

# Cvičenie 1 - Formálne jazyky a gramatiky

Ing. Viliam Hromada, PhD.

C-510  
Ústav informatiky a matematiky  
FEI STU

`viliam.hromada@stuba.sk`

# Úloha č. 1

Nech abeceda  $A = \{a, b, c\}$ , nech reťazce nad touto abecedou sú  $x = aa, y = \varepsilon, z = abc$ .

1. Napíšte obrátený reťazec  $x^R, y^R, z^R$  a ich dĺžky
2. Napíšte výsledok zreťazenia  $xx, xy, xz, yx, yy, yz, zx, zy, zz$  a ich dĺžky
3. Napíšte  $x^i, y^i, z^i$ , pre  $i \in \{0, 1, 2, 3\}$  a ich dĺžky
4. Čo je  $A^*$  ? Čo je  $A^+$ ?



## Úloha č. 1 - odpovede

1.  $x^R = aa, y^R = \varepsilon, z^R = cba$ .  $|x^R| = |x| = 2, |y^R| = |y| = 0, |z^R| = |z| = 3$ .
2.  $xx = aaaa, xy = aa, xz = aaabc, yx = aa, yy = \varepsilon, yz = abc, zx = abcaa, zy = abc, zz = abcabc$ .

Zrejme  $|uv| = |u| + |v|$ , t.j.:

$$|xx| = 2|x| = 4, |xy| = |x| + |y| = 2, |xz| = |x| + |z| = 2 + 3 = 5, |yx| = 2, |yy| = |y| + |y| = 0, |yz| = 3, |zx| = 5, |zy| = 3, |zz| = 6.$$

3.  $x^0 = \varepsilon, x^1 = aa, x^2 = aaaa, x^3 = aaaaaa, y^0 = \varepsilon, y^1 = \varepsilon, y^2 = \varepsilon, y^3 = \varepsilon, z^0 = \varepsilon, z^1 = abc, z^2 = abcabc, z^3 = abcabcabc$

Zrejme  $|u^i| = i|u|$ , t.j.  $|x^0| = 0|x| = 0, |x^2| = 2|x| = 4, |x^3| = 3|x| = 6, \dots, |y^2| = 2|y| = 0, \dots, |z^3| = 3|z| = 9$

4.  $A^* = \{\varepsilon, a, b, c, aa, ab, ac, ba, bb, bc, ca, cb, cc, aaa, aab, \dots\}$   
 $A^+ = \{a, b, c, aa, ab, ac, ba, bb, bc, ca, cb, cc, aaa, aab, \dots\} = A^* \setminus \{\varepsilon\}$ .



## Úloha č. 2

Uvažujme abecedu  $A = \{a, b, c\}$  a nasledovné jazyky nad touto abecedou:

- $L_1 = \{c\}$
- $L_2 = \{aaa, aba, aab, abb, baa, bab, bba, bbb\}$
- $L_3 = \{aw \mid w \in \{a, b, c\}^*\}$
- $L_4 = \{wcw^R \mid w \in \{a, b\}^*\}$

1. Ktoré jazyky  $L_1, L_2, L_3, L_4$  sú konečné a ktoré sú nekonečné? Popíšte slovne jazyky  $L_3, L_4$
2. Uved'te jazyk, ktorý vznikne ako zret'azenie  $L_1L_1, L_1L_2, L_2L_1, L_2L_2$
3. Popíšte jazyk  $L_1^2, L_1^3, L_1^*, L_1^+, L_2^*, L_3^C, L_4^2$
4. Popíšte jazyk  $L_2 \cap L_3, L_2 \cap L_4, L_3 \cap L_4, L_4 \setminus L_1, L_1 \cup L_1^C$



## Úloha č. 2 - odpovede

1.  $L_1, L_2$  sú konečné jazyky, pretože  $|L_1| = 1, |L_2| = 8$ . Jazyky  $L_3, L_4$  sú nekonečné.

$L_3 = \{a, aa, ab, ac, aaa, aab, aac, aba, abb, abc, aca, acb, acc, aaaa, \dots\}$  t.j. všetky reťazce začínajúce  $a$

$L_4 =$

$\{c, aca, bcb, aacaa, abcba, bacab, bbcbb, aaacaaa, aabcbaa, abacaba, \dots\}$  t.j. všetky **palindrómy** nepárnej dĺžky, ktoré majú uprostred symbol  $c$  a ostatné symboly môžu byť len  $a$  a  $b$ .

