

# Bezkontextové a regulárne gramatiky - mierne pokročilejšie

Eduard Baklažán Batmedijn

**Definícia 1.** Symbolom  $\oplus_m$  v tomto riešení značíme operáciu „sčítania modulo  $m$ “ na celých číslach:

$$s \oplus_m t = (s + t) \bmod m.$$

**Úloha 1.** Zostrojte regulárnu gramatiku generujúcu jazyk

$$L = \{w \in \{a, b\}^* \mid \#_a(w) \equiv 1 \pmod{41} \wedge \#_b(w) \equiv 7 \pmod{37}\}.$$

Správnosť svojej konštrukcie *poriadne* dokážte.

*Riešenie.* Naša gramatika  $G = (N, T, P, \sigma)$  bude mať  $41 \cdot 37$  neterminálov, pre jednoduchosť môžeme zobrať  $N = \mathbb{Z}_{41} \times \mathbb{Z}_{37}$ <sup>1</sup>. Pomocou týchto neterminálov si budeme „počítat“ počet doteraz vygenerovaných  $a$ -čok a  $b$ -čok (presnejšie, zvyšok tohto počtu po delení 41, resp. 37). Po vygenerovaní  $s$   $a$ -čok a  $t$   $b$ -čok teda vo vetnej forme chceme mať neterminál  $(s \bmod 41, t \bmod 37)$  (okrem prípadu, že už máme hotové slovo obsahujúce iba terminály). Preto použijeme nasledovnú množinu pravidiel:

$$\begin{aligned} P = & \{(i, j) \rightarrow a(i \oplus_{41} 1, j) \mid i \in \mathbb{Z}_{41} \wedge j \in \mathbb{Z}_{37}\} \cup \\ & \{(i, j) \rightarrow b(i, j \oplus_{37} 1) \mid i \in \mathbb{Z}_{41} \wedge j \in \mathbb{Z}_{37}\} \cup \\ & \{(1, 7) \rightarrow \varepsilon\} \end{aligned}$$

Definíciu gramatiky  $G$  dokončíme položením  $T = \{a, b\}$  a  $\sigma = (0, 0)$ .

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

$\square$ : Dokážeme, že každá vetná forma odvoditeľná v  $G$  je jedného z nasledujúcich tvarov:

- (a)  $u(i, j)$ , kde  $u \in \{a, b\}^*$ ;  $\#_a(u) \equiv i \pmod{41}$  a  $\#_b(u) \equiv j \pmod{37}$
- (b)  $u$ , kde  $u \in \{a, b\}^*$ ;  $\#_a(u) \equiv 1 \pmod{41}$  a  $\#_b(u) \equiv 7 \pmod{37}$

Indukciou na dĺžku odvodenia.

- 1° Nulakrokovým odvodením vieme odvodiť iba vetnú formu  $\sigma = (0, 0)$ , ktorá je tvaru (a) pre  $i := 0, j := 0, u := \varepsilon$ .
- 2° Predpokladajme, že všetky vetné formy odvoditeľné na  $n$  krokov sú jedného z povolených tvarov. Majme ľubovoľnú vetnú formu  $w$  odvoditeľnú na  $n + 1$  krokov (formálne:  $(0, 0) \Rightarrow^{n+1} w$ ). Toto odvodenie sa dá rozdeliť na dve časti:

$$(0, 0) \Rightarrow^n v \Rightarrow w.$$

Vetná forma  $v$  je pritom odvoditeľná na  $n$  krokov, teda z indukčného predpokladu musí byť jedného z povolených tvarov. Ak by bola tvaru (b), nedal by sa z nej urobiť krok odvodenia, pretože by neobsahovala neterminál. Preto musí byť tvaru (a), teda pre nejaké  $u, i, j$  platí  $v = u(i, j)$ , pričom  $\#_a(u) \equiv i \pmod{41} \wedge \#_b(u) \equiv j \pmod{37}$ . Toto slovo má jediný neterminál  $(i, j)$ , pre ktorý existujú nanaajvyš 3 pravidlá:

$$(i) \quad (i, j) \rightarrow a(i \oplus_{41} 1, j)$$

---

<sup>1</sup>teda  $N = \{(i, j) \mid i \in \{0, 1, \dots, 40\}, j \in \{0, 1, \dots, 36\}\}$

(ii)  $(i, j) \rightarrow b(i, j \oplus_{37} 1)$

(iii) Ak  $i = 1$  a  $j = 7$ , tak máme ešte pravidlo  $(i, j) \rightarrow \varepsilon$ .

