3.2.2. Formální jazyky

V této podkapitole se budeme zabývat speciálními jednorozměrnými relačními strukturami - obvykle nazývanými řetězy. Pravidla (produkce) p pro práci s nimi se sice zjednoduší ,zato však vznikne nutnost používání celých systémů pravidel a substitucí. Z propracovaných oblastí teorie popisu reality pomocí aparátu relačních struktur pak přistoupí pojem gramatiky a pojem formálního jazyka generovaného gramatikou.

Def.3.2.2.1.: Je dána abeceda A. Řetěz nad abecedou A definujeme takto :

(i) Prázdný řetěz e (neobsahuje žádný symbol z A) je řetězem nad abecedou A.
(ii) Je-li r řetěz nad abecedou A a symbol $x \in A$, pak rx je řetězem nad abecedou A.
(iii) Mějme $r_1, r_2, \ldots, r_n$ řetězy nad abecedou A.Pak řetěz $r = r_1 r_2 \ldots r_n$ je řetězem nad abecedou A.
(iv) Žádné jiné řetězy,než vytvořené podle (i),(ii),(iii) nad abecedou A neexistují. ∎

(Množinu všech řetězů nad abecedou A označíme A^*.)

Příklad 11.3.:

Mějme symboly a,b,c,d $\in$ A.Podle (i) a (ii) je $ea \in A^*$.(Ve skutečnosti ea = ae = a .) Podobně pak dále pro b,c,d - tedy $ab \in A^*$, $cd \in A^*$, ... atd. tedy i $abcd \in A^*$.Operace zřetězení definovaná v (iii) nahrazuje opakované použití konstrukce (ii).

Def.3.2.2.2.: Gramatikou rozumíme čtveřici G = (N,T,S,P) , kde

N je abeceda tzv. neterminálních symbolů ,

T je abeceda tzv. terminálních symbolů ,

$(N \cap T = \emptyset)$,

$S \in N$ je počáteční symbol gramatiky (někdy také větný nebo startovací symbol) ,

P je množina pravidel (produkcí) typu : p = (L,R) , kde $L,R \in (N \cup T)^*$.(Substituci podle pravidla p budeme pro jednoduchost označovat pouze symbolem $\Rightarrow$.) ∎

Jsou-li symboly abecedy T jednorozměrné a typ pravidel nemá další specifikace k definici 3.2.2.2.,pak gramatika představuje aparát pro generování řetězů $r \in (N \cup T)^*$.
Množinu řetězů $w \in T^*$, které lze vytvořit pomocí gramatiky G,nazýváme množinou slov generovaných gramatikou G.

Formálním jazykem L(G) rozumíme množinu všech slov generovaných gramatikou G .

(Pro řetězy $r \in (N \cup T)^*$ je užíván též termín "větná forma" a místo o slovech w pak hovoříme o "větách".)

Příklad 12.3.:

Gramatika $G = (\{A,S\}, \{0,1\}, S, \{S \Rightarrow 0A1, 0A \Rightarrow 00A1, A \Rightarrow e\})$ generuje jazyk $L(G) = \{0^n 1^n \mid n \geq 1\}$ (skládající se ze slov, v nichž následuje n symbolů "1" po předchozích n symbolech "0").

Při generování slov sledujeme vývoj následujících řetězů :

$S \rightarrow 0A1 \rightarrow 00A11 \rightarrow 0011$,

$S \rightarrow 0A1 \rightarrow 00A11 \rightarrow 000A111 \rightarrow 000111$.

Příklad 13.3.:

Dokažte, že gramatika $G_1 = (\{S\}, \{0,1\}, S, \{S \Rightarrow 0S1, S \Rightarrow 01\})$ generuje tentýž formální jazyk jako gramatika G z příkladu 12.3. (Gramatiky, které generují identické jazyky, nazýváme ekvivalentní.)

Pozn.: Pravidla $p \in P$, která mají stejnou část L budeme dále úsporně zapisovat takto : $L \Rightarrow R_1 \mid R_2 \mid \dots \mid R_k$.

Def.3.2.2.3.: Rozlišujeme následující typy gramatik $G = (N,T,S,P)$:

1. Typ 0 - gramatika neomezená - jestliže vyhovuje definici 3.2.2.2. bez dalších specifikací.
2. Typ 1 - gramatika kontextová - jestliže každé pravidlo z P má tvar $L \Rightarrow R$, kde $L \in (N \cup T)^* \times N \times (N \cup T)^*$, $R \in (N \cup T)^*$ a délka řetězů L je menší nebo rovna délce řetězů R. (Vyjímku tvoří pravidlo $S \Rightarrow e$ v případě, že se S neobjeví na pravé straně žádného pravidla).
3. Typ 2 - gramatika bezkontextová - jestliže $L \in N$ a $R \in (N \cup T)^*$.
4. Typ 3 - gramatika regulární - jestliže $L \in N$ a $R = wB$, kde $w \in T^*$, $B \in N$. (Vyjímku tvoří pravidlo $S \Rightarrow e$ pro případ, že se S neobjeví na pravé straně žádného pravidla.) ∎

Příklad 14.3.:

Gramatika G z příkladu 12.3. je neomezená: Pro pravidlo $A \Rightarrow e$ neplatí podmínka délek L,R a $A \neq S$.
Gramatika G_1 z příkladu 13.3. je bezkontextová.

