

Týden 4

Přednáška - první část

(viz také slidy k této přednášce ...)

Bezkontextové gramatiky

Připomněli jsme si, že

$$(((a \cdot a) + b)^*)$$

je příklad (úplně uzávorkovaného) regulárního výrazu, který reprezentuje jazyk v abecedě $\{a, b\}$. Uvědomili jsme si, že ovšem samotný regulární výraz je prostě řetězcem symbolů abecedy

$$\Sigma = \{a, b, +, \cdot, *, (,)\}.$$

(Pro úplnost bychom měli přidat znaky \emptyset, ϵ , ale zde se bez nich obejdeme.)

Ne každý řetězec v Σ^* je ovšem regulárním výrazem; např. řetězec $a) ++$ (regulárním výrazem není).

Snadno jsme vyvodili, že množina těchto regulárních výrazů, označovaná $RV(\{a, b\})$, není regulárním jazykem. Dá se ale generovat bezkontextovou gramatikou, např.

$$R \longrightarrow a \mid b \mid (R + R) \mid (R \cdot R) \mid (R^*).$$

Jiná gramatika generující uvedený jazyk je

$$\begin{aligned} R &\longrightarrow L \mid (RBR) \mid (RU) \\ L &\longrightarrow a \mid b \\ B &\longrightarrow + \mid \cdot \\ U &\longrightarrow * \end{aligned}$$

Demonstrovali jsme si základní pojmy teorie bezkontextových gramatik. Ukázali jsme si (*levou derivaci* slova $(((a \cdot a) + b)^*)$, příslušný *derivační strom*, apod.

Připomněli jsme definici *bezkontextové gramatiky* jako struktury

$$G = (\Pi, \Sigma, S, P)$$

a *jazyka generovaného gramatikou*

$$L(G) = \{w \in \Sigma^* \mid S \Rightarrow^* w\}.$$

Pak jsme se vrátili k příkladu jazyka $RV(\{a, b\})$ a upravili jej tak, že v řetězcích (regulárních výrazech) vynecháváme tečku pro zřetězení a nemusíme plně uzávorkovávat. Např. výraz $(((a \cdot a) + b)^*)$ můžeme zapsat $(aa + b)^*$. Takto upravený jazyk generuje např. gramatika

$$R \longrightarrow a \mid b \mid R + R \mid R \cdot R \mid RR \mid R^* \mid (R).$$

Všimli jsme si ovšem, že např. slovo $aa+b$ má v této gramatice dva různé derivační stromy; tedy tato gramatika *není jednoznačná*. Příčinou je tady fakt, že naše dohodnutá priorita operátorů není v gramatice reflektována.

Demonstrovali jsme, že v tomto případě lze nalézt ekvivalentní gramatiku (tedy gramatiku generující tentýž jazyk), která jednoznačná je. Úvahami nad strukturou regulárních výrazů jsme postupně došli ke gramatice

$$\begin{aligned} R &\longrightarrow T + R \mid T \\ T &\longrightarrow FT \mid F \\ F &\longrightarrow F^* \mid (R) \mid C \\ C &\longrightarrow a \mid b \end{aligned}$$

Pomohla nám úvaha, že T (Term) reprezentuje ty regulární výrazy, které nejsou ve tvaru $R_1 + R_2$ (pro dva regulární výrazy R_1, R_2), a F (Factor) reprezentuje ty výrazy, které nejsou ve tvaru $R_1 + R_2$ ani R_1R_2 .

Uvedli jsme pojem (*vnitřně*) *jednoznačný bezkontextový jazyk* a několik souvisejících poznámek.

Započali jsme příklad ze cvičení – (jednoznačná) gramatika pro booleovské formule. Pak jsme zkonstruovali gramatiky pro dva jednoduché jazyky:

Promyšlením tvarů slov jsme zkonstruovali gramatiky pro jazyk palindromů $\{w \in \{a, b\}^* \mid w = w^R\}$

$$S \longrightarrow \varepsilon \mid a \mid b \mid aSa \mid bSb$$

a pro jazyk posloupností v abecedě $\{ (,), [,] \}$, které odpovídají správnému uzávorkování

$$S \longrightarrow \varepsilon \mid SS \mid (S) \mid [S].$$

Přednáška - druhá část

Vrátili jsme se k větě

Věta. Regulárními výrazy lze reprezentovat právě regulární jazyky.

Věnovali jsme se dvoucestným (či dvousměrným) konečným automatům.

(Podrobnější poznámky lze nalézt v průběhu výuky v loňském roce ...)

Partie textu k prostudování

Jedná se zejména o části 4.1., 4.2., 4.3. (bezkontextové gramatiky, jednoznačné gramatiky).