Ak sa v poslednom kroku odvodenia použilo pravidlo typu (i), potom  $w = ua(i \oplus_{41} 1, j)$ . Platí:

$$\begin{aligned}\#_a(ua) &\equiv \#_a(u) + 1 \equiv i + 1 \equiv i \oplus_{41} 1 \pmod{41} \\ \#_b(ua) &\equiv \#_b(u) \equiv j \pmod{37}\end{aligned}$$

To znamená, že aj slovo  $w = ua(i \oplus_{41} 1, j)$  je tvaru (a), pre  $u := ua, i := i \oplus_{41} 1, j := j$ .

Ak sa v poslednom kroku použilo pravidlo typu (ii), potom  $w = ub(i, j \oplus_{37} 1)$ . Toto slovo je tiež tvaru (a), pre  $u := ub, i := i, j := j \oplus_{37} 1$ .

Ak sa v poslednom kroku použilo pravidlo typu (iii), potom  $w = u$ . Pritom musí platiť  $i = 1$  a  $j = 7$ . Potom ale

$$\begin{aligned}\#_a(w) &\equiv \#_a(u) \equiv i \equiv 1 \pmod{41} \\ \#_b(w) &\equiv \#_b(u) \equiv j \equiv 7 \pmod{37},\end{aligned}$$

teda vetná forma  $w$  je tvaru (b).

Ukázali sme, že každá vetná forma odvoditeľná v našej gramatike je buď tvaru (a), alebo tvaru (b). Keďže všetky vetné formy tvaru (a) obsahujú neterminál, v jazyku  $L(G)$  budú iba slová tvaru (b), čo sú práve slová z jazyka  $L$ .

$\square$ : Najprv ukážeme, že pre každé slovo  $w \in \{a, b\}^*$  vieme v gramatike  $G$  vygenerovať vetnú formu  $w(q, r)$ , kde  $q = \#_a(w) \pmod{41}$  a  $r = \#_b(w) \pmod{37}$ . Indukciou vzhľadom na dĺžku slova  $w$ :

1° Pre  $w = \varepsilon$  vieme triviálne vygenerovať  $(0, 0)$ .

2° Predpokladajme, že pre všetky slová dĺžky  $n$  naše tvrdenie platí. Nech  $w \in \{a, b\}^{n+1}$ . Slovo  $w$  sa dá zapísať ako  $uc$ , kde  $u \in \{a, b\}^n$  a  $c \in \{a, b\}$ . Číslo  $\#_a(u) \pmod{41}$  označme  $s$  a číslo  $\#_b(u) \pmod{37}$  označme  $t$ . Pre slovo  $u$  platí indukčný predpoklad, teda  $(0, 0) \Rightarrow^* u(s, t)$ . V prípade, že  $c = a$  môžeme ďalej použiť pravidlo  $(s, t) \rightarrow a(s \oplus_{41} 1, t)$ , čím dostávame

$$(0, 0) \Rightarrow^* u(s, t) \Rightarrow ua(s \oplus_{41} 1, t) = w(q, r).$$

Podobne, ak  $c = b$ , použijeme pravidlo  $(s, t) \rightarrow b(s, t \oplus_{37} 1)$ , čím dostaneme

$$(0, 0) \Rightarrow^* u(s, t) \Rightarrow ub(s, t \oplus_{37} 1) = w(q, r).$$

Teraz už ľahko dokážeme, že  $L(G) \supseteq L$ . Majme ľubovoľné  $w \in L$ . Z definície jazyka  $L$  máme  $\#_a(w) \equiv 1 \pmod{41}$  a  $\#_b(w) \equiv 7 \pmod{37}$ . Platí teda

