Automaty a Gramatiky

Poznámky z přednášek

Letní semestr 2020/2021

Viktor Soukup, Lukáš Salak

Verze 0.1
2. března 2021

Obsah

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, λNFA)	$\leftrightarrow$	regulární jazyky

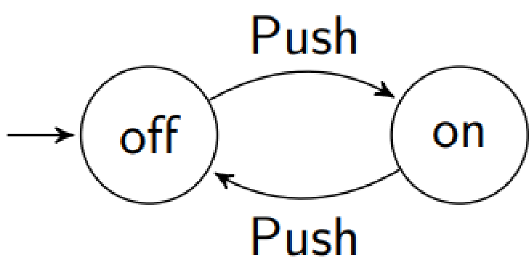
Nejjednodušší jsou nejníž, turingův stroj je nejkomplikovanější. Každá gramatika odpovídá nějaké třídě automatů.

Proč to řešíme?

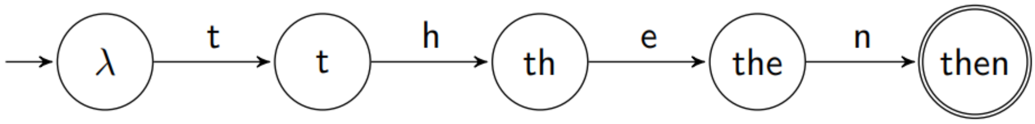
- zpracování přirozeného jazyka,
- překladače (lexikální, syntaktická analýza...),
- návrh, popis, verifikace hardware...
- hledání výskytu slova v textu (grep),
- verifikace systémů s konečně mnoha stavy

Příklad:

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



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



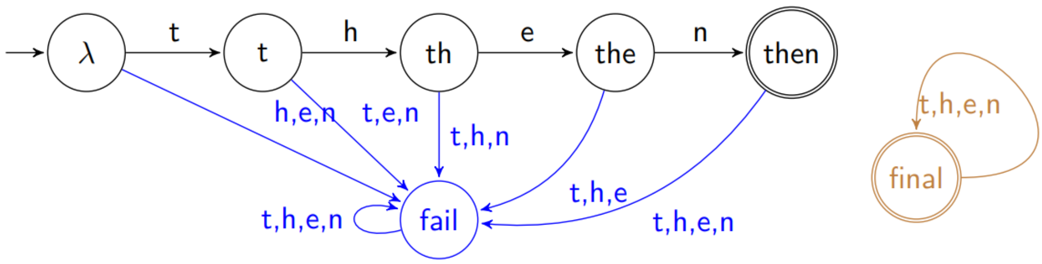
Definice (Deterministický konečný automat (DFA))**:** $A = (Q, \Sigma, \delta, q_0, F)$ sestává z:

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

5. neprázdné **množiny koncových (přijímajících) stavů** (final states) $F \subseteq Q$, označených dvojitým 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 $\forall s \in \Sigma : \delta(final, s) = final$.



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

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 neprádzných 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 $\delta : Q \times \Sigma \to Q$.
Rozšířenou přechodovou funkci δ^*: $Q \times \Sigma^* \to Q$ (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 $L = L(A)$.
- Třídu 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 $a_1a_2a_3 \ldots a_m = w \in L$ délky $m \geq n, a_i \in \Sigma$.*
- *Definujeme: $\forall i p_i = \delta^*(q_0, a_1a_2 \ldots a_i)$. Platí $p_0 = q_0$.*
- *Máme $n + 1 p_i$ a n stavů, některý se opakuje, vezměme první takový, t. j. $(\exists i, j : 0 \leq i < jqleqn \& p_i = p_j)$.*
- *Definujeme $x = a_1a_2 \ldots a_i, y = a_{i+1}a_{i+2} \ldots a_j, z = a_{j+1}a_{j+2} \ldots a_m$, t.j. $w = xyz, y \neq \lambda, |xy| \leq n$.*
- *pak y^k můžeme opakovat libovolněkrát a vstup je také akceptovaný.*

$\square$

Příklad (Aplikace pumping lemmatu)**:** TODO

The End