

Automaty a Gramatiky

Poznámky z přednášek

Letní semestr 2020/2021

Viktor Soukup, Lukáš Salak

Obsah

1	První přednáška	2
2	Druhá přednáška	4

1 První přednáška

Poznámka (Chomského hierarchie): Automaty a gramatiky - dva způsoby popisu:

Turingovy stroje	\leftrightarrow	gramatiky Typu 0
lineárně omezené automaty	\leftrightarrow	kontextové gramatiky, monotónní gramatiky
zásobníkové automaty	\leftrightarrow	bezkontextové gramatiky
konečné automaty (DFA, NFA, λ NFA)	\leftrightarrow	regulární jazyky

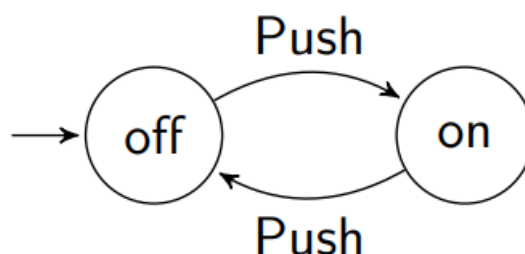
Nejjednodušší jsou nejnižší, turingův stroj je nejkomplikovanější. Každá gramatika odpovídá nějaké třídě automatů.

Proč to řešíme?

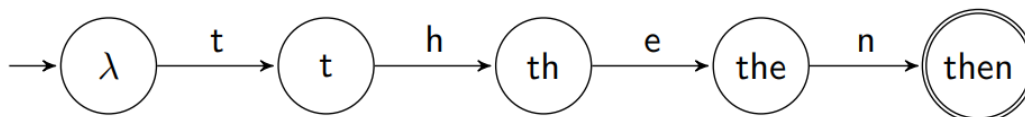
- zpracování přirozeného jazyka,
- překladače (lexikální, syntaktická analýza...),
- návrh, popis, verifikace hardware...
- hledání výskytu slova v textu (grep),
- verifikace systémů s konečně mnoha stavy

Příklad:

1. Návrh a verifikace integrovaných obvodů, např. Konečný automat modelující spínač on/off



2. Lexikální analýza, např. Konečný automat rozpoznávající slovo *then*



Definice (Deterministický konečný automat (DFA)): $A = (Q, \Sigma, \delta, q_0, F)$ sestává z:

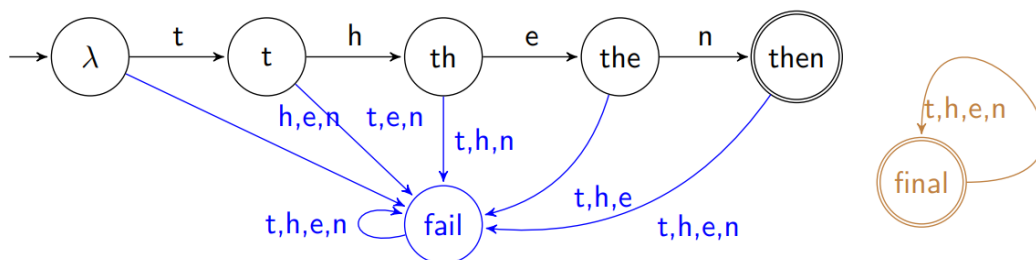
1. konečné množiny **stavů**, zpravidla značíme Q
2. konečné neprázdné množiny **vstupních symbolů (abecedy)**, značíme Σ
3. **přechodové funkce** zobrazení $Q \times X \rightarrow Q$, značíme δ , která bude reprezentovaná hranami grafu
4. **počátečního stavu** $q_0 \in Q$, vede do něj šipka 'odnikud'

5. neprázdné **množiny koncových (přijímajících) stavů** (final states) $F \subseteq Q$, označených dvojíým kruhem či šipkou 'ven'.

Poznámka:

Pokud pro některou dvojici stavu a písmene není definovaný přechod, přidáme nový stav *fail* a přechodovou funkci doplníme na totální přidáním šipek do *fail*.

Pokud je množina F prázdná, přidáme do ní i Q nový stav *final* do kterého vedou jen přechody z něj samotného $\forall s \in \Sigma : \delta(\text{final}, s) = \text{final}$.



Příklad:

Automat A přijímající $L = x01y : x, y \in \{0, 1\}^*$.