2.  $L_1L_1 = \{cc\}$ ,  $L_1L_2 = \{caaa, caba, caab, cabb, cbaa, cbab, cbba, cbbb\}$ ,

$L_2L_1 = \{aaac, abac, aabc, abbc, baac, babc, bbac, bbbc\}$ ,

$L_2L_2 =$

$\{aaaaaa, aaaaba, aaaaab, \dots, aaabbb, abaaaa, abaaba, abaaab, \dots, ababbb, \dots, bbbaaa, \dots, bbbbbb\}$



## Úloha č. 2 - odpovede

3.  $L_1^2 = L_1 L_1 = \{cc\}$ ,  $L_1^3 = L_1 L_1^2 = \{ccc\}$ ,  $L_1^* = \{c\}^* = \{\varepsilon, c, cc, ccc, cccc, \dots\} = \{c^i \mid i \in \{0, 1, 2, 3, \dots\}\}$ ,  $L_1^+ = \{c, cc, ccc, cccc, \dots\} = \{c^i \mid i \in \{1, 2, 3, \dots\}\}$   
 $L_2^* = \{\varepsilon, aaa, aba, \dots, bbb, aaaaaa, \dots, aaaaaaaaaa, \dots, abbbaabbabbbb, \dots\}$

$L_3^C = A^* \setminus L_3$ , t.j. všetky reťazce **nezačínajúce**  $a$ -čkom, t.j.

$$L_3^C = \{\varepsilon\} \cup \{bw \mid w \in \{a, b, c\}^*\} \cup \{cw \mid w \in \{a, b, c\}^*\} =$$

$$L_3^C = \{\varepsilon, b, c, ba, bb, bc, ca, cb, cc, baa, \dots\}$$

$$L_4^2 = L_4 L_4 = \{cc, caca, acac, cbcb, bcbc, acaaca, acabcb, \dots\} = \{xy \mid x = wcw^R, y = ucu^R, w \in \{a, b\}^*, u \in \{a, b\}^*\}.$$



## Úloha č. 2 - odpovede

4.  $L_2 \cap L_3 = \{aaa, aba, aab, abb\}$

$$L_2 \cap L_4 = \emptyset$$

$L_3 \cap L_4 = \{aca, aacaa, abcba, aaacaaa, abacaba, abbcbbba, \dots\} = \{wcw^R \mid w = ax, x \in \{a, b\}^*\}$ , t.j.  $L_3 \cap L_4$  sú palindrómy nepárnej dĺžky majúce uprostred  $c$ , pričom ostatné symboly môžu byť len  $a$  a  $b$  a navyše začínajú  $a$ -čkom.

$L_4 \setminus L_1 = \{aca, bcb, aacaa, \dots\}$   $L_3 \cap L_4$  sú palindrómy nepárnej dĺžky aspoň 3 majúce uprostred  $c$ , pričom ostatné symboly môžu byť len  $a$  a  $b$

$$L_1 \cup L_1^C = A^*.$$



## Úloha č. 3

Je daná gramatika  $G = (N, T, P, S)$ , kde  $N = \{S, A\}$ ,  $T = \{0, 1\}$ , počiatočný neterminál je  $S$ , s pravidlami  $P$ :

$$S \rightarrow 0A$$

$$A \rightarrow 0A \mid 1A \mid 0 \mid 1$$

- určte typ gramatiky (regulárna, bezkontextová, kontextová, frázová)
- nájdite ododenie slov 0101, 0111, 1000 v danej gramatike.
- Určte, aký jazyk gramatika generuje (slovne, formálnym zápisom).





## Úloha č. 3 - odpovede

a) Gramatika je **regulárna** (technicky je aj bezkontextová, kontextová a frázová)

b)  $S \Rightarrow 0A \Rightarrow 01A \Rightarrow 010A \Rightarrow 0101$ , teda  $S \Rightarrow^* 0101$ .

b)  $S \Rightarrow 0A \Rightarrow 01A \Rightarrow 011A \Rightarrow 0111$ , teda  $S \Rightarrow^* 0111$ .

b)  $S \Rightarrow 0A \Rightarrow \dots \Rightarrow 1000$  sa zjavne nedá, teda  $S \not\Rightarrow^* 1000$

c) Vidíme, že gramatika generuje také reťazce z núl a jednotiek, ktoré začínajú nulou a sú dĺžky aspoň 2. Preto by sme mohli písať, že jazyk generovaný gramatikou  $G$  je:

$$L(G) = \{0w \mid w \in \{0, 1\}^+\}$$

teda, že na začiatku obsahujú 0 a za ňou je reťazec z núl a jednotiek dĺžky aspoň 1, teda dokopy tvoria reťazce z 0 a 1 začínajúce nulou dĺžky aspoň 2.

