AUTOMATY A GRAMATIKY

1

Pavel Surynek

Univerzita Karlova v Praze

Matematicko-fyzikální fakulta
Katedra teoretické informatiky a matematické logiky

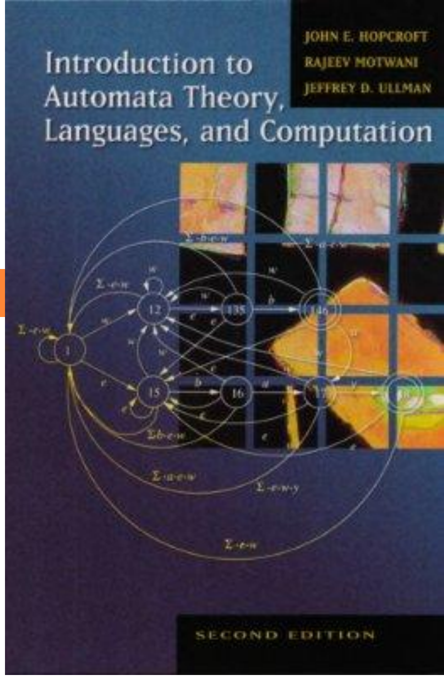
Stručný přehled přednášky

- Automaty
  - Formální jazyky, operace
  - Konečné automaty a jejich varianty
  - Regulární jazyky, regulární výrazy
- Gramatiky
  - Chomského hierarchie jazyků
  - Bezkontextové a kontextové jazyky
  - Uzávěrové vlastnosti
- Souvislosti
  - Turingovy stroje
  - Algoritmicky nerozhodnutelné problémy

Historie

- Teorie formálních jazyků
  - Počátky 50. léta
    - pokus o popis nekonečných objektů konečnými prostředky
  - 60. léta
    - vyšší programovací jazyky – syntaxe
      - FORTRAN, ALGOL, PASCAL
- Biologické procesy (popis buňky)
  - 40. léta
    - umělý neuron, neuronová síť, celulární automat
- Výpočetní modely (theory of computation)
  - snaha o formalizaci algoritmu
- Osobnosti
  - Kleene, Post, Church, Turing, Markov, Huffmann, Shannon, von Neumann, McCulloch, Pitts, Chomsky

Zdroje



- **Původní**
  - John E. Hopcroft, Rajeev Motwani, Jeffrey D. Ullman: Introduction to Automata Theory, Languages, and Computation. Addison-Wesley, 2001.
  - Jeffrey D. Ullman: Introduction to Automata and Complexity Theory, lecture notes. On-line: http://infolab.stanford.edu/~ullman, Stanford University, 2010.
  - Jeffrey D. Ullman: Stanford Automata, video lecture. On-line: https://www.coursera.org/course/automata, Coursera, 2013.

- **Tuzemské**
  - Pavel Surynek: Současná přednáška. On-line: http://ktiml.mff.cuni.cz/~surynek, od 2014.
  - Roman Barták: Automaty a gramatiky, slidy. On-line: http://ktiml.mff.cuni.cz/~bartak, 2005 - 2013.
  - Václav Koubek: Automaty a gramatiky, skripta. On-line: http://ktiml.mff.cuni.cz, 1996.

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

Slova, jazyky

- Abeceda
  - X – konečná neprázdná množina symbolů
- Slovo
  - konečná posloupnost prvků z X
    - $w = x_1x_2 \ldots x_n$, kde $n \in \mathbb{N}_0$ a $x_i \in X$ pro $i = 1,2, \ldots, n$
      - n - délka slova, $|w| = n$
  - prázdná posloupnost pro n = 0
    - prázdné slovo λ (někdy Λ nebo ε)
  - X* - množina všech slov nad X
  - X^+ - množina všech neprázdných slov nad X
    - $X^+ = X^* - \{\lambda\}$
- Jazyk L nad abecedou X
  - (libovolná) množina slov nad X
    - $L \subseteq X^*$, tj. $L \in \mathcal{P}(X^*)$