$$(0, 0) \Rightarrow^* w(1, 7).$$

Použitím pravidla  $(1, 7) \rightarrow \varepsilon$  potom dostávame:

$$(0, 0) \Rightarrow^* w(1, 7) \Rightarrow w,$$

teda  $w \in L(G)$ . □

**Úloha 2.** Nájdite čo možno najjednoduchší „množinový“ zápis pre jazyk generovaný bezkontextovou gramatikou  $G = (N, T, P, \sigma)$ , kde  $N = \{\sigma, \alpha, \beta, \gamma\}$ ,  $T = \{a, b, c\}$  a

$$P = \{\sigma \rightarrow \alpha c \beta \\ \alpha \rightarrow ababa \mid \gamma \\ \beta \rightarrow \beta a \beta b \beta \mid \beta b \beta a \beta \mid \varepsilon \\ \gamma \rightarrow a \gamma \mid b \gamma \mid \varepsilon\}.$$

Svoje tvrdenie *poriadne* dokážte.

*Riešenie.* Na začiatok si urobme nejakú intuíciu o gramatike  $G$  a jazyku, ktorý generuje.

Keď sa pozrieme, aké pravidlá máme z neterminálu  $\gamma$ , môžeme si všimnúť, že sa z neho dajú vygenerovať všetky slová z  $\{a, b\}^*$  (a nič iné).

Z terminálu  $\alpha$  máme pravidlo  $\alpha \rightarrow \gamma$ , teda z  $\alpha$  sa tiež určite dajú vygenerovať všetky slová z  $\{a, b\}^*$ . Okrem toho máme ešte pravidlo  $\alpha \rightarrow ababa$ , ktoré by nám mohlo pomôcť generovať z  $\alpha$  aj nejaké ďalšie slová. V skutočnosti však ani s pomocou tohto pravidla nevieme z  $\alpha$  vygenerovať žiaden iný terminál ako  $a$  a  $b$ , takže žiadne slová mimo  $\{a, b\}^*$  z  $\alpha$  nevygenerujeme.

Pri pohľade na pravidlá z terminálu  $\beta$  si môžeme všimnúť, že z  $\beta$  sa tiež dajú generovať iba slová z terminálov  $a$  a  $b$ , avšak nie nutne všetky. Ďalej si môžeme všimnúť, že počet  $a$ -čok a  $b$ -čok bude v každom slove vygenerovanom z  $\beta$  rovnaký (lebo vždy, keď budeme generovať nejaké  $a$ -čko, vygenerujeme spolu s ním aj jedno  $b$ -čko a naopak). To, či sa z  $\beta$  dajú vygenerovať všetky slová z  $\{a, b\}$  obsahujúce rovnako veľa  $a$ -čok a  $b$ -čok je už ťažšia otázka, ale neskôr ukážeme, že áno.

Jedno odvodenie v gramatike  $G$  teda bude vyzeráť zhruba takto:

1. Prvý krok odvodenia bude  $\sigma \Rightarrow \alpha c \beta$ .
2. Z  $\alpha$  sa vygeneruje nejaké slovo z  $\{a, b\}^*$ .
3. Z  $\beta$  sa vygeneruje nejaké slovo z  $\{a, b\}^*$ , ktoré má rovnako veľa  $a$ -čok a  $b$ -čok.

Časti 2 a 3 sa nemusia nutne diať v takomto poradí, môžu sa ľubovoľne prelínať.

Vyzerá to teda, že gramatika  $G$  generuje jazyk

$$L = \{ucv \mid u, v \in \{a, b\}^* \wedge \#_a(v) = \#_b(v)\}.$$

Podme teraz dokázať, že naozaj  $L(G) = L$ . Rovnosť dokážeme dvoma inklúziami.