## Úloha č. 4

Je daná gramatika popisujúca syntax podmieneného príkazu  $G = (N, T, P, S)$  s pravidlami  $P$ :

1.  $\langle \text{príkaz} \rangle \rightarrow \mathbf{p1}$
2.  $\langle \text{príkaz} \rangle \rightarrow \mathbf{p2}$
3.  $\langle \text{príkaz} \rangle \rightarrow \mathbf{if} \langle \text{podmienka} \rangle \mathbf{then} \langle \text{príkaz} \rangle \langle \text{else časť} \rangle$
4.  $\langle \text{else časť} \rangle \rightarrow \mathbf{else} \langle \text{príkaz} \rangle$
5.  $\langle \text{else časť} \rangle \rightarrow \varepsilon$
6.  $\langle \text{podmienka} \rangle \rightarrow \mathbf{cond}$

Neterminály  $N = \{ \langle \text{príkaz} \rangle, \langle \text{else časť} \rangle, \langle \text{podmienka} \rangle \}$  terminály

$T = \{ \mathbf{p1}, \mathbf{p2}, \mathbf{if}, \mathbf{then}, \mathbf{else}, \mathbf{cond} \}$ ,  $S = \langle \text{príkaz} \rangle$



## Úloha č. 4

Zistite, či nasledovné "fragmenty kódu" spĺňajú uvedenú syntax (t.j. či **majú v gramatike odvodenie**):

1. **p1**
2. **p1 p2**
3. **if cond then p1**
4. **if cond then p1 else**
5. **if cond then p1 else if cond then p1 else p2**



## Úloha č. 4 - odpovede

Riešenie:

1.  $\langle \text{príkaz} \rangle \Rightarrow \mathbf{p1}$
2.  $\langle \text{príkaz} \rangle \not\Rightarrow^* \mathbf{p1 p2}$
3.  $\langle \text{príkaz} \rangle \Rightarrow \mathbf{if \langle \text{podmienka} \rangle \text{ then } \langle \text{príkaz} \rangle \langle \text{else časť} \rangle} \Rightarrow$   
 $\mathbf{if \ cond \ then \ } \langle \text{príkaz} \rangle \langle \text{else časť} \rangle \Rightarrow \mathbf{if \ cond \ then \ p1 \ } \langle \text{else časť} \rangle \Rightarrow$   
 $\mathbf{if \ cond \ then \ p1}$
4.  $\langle \text{príkaz} \rangle \Rightarrow \mathbf{if \langle \text{podmienka} \rangle \text{ then } \langle \text{príkaz} \rangle \langle \text{else časť} \rangle} \Rightarrow$   
 $\mathbf{if \ cond \ then \ } \langle \text{príkaz} \rangle \langle \text{else časť} \rangle \Rightarrow \mathbf{if \ cond \ then \ p1 \ } \langle \text{else časť} \rangle \Rightarrow$   
 $\mathbf{if \ cond \ then \ p1 \ else \ } \langle \text{príkaz} \rangle \not\Rightarrow \mathbf{if \ cond \ then \ p1 \ else}$   
(pretože z neterminálu  $\langle \text{príkaz} \rangle$  nie je možné odvodiť len  $\epsilon$ )

## Úloha č. 4 - odpovede

Riešenie pre **if cond then p1 else if cond then p1 else p2**:

<príkaz> ⇒ **if**<podmienka>**then**<príkaz><else časť> ⇒

**if cond then** <príkaz><else časť> ⇒ **if cond then p1** <else časť> ⇒

**if cond then p1 else**<príkaz> ⇒

**if cond then p1 else if**<podmienka>**then**<príkaz><else časť> ⇒

**if cond then p1 else if cond then**<príkaz><else časť> ⇒

**if cond then p1 else if cond then p1** <else časť> ⇒

**if cond then p1 else if cond then p1 else** <príkaz> ⇒

**if cond then p1 else if cond then p1 else p2**



## Úloha č. 5

Uvažujte abecedu  $A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, -\}$ . Navrhnite gramatiku, ktorá generuje celočíselné konštanty (t.j. čísla z množiny  $\mathbb{Z}$ ), t.j. reťazce:

- $0, 1, 2, 3, 4, \dots, 9, 10, 11, 12, 13, \dots$
- $-1, -2, -3, -4, \dots - 9, -10, \dots$



## Úloha č. 5 - riešenie

Určite terminálne symboly gramatiky budú vlastne abeceda, t.j.

$T = A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, -\}$ . Navrhne sústavu pravidiel, ktoré tieto čísla generujú:

- $S \rightarrow -A|0|1|2|3|4|5|6|7|8|9|1B|2B|3B|4B|5B|6B|7B|8B|9B$
- $A \rightarrow 1|2|3|4|5|6|7|8|9|1B|2B|3B|4B|5B|6B|7B|8B|9B$
- $B \rightarrow 0|1|2|3|4|5|6|7|8|9|1B|2B|3B|4B|5B|6B|7B|8B|9B|0B$

Teda neterminály  $N = \{S, A, B\}$  a  $S$  je počiatočný neterminál. Všimnite si, že naša gramatika:

- Negeneruje také čísla, ktoré by začínali nulami - s výnimkou len nuly samotnej.
- Negeneruje napr.  $-0$ , čiže nula je vždy bez znamienka.
- Je regulárna!