Automat $A = (\{q_0, q_1, q_2\}, 0, 1, \delta, q_0, q_1)$

Reprezentujeme stavovým diagramem (grafem), pomocí tabulky nebo stavovým stromem

Definice (Abeceda, slova, jazyky): Mějme neprázdnou množinu symbolů Σ .

- **Slovo** je konečná (i prázdná) posloupnost symbolů $s \in \Sigma$, **prázdné slovo** se značí λ nebo ϵ
- **Množinu všech slov v abecedě** Σ značíme Σ^*
- množinu všech neprázdných slov v abecedě značíme Σ^+
- **jazyk** $L \subseteq \Sigma^*$ je množina slov v abecedě Σ

Definice (Operace na Σ^*):

1. **zřetězení slov** $u.v$ nebo uv
2. **mocnina** (počet opakování) $u^n (u^0 = \lambda, u^1 = u, u^{n+1} = u^n.u)$
3. **délka slova** $|u| (|\lambda| = 0, |auto| = 4)$
4. **počet výskytů** $s \in \Sigma$ ve slově u značíme $|u|_s (|zmrzlina|_z = 2)$.

Definice (Rozšířená přechodová funkce): Mějme přechodovou funkci $\delta : Q \times \Sigma \rightarrow Q$.

Rozšířenou přechodovou funkci $\delta^* : Q \times \Sigma^* \rightarrow Q$ (tranzitivní uzávěr δ) definujeme induktivně:

1. $\delta^*(q, \lambda) = q$,
2. $\delta^*(q, wx) = \delta(\delta^*(q, w), x)$ pro $x \in \Sigma, w \in \Sigma^*$.

Poznámka: Pokud se v textu objeví δ aplikované na slova, míní se tím δ^* .

Definice (Jazyk rozpoznávaný(přijímaný, akceptovaný) konečným automatem): Jazykem rozpoznávaným konečným automatem $A = (Q, \Sigma, \delta, q_0, F)$ nazveme jazyk $L(A) = \{w : w \in \Sigma^* \& \delta^*(q_0, w) \in F\}$.

- 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 $L = L(A)$.
- Třidu jazyků rozpoznatelných konečnými automaty označíme \mathcal{F} , nazveme **regulární jazyky**.

Věta (Iterační (pumping) lemma pro regulární jazyky): *Mějme regulární jazyk L . Pak existuje konstanta $n \in \mathbb{N}$ (závislá na L) tak, že každé $w \in L; |w| \geq n$ můžeme rozdělit na tři části, $w = xyz$, že:*

1. $y \neq \lambda$
2. $|xy| \leq n$
3. $\forall k \in \mathbb{N}_0$, slovo xy^kz je také v L .

Důkaz:

- Mějme regulární jazyk L , pak existuje DFA A s n stavy, že $L = L(A)$.
- Vezměme libovolné slovo $a_1a_2a_3 \dots a_m = w \in L$ délky $m \geq n, a_i \in \Sigma$.
- Definujeme: $\forall i p_i = \delta^*(q_0, a_1a_2 \dots a_i)$. Platí $p_0 = q_0$.
- Máme $n + 1 p_i$ a n stavů, některý se opakuje, vezměme první takový, t. j. $(\exists i, j : 0 \leq i < j \leq n \wedge p_i = p_j)$.
- Definujeme $x = a_1a_2 \dots a_i, y = a_{i+1}a_{i+2} \dots a_j, z = a_{j+1}a_{j+2} \dots a_m$, t.j. $w = xyz, y \neq \lambda, |xy| \leq n$.
- pak y^k můžeme opakovat libovolněkrát a vstup je také akceptovaný.

□

Příklad (Aplikace pumping lemmatu): TODO

2 Druhá přednáška

Definice (Kongruence): Mějme konečnou abecedu Σ a relaci ekvivalence \sim na Σ^* (reflexivní, symetrická, tranzitivní). Potom:

1. \sim je pravá kongruence, jestliže $(\forall u, v, w \in \Sigma^*) u \sim v \implies uw \sim vw$.
2. je konečného indexu, jestliže rozklad Σ^* / \sim má konečný počet tříd.
3. Třidu kongruence \sim obsahující slovo u značíme $[u]_\sim$, resp. $[u]$.