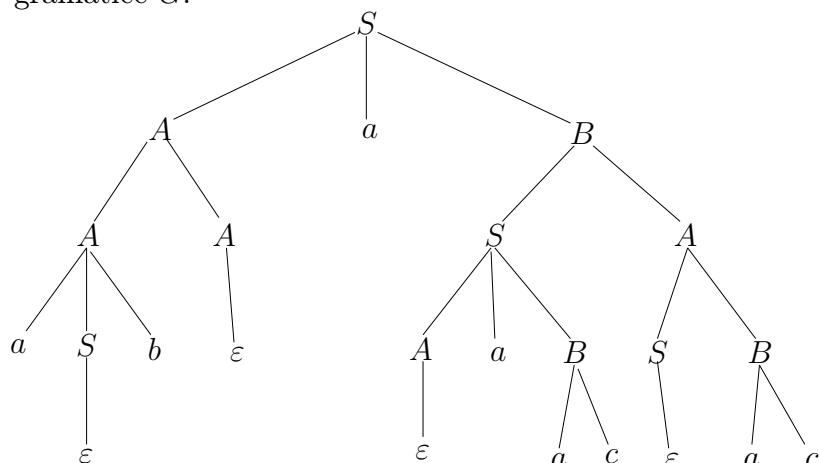
(Máte si udělat přinejmenším dobrou první představu a zamyslet se nad příklady, speciálně těmi plánovanými na cvičení, ať se můžete na cvičení aktivně účastnit a případné problémy si tam objasnit.)

Cvičení

Prezentace referátů 5 a 6

Příklad 4.1

Na obrázku je derivační strom pro slovo $w = abaaacac$ odpovídající jisté bezkontextové gramatice G .



- Vypište všechna pravidla G , jejichž existenci můžete vyvodit z daného derivačního stromu.
- Napište levé odvození (levou derivaci) slova w podle gramatiky G .
- Najděte *menší* derivační strom pro slovo $abaaacac$ a zakreslete jej tak, že všechny listy budou na stejné úrovni (tedy odvozené slovo bude celé na „jednom řádku“).
- Najděte nejlevější větev (v onom menším stromě), která obsahuje dva výskyty neterminálu B . Využijte to k důkazu, že gramatika generuje také slovo $abaac$. Pak ukažte, že gramatika také generuje slova $aba(a)ac(ac)$, $aba(a)^2ac(ac)^2$, $aba(a)^3ac(ac)^3$, \dots
- Lze z dostupné informace zjistit něco ohledně jednoznačnosti gramatiky G ?

Příklad 4.2

(podle plánu začato na přednášce)

Uvažujme jazyk sestávající ze všech booleovských formulí s proměnnými x_1, x_2, \dots a logickými spojkami \neg, \wedge, \vee ; mohou se v nich používat závorky $(,)$, ale není nutné plně závkovat. Každá taková formule je tedy řetězcem v abecedě

$$\Sigma = \{ x, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, \neg, \wedge, \vee, (,) \};$$

jako příklad může sloužit řetězec $(\neg x_{15} \vee x_2 \wedge x_5) \wedge \neg x_{21} \vee \neg(x_2 \vee x_5)$, který do jazyka patří. (Samozřejmě zde můžeme preferovat přehlednější zápis $(\neg x_{15} \vee x_2 \wedge x_5) \wedge \neg x_{21} \vee \neg(x_2 \vee x_5)$, ale to není podstatné.)

Navrhněte co nejjednodušší bezkontextovou gramatiku generující uvedený jazyk.

Takto navržená (jednoduchá) gramatika asi není jednoznačná; ověřte. Zkonstruujte pak pro stejný jazyk jednoznačnou gramatiku, u níž derivační stromy přirozeně odpovídají obvyklé prioritě operátorů: negace váže silněji než konjunkce a konjunkce váže silněji než disjunkce.

Příklad 4.3

Uvažujme jazyk $L = \{ w \in \{a, b\}^* \mid |w| \geq 1 \text{ a } |w|_a = |w|_b \}$.

Charakterizujte slova z $L^2 = L \cdot L$. Je pravda, že $L = L^2$? Platí případně alespoň jedna z inkluzí $L \subseteq L^2$, $L^2 \subseteq L$?

Charakterizujte slova z $L - L^2$.

Na základě předešlých úvah navrhněte bezkontextovou gramatiku generující L .

Příklad 4.4

Snažte se co nejdůležitěji charakterizovat jazyk generovaný gramatikou

$$S \longrightarrow bSS \mid a$$

Příklad 4.5

(V případě nedostatku času dokončit příště.)

Navrhněte bezkontextové gramatiky generující následující jazyky:

- $L_1 = \{ w \in \{a, b\}^* \mid w \text{ obsahuje podslovo } baab \}$
- $L_2 = \{ w \in \{a, b\}^* \mid |w|_b \bmod 3 = 0 \}$
- $L_3 = \{ ww^R \mid w \in \{a, b\}^* \}$
- $L_4 = \{ 0^n 1^m 0^n \mid m, n \geq 0 \}$
- $L_5 = \{ 0^n 1^m \mid 1 \leq n \leq m \leq 2n \}$