## Úloha č. 6

Nájdite gramatiku, ktorá generuje také reťazce zo symbolov  $\{a, b\}$  (t.j. nad abecedou  $A = \{a, b\}$ ), v ktorých je rovnaký počet symbolov  $a$  a  $b$ , t.j. gramatiku, ktorá generuje jazyk:

$$L = \{w \in \{a, b\}^* \mid \#_a(w) = \#_b(w)\},$$

kde  $\#_a(w)$  označuje počet symbolov  $a$  v reťazci  $w$  (analogicky  $\#_b(w)$ )





## Úloha č. 6 - zamyslenie

Aké reťazce patria do jazyka  $L$ ?

- $\{\varepsilon\}$ , lebo  $\#_a(\varepsilon) = \#_b(\varepsilon) = 0$
- $\{ab, ba\}$ , lebo pre ne  $\#_a(ab) = \#_b(ab) = 1$ , resp.  $\#_a(ba) = \#_b(ba) = 1$
- $\{aabb, abab, abba, baba, baab, bbaa\}$ , lebo sa v nich  $a$  aj  $b$  vyskytuje 2krát
- ...

Aké reťazce nepatria do jazyka  $L$ ?

- $\{a, aab, aba, baa, aa, aaab, aaba, \dots\}$  lebo je v nich viac  $a$  než  $b$
- $\{b, bba, bab, abb, bb, bbba, bbab, \dots\}$  lebo je v nich viac  $b$  než  $a$



## Úloha č. 6 - pokus o riešenie

Prvý pokus o riešenie - nech gramatika  $G$  by bola:

- $S \rightarrow aS \mid bS \mid \varepsilon$

Táto gramatika určite dokáže generovať také reťazce, v ktorých je počet  $a$  a  $b$  rovnaký. Totižto, táto gramatika generuje **všetky reťazce** z písmen  $\{a, b\}$  a určite platí, že  $L \subseteq L(G)$ .

**Problém je**, že generuje aj reťazce **navyše**, napr.  $a$  alebo  $aaa$ , teda také, ktoré **nepatria do**  $L$ , lebo v nich nie je rovnaký počet  $a$  a  $b$ . Teda  $L(G) \not\subseteq L$ . Táto gramatika **nie je správna**.



## Úloha č. 6 - pokus o riešenie

Druhý pokus o riešenie - nech gramatika  $G$  by bola:

1.  $S \rightarrow abS$
2.  $S \rightarrow \varepsilon$

Táto gramatika funguje tak, že každá aplikácia pravidiel č. 1 vyrobí vo výslednom reťazci časť  $ab$ , a teda určite **všetko**, čo  $G$  odvodí má tú vlastnosť, že počet  $a$  a  $b$  je rovnaký, teda  $L(G) \subseteq L$ .

Problém je, že **nie každý reťazec** s rovnakým počtom  $a$  a  $b$  má v danej gramatike odvodenie, napr.  $abba$ , sa v nej odvodit' nedá! Preto  $L \not\subseteq L(G)$ . Táto gramatika **nie je správna**.



## Úloha č. 6 - pokus o riešenie

Tretí pokus o riešenie - nech gramatika  $G$  by bola:

- $S \rightarrow SaSbS \mid SbSaS \mid \varepsilon$

Táto gramatika má nasledovnú logiku. Uvažujme reťazec s rovnakým počtom  $a$  a  $b$ . Potom v ňom nastane jedna z 2 situácií:

- Alebo v ňom **musia existovať**, také  $a$  a  $b$ , že  $a$  je pred  $b$  a navyše platí, že aj v podreťazci pred  $a$  je rovnaký počet  $a, b$ ; v podreťazci medzi  $a, b$  je rovnaký počet  $a, b$ ; v podreťazci za  $b$  je rovnaký počet  $a, b$ .

*abba***ab***bba*

- Alebo v ňom **musia existovať**, také  $b$  a  $a$ , že  $b$  je pred  $a$  a navyše platí, že aj v podreťazci pred  $b$  je rovnaký počet  $a, b$ ; v podreťazci medzi  $b, a$  je rovnaký počet  $a, b$ ; v podreťazci za  $a$  je rovnaký počet  $a, b$ .

*babab***a**



## Úloha č. 6 - pokus o riešenie

Pre gramatiku  $G$

- $S \rightarrow SaSbS \mid SbSaS \mid \varepsilon$

teda platí:

- Určite všetko čo generuje sú reťazce, ktoré majú rovnaký počet  $a$  a  $b$ , teda  $L(G) \subseteq L$ .
- Nebudeme to dokazovať, ale platí aj, že **každý** reťazec s rovnakým počtom  $a$  a  $b$  má v gramatike odvodenie, teda  $L \subseteq L(G)$ .
- A teda  $L = L(G)$ , čiže **naša gramatika je správna**.
- Naša gramatika je **bezkontextová**.