Věta (Myhill-Nerodova Věta): *Nechť L je jazyk nad konečnou abecedou Σ . Potom následující tvrzení jsou ekvivalentní:*

1. L je rozpoznatelný konečným automatem,
2. \exists pravá kongruence \sim konečného indexu nad Σ^* tak, že L je sjednocením jistých tříd rozkladu Σ^* / \sim .

Důkaz:

1. \implies 2.; t.j. automat \implies pravá kongruence konečného indexu
 - definujeme $u \sim v \equiv \delta^*(q_0, u) = \delta^*(q_0, v)$.

- je to ekvivalence (reflexivní, symetrická, tranzitivní)
- je to pravá kongruence (z definice δ^*)
- má konečný index (konečně mnoho stavů)

$$L = \{w | \delta^*(q_0, w) \in F\} = \bigcup_{q \in F} \{w | \delta^*(q_0, w) = q\} = \bigcup_{q \in F} [w | \delta^*(q_0, w) = q]_{\sim}.$$

2. \implies 1.; t.j. pravá kongruence konečného indexu \implies automat

- abeceda automatu nazveme Σ
- za stavy Q volíme třídy rozkladu Σ^* / \sim
- počáteční stav $q_0 \equiv [\lambda]_{\sim}$
- koncové stavy $F = \{c_1, \dots, c_n\}$, kde $L = \bigcup_{i=1, n} c_i$
- přechodová funkce $\delta([u], x) = [ux]$ (je korektní z def. pravé kongruence).
- $L(A) = L$

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

□

Příklad: Sestrojte automat přijímající jazyk

$L = \{w | w \in a, b^* \& |w|_a = 3k + 2\}$, t. j. obsahuje $3k + 2$ symbolů a .

1. $|u|_x$ značí počet symbolů x ve slově u
2. definujeme $u \sim v \equiv (|u|_a \bmod 3 = |v|_a \bmod 3)$
3. třídy ekvivalence $0, 1, 2$
4. L odpovídá třídě 2
5. a - přechody do následující třídy
6. b - přechody zachovávající třídu

28. slide, doplň obrázek

Příklad (Neregulární pumpovatelný jazyk): Ne-regulární jazyk, který lze pumpovat

Jazyk $L = \{u | u = a^+ b^i c^i \vee u = b^i c^j\}$ není regulární (Myhill-Nerodova věta), ale vždy lze pumpovat první písmeno.

1. Předpokládejme, že L je regulární
2. \implies pak \exists pravá kongruence \sim_L konečného indexu m , L je sjednocení některých tříd Σ^* / \sim_L
3. vezmeme množinu slov $S = \{ab, abb, abbb, \dots, ab^{m+1}\}$
4. existují dvě slova $i \neq j$, která padnou do stejné třídy

$i \neq j$	$ab^i \sim ab^j$	
přidáme c^i	$ab^i c^i \sim ab^j c^i$	\sim je kongruence
spor	$ab^i c^i \in L \& ab^j c^i \notin L$	$s' L$ je sjednocení některých tříd Σ^* / \sim_L .

Definice (Dosažitelný stav): Mějme DFA $A = (Q, \Sigma, \delta, q_0, F)$ a $q \in Q$. Řekneme, že stav je dosažitelný, jestliže $\exists w \in \Sigma^* : \delta^*(q_0, w) = q$.

Příklad: Algoritmus na hledání dosažitelných stavů : DFS (důkaz asi není nutný)

Definice (Automatový homomorfismus): Necht' A_1, A_2 jsou DFA. Řekneme, že zobrazení $h : Q_1 \rightarrow Q_2$ na Q_2 je (automatovým) homomorfismem, jestliže:

$$\begin{array}{ll} h(q_{0_1}) = q_{0_2} & \text{'stejný' počáteční stav} \\ h(\delta_1(q, x)) = \delta_2(h(q), x) & \text{'stejný' přechodová funkce} \\ q \in F_1 \Leftrightarrow h(q) \in F_2 & \text{'stejný' koncové stavy.} \end{array}$$

Homomorfismus prostý a na nazýváme isomorfismus.

Definice (Ekvivalence automatů): Dva konečné automaty A, B nad stejnou abecedou Σ jsou ekvivalentní, jestliže rozpoznávají stejný jazyk, t. j. $L(A) = L(B)$.

