stav*)

$F \subseteq Q$ (*množina koncových stavů*)

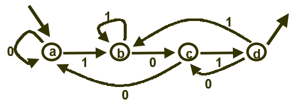


Popis konečného automatu

Stavový diagram (graf)

vrcholy = stavy

hrany = přechody



Tabulka

řádky = stavy+přechody

sloupce = písmena

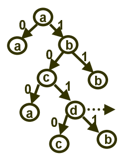
	0	1
→ a	a	b
b	c	b
c	a	d
← d	c	b

Stavový strom

vrcholy = stavy

hrany = přechody

pouze dosažitelné stavy!



Abeceda, slova, jazyky

abeceda X = konečná neprázdná množina symbolů

slovo = konečná posloupnost symbolů (i prázdná)

prázdné slovo λ (e, ε, ...)

X^* = množina všech slov v abecedě X

X^+ = množina všech neprázdných slov

$X^* = X^+ \cup \{\lambda\}$

jazyk $L \subseteq X^*$ (množina slov v abecedě X)

Základní operace se slovy:

zřetězení slov u.v, uv

mocnina u^n ($u^0 = \lambda$, $u^1 = u$, $u^{n+1} = u^n.u$)

délka slova $|u|$ ($|\lambda| = 0$)

Rozšířená přechodová funkce

přechodová funkce δ: $Q \times X \to Q$

rozšířená přechodová funkce δ^* : $Q \times X^* \to Q$

tranzitivní uzávěr δ

induktivní definice

$\delta^*(q,\lambda) = q$

$\delta^*(q,wx) = \delta(\delta^*(q,w),x)$, $\quad x \in X$, $w \in X^*$

úmluva:

δ^* budeme někdy označovat také jako δ

Jazyky rozpoznatelné konečnými automaty

Jazykem rozpoznávaným (akceptovaným, přijímaným) konečným automatem $A = (Q,X,\delta,q_0,F)$ nazveme jazyk:

$L(A) = \{w \mid w \in X^* \ \& \ \delta^*(q_0,w) \in F\}$.

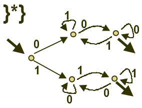
Slovo w je *přijímáno* automatem A, právě když $w \in L(A)$.

Jazyk L je *rozpoznatelný* konečným automatem, jestliže existuje konečný automat A takový, že $L=L(A)$.

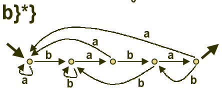
Třídu jazyků rozpoznatelných konečnými automaty značíme $\mathcal{F}$, tzv. *regulární jazyky*.

Příklady regulárních jazyků

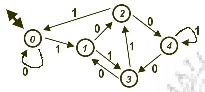
$L= \{ w \mid w \in \{0,1\}^* \ \& \ w=xux, x\in\{0,1\}, u\in\{0,1\}^*\}$



$L= \{ w \mid w \in \{a,b\}^* \ \& \ w=ubaba, u\in\{a,b\}^*\}$

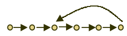


$L= \{ w \mid w \in \{0,1\}^*$ & w je binární zápis čísla dělitelného 5}



$L = \{0^n1^n \mid n\geq 1\}$

není regulární jazyk!



Kongruence

Jak charakterizovat regulární jazyky?

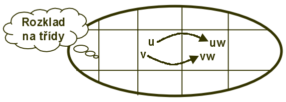
Jak zjistit, že jazyk není rozpoznatelný konečným automatem?

Kongruence

Nechť X je konečná abeceda, ~ je relace ekvivalence (reflexivní, symetrická, transitivní) na X^*. Potom:

a) ~ je *pravá kongruence*, jestliže
$\forall u,v,w\in X^* \quad u\sim v \Rightarrow uw\sim vw$

b) je *konečného indexu*, jestliže
rozklad $X^*/\sim$ má konečný počet tříd

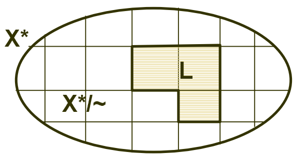


Nerodova věta

Nechť L je jazyk nad konečnou abecedou X. Potom následující tvrzení jsou ekvivalentní:

a) L je rozpoznatelný konečným automatem,

b) existuje pravá kongruence konečného indexu na X^* tak, že L je sjednocením jistých tříd rozkladu $X^*/\sim$.