$\subseteq$ : Dokážeme, že každá vetná forma odvoditeľná v  $G$  je jedného z nasledujúcich tvarov:

- (a)  $\sigma$
- (b)  $xcy$ , kde  $x \in \{a, b, \alpha, \gamma\}^*$  a  $y \in \{a, b, \beta\}^*$ , pričom  $\#_a(y) = \#_b(y)$ .

Indukciou vzhľadom na dĺžku odvodenia.

- 1° Nulakrokovým odvodením vieme odvodiť iba formu  $\sigma$ , ktorá je tvaru (a).
- 2° Predpokladajme, že všetky vetné formy odvoditeľné na  $n$  krokov sú jedného z povolených tvarov. Nech  $w$  je taká vetná forma, že  $\sigma \Rightarrow^{n+1} w$ .

Potom existuje taká vetná forma  $v$ , že

$$\sigma \Rightarrow^n v \Rightarrow w.$$

Vetná forma  $v$  je pritom (z indukčného predpokladu) jedného z povolených tvarov.

Ak je tvaru (a) (teda  $v = \sigma$ ), potom pri kroku odvodenia  $v \Rightarrow w$  vieme použiť iba pravidlo  $\sigma \rightarrow \alpha c \beta$ , teda musí platiť  $w = \alpha c \beta$ , čo je tvaru (b).

Ak je forma  $v$  tvaru (b), potom platí  $v = xcy$  pre nejaké  $x \in \{a, b, \alpha, \gamma\}^*$  a  $y \in \{a, b, \beta\}^*$ , pričom  $\#_a(y) = \#_b(y)$ . Keďže slovo  $v$  neobsahuje neterminál  $\sigma$ , pri kroku odvodenia  $v \Rightarrow w$  sa muselo použiť niektoré z pravidiel

- (i)  $\alpha \rightarrow ababa$
- (ii)  $\alpha \rightarrow \gamma$
- (iii)  $\beta \rightarrow \beta a \beta b \beta$
- (iv)  $\beta \rightarrow \beta b \beta a \beta$
- (v)  $\beta \rightarrow \varepsilon$
- (vi)  $\gamma \rightarrow a\gamma$
- (vii)  $\gamma \rightarrow b\gamma$
- (viii)  $\gamma \rightarrow \varepsilon$

Ak sa v poslednom kroku odvodenia použilo niektoré z pravidiel (i) alebo (ii), prepísala sa tým nejaká  $\alpha$  zo slova  $v$ . Táto  $\alpha$  sa pritom musí nachádzať v slove  $x$  (keďže  $\alpha \neq c$  a slovo  $y$  žiadnu  $\alpha$  neobsahuje). Preto sa slovo  $x$  dá zapísať ako  $x = u\alpha z$ , kde  $u, z \in \{a, b, \alpha, \gamma\}^*$  (a  $\alpha$  medzi  $u$  a  $z$  je tá, ktorá sa prepíše pri kroku  $v \Rightarrow w$ ). Ak bolo použité pravidlo (i), posledný krok odvodenia vyzeral takto:

$$v = u\alpha zcy \Rightarrow uababazcy = w.$$

V takom prípade je  $w$  tvaru (b) pre  $x := uababaz$  a  $y := y$ . Ak bolo použité pravidlo (ii), posledný krok odvodenia bol

$$v = u\alpha zcy \Rightarrow u\gamma zcy = w.$$

V takom prípade je  $w$  tvaru (b) pre  $x := u\gamma z$  a  $y := y$ .

Ak sa v poslednom kroku použilo nejaké z pravidiel (iii), (iv) alebo (v), slovo  $v$  musí byť tvaru  $v = xcu\beta z$  pre nejaké  $u, z \in \{a, b, \beta\}^*$ , pričom  $\#_a(uz) = \#_b(uz)$  a v kroku odvodenia  $v \Rightarrow w$  sa prepíše  $\beta$ , ktorá je medzi  $u$  a  $z$ . Ak bolo použité pravidlo (iii), posledný krok bol