Pre úplnosť v našej gramatike  $G$ :  $N = \{S\}$ ,  $T = \{a, b\}$ .



## Úloha č. 7

**Príklad:** Nájdite gramatiku generujúcu všetky "správne" ozátvorkované výrazy, t.j.:

$$L = \{\varepsilon, (), (()), ()(), ((())), (())(), ()()(), \dots\}.$$

t.j. ku každej ľavej zátvorke existuje pravá zátvorka.

Musí platiť, že  $L(G) = L$ , t.j.:

- $L(G) \subseteq L$ , (gramatika negeneruje "**nič navyše**")
- $L \subseteq L(G)$ . (každé slovo z jazyka  $L$  sa dá v gramatike **odvodit'**)

Ak platia oba body, z toho logicky vyplýva, že  $L(G) = L$ .



**Príklad (pokr.)** Nech gramatika je napr.:  $G = (\{S, A\}, \{(, )\}, P, S)$ .

Pravidlá:

- $S \rightarrow AS$
- $S \rightarrow \varepsilon$
- $A \rightarrow (S)$ .

Je zrejmé, že gramatika generuje správne ozátvorkované výrazy, t.j.  $L(G) \subseteq L$  a **nič navyše**. Otázkou ale zostáva, či generuje **všetky** správne ozátvorkované výrazy, resp. či sa dá každý správne ozátvorkovaný výraz vygenerovať pomocou danej gramatiky (t.j. či  $L \subseteq L(G)$ ).



**Príklad (pokr.)** Dôkaz matematickou indukciou: Uvažujme "stupeň vnorenia"  $l$ :

- $l = 0$ :  $\varepsilon$
  - $l = 1$ :  $()$ ,  $()()$ , ...
  - $l = 2$ :  $(( ))$ ,  $(( ))()$ , ...
1. Pre  $l = 0$  je triviálne ukázať, že gramatika takéto výrazy generuje.
  2. Predpokladajme, že gramatika generuje výrazy pre  $l \leq k$ . Majme výraz  $w$  so stupňom vnorenia  $k + 1$ . Musíme dokázať, že aj takýto výraz gramatika vie vygenerovať.





**Príklad (pokr.)** Výraz  $w$  si rozdelíme na správne-ozátvorkované segmenty:  $w = w_1 w_2 \dots w_n$ , z ktorých aspoň 1 má stupeň vnorenia  $k + 1$ . Predpokladajme, že je to  $w_1$ . Slovo  $w$  vieme odvodiť:

$$S \Rightarrow AS \Rightarrow \dots \Rightarrow \underbrace{A \dots A}_n \Rightarrow (S) \underbrace{A \dots A}_{n-1} \Rightarrow \dots \Rightarrow w_1 \underbrace{A \dots A}_{n-1} \Rightarrow \dots \Rightarrow w_1 w_2 \dots w_n.$$

Odstránením "vonkajšej" zátvorky (keď sme použili pravidlo  $A \rightarrow (S)$ ) sme dostali výraz so stupňom vnorenia  $k$ , ktorý podľa indukčného predpokladu odvodiť vieme. Vieme teda odvodiť aj výraz so stupňom vnorenia  $k + 1$ .

Samozrejme, toto nemusí byť **jediná** gramatika, ktorá daný jazyk generuje.



## Úloha č. 8

Nájdite gramatiku pre jazyk  $L = \{a^n b^n c^n \mid n \in \{1, 2, 3, 4, \dots\}\}$ .



## Úloha č. 8 - zamyslenie

Reťazce patriace od jazyka  $L$ :

1.  $abc$
2.  $aabbcc$
3.  $aaabbbccc$
4.  $aaaabbbbccccc$

Teda reťazce majú **rovnaký počet**  $a, b, c$  a zároveň sú symboly **usporiadané**, t.j. najprv  $a$ , potom  $b$  a nakoniec  $c$ . Gramatika teda musí zabezpečiť, aby sa pri generovaní **sledoval počet symbolov**, aby bola zaručená ich rovnosť po vygenerovaní reťazca terminálov. Navyše musí zaručiť ich usporiadanie.



## Úloha č. 8 - pokus o riešenie

Skúske navrhnúť nasledovné pravidlá:

1.  $S \rightarrow aSBC$
2.  $S \rightarrow aBC$

Naša gramatika vie teraz generovať takúto vetnú formu:

$$S \Rightarrow^n a^n (BC)^n = a^n BC (BC)^{n-1}$$

Ak by sme v nej nahradili každé  $B$  terminálom  $b$  a  $C$  terminálom  $c$ , dostali by sme určite reťazec s rovnakým počtom  $a, b, c$ , avšak **nie usporiadaný!** Preto využijeme **možnosti kontextových gramatík**, ktoré dovoľujú meniť symboly vo vetných formách.