Důkaz Nerodovy věty

a) ⇒ b)

automat ⇒ pravá kongruence konečného indexu

definujme $u\sim v \equiv \delta^*(q_0,u) = \delta^*(q_0,v)$

je to ekvivalence (reflexivní, symetrická, transitivní)

je to pravá kongruence (z definice δ^*)

má konečný index (konečně mnoho stavů)

$L= \{ w \mid \delta^*(q_0,w) \in F\} = \cup_{q \in F} \{ w \mid \delta^*(q_0,w) = q\}$

Pozorování:

stavy odpovídají třídám ekvivalence

Důkaz Nerodovy věty - pokračování

b) ⇒ a)

pravá kongruence konečného indexu ⇒ automat

označme [u] třídu rozkladu obsahující slovo u

Jak sestrojíme konečný automat A?

abeceda X dána

stavy Q - třídy rozkladu $X^*/\sim$

stav $q_0 = [\lambda]$

koncové stavy $F = \{c_1,..,c_n\}$, kde $L= \cup_{i=1..n} c_i$

přechodová funkce $\delta([u],x) = [ux]$

přechodová funkce je korektní (z definice pravé kongruence)

Ještě L(A) = L?

$w\in L \Leftrightarrow w \in \cup_{i=1..n} c_i \Leftrightarrow w\in c_1 \vee \dots \vee w\in c_n \Leftrightarrow [w]= c_1 \vee \dots \vee [w]= c_n \Leftrightarrow [w]\in F \Leftrightarrow w \in L(A)$

$\delta^*([\lambda],w) = [w]$

Použití Nerodovy věty

Konstrukce automatů

Příklad:

Sestrojte automat přijímající jazyk

$L = \{w \mid w \in \{a,b\}^* \text{ \& } w \text{ obsahuje } 3k+2 \text{ symbolů } a\}$

označme $|u|_x$ počet symbolů x ve slově u

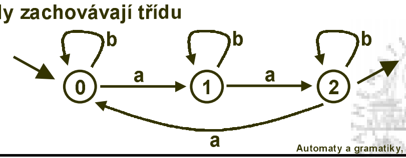
definujme $u \sim v \equiv (\,|u|_a \bmod 3 = |v|_a \bmod 3\,)$

tři třídy ekvivalence 0,1,2 (zbytky po dělení 3)

L odpovídá třídě 2

a-přechody přesouvají do další třídy

b-přechody zachovávají třídu



Použití Nerodovy věty - pokračování

Důkaz neregulárnosti jazyka!

Příklad:

Rozhodněte zda následující jazyk je regulární

$L = \{0^n1^n \mid n \geq 1\}$.

Předpokládejme, že jazyk je regulární

$\Rightarrow$ existuje pravá kongruence konečného indexu m, L je sjednocením tříd

vezmeme slova 0, 00, …, 0^{m+1}

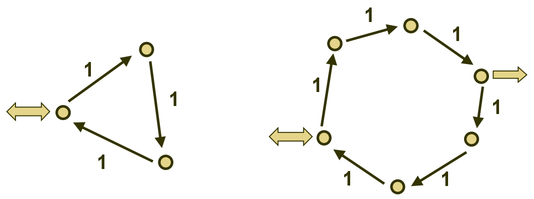
dvě slova padnou do stejné třídy (krabičkový princip)

$i \neq j$	$0^i \sim 0^j$	
přidejme 1^i	$0^i1^i \sim 0^j1^i$	(pravá kongruence)
spor	$0^i1^i \in L$ & $0^j1^i \notin L$	

Jazyk a přijímající automaty

Je automat pro daný jazyk určen jednoznačně?

NENÍ!



$L = \{w \mid w=1^* \text{ \& } |w|=3k\}$

Ekvivalence automatů a homomorfismus

Jak zjistit, zda dva automaty přijímají stejný jazyk?

Říkáme, že konečné automaty A a B jsou *ekvivalentní*, jestliže rozpoznávají stejný jazyk, tj. L(A)=L(B).

Nechť A_1 a A_2 jsou konečné automaty. Řekneme, že zobrazení $h: Q_1 \to Q_2$ je (automatovým) *homomorfismem*, jestliže:

1) $h(q_1) = q_2$ „stejné“ počáteční stavy
2) $h(\delta_1(q,x)) = \delta_2(h(q),x)$ „stejné“ přechodové funkce
3) $q \in F_1 \Leftrightarrow h(q) \in F_2$ „stejné“ koncové stavy

Homomorfismus prostý a na nazýváme *isomorfismus*.

Věta o ekvivalenci automatů

Existuje-li homomorfismus konečných automatů A_1 do A_2, pak A_1 a A_2 jsou ekvivalentní.

Důkaz:

konečnou iterací

$h(\delta_1(q,w)) = \delta_2(h(q),w) \quad w \in X^*$

$$\begin{aligned} w \in L(A_1) \quad & \Leftrightarrow \delta_1(q_1,w) \in F_1 \\ & \Leftrightarrow h(\delta_1(q_1,w)) \in F_2 \\ & \Leftrightarrow \delta_2(h(q_1),w) \in F_2 \\ & \Leftrightarrow \delta_2(q_2,w) \in F_2 \\ & \Leftrightarrow w \in L(A_2) \end{aligned}$$

Homomorfismus automatů

$L = \{w \mid w=1^* \text{ \& } |w|=3k\}$