THE ALGORITHM FOR REDUCING THE NUMBER OF THE NONTERMINAL SYMBOLS IN FEOL SYSTEMS

Ivana RUDOLFOVÁ, Master Degree Programme (5)
Dept. of Information Systems, FIT, BUT
E-mail: xrudol00@stud.fit.vutbr.cz

Supervised by: Dr. Alexander Meduna

ABSTRACT

The paper introduces FETOL grammars with permitting conditions. The productions of these grammars do not have only the forbidding conditions but also another attached strings, called permitting conditions. The derivation step in these grammars can proceed, when the sentential form contains all substrings from the multiset of permitting conditions. This multiset is created in each derivation step according to the productions, which are used for this derivation. The paper also presents the algorithm for reducing the number of the nonterminal symbols in FEOL (FETOL) systems. This algorithm produces a new FEOL (FETOL) grammar with permitting conditions, which generates the same language as the original system.

1 ÚVOD

ETOL systémy se zakazujícími podmínkami (FETOL grammars) tvoří významnou skupinu paralelních gramatik. Umožňují rychle a efektivně popsat různé formální jazyky a jejich použití je velmi intuitivní. Tyto gramatik však často obsahují velké množství nontreminálních symbolů, což muže být například při strojovém zpracování této gramatiky nevýhodné. Algoritmus uvedený v tomto článku umožňuje zredukovat počet nonterminálních symbolů na pět (0, 1, 2, 0' ,1'). Pro výslednou gramatiku by bylo možné použít EOL systém pouze se zakazujícími podmínkami, ale většina pravidel tohoto systému by obsahovala velké množství nových zakazujících podmínek. Vzhledem k tomu, že tyto nové zakazující podmínky lze nahradit pouze jedinou povolující podmínkou, je pro výslednou gramatiku použit FEOL systém s povolujícími podmínkami.

2 DEFINICE

ETOL systém je t+3-tice, $G=(V,T,P_1,\dots,P_t,S)$, $t \geq 1$, kde V je konečná abeceda všech symbolů, T je abeceda terminálních symbolů $(T \subset V)$ a S je axiom systému $S \in (V-T)^+$. Všechna P_i jsou konečné množiny přepisujících pravidel tvaru: $a \rightarrow x$, kde $a \in V$ a $x \in V^*$.

Nechť $u,v \in V^*$, $u = a_1 a_2 ... a_q$, $v = v_1 v_2 ... v_q$, $q = |u|$, $a_j \in V$, $v_j \in V^*$ a $p_1, p_2, ..., p_q$ je sekvence přepisovacích pravidel tvaru $p_j = a_j \to v_j \in P_i$ pro všechna $j = 1,...,q$ a nějaké $i \in \{1,...,t\}$. Potom řetězec *v* je přímo derivován z řetězce *u* za použití pravidel $p_1, p_2, ..., p_q$: $u \Rightarrow_G v\lfloor p_1, p_2, ..., p_q \rfloor$. Jazyk generovaný touto gramatikou L(G) je definován takto: $L(G) = \{w \in T^* : S \Rightarrow_G^* w\}$.

FETOL systém s povolujícími podmínkami je t+4-tice, $G = (V, T, P_1, ..., P_t, S, Z)$, $t \geq 1$, kde *V* je konečná abeceda všech symbolů, *T* je abeceda terminálních symbolů $(T \subset V)$ a S je axiom systému $S \in (V - T)^+$. Všechna P_i jsou konečné množiny přepisujících pravidel tvaru: $(a \to x, F, D)$, kde $a \in V$, $x \in V^*$ a konečné množiny $F \subseteq V^+$, $D \subseteq V^+$. Z je multimnožina povolujících podmínek. Nechť $u,v \in V^*$, $u = a_1 a_2 ... a_q$, $v = v_1 v_2 ... v_q$, $q = |u|$, $a_j \in V$, $v_j \in V^*$ a $p_1, p_2, ..., p_q$ je sekvence přepisovacích pravidel $p_j = (a_j \to v_j, F_j, D_j) \in P_i$, pro všechna $j = 1,...,q$ a nějaké $i \in \{1,...,t\}$, taková že $sub(u) \cap \bigcup_{j=1}^{q} F_j = \emptyset$, $Z = \bigcup_{j=1}^{q} D_j$. Potom řetězec *v* je přímo derivován z řetězce *u*, jestliže $sub(u) \cap Z = Z$, jinak řetězec *u* není derivován danou gramatikou. Jazyk generovaný touto gramatikou *L(G)* je definován takto: $L(G) = \{x \in T^* : S \Rightarrow_G^* x\}$.