## Úloha č. 8 - pokus o riešenie

Doplňme pravidlo č. 3:

1.  $S \rightarrow aSBC$
2.  $S \rightarrow aBC$
3.  $CB \rightarrow BC$

Toto pravidlo umožní "preusporiadať" neterminály  $B, C$  vo vetnej forme tak, aby sme vedeli dostať:

$$S \Rightarrow^n a^n(BC)^n = a^n BC(BC)^{n-1} \Rightarrow^* a^n B^n C^n$$

Teraz môžeme pristúpiť k prepisovaniu neterminálov na terminály - ale spravíme to špeciálnym spôsobom.



## Úloha č. 8 - pokus o riešenie

1.  $S \rightarrow aSBC$
2.  $S \rightarrow aBC$
3.  $CB \rightarrow BC$
4.  $aB \rightarrow ab$
5.  $bB \rightarrow bb$
6.  $bC \rightarrow bc$
7.  $cC \rightarrow cc$

Teraz už vieme každú požadovanú deriváciu úspešne dokončiť:

$$S \Rightarrow^n a^n(BC)^n = a^n BC(BC)^{n-1} \Rightarrow^* a^n B^n C^n \Rightarrow^n a^n b^n C^n \Rightarrow^n a^n b^n c^n$$

Pre úplnosť v gramatike  $N = \{S, A, B, C\}$ ,  $T = \{a, b, c\}$ ,  $S$  je počiatočný neterminál



## Úloha č. 8 - upozornenie

POZOR!!! V predchádzajúcom slajde **nesmieme len pridať pravidlá**

$B \rightarrow b, C \rightarrow c$ , t.j.

1.  $S \rightarrow aSBC$

2.  $S \rightarrow aBC$

3.  $CB \rightarrow BC$

4.  $B \rightarrow b$

5.  $C \rightarrow c$

Lebo by sme vedeli urobiť napríklad deriváciu:

$$S \Rightarrow aSBC \Rightarrow aaBCBC \Rightarrow^* aabcabc$$

čo nie je reťazec patriaci do daného jazyka  $L$ .



K týmto úlohám uvádzam už len možné riešenia bez ďalšieho komentára...

**Nájdite gramatiky pre uvedené jazyky (abecedy sú  $A = \{a, b\}$  pre jazyky číslo 1,2,3,4,6,7,9) a  $A = \{a, b, c\}$  pre jazyky číslo 5,8**

- $L_1 = \{wba \mid w \in \{a, b\}^*\}$ ,
- $L_2 = \{abw \mid w \in \{a, b\}^*\}$ ,
- $L_3 = \{xaby \mid x, y \in \{a, b\}^*\}$ ,
- $L_4 = \{w \mid \#_a(w) \equiv \#_b(w) \pmod{2}, w \in \{a, b\}^*\}$ ,
- $L_5 = \{wcw^R \mid w \in \{a, b\}^*\}$ ,
- $L_6 = \{ww^R \mid w \in \{a, b\}^*\}$ ,
- $L_7 = \{a^n b^n \mid n \in \{1, 2, \dots\}\}$ ,
- $L_8 = \{a^n b^n c^k \mid n, k \in \{0, 1, 2, \dots\}\}$ ,
- $L_9 = \{ww \mid w \in \{a, b\}^*\}$ ,





$$L_1 = \{wba \mid w \in \{a, b\}^*\}$$

Reťazce nad  $\{a, b\}$  končiace príponou  $ba$ .

Regulárna gramatika  $G = (N, T, P, S)$ ,  $N = \{S, A\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatočný neterminál. Pravidlá  $P$ :

- $S \rightarrow aS \mid bS \mid bA$
- $A \rightarrow a$

Dá sa aj bezkontextová gramatika:  $G = (N, T, P, S)$ ,  $N = \{S, A, B\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatočný neterminál. Pravidlá  $P$ :

- $S \rightarrow AB$
- $A \rightarrow \varepsilon \mid aA \mid bA$
- $B \rightarrow ba$

A samozrejme mnoho ďalších ekvivalentných gramatík...



$$L_2 = \{abw \mid w \in \{a, b\}^*\}$$

Reťazce nad  $\{a, b\}$  začínajúce predponou  $ab$ .