Příklad 15.3.:

Gramatika $G_2 = (\{S,A\}, \{0,1\}, S, \{S \Rightarrow 0A1, S \Rightarrow 01, 0A \Rightarrow 00A1, A \Rightarrow 01\})$ je kontextová.

Příklad 16.3.:

Gramatika $G_3 = (\{S,X,Y,Z\}, \{a,b,c\}, S, P)$, kde P tvoří pravidla :

p1: $S \Rightarrow abcX \mid baY$,
p2: $Y \Rightarrow bcY \mid aZ \mid a$,
p3: $X \Rightarrow Z$,
p_4: $Z \Rightarrow b$,

je regulární.

Pozn.: Uvedená klasifikace gramatik je pouze základní , zachycuje pouze tzv. pravé gramatiky (řetězy rostou doprava) a opírá se o typ pravidel.

<u>Věta 3.2.2.1.</u>: Každou pravou regulární gramatiku $G = (N,T,S,P)$ lze transformovat na pravou regulární gramatiku $G' = (N',T,S,P')$, která bude s G ekvivalentní , přičemž množina P' bude obsahovat pouze pravidla typu : $L \in N$, R = (aB nebo e) ,kde $B \in N$, $a \in T$.

<u>Důkaz</u> (popisuje zároveň konstrukci transformace $G \longrightarrow G'$) :

a) Všechna pravidla typu $A \Rightarrow aB$ nebo $A \Rightarrow e$ gramatiky G zařadíme do P'.

b) Každé pravidlo typu $A = a_1 a_2 \dots a_n B$ z P ,kde $a_1,\dots,a_n \in T$, $A,B \in N$ nahradíme soustavou pravidel

$$\begin{aligned} A &\Rightarrow a_1 A_1 \, , \\ A_1 &\Rightarrow a_2 A_2 \, , \\ &\vdots \\ A_{n-1} &\Rightarrow a_n B \, , \end{aligned}$$

která zařadíme do P' a všechny neterminály $A_1,\dots,A_{n-1}$ zařadíme do N'.

c) Každé pravidlo typu $A \Rightarrow a_1 a_2 \dots a_n$ z P , kde $a_1,\dots,a_n \in T$, $A \in N$ nahradíme soustavou pravidel

$$\begin{aligned} A &\Rightarrow a_1 B_1 \, , \\ B_1 &\Rightarrow a_2 B_2 \, , \\ &\vdots \\ B_{n+1} &\Rightarrow e \, , \end{aligned}$$

která zařadíme do P' a všechny terminály $B_1,\dots,B_{n+1}$ připojíme k N'.

d) Ke každému pravidlu typu $A \Rightarrow B$, $A,B \in N$, sestrojíme množiny M_A, M_B pravidel z P' tak,že :

$$M_A = \{ p \in P' \mid p\colon L \Rightarrow xA \; ; \; L \in N', \; A \in N \, , \; x \in T \} \, ,$$

$$M_B = \{ p \in P' \mid p\colon B \Rightarrow yY \; ; \; Y \in N', \; B \in N \, , \; y \in T \} \, .$$

Dále sestrojíme množiny M_A' a M_B' tak,že každé pravidlo v M_A nahradíme pravidlem $p{:}L \Rightarrow xB$ a každé pravidlo v M_B nahradíme pravidlem $p\colon A \Rightarrow yY$, viz. obr. 8.3.

Množiny pravidel M_A' , M_B' připojíme k P' , množiny pravidel M_A,M_B v P' ponecháme , ale pravidlo typu $A \Rightarrow B$ do P' <u>nezařadíme</u> .

Všechna pravidla , kde L=S z G zpracujeme podle bodů a)...d) a S připojíme do G jako počáteční symbol.

Konec důkazu.(Jiná pravidla již v P nejsou.) ∎

<u>Příklad 17.3.</u>:

Uvažujeme gramatiku G_3 z příkladu 16.3.Sestrojme gramatiku G_3' vyhovující podmínkám věty 3.2.2.1. !