$$v = xcu\beta z \Rightarrow xcu\beta a\beta b\beta z = w.$$

Vtedy  $w$  je tvaru (b) pre  $x := x$  a  $y := u\beta a\beta b\beta z$  (ľahko overíme, že podmienka na rovnaký počet  $a$ -čok a  $b$ -čok v  $y$  platí). Ak bolo použité pravidlo (iv), posledný krok bol

$$v = xcu\beta z \Rightarrow xcu\beta b\beta a\beta z = w$$

a  $w$  je tvaru (b) pre  $x := x$  a  $y := u\beta b\beta a\beta z$ . Ak bolo použité (v), posledný krok bol

$$v = xcu\beta z \Rightarrow xcuz = w$$

a  $w$  je tvaru (b) pre  $x := x$  a  $y := uz$ .

Nakoniec, ak sa v poslednom kroku použilo niektoré z pravidiel (vi), (vii), (viii), slovo  $v$  je tvaru  $u\gamma zcy$  pre  $u, z \in \{a, b, \alpha, \gamma\}^*$  a v kroku výpočtu  $v \Rightarrow w$  sa prepísala  $\gamma$  medzi  $u$  a  $z$ . Ak bolo v poslednom kroku použité (vi), tento krok vyzeral ako

$$v = u\gamma zcy \Rightarrow ua\gamma zcy = w$$

a  $w$  je tvaru (b) pre  $x := ua\gamma z$  a  $y := y$ . Ak bolo použité (vii), posledný krok bol

$$v = u\gamma zcy \Rightarrow ub\gamma zcy = w$$

a  $w$  je tvaru (b) pre  $x := ub\gamma z$  a  $y := y$ . Ak bolo použité (viii), posledný krok bol

$$v = u\gamma zcy \Rightarrow uzcy = w$$

a  $w$  je tvaru (b) pre  $x := uz$  a  $y := y$ .

Ukázali sme, že každá vetná forma odvoditeľná v našej gramatike je buď tvaru (a), alebo tvaru (b). Nech  $w$  je ľubovoľné slovo z  $L(G)$ . Nemôže byť tvaru (a), lebo  $\sigma$  nie je terminál. Slovo  $w$  teda musí byť tvaru (b), teda  $w = xcy$ , kde  $x \in \{a, b, \alpha, \gamma\}^*$ ,  $y \in \{a, b, \beta\}^*$  a  $\#_a(y) = \#_b(y)$ . Keďže  $w$  neobsahuje neterminály, bude dokonca platiť  $x \in \{a, b\}^*$  a  $y \in \{a, b\}^*$ . To ale znamená, že  $w \in L$ .

$\square$ : Najprv ukážeme, že pre každé slovo  $w \in \{a, b\}^*$  vieme v gramatike  $G$  vygenerovať vetnú formu  $w\gamma c\beta$ . Indukciou vzhľadom na dĺžku slova  $w$ :

1° Pre  $w = \varepsilon$  máme

$$\sigma \Rightarrow \alpha c\beta \Rightarrow \gamma c\beta.$$

2° Predpokladajme, že pre všetky slová dĺžky  $n$  naše tvrdenie platí. Nech  $w \in \{a, b\}^{n+1}$ . Potom  $w = ud$  pre nejaké  $u \in \{a, b\}^n$  a  $d \in \{a, b\}$ . Z indukčného predpokladu máme

$$\sigma \Rightarrow^* u\gamma c\beta.$$

Vďaka pravidlám  $\gamma \rightarrow a$  a  $\gamma \rightarrow b$  určite platí aj

$$u\gamma c\beta \Rightarrow ud\gamma c\beta,$$

teda  $\sigma \Rightarrow^* ud\gamma c\beta = w\gamma c\beta$ .

Ďalej, pre každé slovo  $w \in \{a, b\}^*$  vieme vygenerovať aj  $wc\beta$ : vďaka  $\gamma \rightarrow \varepsilon$  platí  $w\gamma c\beta \Rightarrow wc\beta$ .

Ostáva nám ešte posledná časť dôkazu: ukázať, že z  $\beta$  sa dajú vygenerovať všetky slová z  $\{a, b\}^*$ , ktoré majú rovnako veľa  $a$ -čok a  $b$ -čok.