Regulárna gramatika  $G = (N, T, P, S)$ ,  $N = \{S, A, B\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatkový neterminál. Pravidlá  $P$ :

- $S \rightarrow aA$
- $A \rightarrow bB \mid b$
- $B \rightarrow aB \mid bB \mid a \mid b$

Alebo bezkontextová gramatika:  $G = (N, T, P, S)$ ,  $N = \{S, A, B\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatkový neterminál. Pravidlá  $P$ :

- $S \rightarrow AB$
- $A \rightarrow ab$
- $B \rightarrow \varepsilon \mid aB \mid bB$

A samozrejme mnoho ďalších ekvivalentných gramatík...



$$L_3 = \{xaby \mid x, y \in \{a, b\}^*\}$$

Reťazce nad  $\{a, b\}$  obsahujúce  $ab$  ako podreťazec.

Regulárna gramatika  $G = (N, T, P, S)$ ,  $N = \{S, A, B\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatkový neterminál. Pravidlá  $P$ :

- $S \rightarrow aS \mid bS \mid aA$
- $A \rightarrow bB \mid b$
- $B \rightarrow aB \mid bB \mid a \mid b$

Alebo bezkontextová gramatika:  $G = (N, T, P, S)$ ,  $N = \{S, A, B, C\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatkový neterminál. Pravidlá  $P$ :

- $S \rightarrow ABC$
- $A \rightarrow \varepsilon \mid aA \mid bA$
- $B \rightarrow ab$
- $C \rightarrow \varepsilon \mid aC \mid bC$

A samozrejme mnoho ďalších ekvivalentných gramatík...

$$L_4 = \{w \mid \#_a(w) \equiv \#_b(w) \pmod{2}, w \in \{a, b\}^*\}$$

Reťazce nad  $\{a, b\}$  v ktorých je počet  $a$  a  $b$  rovnakej parity (t.j. alebo je tam nepárny počet  $a$  a nepárny počet  $b$ , alebo párný počet  $a$  a párný počet  $b$ ).

Regulárna gramatika  $G = (N, T, P, S)$ ,  $N = \{S, A\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatočný neterminál. Pravidlá  $P$ :

- $S \rightarrow \varepsilon \mid aA \mid bA$
- $A \rightarrow aS \mid bS$

(mimochodom, zároveň sú to práve všetky tie reťazce nad  $\{a, b\}$ , ktoré sú párnej dĺžky)



$$L_5 = \{wcw^R \mid w \in \{a, b\}^*\}$$

Palindrómy nad  $\{a, b, c\}$ , kde  $c$  je uprostred a ostatné symboly môžu byť len  $a$  alebo  $b$ .  $L_5 = \{c, aca, bcb, aacaa, abcba, bacab, bbcbb, \dots\}$

Bezkontextová gramatika  $G = (N, T, P, S)$ ,  $N = \{S\}$ ,  $T = \{a, b, c\}$ ,  $S$  je počiatočný neterminál. Pravidlá  $P$ :

- $S \rightarrow c \mid aSa \mid bSb$

(k tomuto jazyku sa **nedá** zostrojiť regulárna gramatika)



$$L_6 = \{ww^R \mid w \in \{a, b\}^*\}$$

Palindrómy nad  $\{a, b\}$  párnej dĺžky,  $L_6 = \{\varepsilon, aa, bb, aaaa, abba, baab, bbbb, \dots\}$

Bezkontextová gramatika  $G = (N, T, P, S)$ ,  $N = \{S\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatočný neterminál. Pravidlá  $P$ :

- $S \rightarrow \varepsilon \mid aSa \mid bSb$

(k tomuto jazyku sa **nedá** zostrojiť regulárna gramatika)



$$L_7 = \{a^n b^n \mid n \in \{1, 2, \dots\}\}$$

Reťazce nad  $\{a, b\}$  v ktorých je rovnaký počet  $a$  a  $b$  a symboly sú navyše usporiadané tak, že na začiatku sú všetky  $a$  a potom všetky  $b$ . Navyše musia obsahovať aspoň 1  $a$ , resp.  $b$ .  $L_7 = \{ab, aabb, aaabbb, aaaabbbb, \dots\}$

Bezkontextová gramatika  $G = (N, T, P, S)$ ,  $N = \{S\}$ ,  $T = \{a, b\}$ ,  $S$  je počiatočný neterminál. Pravidlá  $P$ :

- $S \rightarrow ab \mid aSb$

(k tomuto jazyku sa **nedá** zostrojiť regulárna gramatika)



$$L_8 = \{a^n b^n c^k \mid n, k \in \{0, 1, 2, \dots\}\}$$

Reťazce nad  $\{a, b, c\}$  v ktorých je rovnaký počet  $a$  a  $b$  - počet  $c$  je ľubovoľný - a symboly sú navyše usporiadané tak, že na začiatku sú všetky  $a$ , potom všetky  $b$  a na záver všetky  $c$ .

$$L_8 = \{\varepsilon, c, ab, abc, aabb, aabbc, aabbcc, aabbccc, aaabbb, aaabbbc, \dots\}$$

