

Jazyk, operace s jazyky

1. Abeceda A obsahuje všechna malá písmena češtiny a angličtiny. Uvažujme jazyky L_1 a L_2 nad abecedou A . L_1 je množina všech názvů měsíců v roce v češtině, L_2 je množina názvů všech měsíců v roce v angličtině. Rozhodněte o pravdivosti následujících tvrzení:

- a) $L_1 = L_2$. **NE**
- b) Mohutnost L_1 a L_2 je stejná. **ANO**
- c) Průnik L_1 a L_2 je prázdný. **ANO**
- d) Sjednocení L_1 a L_2 obsahuje slovo "červenjune". **NE**
- e) Zřetězení L_1 a L_2 obsahuje slovo "červenjune". **ANO**
- f) Mohutnost zřetězení L_1 a L_2 je menší než 100. **NE**
- g) Doplněk L_1 obsahuje slovo "červenjune". **ANO**
- h) Doplněk L_2 obsahuje slovo "monday". **ANO**
- i) Druhá mocnina L_1 obsahuje slovo "ledenúnorleden". **NE**
- j) Iterace L_1 obsahuje slovo "ledenúnorleden". **ANO**

Odpovědi

- a) $L_1 = \{\text{leden, únor, březen, duben, květen, červen, červenec, srpen, září, říjen, listopad, prosinec}\}$. $L_2 = \{\text{january, february, march, april, may, june, july, august, september, october, november, december}\}$. Zřejmě se jedná o různé množiny, jejich prvky jsou celá slova, nikoli písmena apod.
- b) Ano, obě množiny obsahují dvanáct prvků, tj. dvanáct slov.
- c) Všechna slova v obou množinách jsou různá, tudíž zřejmě množiny L_1 a L_2 mají prázdný průnik.
- d) Sjednocení obou jazyků je $L_1 \cup L_2 = \{\text{leden, únor, březen, duben, květen, červen, červenec, srpen, září, říjen, listopad, prosinec, january, february, march, april, may, june, july, august, september, october, november, december}\}$.
- e) Slovo je prvkem zřetězení dvou jazyků, pokud jeho první část je slovem prvního jazyka a jeho zbytek až do konce slovem druhého jazyka. To pro náš případ se slovem "červenjune" platí.
- f) Libovolné slovo jednoho jazyka můžeme zřetězit s libovolným slovem druhého jazyka a vyjde nám slovo náležející do zřetězení obou jazyků. V našem případě mají oba jazyky 12 slov, jejich zřetězení obsahuje všechny možné takto vytvořené nápisy a těch je celkem $12 \times 12 = 144$.
- g) Doplněk daného jazyka obsahuje všechny možné a myslitelné řetězce nad danou abecedou, jen jedině ne ty, které jsou prvky daného jazyka. Proto doplněk L_1 obsahuje i slovo "červenjune".
- h) Slovo "monday" je vytvořeno nad abecedou A , není prvkem L_2 , proto je prvkem doplňku L_2 .
- i) Druhá mocnina jazyka obsahuje všechna slova vzniklá zřetězením právě dvou slov původního jazyka. Slovo "ledenúnorleden" nelze rozdělit na dvě slova jazyka L_1 , nepatří tedy do jeho druhé mocniny.
- j) Iterace jazyka obsahuje všechna slova vzniklá zřetězením libovolného počtu slov původního jazyka. Zřejmě slovo "ledenúnorleden" je takovým zřetězením.

Gramatika

2. Považte gramatiku G_1 s množinou neterminálů $N = \{S, A, B\}$, množinou terminálů $T = \{0,1\}$, startovním symbolem S a s pravidly:

$S \rightarrow 0A1 \mid 1B0$

$A \rightarrow 1A0 \mid 1$

$B \rightarrow 1B \mid 1$.

Označme L_3 jazyk generovaný gramatikou G_1 , $L_3 = L(G_1)$. Rozhodněte o pravdivosti následujících tvrzení:

- a) L_3 je konečný jazyk. **NE**
- b) L_3 obsahuje prázdné slovo. **NE**
- c) L_3 obsahuje právě dvě slova délky nejvýše 3. **ANO**
- d) L_3 obsahuje pouze slova liché délky. **NE**
- e) Každé slovo L_3 obsahuje znak 0. **ANO**

- f) Každé slovo L_3 obsahuje alespoň dva znaky 0. **NE**
- g) Žádné slovo L_3 neobsahuje souvislý řetězec 01010. **ANO**, neobsahuje
- h) Existuje slovo L_3 , které obsahuje řetězec 11011. **NE**
- i) Pokud dvě slova L_3 mají stejnou délku, pak začínají stejným symbolem. **NE**
- j) Pokud dvě slova L_3 začínají stejným symbolem, pak mají stejnou délku. **NE**