Př.: X = {a, b, c}

Př.: abbcaa, |abbcaa| = 6

Př.: L = {abc, abcc, abccc, … }

Pozn.: X* spočetná
$\mathcal{P}(X^*)$ nespočetná

Operace se slovy a jazyky

- Konkatenace – zřetězení slov
  - $u,v \in X^*$, kde $u = x_1x_2 \dots x_n$ a $v = y_1y_2 \dots y_m$
  - $u.v = x_1x_2 \dots x_n y_1 y_2 \dots y_m$
    - Př.: u = abba, v = cba, pak u.v = abbacba
- Mocnina slova
  - $u \in X^*$
  - $u^0 = \lambda$, $u^1 = u$, $u^{i+1} = u.u^i$
- Prefix, sufix, zrcadlový obraz
  - $w \in X^*$, $u \in X^*$ je prefix w, jestliže $\exists v \in X^*$, že $w = u.v$
  - $w \in X^*$, $u \in X^*$ je sufix w, jestliže $\exists v \in X^*$, že $w = v.u$
  - $w \in X^*$, kde $w = z_1z_2 \dots z_n$, pak $w^R = z_n z_{n-1} \dots z_1$ je zrcadlový obraz w
- Množinové operace s jazyky
  - $K, L \subseteq X^*$
  - $K \cap L$, $K \cup L$, $K - L$ (rozdíl), $-L = X^* - L$ (doplněk)
  - K/L pravý kvocient K podle L
    - $K/L = \{u \mid (\exists v \in L)\ u.v \in K\}$
  - L\K levý kvocient K podle L
    - $L \backslash K = \{u \mid (\exists v \in L)\ v.u \in K\}$

Př.: **w =** abbcaa
abb je *prefix* w
bcaa je *sufix* w

Př.: $K = \{ 01w \mid w \in \{0,1\}^* \}$
$L = \{ 0w1 \mid w \in \{0,1\}^* \}$
$K/L = K \cup \{\lambda\}$
$L \backslash K = \{0, 1\}^*$

Konečný automat

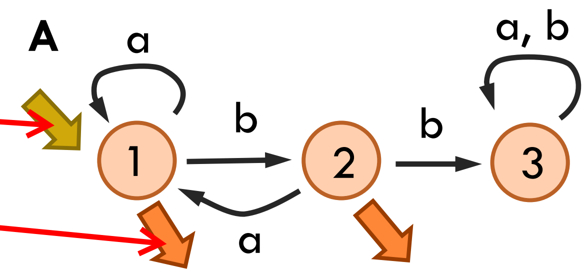
- **Konečný automat** $A = (Q, X, \delta, q_0, F)$
  - Q – konečná neprázdná množina stavů kde se necházíme ve stavu
  - X – konečná neprázdná množina symbolů to co čteme
    - *abeceda* (prvky *písmena*)
  - $\delta: Q \times X \longrightarrow Q$ přechodová funkce "tabulka přechodů"
  - $q_0 \in Q$ – počáteční stav
  - $F \subseteq Q$ – množina přijímajících stavů ty co nás zajímají na konci výpočtu
- Výpočet automatu A nad $w \in X^*$
  - rozšířená přechodová funkce $\delta^*: Q \times X^* \longrightarrow Q$
    - $\delta^*(q, \lambda) = q \qquad \forall q \in Q$
    - $\delta^*(q, wx) = \delta(\delta^*(q, w), x) \qquad \forall q \in Q, \forall x \in X, \forall w \in X^*$

    delta bez hvězdy je pro písmeno
    delta* je pro slovo
  - slovo w je přijímáno automatem A, jestliže $\delta^*(q_0, w) \in F$
    - $L(A) = \{ w \mid w \in X^* \text{ a } \delta^*(q_0, w) \in F\}$
- Regulární jazyk
  - jazyk L se nazývá regulární, jestliže existuje konečný automat A, že L(A) = L

Popis konečného automatu

Př.: L(A) = {w ∈ {a,b}* | w neobsahuje bb}

- **Stavový diagram** (ohodnocený graf)
  - vrcholy odpovídají stavům Q
  - ohodnocené orientované hrany odpovídají přechodové funkci δ
  - speciálně vyznačený počáteční stav q_0 a přijímající stavy F



