Chomského normální forma

Posledním kanonickým tvarem, který probereme, je Chomského normální forma (zkráceně CNF). Nejdříve si uvedeme definici.

Definice – Chomského normální forma

Řekněme, že bezkontextová gramatika $G = (N, \Sigma, P, S)$ je v Chomského normální formě, právě když všechna pravidla v P jsou následujícího tvaru:

1. $A \to a,\ a \in \Sigma$
2. $A \to BC,\ B, C \in N$
3. pokud $\varepsilon \in L(G)$, pak $S \to \varepsilon \in P$ za podmínky, že se S nevyskytuje na pravé straně žádného pravidla.

Poznámka.

Buď $G = (N, \Sigma, P, S)$ bezkontextová gramatika v CNF. Derivační strom s výsledkem $w \in L(G)$ je speciální binární strom, pro jehož libovolný vnitřní uzel u platí:

1. má-li u právě dva následníky, pak se jedná o vnitřní uzly s návěštím z N.
2. má-li u právě jednoho následníka, pak je to list s návěštím ze Σ.

U takto definovaného derivačního stromu lze ukázat souvislost mezi max. délkou nějaké cesty od kořene k listu a max. počtem všech listů. Této skutečnosti se podstatně využívá v důkazové technice zvané Pumping lemma pro bezkontextové jazyky.

Příklad 1.

Buď dána bezkontextová gramatika $G = (\{S, A, B\}, \{a, b\}, P, S)$, kde množina pravidel $P = \{$

$$\begin{aligned} S &\to a \mid ba \mid AB \\ A &\to bA \mid Ba \\ B &\to b \mid BAB\} \end{aligned}$$

Chceme-li mít uvedenou gramatiku v Chomského normální formě, musíme ji upravit. Naznačíme si postupně 3 možné „úpravy“, které poté formalizujeme v algoritmu pro převod do CNF.

1. Některá pravidla ponecháme tak, jak jsou, protože odpovídají CNF. Jsou to $S \to a \mid AB,\ B \to b$.
2. U pravidel $A \to \alpha,\ |\alpha| \geq 2$ obsahujících terminální symboly provedeme jejich „očárkování“, tj. libovolný terminál $x \in \Sigma$ nahradíme za nový neterminál x' a přidáme pravidlo $x' \to x$. Takto si můžeme pomoci u pravidel $S \to ba,\ A \to bA \mid Ba$, která nahradíme

   $$\begin{aligned} S &\to b'a', \\ A &\to b'A \mid Ba', \\ a' &\to a, \\ b' &\to b. \end{aligned}$$

   Je patrné, že nová pravidla již vyhovují CNF.

3. V předchozím případě jsme „očárkováním“ zajistili splnění CNF u pravidel, jejichž pravá strana má délku právě 2. Co však pravidla $A \to \beta$, kde délka řetězce β je více než 2? Toto kritérium splňuje pravidlo $B \to BAB$. I zde je však pomoc snadná. Místo $B \to BAB$ zapíšeme $B \to BX$ a přidáme $X \to AB$.

   Tento postup navíc můžeme opakovat libovolně krát dlouho. Tedy: máme-li pravidlo $A \to \beta_1\beta_2 \ldots \beta_n$, pak jej postupně nahradíme pravidly:

$$\begin{array}{rcl} A & \to & \beta_1 X_1, \\ X_1 & \to & \beta_2 X_2, \\ X_2 & \to & \beta_3 X_3, \\ & \ldots & \\ X_{n-2} & \to & \beta_{n-1}\beta_n \end{array}$$

Algoritmus (převod do Chomského normální formy)

Vstup: Bezkontextová gramatika $G = (N, \Sigma, P, S)$ bez jednoduchých pravidel.
Výstup: Bezkontextová gramatika $G' = (N', \Sigma, P', S)$ v Chomského normální formě.

$P' := \emptyset$
$N' := N$
foreach pravidlo $A \to \alpha$ **do**

1. je-li $\alpha = a$ nebo $\alpha = BC$, kde $a \in \Sigma,\ B, C \in N$, pak $P' := P' \cup \{A \to \alpha\}$.
   (Tj. přidej do P' pravidla, která splňují CNF.)
2. je-li $|\alpha| = |\alpha_1\alpha_2| = 2$ a alespoň jeden symbol $\alpha_i \in \Sigma$, přidej do P' pravidla
   - $A \to \alpha_1'\alpha_2,\ \alpha_1' \to \alpha_1$, je-li pouze $\alpha_1 \in \Sigma$;
     $N' := N' \cup \{\alpha_1'\}$
   - $A \to \alpha_1\alpha_2',\ \alpha_2' \to \alpha_2$, je-li pouze $\alpha_2 \in \Sigma$;
     $N' := N' \cup \{\alpha_2'\}$
   - $A \to \alpha_1'\alpha_2',\ \alpha_1' \to \alpha_1,\ \alpha_2' \to \alpha_2$, jsou-li oba $\alpha_1, \alpha_2 \in \Sigma$;
     $N' := N' \cup \{\alpha_1', \alpha_2'\}$
3. Pokud $\alpha = \alpha_1\alpha_2 \ldots \alpha_n$, kde $n > 2$, pak do P' přidej následující sadu pravidel:

$$\begin{array}{rcl} A & \to & \beta_1 X_1, \\ X_1 & \to & \beta_2 X_2, \\ X_2 & \to & \beta_3 X_3, \\ & \ldots, & \\ X_{n-2} & \to & \beta_{n-1}\beta_n, \end{array}$$

   kde

   - $\beta_i = \alpha_i$, je-li $\alpha_i \in N$.
   - $\beta_i = \alpha_i'$, je-li $\alpha_i \in \Sigma$. V tomto případě přidej do P' pravidlo $\alpha_i' \to \alpha_i$ a do množiny neterminálů N' doplň α_i'.

   $N' := N' \cup \{X_1, \ldots, X_{n-2}\}$

done

Poznámka.

Zmíněný algoritmus prochází všechna pravidla a u každého z nich provede jeden ze tří následujících kroků:

1. ponechá pravidlo ve stávajícím tvaru, protože splňuje Chomského normální formu.

2. Má-li pravá strana pravidla pouze dva symboly a nesplňuje CNF, je zřejmé, že alespoň jeden ze symbolů musí být terminál. V takovém případě očárkujeme terminály x na pravé straně pravidla a přidáme $x' \to x$ do P'.

3. Má-li pravá strana pravidla $A \to \alpha_1 \ldots \alpha_n$ více než dva symboly, pak ji nahrazujeme řetězcem $\alpha_1 X_1$, kde X_1 je nový neterminál. Je-li navíc symbol $\alpha_1 \in \Sigma$, pak jej očárkujeme a do P' přidáme $\alpha_1' \to \alpha_1$.

   Za neterminál X_1 jsme „ukryli" zbylou část řetězce $\alpha_2\alpha_3 \ldots \alpha_n$. Jakým způsobem přidáme pravidla generující zmíněný řetězec tak, aby byla v CNF? Provedeme stejný krok jako v případě A-pravidla. Přidáme $X_1 \to \alpha_2 X_2$, je-li $\alpha_2 \in \Sigma$, „očárkujeme" α_2 a do P' vložíme $\alpha_2' \to \alpha_2$. A takto pokračujeme, dokud se pod X_i neskrývají pouze dva symboly...

Příklad 2.

Je dána bezkontextová gramatika $G = (\{S, A, B\}, \{a, b\}, P, S)$ s pravidly $P = \{$

$$
\begin{array}{lcl}
S & \to & SaSbS \mid aAa \mid bBb \\
A & \to & aA \mid aaa \\
B & \to & Bb \mid bb \mid b\}
\end{array}
$$

Převeďte gramatiku G do Chomského normální formy.

Řešení.

- $S \to SaSbS$ nahradíme (dle kroku 3.) postupně pravidly $S \to SS_1,\ S_1 \to a'S_2,\ S_2 \to SS_3,\ S_3 \to b'S$ a přidáme ještě $a' \to a,\ b' \to b$.
- $S \to aAa$ nahradíme (dle kroku 3.) pravidly $S \to a'S_4,\ S_4 \to Aa'$.
- $S \to bBb$ nahradíme (dle kroku 3.) pravidly $S \to b'S_5,\ S_5 \to Bb'$.
- $A \to aA$ nahradíme (dle kroku 2.) pravidlem $A \to a'A$.
- $A \to aaa$ nahradíme (dle kroku 3.) pravidly $A \to a'A_1,\ A_1 \to a'a'$.
- $B \to Bb$ nahradíme (dle kroku 2.) pravidlem $B \to Bb'$.
- $B \to bb$ nahradíme (dle kroku 2.) pravidlem $B \to b'b'$.
- $B \to b$ ponecháme (dle kroku 1.) beze změny.

Výsledná gramatika $G' = (\{S, A, B, S_1, S_2, S_3, S_4, S_5, A_1, a', b'\}, \{a, b\}, P', S)$ s pravidly $P' = \{$

$$
\begin{array}{rcl}
S & \to & SS_1 \mid a'S_4 \mid bS_5 \\
A & \to & a'A \mid a'A_1 \\
B & \to & Bb' \mid b'b' \mid b \\
S_1 & \to & a'S_2 \\
S_2 & \to & SS_3 \\
S_3 & \to & b'S \\
S_4 & \to & Aa' \\
S_5 & \to & Bb' \\
A_1 & \to & a'a' \\
a' & \to & a \\
b' & \to & b\}
\end{array}
$$

je v Chomského normální formě.

Příklad 3.

Buď dána bezkontextová gramatika $G = (\{S, A, B, C\}, \{a, b, c\}, P, S)$ s pravidly $P = \{$

$$
\begin{array}{rcl}
S & \to & ABc \mid \varepsilon \\
A & \to & AbAa \mid BC, \\
B & \to & bB \mid CAa \mid b \\
C & \to & cCaA \mid abc\}
\end{array}
$$

Převeďte gramatiku G do Chomského normální formy.