Bezkontextová gramatika  $G = (N, T, P, S)$ ,  $N = \{S, A, B\}$ ,  $T = \{a, b, c\}$ ,  $S$  je počiatočný neterminál. Pravidlá  $P$ :

- $S \rightarrow AB$
- $A \rightarrow \varepsilon \mid aAb$
- $B \rightarrow \varepsilon \mid cB$

(k tomuto jazyku sa **nedá** zostrojiť regulárna gramatika)





$$L_9 = \{ww \mid w \in \{a, b\}^*\}$$

Reťazce nad  $\{a, b\}$  ktoré pozostávajú z 2 rovnakých podreťazcov.

$$L_9 = \{\varepsilon, aa, bb, aaaa, abab, baba, bbbb, aaaaaa, abaaba, aabaab, \dots\}$$

Kontextová gramatika  $G = (N, T, P, S')$ ,  $N = \{S', S, X, Z, A, B\}$ ,  $T = \{a, b\}$ ,  $S'$  je počiatočný neterminál. Pravidlá  $P$  sú na ďalšom slajde (sem sa už nezmestili :)

(k tomuto jazyku sa **nedá** zostrojiť ani regulárna, ani bezkontextová gramatika).



## $L_9$ pravidlá

1.  $S' \rightarrow SZ$
2.  $S \rightarrow aSA$
3.  $S \rightarrow bSB$
4.  $S \rightarrow X$
5.  $XA \rightarrow Xa$
6.  $XB \rightarrow Xb$
7.  $aA \rightarrow Aa$
8.  $aB \rightarrow Ba$
9.  $aZ \rightarrow Za$
10.  $bA \rightarrow Ab$
11.  $bB \rightarrow Bb$
12.  $bZ \rightarrow Zb$
13.  $XZ \rightarrow \varepsilon$

## $L_9$ vysvetlenie na príklade

Nech  $ww = abab$ . Ako by sme ho odvodili v gramatike:

1. Z počiatočného neterminálu  $S'$  odvodíme vetnú formu, ktorá končí neterminálom  $Z$  - ten označuje koniec vetnej formy.

$$S' \Rightarrow SZ$$

2. Slovo, ktoré chceme odvodit' má mať tvar  $ww = abab$ . Pomocou pravidiel č. 2 a 3 odvodíme vo vetnej forme na začiatku reťazec  $w = ab$ , za ktorým je neterminál  $S$ , za ktorým je postupnosť neterminálov  $A, B$  taká, že je zrkadlovým obrazom reťazca  $w$ , teda  $BA$ . Na konci vetnej formy je stále neterminál  $Z$ .

$$S' \Rightarrow SZ \Rightarrow aSAZ \Rightarrow abSBAZ$$

3. Pravidlom č. 4 prepíšeme  $S$  na  $X$  v momente, keď je na začiatku vetnej formy požadovaný reťazec  $w$ :

$$S' \Rightarrow SZ \Rightarrow aSAZ \Rightarrow abSBAZ \Rightarrow abXBAZ$$



## $L_9$ vysvetlenie na príklade

Nech  $ww = abab$ . Ako by sme ho odvodili v gramatike:

4. Teraz vieme, že potrebujeme zrkadlovo otočiť reťazec z neterminálov  $A, B$  vo vetnej forme medzi neterminálmi  $X$  a  $Z$ . Na to slúžia pravidlá č. 5 - 12. Vezmeme prvý neterminál, zameníme ho za zodpovedajúci terminál (pravidlo 5, resp. 6) a presunieme ho na koniec vetnej formy - za neterminál  $Z$  (pravidlá 7 - 9, resp. 10 - 12).

$$abXBAZ \Rightarrow abXbAZ \Rightarrow abXAbZ \Rightarrow abXAZb$$

5. Toto opakujeme pre všetky neterminály medzi  $X$  a  $Z$ :

$$abXAZb \Rightarrow abXaZb \Rightarrow abXZab$$



## $L_9$ vysvetlenie na príklade

Nech  $ww = abab$ . Ako by sme ho odvodili v gramatike:

7. V momente, keď nám vo vetnej forme stoja za sebou  $XZ$  vieme, že sme skončili a túto skupinu neterminálov odstránime prepísaním na  $\varepsilon$  (pravidlo č. 13)

$$abXZab \Rightarrow abab$$

Spolu teda dostávame deriváciu reťazca  $abab$ :

$$\begin{aligned} S' &\Rightarrow SZ \Rightarrow aSAZ \Rightarrow abSBAZ \Rightarrow abXBABZ \Rightarrow abXbAZ \Rightarrow abXAbZ \Rightarrow \\ &\Rightarrow abXAZb \Rightarrow abXaZb \Rightarrow abXZab \Rightarrow abab \end{aligned}$$