- **Tabulka**
  - řádky odpovídají stavům Q
  - sloupce symbolům X
  - místo na řádku q∈Q a ve sloupci x∈X obsahuje δ(q, x)
  - vyznačení počátečního a přijímajících stavů

δ(1,a) = 1

A	a	b
1	1	2
2	1	3
3	3	3

Výpočetní síla

- Charakterizace regulárních jazyků
  - Kongruence
    - Ekvivalence (tj. reflexivní, symetrická a tranzitivní relace) ~ nad X* se nazývá *pravá kongruence*, jestliže:
      $\forall u, v, w \in X^*\ u \sim v \Rightarrow uw \sim vw$
      slučitelnost s operací: přidání slova zprava
    - ~ rozkládá X* na třídy ekvivalence
      - $[u]_\sim$ - třída ekvivalence určená slovem $u \in X^*$
      - $v \in X^*$, $v \in [u]_\sim$, jestliže $v \sim u$
      - $X^*/_\sim$ množina tříd rozkladu (tj. $[u]_\sim \in X^*/_\sim$)
    - ~ je *konečného indexu*, jestliže rozkládá X* na konečně mnoho ekvivalenčních tříd
  - **Myhill-Nerodova věta**
    - Jazyk L lze přijímat konečným automatem, právě když existuje pravá kongruence ~ konečného indexu, že L je sjednocením některých jejích tříd.

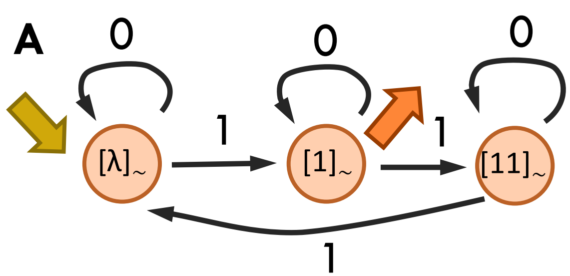
Důkaz Myhill-Nerodovy věty

- $\Rightarrow$
  - máme KA $A = (Q, X, \delta, q_0, F)$, že $L(A) = L$
  - pro $u,v \in X^*$ definujeme $u \sim v$, jestliže $\delta^*(q_0, u) = \delta^*(q_0, v)$
    - $\sim$ je ekvivalence, tj. má smysl uvažovat o $X^*/_\sim$
    - Q je konečná $\Rightarrow X^*/_\sim$ je konečná
    - $\forall u,v,w \in X^*$ když $\delta^*(q_0, u) = \delta^*(q_0, v)$, pak $\delta^*(\delta^*(q_0, u), w) = \delta^*(\delta^*(q_0, v), w)$, tedy $\sim$ je pravá kongruence
  - $L(A) = \{ w \mid w \in X^* \text{ a } \delta^*(q_0, w) \in F \} = \bigcup_{f \in F}\{w \mid \delta^*(q_0, w)=f\}$ tedy to jsou třídy ekvivalence
- $\Leftarrow$
  - máme pravou kongruenci $\sim$
  - položíme $Q = X^*/_\sim$
    - $q_0 = [\lambda]_\sim$
  - pro $x \in X$ a $w \in X^*$ položíme $\delta([w]_\sim, x) = [wx]_\sim$
    - pro $u,v \in X^*$ by mělo platit, že $\delta([u]_\sim, x) = \delta([v]_\sim, x)$, pokud $u \sim v$
    - $ux \sim vx$ je vlastnost pravé kongruence, tedy $[ux]_\sim = [vx]_\sim$
  - F = třídy z $X^*/_\sim$ tvořící L
    - $w \in L$, právě když $[w]_\sim \in F \Leftrightarrow \delta^*([\lambda]_\sim, w) = [w]_\sim$, což je, právě když $w \in L(A)$