Odpovědi

- a) L_3 není konečný jazyk. Po použití pravidla $S \rightarrow 1B0$ můžeme libovolněkrát použít pravidlo $B \rightarrow 1B$ a nakonec jedenkrát pravidlo $B \rightarrow 1$. Tím můžeme vytvořit každý prvek nekonečně velké množiny řetězců $\{110, 1110, 1110, 11110, \dots\}$.
- b) Při jakékoli expanzi startovního symbolu S (t.j. při jakémkoli použití pravidla se startovním symbolem S na levé straně) se v generovaném řetězci octne 1 i 0. Prázdné slovo tedy nelze ze startovního symbolu vygenerovat.
- c) Jediné dva způsoby, jak vygenerovat slovo délky 3 v L_3 , jsou: (1) Použitím pravidel $S \rightarrow 0A1$, $A \rightarrow 1$ resp. (2) Použitím pravidel $S \rightarrow 1B0$, $B \rightarrow 1$. Tím získáme slovo 011 resp. 110.
- d) Použitím pravidla $S \rightarrow 1B0$ a dále opakovaným používáním pravidla $B \rightarrow 1B$ můžeme v každém kroku zvětšit délku odvozaného řetězce o 1, můžeme proto tvořit slova libovolné délky větší než 2.
- e) Obě pravidla, která mají na levé straně startovní symbol, generují řetězec obsahující 0. Každé slovo L_3 tedy tento symbol bude obsahovat.
- f) Viz c).
- g) Když začneme pravidlem $S \rightarrow 0A1$, můžeme dále generovat pomocí pravidel $A \rightarrow 1A0 \mid 1$ už jen řetězce 011, 01101, 0111001, 011110001, 01111100001, ... Ty daný řetězec 01010 neobsahují. Když začneme pravidlem $S \rightarrow 1B0$, můžeme dále generovat pomocí pravidel $B \rightarrow 1B \mid 1$ už jen řetězce 110, 1110, 11110, 111110, ... Ty daný řetězec 01010 také neobsahují.
- h) Odpověď viz g).
- i) Z odpovědi g) vidíme, že slova stejné délky mohou začínat různými symboly.
- j) Z odpovědi g) vidíme, že slova různé délky mohou začínat stejným symbolem.

3. Odpovězte na stejné otázky a)– j) pro jazyk L_4 generovaný gramatikou $G_2: N = \{S, A, B\}, T = \{0,1\}$, s pravidly

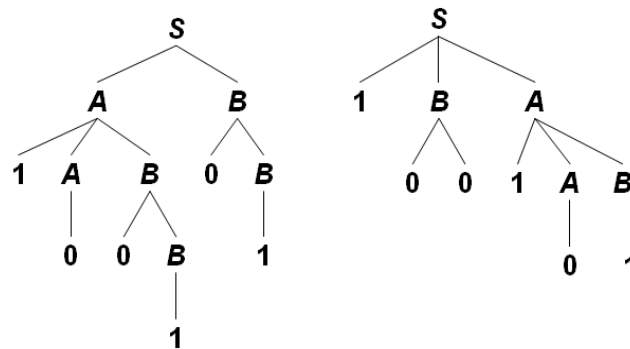
$$\begin{aligned} S &\rightarrow AB \mid 1BA \\ A &\rightarrow 1AB \mid 0 \\ B &\rightarrow 0B \mid 1 \mid 00. \end{aligned}$$

Odpovědi

- a) NE, podobně jako ve 3a).
- b) NE, podobně jako ve 3b).
- c) NE, L_4 obsahuje slova 01, 000, 101.
- d) NE, argument je podobný 3d).
- e) ANO, během generování libovolného slova musíme vždy alespoň jednou použít neterminál A . Ten však lze expandovat na terminál pouze pomocí pravidla $A \rightarrow 0$. Proto symbol 0 je obsažen v každém slově L_4 .
- f) NE, L_4 obsahuje slova 1011, 110111, 11101111, 1111011111, ... To snadno ověříme, když se při generování slov omezíme pouze na pravidla použitá v následujícím pořadí: $S \rightarrow AB$ (jednou), $A \rightarrow 1AB$ (k -krát), $A \rightarrow 0$ (jednou), $B \rightarrow 1$ ($(k+1)$ -krát), $k > 0$.
- g) ANO, použijeme pravidlo $S \rightarrow AB$, pak pravidlo $A \rightarrow 1AB$ dvakrát za sebou, získáme tak řetězec 11ABBB, dále použijeme třikrát pravidlo $B \rightarrow 0B$, získáme řetězec 11A0B0B0B a konečně pomocí pravidla $B \rightarrow 1$ získáme řetězec 11A010101. Dále můžeme pokračovat libovolnou expanzí neterminálu A .
- h) ANO, viz odpověď f).
- i) NE, viz odpověď c).
- j) NE, viz odpověď f).

Derivační strom

4. Jazyk L_4 obsahuje slovo 100101. Nakreslete dva různé derivační stromy tohoto slova.



Regulární jazyk

5. Ověřte, že jazyk $(10)^+$ lze generovat například regulární gramatikou, která obsahuje pravidla

$$S \rightarrow 1A$$

$$A \rightarrow 0B \mid 0$$

$$B \rightarrow 1A.$$

Odpověď

Stačí nejprve aplikovat pravidlo $S \rightarrow 1A$, poté libovolněkrát opakovat aplikaci dvojice pravidel $A \rightarrow 0B$, $B \rightarrow 1A$ (v tomto pořadí) a nakonec aplikovat pravidlo $A \rightarrow 0$.

6. Sestrojte regulární gramatiku, která generuje jazyk

a) $(10)^*$.

b) $(110)^*$.

c) $(110)^*$.

Odpovědi

a) Je to zcela podobné úloze 5, jen je třeba přidat prázdné slovo:

$$S \rightarrow 1A \mid \varepsilon$$

$$A \rightarrow 0B \mid 0$$

$$B \rightarrow 1A.$$

b) Je to podobné úloze 5, jen je třeba přidat další neterminál, který zajistí opakování dvou jedniček :

$$S \rightarrow 1A_1$$

$$A_1 \rightarrow 1A_2$$

$$A_2 \rightarrow 0B \mid 0$$

$$B \rightarrow 1A_1.$$

c) Podobně jako a) a b).

Úlohu 6 lze řešit s použitím menšího počtu neterminálů:

a) $S \rightarrow 1A \mid \varepsilon$
 $A \rightarrow 0S \mid 0$

b) $S \rightarrow 1A_1$
 $A_1 \rightarrow 1A_2$
 $A_2 \rightarrow 0S \mid 0$

c) Opět podobně.