Věta (Věta o ekvivalenci automatů): *Existuje-li homomorfismus konečných automatů A_1 do A_2 , pak jsou A_1 a A_2 ekvivalentní.*

Důkaz:

1. Pro libovolné slovo $w \in \Sigma^*$ konečnou iterací

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

2. dále

$$\begin{aligned} w \in L(A_1) &\Leftrightarrow \delta_1^*(q_{0_1}, w) \in F_1 \\ &\Leftrightarrow h(\delta_1^*(q_{0_1}, w)) \in F_2 \\ &\Leftrightarrow \delta_2^*(h(q_{0_1}, w)) \in F_2 \\ &\Leftrightarrow \delta_2^*(q_{0_2}, w) \in F_2 \\ &\Leftrightarrow w \in L(A_2) \end{aligned}$$

□

Definice (Ekvivalence stavů): Říkáme, že stavy $p, q \in Q$ konečného automatu A jsou ekvivalentní, pokud:

1. Pro všechna vstupní slova $w : \delta^*(p, w) \in F \Leftrightarrow \delta^*(q, w) \in F$.

Pokud dva stavy nejsou ekvivalentní, říkáme že jsou rozlišitelné.

Příklad: Ten example je na slide 36, nejlepší s tím obrázkem

Definice (Algoritmus hledání rozpoznatelných stavů v DFA): Následující algoritmus nalezne rozlišitelné stavy:

1. Základ: Pokud $p \in F$ (přijímající) a $q \notin F$, pak je dvojice $\{p, q\}$ rozlišitelná.
2. Indukce: Necht' $p, q \in Q, a \in \Sigma$ a o dvojici $r, s : r = \delta(p, a), s = \delta(q, a)$ víme, že jsou rozlišitelné. Pak i $\{p, q\}$ jsou rozlišitelné.
3. opakuj dokud \exists nová trojice $p, q \in Q, a \in \Sigma$.

Doplň obrázky zo slidov, 37/38

Věta: *Pokud dva stavy nejsou odlišeny předchodzím algoritmem, pak jsou tyto stavy ekvivalentní.*

Důkaz: Korektnost algoritmu

1. Uvažujme špatné páry stavů, které jsou rozlišitelné a algoritmus je nerozlišil.
2. Vezměme z nich pár p, q rozlišitelný nejkratším slovem $w = a_1 \dots a_n$.
3. Stavy $r = \delta(p, a_1), s = \delta(q, a_1)$ jsou rozlišitelné kratším slovem $a_2 \dots a_n$, takže pár není mezi špatnými.
4. Tedy jsou 'vyškrtnuté' algoritmem.
5. Tedy v příštím kroku algoritmus rozliší i p, q .

□

Čas výpočtu je polynomiální vzhledem k počtu stavů.

1. V jednom kole uvažujeme všechny páry, t.j. $O(n^2)$.
2. Kol je maximálně $O(n^2)$, protože pokud nepřidáme křížek, končíme.
3. Dohromady $O(n^4)$.

Algoritmus lze zrychlit na $O(n^2)$ pamatováním stavů, které závisí na páru $\{r, s\}$ a sledováním těchto seznamů 'zpátky'.

Definice (Redukovaný DFA): Deterministický konečný automat je redukovaný, pokud

1. nemá dosažitelné stavy,
2. žádné dva stavy nejsou ekvivalentní.

Definice (Redukt): Konečný automat B je redukem automatu A, jestliže:

1. B je redukovaný,
2. A a B jsou ekvivalentní

ADD PICS PLS (*don't shout pls*)

Věta (Algoritmus na nalezení reduktu DFA A):

1. Ze vstupního DFA A eliminujeme stavy nedosažitelné z počátečního stavu.
2. Najdeme rozklad zbylých stavů na třídy ekvivalence.
3. Konstruujeme DFA B na třídách ekvivalence jakožto stavech.
Přechodová funkce B γ , mějme $S \in Q_B$. Pro libovolné $q \in S$ označíme T třídu ekvivalence $\delta(q, a)$ a definujeme $\gamma(S, a) = T$. Tato třída musí být stejná pro všechna $q \in S$.
4. Počáteční stav B je třída obsahující počáteční stav A.
5. Množina přijímajících stavů B jsou bloky odpovídající přijímacím stavům A.

The End