**Definícia 2.** Majme ľubovoľné slovo  $w = a_1a_2 \dots a_k$ . Slová  $\overleftarrow{w}$ ,  $\overleftarrow{w}$  a  $\overrightarrow{w}$  definujeme ako:

$$\begin{aligned}\overleftarrow{w} &= \beta a_1 \beta a_2 \beta a_3 \beta \dots \beta a_k \beta \\ \overleftarrow{w} &= \beta a_1 \beta a_2 \beta a_3 \beta \dots \beta a_k \\ \overrightarrow{w} &= a_1 \beta a_2 \beta a_3 \beta \dots \beta a_k \beta.\end{aligned}$$

Teraz ukážeme, že pre každé  $w \in \{a, b\}^*$  také, že  $\#_a(w) = \#_b(w)$  platí  $\beta \Rightarrow^* \overleftarrow{w}$ . Indukciou vzhľadom na dĺžku slova  $w$ :

1° Pre  $w = \varepsilon$  máme  $\overleftarrow{w} = \beta$  a  $\beta \Rightarrow^* \beta$  platí triviálne.

2° Nech tvrdenie platí pre všetky slová dlhé najviac  $n$  a slovo  $w$  má dĺžku  $n+1$ . V slove  $w$  sa určite niekde vyskytujú písmená  $a$  a  $b$  bezprostredne vedľa seba (ak  $w$  začína písmenom  $a$ , potom prvý výskyt písmena  $b$  má tesne pred sebou  $a$ . Ak  $w$  začína písmenom  $b$ , tak prvé  $a$ -čko má tesne pred sebou  $b$ ). To znamená, že je tvaru  $w = xaby$ , alebo  $xbay$  pre nejaké  $x, y \in \{a, b\}^*$ .

Potom

$$\overleftarrow{w} = \overleftarrow{x} \beta a \beta b \beta \overrightarrow{y}$$

alebo

$$\overleftarrow{w} = \overleftarrow{x} \beta b \beta a \beta \overrightarrow{y}.$$

Vďaka pravidlám  $\beta \rightarrow \beta a \beta b \beta$  a  $\beta \rightarrow \beta b \beta a \beta$  platí

$$\overleftarrow{xy} = \overleftarrow{x} \beta \overleftarrow{y} \Rightarrow \overleftarrow{x} \beta a \beta b \beta \overleftarrow{y}$$

a

$$\overleftarrow{xy} = \overleftarrow{x} \beta \overleftarrow{y} \Rightarrow \overleftarrow{x} \beta b \beta a \beta \overleftarrow{y},$$

teda určite platí

$$\overleftarrow{xy} \Rightarrow \overleftarrow{w}.$$

Slovo  $xy$  je pritom kratšie než  $w$  a platí  $xy \in \{a, b\}^*$  aj

$$\#_a(xy) = \#_a(w) - 1 = \#_b(w) - 1 = \#_b(xy),$$

teda naň môžeme použiť indukčný predpoklad. Dostávame:

$$\beta \Rightarrow^* \overleftarrow{xy} \Rightarrow \overleftarrow{w}.$$

Vezmime si teraz ľubovoľné slovo  $w \in L$ . To sa dá zapísať ako  $ucv$ , kde  $u, v \in \{a, b\}^*$  a  $\#_a(v) = \#_b(v)$ . Už sme ukázali, že  $\sigma \Rightarrow^* uc\beta$  aj že  $\beta \Rightarrow^* \overleftarrow{v}$ . Potom ale aj  $uc\beta \Rightarrow^* uc\overleftarrow{v}$ . Nakoniec, vďaka pravidlu  $\beta \rightarrow \varepsilon$  zjavne platí  $\overleftarrow{v} \Rightarrow^* v$ . Spojením všetkého dokopy dostávame

$$\sigma \Rightarrow^* uc\beta \Rightarrow^* uc\overleftarrow{v} \Rightarrow^* ucv = w.$$

□