(0,5 bodu) Navrhněte bezkontextovou gramatiku pro následující bezkontextový jazyk:

$$L_1 = \{w : w \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}^* \wedge w = w^{\mathrm{R}}\} \cap \{u : u \in \{\mathtt{a}, \mathtt{b}\}^* \wedge |u|_{\mathtt{a}} \bmod 3 = 0\}$$

Poznámka: Pomocí zápisu $|x|_a$ značíme počet symbolů a v řetězci x.

ŘEŠENÍ:

Můžeme si všímnout, že $L_2 := \{w : w \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}^* \wedge w = w^{\mathrm{R}}\}$ obsahuje všechny palindromy, tj:

$$L_2 = \{\epsilon, a, b, c, aba, bab, cac, abbcbba, ...\}.$$

$L_3 := \{u : u \in \{\mathtt{a}, \mathtt{b}\}^* \wedge |u|_{\mathtt{a}} \bmod 3 = 0\}$ obsahuje všechny řetězce které mají počet symbolů a takový, že nemá žádný zbytek po dělení 3, tj. řetězec musí obsahovat 0, 3, 6, 9 atd. symbolů a.

$$L_3 = \{\epsilon, b, bb, bbb, bbbb, ..., aaa, abaa, baaabbbbaaba, ...\}$$

Průnik L_2 a L_3 musí tedy obsahovat všechny řetězce w, pro které platí, že w je palindrom $\wedge$ je nad abecedou $\{a, b\}$ (L_3 vylučuje všechny řetězce které mají alespoň jeden symbol c) $\wedge$ počet symbolů a má žádný zbytek po dělení 3

Tj:

$$L_1 = L_2 \cap L_3 := \{w : w \in \{a, b\}^* \wedge w = w^R \wedge |w|_a \mod 3 = 0\}$$

Například:

$$L_1 = \{b, bb, bbb, ..., aaa, ababa, baaab, aaabaaa, bbaababababaabb, ...\}$$

$G = (N, \Sigma, P, S)$ je bezkontextová gramatika, která právě generuje takový jazyk, tj. $L(G) = L_1$, kde:

$$
\begin{aligned}
& N := \{S, B, C\} \\
& \Sigma := \{a, b\} \\
& P := \{ \\
& \qquad S \to \epsilon \mid aBa \mid bSb \mid b, && |w|_a \bmod 3 = 0 \\
& \qquad B \to a \mid bBb \mid aCa, && |w|_a \bmod 3 = 2 \\
& \qquad C \to aa \mid bCb \mid aSa && |w|_a \bmod 3 = 1 \\
& \}
\end{aligned}
$$

Příklady derivací:

$$
\begin{aligned}
& S \Rightarrow \epsilon \\
& S \Rightarrow b \\
& S \Rightarrow aBa \Rightarrow aaa \\
& S \Rightarrow aBa \Rightarrow abBba \Rightarrow ababa \\
& S \Rightarrow aBa \Rightarrow abBba \Rightarrow abbBbba \Rightarrow abbabba \\
& S \Rightarrow aBa \Rightarrow abBba \Rightarrow abbBbba \Rightarrow abbaCabba \Rightarrow abbaaaabba \\
& S \Rightarrow aBa \Rightarrow abBba \Rightarrow abbBbba \Rightarrow abbaCabba \Rightarrow abbabCbabba \Rightarrow abbabaababba \\
& S \Rightarrow aBa \Rightarrow abBba \Rightarrow abbBbba \Rightarrow abbaCabba \Rightarrow abbaaSaabba \Rightarrow abbaabSbaabba \Rightarrow abbaabbbaabba
\end{aligned}
$$

(0,5 bodu) Uvažujte následující dvě tvrzení. U každého tvrzení rozhodněte, zdali je pravdivé, nebo nepravdivé. Následně svou odpověď řádně dokažte.

Tvrzení 1 Každý konečný jazyk je regulární.

ŘEŠENÍ:

Pravdivost tohoto tvrzení dokážeme pomocí indukce:

Nechť máme konečný jazyk L, který má n řetězců, $n \in \mathbb{N}_0$.

ZK_1 :

Pro $n = 0$:

Máme prázdný jazyk, který se dá vygenerovat regulární gramatikou:

$$G = (\{S\}, \{\}, \{\}, S)$$

a tedy je to regulární jazyk.

Pro $n = 1$:

Máme jazyk, který obsahuje pouze jeden řetězec. Dokážeme, že takový jazyk je regulární:

Důkaz provedeme zase indukcí:

Potřebujeme, pomocí regularní gramatiky vygenerovat jeden řetězec delky $m \in \mathbb{N}_0$, který dle definice je konečná posloupnost symbolů abecedy.

