Část I

Regulární jazyky

1 Úvod do regulárních jazyků

1.1 Základní pojmy z oblasti formálních jazyků

- symbol
- abeceda
- řetězec
  - délka řetězce
  - prázdný jazyk
- formální jazyk nad abecedou
- Konkatenace nad řetězci (symbol $\cdot$)
  - signatura operace $\Sigma^* \times \Sigma^* \to \Sigma^*$
  - $w, w' \in \Sigma^*$:
  - $w = a_1 a_2 ... a_n$,
  - $w' = a'_1 a'_2 ... a'_m$, $n, m \geq 0$
  - $w \cdot w' = a_1 a_2 ... a_n a'_1 a'_2 ... a'_m$
  - Vlastnosti: asociativní operace, neutrální prvek ε vzhledem k operaci $\cdot$
- Další pojmy: prefix, sufix, podřetězec, podposloupnost,
- mocnina řetězce a^i
- operace nad jazyky:
  - jsou definovány množinové operace (sjednocení, průnik, doplněk, rozdíl, ...)
  - konkatenace jazyků L_1 a L_2
    - signatura operace: $2^{\Sigma^*} \times 2^{\Sigma^*} \to 2^{\Sigma^*}$
    - vstup: L_1 nad Σ_1 a L_2 nad Σ_2

    * výstup: jazyk L nad $\Sigma_1 \cup \Sigma_2$
    * $L_1 \cdot L_2 = \{xy \mid x \in L_1 \wedge y \in L_2\}$
  - iterace jazyků:
    * $L^0 = \{\varepsilon\}$
    * $L^n = L \cdot L^{n-1}$
    * $L^* = \bigcup_{n \geq 0} L^n$
    * $L^+ = \bigcup_{n \geq 1} L^n$

- Gramatika $G = (N, \Sigma, P, S)$, $N \cap \Sigma = \emptyset$
  - P je množina pravidel ve tvaru $(\alpha, \beta) \in P$, kde α obsahuje alespoň jeden nonterminál a *beta* je libovolná posloupnost terminálů a nonterminálů
  - zápis pomocí relace: $(N \cup \Sigma)^* N (N \cup \Sigma)^* \times (N \cup \Sigma)^*$
  - alternativně: zápis pomocí funkce $(N \cup \Sigma)^* N (N \cup \Sigma)^* \rightarrow (N \cup \Sigma)^*$

- Relace přímé derivace $\Rightarrow$ je definována na množině $(N \cup \Sigma)^*$
  - $\lambda \Rightarrow \mu \stackrel{def}{\Longleftrightarrow} \lambda = \gamma\alpha\delta \;\wedge\; \mu = \gamma\beta\delta \;\wedge\; (\alpha, \beta) \in P$

- Relace derivace $\Rightarrow^+$ je tranzitivní uzávěr relace přímé derivace
  - Pokud platí $\lambda \Rightarrow^+ \mu$, pak existuje posloupnost derivací délky $n \geq 1$, přičemž derivace délky n obsahuje $n+1$ prvků z množiny $(N \cup \Sigma)^*$

- $\Rightarrow^*$ je potom tranzitivní a relflexivní uzávěr relace přímé derivace

- Větná forma: $S \Rightarrow^* \alpha$, kde $\alpha \in (N \cup \Sigma)^*$

- Věta: $S \Rightarrow^* \alpha$, kde $\alpha \in \Sigma^*$

- jazyk generovaný gramatikou G: $L(G) = \{w \in \Sigma^* \mid S \Rightarrow^* w\}$

- Chomského hierarchie:
  - Typ 0 - neomezené gramatiky: $(N \cup \Sigma)^* N (N \cup \Sigma)^* \rightarrow (N \cup \Sigma)^*$
  - Typ 1 - kontextové gramatiky:
  - Typ 2 - bezkontextové gramatiky: $A \rightarrow \alpha$
  - Typ 3 - regulární gramatiky: $A \rightarrow xB|x|\varepsilon$

1.2 Konečné automaty a regulární jazyky

- NKA $M = (Q, \Sigma, \delta, q_0, F)$
  - δ: $Q \times \Sigma \to 2^Q$
  - pokud $\forall q \in Q\ \forall a \in \Sigma$:$|\delta(q, a)| \leq 1$, potom M je DKA
- lze definovat DKA s upravenou přechodovou funkcí δ: $Q \times \Sigma \to Q$
- lze definovat RKA s rozšířenou přechodovou funkcí δ: $Q \times (\Sigma \cup \{\varepsilon\}) \to 2^Q$
- konfigurace je prvek (q, w) z $Q \times \Sigma^*$
  - q je aktuální stav
  - w je doposud nezpracovaná část vstupního řetězce
- počáteční konfigurace: $(q_0, a_1 a_2 \dots a_n)$
- koncová konfigurace: (q_F, ε), kde $q_F \in F$
- přechodová relace automatu M je binární relace $\underset{M}{\vdash}$ na množině konfigurací $(Q \times \Sigma^*)$, $\underset{M}{\vdash} \subseteq (Q \times \Sigma^*) \times (Q \times \Sigma^*)$
  - $(q, w) \underset{M}{\vdash} (q', w') \overset{def}{\iff} w = aw' \wedge q' \in \delta(q, a)$, pro $q, q' \in Q, a \in \Sigma, w, w' \in \Sigma^*$
- řetězec přijímaný NKA je definován: $(q_0, w) \overset{*}{\underset{M}{\vdash}} (q, \varepsilon),\ q \in F$
- epsilon uzávěr je funkce, ε-uzávěr$(q) = \{p \mid \exists w \in \Sigma^* : (q, w) \overset{*}{\vdash} (p, w)\}$
- zobecnění pro libovolnou množinu $T \subseteq Q$: ε-uzávěr$(T) = \bigcup_{s \in T} \varepsilon$-uzávěr$(s)$

1.3 Regulární množiny a výrazy

Následující množiny jsou regulární množiny nad Σ a žádné jiné:

- $\emptyset$
- $\{\varepsilon\}$
- $\{a\}$

- je-li R a S regulární množina, potom:
  - $R \cup S$
  - $R \cdot S$
  - R^*

Regulární výrazy označují regulární množiny, tak, že:

- $\emptyset$ je RV značící regulární množinu $\emptyset$
- ε RV značící regulární množinu $\{\varepsilon\}$
- a RV značící regulární množinu $\{a\}$, pro všechna $a \in \Sigma$
- je-li r a s RV značící regulární množiny R a S, pak:
  - $(r + s)$ je RV značící regulární množinu $R \cup S$
  - (rs) je RV značící regulární množinu $R \cdot S$
  - (r^*) je RV značící regulární množinu R^*