Postupujeme podle bodů a) až d) věty :

a) Do P' zařadíme z P jediné pravidlo : $Y \Rightarrow aZ$ bez dalších změn.

b) Pravidlo $S \Rightarrow abcX$ nahradíme soustavou pravidel

$$S \Rightarrow aA_1 \quad ,$$
$$A_1 \Rightarrow bA_2 \quad ,$$
$$A_2 \Rightarrow cX \quad ,$$

podobně pravidlo $S \Rightarrow baY$ nahradíme soustavou pravidel

$$S \Rightarrow bB_1 \quad ,$$
$$B_1 \Rightarrow aY \quad ,$$

a konečně - pravidlo $Y \Rightarrow bcY$ nahradíme soustavou pravidel

$$Y \Rightarrow bC_1 \quad ,$$
$$C_1 \Rightarrow cY \quad .$$

Všechna nově vytvořená pravidla zařadíme do P' a všechny nově zavedené neterminály připojíme k N'.

c) Pravidlo $Y \Rightarrow a$ nahradíme dvěma pravidly

$$Y \Rightarrow aD_1 \quad ,$$
$$D_1 \Rightarrow e \quad , \quad ,$$

a pravidlo $Z \Rightarrow b$ nahradíme dvěma pravidly

$$Z \Rightarrow bE_1 \quad ,$$
$$E_1 \Rightarrow e \quad .$$

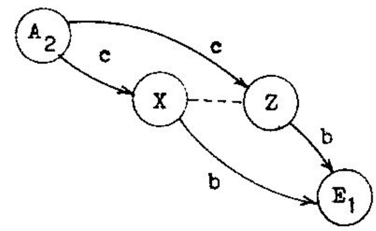
Všechna nová pravidla zařadíme do P' a všechny nové neterminály připojíme k N'.

d) K pravidlu $X \Rightarrow Z$ sestrojíme množiny M_X , M_Z ;

$$M_X = \{ A_2 \Rightarrow cX \} \quad ,$$
$$M_Z = \{ Z \Rightarrow bE_1 \} \quad .$$

Dále sestrojíme množiny M'_X a M'_Z (viz. obr. 8.3.)

$$M'_X = \{ A_2 \Rightarrow cZ \} \quad ,$$
$$M'_Z = \{ X \Rightarrow bE_1 \} \quad .$$



Obr.8.3.

Gramatika $G'_3 = (\{ S,X,Y,Z,A_1,A_2,B_1,C_1,D_1,E_1 \} , \{ a,b,c \} , S , P')$, kde P' obsahuje pravidla :

$p'_1 : S \Rightarrow aA_1 \mid bB_1$,

$p'_2 : Y \Rightarrow bC_1 \mid aD_1$,

$p'_3 : X \Rightarrow bE_1$,

$p'_4 : Z \Rightarrow bE_1$,

$p'_5 : A_1 \Rightarrow bA_2$,

$p'_6 : A_2 \Rightarrow cX \mid cZ$,

$p'_7 : B_1 \Rightarrow aY$,

$p'_8 : C_1 \Rightarrow cY$,

$p'_9 : D_1 \Rightarrow e$,

$p'_{10} : E_1 \Rightarrow e$.

Kontrolní otázky a úlohy :

1. Kdy je gramatika bezkontextová zároveň gramatikou regulární ?
2. Jak dokázat , že gramatiky G a G' jsou ekvivalentní aniž bychom sestrojovali jazyky $L(G)$ a $L(G')$?
3. Uvažujme určité slovo w jistého formálního jazyka $L(G)$. Za jakých podmínek by byla konstrukce slova w jediná a za jakých podmínek by byl postup generování slova w deterministický ?
4. Je dána gramatika $G = (\{ A,B,C,X \}, \{ y \} , A , P)$, kde P obsahuje pravidla

   $p_1 : A \Rightarrow BC \mid e$,

   $p_2 : B \Rightarrow yXA \mid A$,

   $p_3 : C \Rightarrow y$.

   Určete typ gramatiky a přesvědčte se,zda $L(G)$ není neprázdný !
5. Je dána pravá regulární gramatika $G = (\{S,X,Y\} , \{ a,b,c \} , S , P)$, kde P obsahuje pravidla :

   $p_1 : S \Rightarrow ab \mid Y \mid e$,

   $p_2 : Y \Rightarrow bY \mid cbX \mid X$,

   $p_3 : X \Rightarrow c$.

   Sestrojte gramatiku G' ekvivalentní s danou gramatikou G tak,aby P' obsahovala pouze pravidla typu : $R = aB$ nebo e , $a \in T$, $B \in N$.
6. Je dána gramatika $G = (\{ B,S,X \} , \{ a,b \} , S , P)$,kde P obsahuje pravidla :

   $p_1 : S \Rightarrow aaB$, $p_2 : B \Rightarrow XSb$, $p_3 : X \Rightarrow SB \mid b$.

   Určete typ gramatiky a zjistěte,zda $L(G)$ není neprázdný !
7. Je dána gramatika $G = (\{ E,F,S,T \} , \underbrace{\{ (,+,*,),a \}}_{T} , S , P)$, kde P obsahuje pravidla :

   $p_1 : S \Rightarrow (E)$, $p_2 : E \Rightarrow E + T \mid T$, $p_3 : T \Rightarrow T * T \mid F$, $p_4 : F \Rightarrow S \mid a$.

   Určete typ gramatiky a zkonstruujte slova : $w_1 = a + a$, $w_2 = a * (a + a)$, $w_3 = a + a * (a + a)$, $w_4 = (a + a + a) * (a + a)$.