Aplikace Myhill-Nerodovy věty

- Konstrukce konečného automatu
  - $L = \{w \mid w \in \{0,1\}^* \wedge (\exists k \in \mathbb{N}_0) |w|_1 = 3k+1\}$
    - definujeme $u \sim v$, jestliže $|u|_1 \bmod 3 = |v|_1 \bmod 3$
    - jedná se o pravou kongruenci
      - třídy $[\lambda]_\sim, [1]_\sim, [11]_\sim$
      - $L = [1]_\sim$

zvětším počet jedniček na obou stranách stejně, takže se mod zachová



- Důkaz neregularity jazyka
  - $L = \{0^n 1^n \mid n \in \mathbb{N}_0\}$ není kam si poznamenat kolik začlo nul
    - předpokládejme, že L je regulární
      - máme pravou kongruenci $\sim$ konečného indexu, nechť k je index
      - L je sjednocením některých jejích tříd
    - volme slova $0, 00, \ldots, 0^k, 0^{k+1}$
      - existují $i,j \in \{1, \ldots, k+1\}$, $i \neq j$, že $0^i \sim 0^j$
      - přidáme 1^i, z vlastnosti pravé kongruence je $0^i1^i \sim 0^j1^i$
      - $0^i1^i \in L$, ale $0^j1^i \notin L$, přitom 0^i1^i a 0^j1^i jsou ve stejné ekvivalenční třídě

Pumping (iterační) lemma

- Pumping lemma
  - Nechť L je regulární jazyk, pak existuje $n \in \mathbb{N}$, že libovolné slovo $z \in L$ takové, že $|z| \geq n$, lze napsat ve tvaru $z = u.v.w$, kde $|u.v| \leq n$, $|v| \geq 1$ a $u.v^i.w \in L$ pro všechna $i \in \mathbb{N}_0$.
  - Více logicky
    - Nechť L je regulární jazyk, pak existuje $n \in \mathbb{N}$, že $(\forall z \in L)[|z| \geq n \Rightarrow (\exists u,v,w \in X^*)(z = u.v.w \wedge |u.v| \leq n \wedge |v| \geq 1 \wedge (\forall i \in \mathbb{N}_0) u.v^i.w \in L)]$.

Pavel Surynek, 2015

Důkaz pumping lemmatu

- Je-li L regulární, pak existuje KA A, že $L(A) = L$
  - n = počet stavů automatu A
  - výpočet nad slovem z, kde $|z| \geq n$, navštíví některý stav aspoň dvakrát, nechť první takový stav je p
    - při první návštěvě p byl přečten prefix u
      - $\delta^*(q_0, u) = p$
    - při druhé návštěvě p byl přečten prefix uv
      - $\delta^*(q_0, uv) = p$
    - $|uv| \leq n$
      - byl uvažován první opakující se stav
    - $|v| \geq 1$
      - návrat do p se uskutečnil čtením aspoň jednoho písmena
    - $\delta^*(q_0, uvw) = f \in F$, pak $\delta^*(q_0, uw) = f$ a $\delta^*(q_0, uv^iw) = f$ pro $i = 2, 3, \ldots$

vynecháním v, můžu pokračovat

Použití pumping lemmatu

- Vyloučení, že daný jazyk L je regulární
  - dívejme se na pumping lemma jako na implikaci
    - regulární L ⇒ pro L platí pravá strana lemmatu
    - pro L neplatí pravá strana lemmatu ⇒ L není regulární
  - neplatí pravá pumping lemmatu
    - využijeme logické vyjádření, vytvoříme negaci
      - $\forall n \in \mathbb{N}\ (\exists z \in L)[|z| \geq n \wedge (\forall u,v,w \in X^*)((z = u.v.w \wedge |u.v| \leq n \wedge |v| \geq 1) \Rightarrow (\exists i \in \mathbb{N}_0) u.v^i.w \notin L)]$.
- $L=\{0^k1^k \mid k \in \mathbb{N}_0\}$
  - n (od nepřítele, tedy libovolné)
    - pro n vezmeme slovo $z = 0^n1^n$
    - jistě $|0^n1^n| \geq n$, pro libovolný rozklad splňující $z = u.v.w \wedge |u.v| \leq n \wedge |v| \geq 1$ je $v = 0^j$ pro $j \in \mathbb{N}$, $j \geq 1$
    - zvolme i = 2 a dostáváme, že $u.v^2.w = 0^{n+j}1^n \notin L$
- Jedná se nutnou podmínku, nikoli postačující (lemma je implikace).
  - $L = \{w \in \{a,b,c\}^* \mid w = a^+b^ic^i \vee w = b^ic^j\}$ není regulární, pravá strana platí