Bezkontextové gramatiky - základy

Šimon Sádovský

Poznámka 1. Je úplne **v poriadku** pokiaľ sa **na začiatku štúdia FoJákov** trocha „strácate vo formalizmoch“ a máte problémy čítať a písať FoJu poriadne. Avšak **nie je vôbec v poriadku**, ak to tak ostane **do konca semestra**. Písať matematický text korektne a zrozumiteľne vyžaduje tréning, jednoducho si to treba odmakať.

Ako si môžete všimnúť v uvedenom príklade, prelína sa tam bežne čitateľný text a matematické formalizmy. Nakoľko kedy zvoliť bežný text a kedy formalizmus je tak trochu kumšt. **Cieľom formalizmov je matematický text sprehľadniť, zjednodušiť jeho čitateľnosť a zároveň napomôcť jeho jednoznačnosti.**

Taktiež uvádzam zopár ekvivalentných štandardných formulácií, ktoré sa pri textoch ohľadom (bezkontextových) gramatík často vysktujú. Predpokladáme, že pracujeme s bezkontextovou gramatikou $G = (N, T, P, \sigma)$.

- **Ak píšem** „Nech $u \in (N \cup T)^*$ také, že $\sigma \Rightarrow_G^* u$“ **je to to isté ako** „Nech u je vetná forma gramatiky G“. **Pre lepšiu zrozumiteľnosť môžem použiť aj formuláciu** „Nech u je vetná forma, ktorú dokážem odvodiť v gramatike G.“ **Ak je z kontextu zjavné, že pracujeme práve s gramatikou G a nie inou, môžem použiť** „Nech $u \in (N \cup T)^*$ také, že $\sigma \Rightarrow^* u$“.
- **Ak píšem** „$\sigma \Rightarrow_G^k u$“, **je to to isté ako** „u je vetná forma gramatiky G odvodená na k krokov“. **Ak je z kontextu zjavné, že pracujeme práve s gramatikou G a nie inou, môžem použiť aj** „$\sigma \Rightarrow^k u$“, **čo je to isté ako** „u je vetná forma odvodená na k krokov“.
- **Ak píšem** „$\sigma \Rightarrow_G^* w$“, **je to to isté ako** „existuje odvodenie slova w v gramatike G“. **Ak je z kontextu zjavné, že pracujeme práve s gramatikou G a nie inou, môžem použiť aj** „$\sigma \Rightarrow^* w$“ **čo je to isté ako** „existuje odvodenie slova w“.
- **Ak píšem** $w \in T^*$, **je to to isté ako** „w je terminálne slovo gramatiky G“. **Ak je z kontextu zjavné, že pracujeme práve s gramatikou G a nie inou, môžem použiť aj** „w je terminálne slovo“.
- **Ak píšem** $\sigma \to \varepsilon \in P$, **je to to isté ako** „gramatika G obsahuje pravidlo $\sigma \to \varepsilon$“. **Ak je z kontextu zjavné, že pracujeme práve s gramatikou G a nie inou, môžem použiť aj** „ Existuje pravidlo $\sigma \to \varepsilon$“.

Zároveň upozorňujem, že tento zoznam „konverzií formalizmov na slovenčinu“ nie je zďaleka úplný. Navyše aj horeuvedené formálne zapísané formulácie sa v slovenčine dajú povedať zrozumiteľne a jednoznačne aj iným spôsobom, ako uvedeným.

Úloha 1. Zostrojte bezkontextovú gramatiku generujúcu jazyk

$$L = \{a^n b^n \mid n \in \mathbb{N}\}.$$

Správnosť svojej konštrukcie *poriadne* dokážte.

Dôkaz. Definujeme bezkontextovú $G = (N, T, P, \sigma)$, kde $N = \{\sigma\}$, $T = \{a, b\}$ a

$$P = \{\sigma \quad \to \quad a\sigma b \mid \varepsilon\}.$$

Teraz potrebujeme dokázať, že platí $L(G) = L$. Rovnosť dokážeme dvoma inklúziami.

$\boxed{\subseteq}$: Najprv dokážeme nasledujúce pomocné tvrdenie.

Tvrdenie 1. *Každá vetná forma, ktorú dokážeme v gramatike G odvodiť je jedného z nasledujúcich tvarov:*

(i) $a^n \sigma b^n$ *kde* $n \in \mathbb{N}$

(ii) $a^n b^n$ *kde* $n \in \mathbb{N}$

Indukciou na dĺžku odvodenia.

1°: Nulakrokovým odvodením dokážeme odvodiť iba vetnú formu σ. Táto vetná forma je tvaru (i) pre $n = 0$.