3 ALGORITMUS PRO REDUKCI POČTU NONTERMINÁLNÍCH SYMBOLŮ V FEOL (FETOL) SYSTÉMECH

- **Vstup:** libovolný FEOL systém G=(V, T, P, S) s min. 6 nonterminály (aby redukce měla smysl)
- **Výstup:** FEOL system s povolujícími podmínkami G’=(V’, T, P’, S’, Z) s 5 nonterminály {0,1,2,0’,1’)
- **Postup:**
  for all $a \in (V - T)$: find a shortest unique binary code *h(a).*
  for all $p \in P$, *p* is of a form $p = (a \to v, F, D)$:
    if $a \in T$ then add $(a \to h(v),\ h(F),\ \emptyset)$
      else add $(2 \to h(a)^{|} \cdot h(v),\ h(F),\ h(a)^{|} \cdot h(a)^{|})$
      where $v = v_1 v_2 ... v_q$, $(h(v_i) : v_i \in T) = v_i$ and $(h(v_i) : v_i \in (V - T)) = h(v_i) \cdot 2$,
        $F = f_1 f_2 ... f_n$, $(h(f_i) : f_i \in T) = f_i$ and $(h(f_i) : f_i \in (V - T)) = h(f_i) \cdot 2$,
  add $0 \to 0'$, $1 \to 1'$, $0' \to \varepsilon$, $1' \to \varepsilon$ to *P’*.
  $S' = h(S)$ where $S = s_1 s_2 ... s_n$, $(h(s_i) : s_i \in T) = s_i$ and $(h(s_i) : s_i \in (V - T)) = h(s_i) \cdot 2$,
- **Popis postupu:**

1. Pro všechny nonterminální symboly vstupního systému sestavíme binární kódy. Pro kódování použijeme libovolné injektivní zobrazení h(a) tak, aby délka kódu byla co nejmenší. Délka kódu: $l = \lceil \log_2 N \rceil$
2. Vytvoříme nová pravidla systému takto:

- pravidla pro terminální symboly: levá strana zůstane zachována, pravou stranu upravíme tak, že nonterminální symboly nahradíme jejich binárními kódy doplněnými zprava symbolem 2 (h(a).2).
- pravidla pro nonterminální symboly nahradíme pravidly pro symbol 2, kde pravá strana pravidla bude obsahovat konkatenaci kódu nonterminálního symbolu, který byl na levé straně pravidla, upravený do podoby h(a)` (0≈0`, 1≈1`), a pravé strany pravidla, kde nonterminální symboly budou nahrazeny jejich binárními kódy doplněnými z pravé strany symbolem 2.

3. přidáme pravidla: 0→0', 1→1', 0'→ε, 1'→ε.
4. Vytvoříme nový axiom systému: nonterminály v původním axiomu nahradíme jejich binárními kódy konkatenovanými se symbolem 2.
5. Upravíme zakazující podmínky:

- nonterminální symboly v původních zakazujících podmínkách nahradíme jejich kódy konkatenovanými se symbolem 2 (h(a).2).

6. Vytvoříme povolující podmínky k novým pravidlům pro symbol 2:

- podmínkou bude dvojitá konkatenace kódu nonterminálního symbolu, který byl na levé straně původního pravidla, upraveného na $h(a)'$: $h(a)'.\ h(a)'$

4 ZÁVĚR

Uvedený algoritmus je použitelný pro libovolný FEOL (FETOL) systém s více než 5 nonterminálními symboly, aby transformace měla smysl. Obecně je tento algoritmus použitelný i pro EOL systémy, ale pro systémy se zakazujícími podmínkami má větší význam, protože tyto systémy obvykle obsahují více nonterminálních symbolů. Výhodou uvedeného algoritmu je to, že neprodlužuje kontext původního systému. Správné použití jednotlivých pravidel (kontext) je zajištěno pomocí povolujících podmínek. Jestliže pravidla nejsou použita tak, aby gramatika generovala požadované řetězce, derivace se zablokuje, protože derivovaný řetězec neobsahuje podřetězce odpovídající požadovaným povolujícím podmínkám.

LITERATURA

[1] Rozenberg, G., Salomaa, A.: The mathematical theory of L systems, Academic Press , London 1980, ISBN/ISSN 0125971400

[2] Meduna, A., Švec, M.: Forbidding ETOL Grammars, Elsevier Science (2003)