

Automaty a formální jazyky

Přednáška IV.

Gramatiky

Gramatiky - motivace

- Sada pravidel pro vytváření slov (řetězců), jež patří do daného jazyka.
- Původně: přirozené jazyky
- Formalizováno pro studium přirozených jazyků a popis jazyků umělých
- V IT: popis programovacích jazyků (C, Java, ...)

Jak to vypadá?

- Definujeme **sadu pravidel** pro generování jazyka
- Vycházíme z počátečního symbolu a ten postupně rozvíjíme
- Končíme, když celý řetězec obsahuje znaky z abecedy

Neformální příklad – zápis matematického výrazu:

V: $A * B$

A: celé číslo

B: celé číslo

Produkční (přepisovací) systém

- Produkční systém R je dvojice (V, P) kde
 - V je konečná abeceda
 - P je konečná množina produkčních (přepisovacích) pravidel
 - **Pravidlo** je uspořádaná dvojice (u, v) , kde $u, v \in V^*$, zapisujeme zpravidla $u \rightarrow v$
- Příklad:
 - $V = \{0, 1, P\}$
 - $P \rightarrow 0P0$
 - $P \rightarrow 1$Opakovaným používáním přepisovacích pravidel získáváme jazyk, jenž obsahuje řetězce:
010 00100 0001000 0P0 00P00

Přímé a nepřímé přepsání

- **Definice:** Slovo u se **přímo přepisuje** na slovo v , pokud existují taková slova $u, v, w, x, y, z \in V^*$, pro která platí:
 $v = xyz$, $u = xwz$, $y \rightarrow w$ je přepisovací pravidlo.
- **Definice:** Slovo u se **přepíše** na v , pokud existují slova $u_1, u_2, \dots, u_n \in V^*$ takové, že
 $u \rightarrow u_1 \rightarrow u_2 \rightarrow \dots \rightarrow u_n \rightarrow v$
Značíme $u \rightarrow^* v$
- Posloupnost u_1, u_2, \dots, u_n nazýváme **odvozením (derivací)**
- Minimální derivace je taková derivace, kde žádné dva prvky nejsou shodné.

Generativní gramatika - formálně

- **Definice: Generativní gramatika** je uspořádaná čtveřice $G = (\Pi, \Sigma, S, P)$ kde:
 - Π je konečná množina (neterminály)
 - Σ je konečná množina (terminály)
 - $S \in \Pi$ je počáteční symbol
 - P je množina produkčních pravidel $u \rightarrow v$:
 $u, v \in (\Sigma \cup \Pi)^*$
řetězec u obsahuje alespoň jeden neterminální symbol.

Co to znamená?

- **Terminály** = tvořeny z „cílové“ abecedy, přepisování zde končí
- **Neterminály** = pomocné symboly pro sestavení odvození
- Příklad: Jazyk obsahuje kombinace nul a jedniček, slovo nezačíná nulou.

$\Sigma = \{0,1\}$, $\Pi = \{A,B,C\}$, poč. symbol A

Produkční pravidla:

A \rightarrow BC

B \rightarrow 0

C \rightarrow C

C \rightarrow 1

C \rightarrow 0

Domluva pro zápis

- **Neterminály:**
 - Velká písmena v jednodušším zápisu
 - Uzavřeno do závorek $\langle \rangle$, použijeme-li slovo, abychom zápisu rozuměli: $\langle \text{číslo} \rangle$
- **Terminály** jsou prvky abecedy, která tvoří generovaný jazyk, zde žádná dohoda nemůže být.
- Přepisovací pravidla se stejnou levou stranou se slučují s pomocí symbolu $|$:
 - $A \rightarrow BC$
 - $B \rightarrow 0 \mid 1 \mid C$

Gramatika a jazyk

- **Definice:** Jazyk $L(G)$ je generován gramatikou G právě tehdy, když všechna slova jazyka lze odvodit z počátečního symbolu gramatiky.

$$L(G) = \{w; w \in \Sigma^*, S \rightarrow^* w\}$$

- **Definice:** Gramatiky G_1 a G_2 jsou si ekvivalentní právě tehdy, když $L(G_1) = L(G_2)$.

Chomského hierarchie

- Rozdělení gramatik do 4 skupin podle jejich vlastností (záleží na tvaru produkčních pravidel).
- **Třída 0: Gramatiky typu 0** mají obecná produkční pravidla.
- **Třída 1: Gramatiky typu 1 – kontextové jazyky** mají produkční pravidla ve tvaru
 1. $\alpha X \beta \rightarrow \alpha Y \beta$
kde $X \in \Pi$, $\alpha, \beta \in (\Pi \cup \Sigma)^*$, $Y \in (\Pi \cup \Sigma)^+$
 1. Smí obsahovat pravidlo tvaru $S \rightarrow \varepsilon$; ale S se pak nesmí vyskytovat na pravé straně žádného pravidla.

Chomského hierarchie II.

- **Třída 2: Gramatiky typu 2** – bezkontextové jazyky mají produkční pravidla pouze ve tvaru $X \rightarrow w$, kde $X \in \Pi$, $w \in (\Pi \cap \Sigma)^*$
- **Třída 3: Gramatiky typu 3** – regulární jazyky mají pouze pravidla ve tvaru:
 $X \rightarrow wY$, $X \rightarrow w$, kde $X, Y \in \Pi$, $w \in \Sigma^*$

Vztah mezi třídami CH. H.

- **Věta:**

$$L(G_0) \supset L(G_1) \supset L(G_2) \supset L(G_3)$$

(Indexy u gramatik označují třídu jednotlivých gramatik v Chomského hierarchii.)