

Chomského hierarchie gramatik a jazyků

Lingvista Noam Chomsky rozdělil gramatiky do čtyř skupin (typů) na základě různých omezení na tvar pravidel. Jeho práce (z konce 50. let) byla původně motivována úvahami o struktuře přirozeného jazyka, z dnešního pohledu má však především význam jako základní (a neobyčejně výhodné) rozdělení gramatik podle jejich popisné síly. Chomského hierarchie rozlišuje následující čtyři typy gramatik:

typ 0: Gramatika typu 0 nemá žádná omezení na tvar pravidel (frázová gramatika)

typ 1: Gramatika je typu 1 (kontextová gramatika) \iff každé její pravidlo $\alpha \rightarrow \beta$ [$\alpha, \beta \in (N \cup \Sigma)^*$] má omezení $|\alpha| \leq |\beta|$ s eventuální výjimkou $S \rightarrow \varepsilon$, pokud se S nevyskytuje na pravé straně žádného jiného pravidla.

typ 2: Gramatika je typu 2 (bezkontextová gramatika) \iff všechna její pravidla jsou tvaru $A \rightarrow \beta$ [$A \in N, \beta \in (N \cup \Sigma)^*$] a mají omezení $|\beta| \geq 1$ opět s eventuální výjimkou $S \rightarrow \varepsilon$, pokud se S nevyskytuje na pravé straně žádného jiného pravidla.

typ 3: Gramatika je typu 3 (regulární gramatika) \iff každé její pravidlo je tvaru $A \rightarrow a$ nebo $A \rightarrow aB$, kde $A, B \in N, a \in \Sigma$ s eventuální výjimkou $S \rightarrow \varepsilon$, pokud se S nevyskytuje na pravé straně žádného jiného pravidla.

Poznámka.

Chomského hierarchie gramatik určuje též třídy (množiny) jazyků. Jazyk L je

1. regulární, když existuje regulární gramatika G (typu 3), která jej generuje.
2. bezkontextový, když existuje bezkontextová gramatika G (typu 2), která jej generuje.
3. kontextový, když existuje kontextová gramatika G (typu 1), která jej generuje.
4. typu 0, když existuje frázová gramatika G (typu 0), která jej generuje.

Množinový vztah jednotlivých tříd se dá zapsat takto:

$$\text{typ 3} \subset \text{typ 2} \subset \text{typ 1} \subset \text{typ 0}$$

Je tedy důležité poznamenat, že to, že nalezneme k nějakému jazyku L bezkontextovou gramatiku, ještě neznamená, že je jazyk L bezkontextový. Třeba jsme jen hledali špatně, a regulární gramatika pro jazyk L existuje.

Příklad 1.

Bud' $G = (\{S, A, B, C\}, \{a, b, c, d\}, P, S)$, kde $P = \{$

$$\begin{aligned} S &\rightarrow aSb \mid cAd \\ cA &\rightarrow aB \mid Ca \\ Bd &\rightarrow Sb \mid A \\ Cad &\rightarrow ab \mid \varepsilon \} \end{aligned}$$

Jaký jazyk generuje gramatika G a jakého je typu?

Řešení. Gramatika je typu 0 protože pravidla $Bd \rightarrow A$, $Cad \rightarrow ab$ a $Cad \rightarrow \varepsilon$ porušují omezení gramatiky typu 1 říkající, že každé její pravidlo $\alpha \rightarrow \beta$ musí splňovat podmínku $|\alpha| \leq |\beta|$.

Gramatika generuje následující slova:

- $S \Rightarrow cAd \Rightarrow Cad \Rightarrow ab$
- $S \Rightarrow cAd \Rightarrow Cad \Rightarrow \varepsilon$
- $S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aacAdbb \Rightarrow aaCadbb \Rightarrow aaabbb$
- $S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aacAdbb \Rightarrow aaCadbb \Rightarrow aabb$
- $S \Rightarrow^3 aaaSbbb \Rightarrow aaacAdbbb \Rightarrow aaaaBdbbb \Rightarrow aaaaSbbbb \Rightarrow aaaacAdbbbb \Rightarrow aaaaCadbbbb \Rightarrow aaaaabbbbb$
- $S \Rightarrow^3 aaaSbbb \Rightarrow aaacAdbbb \Rightarrow aaaaBdbbb \Rightarrow aaaaSbbbb \Rightarrow aaaacAdbbbb \Rightarrow aaaaCadbbbb \Rightarrow aaaaabbbbb$

Generovaná slova můžeme též zapsat následujícím způsobem:

$$L(G) = \{a^i b^i, i \in \mathbb{N}_0\}$$

Příklad k samostatnému řešení.

V předchozím **Příkladu 1** jsme měli gramatiku typu 0 generující jazyk $L = \{a^i b^i \mid i \in \mathbb{N}_0\}$. Zkuste najít bezkontextovou gramatiku pro tentýž jazyk L .

Příklad 2.

Bud' $G = (\{S, A\}, \{a, b, c\}, P, S)$, kde $P = \{$

$$\begin{aligned} S &\rightarrow bS \mid cS \mid aA \\ A &\rightarrow aA \mid bA \mid cA \mid \varepsilon \} \end{aligned}$$

Jaký jazyk gramatika G generuje a jakého je typu?

Řešení. Gramatika je typu 0 protože pravidlo $A \rightarrow \varepsilon$ porušuje omezení gramatiky typu 1 říkájící, že každé její pravidlo $\alpha \rightarrow \beta$ musí splňovat podmínku $|\alpha| \leq |\beta|$.

Generovaná slova můžeme též zapsat následujícím způsobem:

$$L(G) = \{w \in \{a, b, c\}^*, \#_a(w) \geq 1\}$$

Příklad 3.

Navrhněte regulární gramatiku pro následující jazyk $L: L = \{a, b, c, d\}^2 \cdot \{a, b, c, d\}^*$.

Řešení. $G = (\{S, A\}, \{a, b, c, d\}, P, S)$, kde $P = \{$

$$\begin{aligned} S &\rightarrow aA \mid bA \mid cA \mid dA \\ A &\rightarrow a \mid b \mid c \mid d \mid aA \mid bA \mid cA \mid dA \} \end{aligned}$$

Poznámka.

I následující gramatika G_2 generuje jazyk L (zadaný v **Příkladu 3**), ale už není regulární: $G = (\{S, A, B\}, \{a, b, c, d\}, P, S)$, kde $P = \{$

$$\begin{aligned} S &\rightarrow aA \mid bA \mid cA \mid dA \\ A &\rightarrow aB \mid bB \mid cB \mid dB \\ B &\rightarrow \varepsilon \mid aB \mid bB \mid cB \mid dB \} \end{aligned}$$

Bohužel navíc je tam pravidlo $B \rightarrow \varepsilon$.