ZK_2 :

Pro $m = 0$, máme prázdný řetězec ϵ, který vygeneruje regulární gramatika:

$$G = (\{S\}, \{\}, \{S \to \epsilon\}, S)$$

tj. $L(G) = \{\epsilon\}$ je regulární jazyk

Pro m = 1, máme vygenerovat řetezec $w = w_1$. To umíme pomoci regulární gramatiky:

$$G = (\{W_1\}, \{w_1\}, \{W_1 \to w_1\}, W_1)$$

Pro m = 2, máme vygenerovat řetězec $w = w_1 w_2$. To zase umíme pomoci regulární gramatiky:

$$G = (\{W_1, W_2\}, \{w_1, w_2\}, \{W_1 \to w_1 W_2, W_2 \to w_2\}, W_1)$$

Pro m = 3, máme vygenerovat řetězec $w = w_1 w_2 w_3$. To zase umíme pomoci regulární gramatiky:

$$G = (\{W_1, W_2, W_3\}, \{w_1, w_2, w_3\}, \{W_1 \to w_1 W_2, W_2 \to w_2 W_3, W_3 \to w_3\}, W_1)$$

IK_2 :

Pro m + 1, máme vygenerovat řetězec $w = w_1 w_2 w_3 ... w_m w_{m+1}$. To uděláme následovně:

Dle indukčního předpokladu mějme regulární gramatiku G, která generuje jeden řetězec délky m (a tedy i regulární jazyk obsahující tento řetězec):

$$\begin{aligned}
&G = (\\
&\{W_1, W_2, W_3, W_4, ..., W_m\}, \\
&\{w_1, w_2, w_3, ...w_m\}, \\
&\{ \\
&\quad W_1 \to w_1 W_2, \\
&\quad W_2 \to w_2 W_3, \\
&\quad W_3 \to w_3 W_4, \\
&\quad ... \\
&\quad W_{m-1} \to w_{m-1} W_m, \\
&\quad W_m \to w_m \\
&\}, \\
&W_1 \\
&)
\end{aligned}$$

Symbol na pozici $m + 1$ vygenerujeme jednoduše tím, že rozšíříme poslední přepisovací pravidlo, a přídáme ještě jedno pravidlo, které vygeneruje symbol w_{m+1} na posledním místě:

$G = ($
$\{W_1, W_2, W_3, W_4, ..., W_m\} \cup \{W_{m+1}\},$
$\{w_1, w_2, w_3, ...w_m\} \cup \{w_{m+1}\},$
$\{$

$W_1 \to w_1 W_2,$
$W_2 \to w_2 W_3,$
$W_3 \to w_3 W_4,$
$...$
$W_{m-1} \to w_{m-1} W_m,$
$W_m \to w_m W_{m+1},$
$W_{m+1} \to w_{m+1}$

$\},$
W_1
$)$

Výsledná gramatika pro $m+1$ symbolů je zase regulární, a tedy i generovaný jazyk je taky regulární.

Tím jsme dokázali, že libovolný jazyk, který má právě jeden řetězec libovolné delky je vždycky regulární.

IK_1 :

Dokážeme teď, že jazyk, který má n + 1 řetězců, je taky regulární: Víme, že regulární jazyky jsou uzavřené na operaci sjednocení. Ňechť L_{n+1} je jazyk který má $n+1$ řetězců. Potom platí:

$$L_{n+1} = L_n \cup L_1$$

kde L_n je jazyk který má n řetězců a L_1 je jazyk který má 1 řetězec.

Z indkučního předpokladu platí, že L_n je regulární jazyk. Ze zakladního kroku ZK_1 víme, že i L_1 je regulární. Potom i sjednocení L_n a L_1 bude jazyk regulární, tj. jazyk L_{n+1} je regulární.

Tím jsme dokázali, že pro každý konečný jazyk platí, že je regulární.

Tvrzení 2 Každý regulární jazyk je konečný.

ŘEŠENÍ:

Dokážeme sporem, že tvrzení neplatí:

Stačí, abychom našli regulární gramatiku, která bude generovat nekonečně mnoho řetězců:

$$G = (\{S\}, \{a\}, \{S \to a \mid aS\ \}, S)$$

Potom:

$$L(G) = \{a, aa, aaa, aaaa, aaaaa, aaaaaa, ...\}$$

obsahuje nekonečně mnoho řetězců obsahujících pouze symboly a, a tedy tento regulární jazyk generovaný regulární gramatikou není konečný, což je ve sporu s tvrzením, že každý regulární jazyk je konečný.