2°: Nech $k \in \mathbb{N}$. Predpokladajme, že každá vetná forma gramatiky G odvodená na nanajvýš k krokov je jedného z tvarov (i) alebo (ii). Nech pre $u \in (N \cup T)^*$ platí $\sigma \Rightarrow^{k+1} u$. Potom musí existovať $v \in (N \cup T)^*$ také, že platí $\sigma \Rightarrow^k v \Rightarrow u$, *neformálne* v *je „vetná forma odvodená tesne pred* u*“*. Keďže vetná forma v je odvodená na k krokov, platí o nej indukčný predpoklad, teda musí byť tvaru (i) alebo (ii). Ak by v bola tvaru (ii), tak by neobsahovala žiaden neterminál a teda by sa z nej nedal spraviť ďalší krok odvodenia. Teda v musí byť tvaru (i). To znamená, že existuje $m \in \mathbb{N}$ také, že $v = a^m \sigma b^m$. Teraz skúmajme posledný krok odvodenia $v \Rightarrow u$. Máme dve možnosti aké pravidlo v ňom použiť, postupne ich rozoberieme:

- Použijeme pravidlo $\sigma \to a\sigma b$. Teda posledný krok odvodenia vyzerá takto: $a^m \sigma b^m \Rightarrow a^{m+1} \sigma b^{m+1}$. V tomto prípade teda dostávame, že platí $u = a^{m+1} \sigma b^{m+1}$. To znamená, že u je tvaru (i) pre $n = m + 1$.
- Použijeme pravidlo $\sigma \to \varepsilon$. Teda posledný krok odvodenia vyzerá takto: $a^m \sigma b^m \Rightarrow a^m b^m$. V tomto prípade teda dostávame, že platí $u = a^m b^m$ pre $n = m$. To znamená, že u je tvaru (ii).

Kedže iné pravidlá sme použiť nemohli a v každom z možných prípadov sme dokázali, že vetná forma u je jedného z požadovaných tvarov, tak dôkaz Tvrdenia 1 je hotový.

Teraz prejdime k dôkazu $L(G) \subseteq L$. Nech $w \in L(G)$. Teda $\sigma \Rightarrow^* w$. Potom podľa Tvrdenia 1 musí platiť, že w je jedného tvarov (i) alebo (ii). Keďže $w \in T^*$ a vetné formy tvaru (i) obsahujú neterminál σ, tak nutne w je tvaru (ii). Teda existuje $n \in \mathbb{N}$ také, že $w = a^n b^n$. Teraz už zjavne $w \in L$.

$\boxed{\supseteq}$: Najprv dokážeme nasledujúce pomocné tvrdenie.

Tvrdenie 2. *Nech* $n \in \mathbb{N}$. *Potom* $\sigma \Rightarrow^* a^n \sigma b^n$.

Indukciou na n.

1°: Nech $n = 0$. Potom triviálne platí $\sigma \Rightarrow^* \sigma$.

2°: Nech $n \in \mathbb{N}$ také, že $n \geq 1$. Podľa IP platí $\sigma \Rightarrow^* a^{n-1} \sigma b^{n-1}$. Keďže gramatika G obsahuje pravidlo $\sigma \to a\sigma b$, tak platí $a^{n-1} \sigma b^{n-1} \Rightarrow a^n \sigma b^n$. Dajúc dokopy predchádzajúce dostávam $\sigma \Rightarrow a^{n-1} \sigma b^{n-1} \Rightarrow a^n \sigma b^n$, teda platí $\sigma \Rightarrow a^n \sigma b^n$.

Teraz prejdime k dôkazu $L \subseteq L(G)$. Nech $w \in L$. Teda existuje $n \in \mathbb{N}$ také, že platí $w = a^n b^n$. Podľa Tvrdenia 2 platí $\sigma \Rightarrow^* a^n \sigma b^n$. Kedže $\sigma \to \varepsilon \in P$, tak platí $a^n \sigma b^n \Rightarrow a^n b^n$. Dajúc dokopy predchádzajúce dostávame $\sigma \Rightarrow^* a^n \sigma b^n \Rightarrow a^n b^n$, čo znamená, že $w \in L(G)$ a dôkaz je hotový. □

Poznámka 2. Tvrdenia typu $L = L(G)$ štandardne dokazujeme pomocou dvoch inklúzií ako sme aj videli na príklade. Je dobré si uvedomiť nasledovné.

V prípade inklúzie $L(G) \subseteq L$ dokazujeme, že gramatika nám nevygeneruje nič zlé vzhľadom na to, čo by sme od nej chceli. Pozitívne povedané, všetko čo gramatika generuje je OK. To sme spravili tak, že sme nejak popísali všetky možné odvodenia, ktoré táto gramatika dokáže spraviť a ukázali sme, že každé z nich je OK.

V prípade inklúzie $L \subseteq L(G)$ musíme ukázať, že každé slovo, ktoré jazyk L obsahuje, dokážeme v gramatike G vygenerovať. To sme spravili tak, že pre každé slovo z jazyka L sme zostrojili odvodenie v gramatike G.