

TREE CONTROLLED REGULATED GRAMMARS

Martin Kunštátský

Bachelor Degree Programme (1), FIT BUT

E-mail: xkunst01@stud.fit.vutbr.cz

Supervised by: Jiří Koutný

E-mail: ikoutny@fit.vutbr.cz

ABSTRACT

Regulated grammars enable us to generate languages of classes stronger than context-free, using all the comfort of context-free grammars. An example of regulated context-free grammars are grammars regulated by restriction placed on its derivation tree. There are many ways of controlling context-free grammars by its derivation tree, most of them result from two basic principles of such control: horizontal control and vertical control.

1 ÚVOD

Řízené gramatiky jsou formalismem, který si získává v počítačové praxi stále silnější pozice. Hlavním důvodem je skutečnost, že umožňují pouhým rozšířením bezkontextových gramatik o možnost omezovat nějakým způsobem použití derivačních pravidel generovat až jazyky rekurzivně spočetné. Navíc nemají nevýhody ostatních gramatik generujících tyto jazyky (neomezených gramatik, případně i kontextových gramatik) spočívající zejména v neexistenci vhodné dostatečně jednoduché grafové struktury popisující derivace dané gramatiky pro konkrétní řetězec.

Jednou ze skupin řízených gramatik jsou gramatiky stromové. Jedná se o řízené gramatiky, u nichž dochází k omezování jejich derivačního stromu.

Poznámka: V textu předpokládám znalost základních pojmů z oblasti teorie formálních jazyků, jako jsou termíny *řízená gramatika (regulated grammar)*, *derivační strom*, ... Vysvětlení významu těchto pojmů naleznete mimo jiné v referencované literatuře.

Má bakalářská práce spočívá v nastudování a shrnutí teorie gramatik řízených derivačním stromem a v implementaci aplikace demonstrující jejich vlastnosti. Kvůli omezením na rozsah příspěvku se zde zabývám pouze teoretickou částí.

2 ROZBOR

Jednou z možností, jak řídit bezkontextové gramatiky, je doplnit je o možnost omezování jejich derivačního stromu. V zásadě existují dva základní principy takové kontroly:

- horizontální kontrola derivačního stromu, tj. kontrola vlastností řetězců vzniklých zřetěžením jednotlivých symbolů na ležících jedné *úrovni* derivačního stromu.

- vertikální kontrola derivačního stromu, tj. kontrola vlastností řetězců vzniklých zřetěžením jednotlivých symbolů ležících *na cestě* mezi kořenovým uzlem a terminálem.

Kontrola řetězců získaných z derivačního stromu obvykle spočívá v ověření určité vlastnosti daného řetězce, obvyklá požadovaná vlastnost je, aby daný řetězec patřil do určitého konkrétně definovaného (například bezkontextového) jazyka.

2.1 HORIZONTÁLNÍ KONTROLA DERIVAČNÍHO STROMU

Horizontální kontrola derivačního stromu spočívá v kontrole úrovní v derivačním stromu. Nabízí se zde množství možností, jak konkrétně tuto kontrolu provádět. Je možné kontrolovat řetězce pro každou úroveň derivačního stromu zvlášť, případně je možné zřetězit všechny řetězce pro jednotlivé úrovně do jednoho a poté kontrolovat tento jediný řetězec.

Příkladem řízených gramatik s horizontální kontrolou derivačního stromu jsou *stromem řízené gramatiky* navržené v práci *Tree Controlled Grammars* [2] K. Culikem a H. Maurerem.

Stromem řízená gramatika H je dvojice $H = (G, R)$, kde $G = (N, T, P, S)$ je bezkontextová gramatika (tzv. *bázová gramatika*) a $R \subset (N \cup T)^*$ je regulární množina (*řídící množina*).

Tato gramatika generuje řetězec x právě tehdy, když existuje derivační strom gramatiky G pro řetězec x takový, že pro každou úroveň tohoto stromu s výjimkou poslední platí, že řetězec získaný zřetěžením všech symbolů na dané úrovni zleva doprava je obsažen v řídící množině R .

Lze dokázat, že každý rekurzivně spočetný jazyk může být generován stromem řízenou gramatikou ([2]).

Příklad stromem řízené gramatiky: $H = (G, R)$,

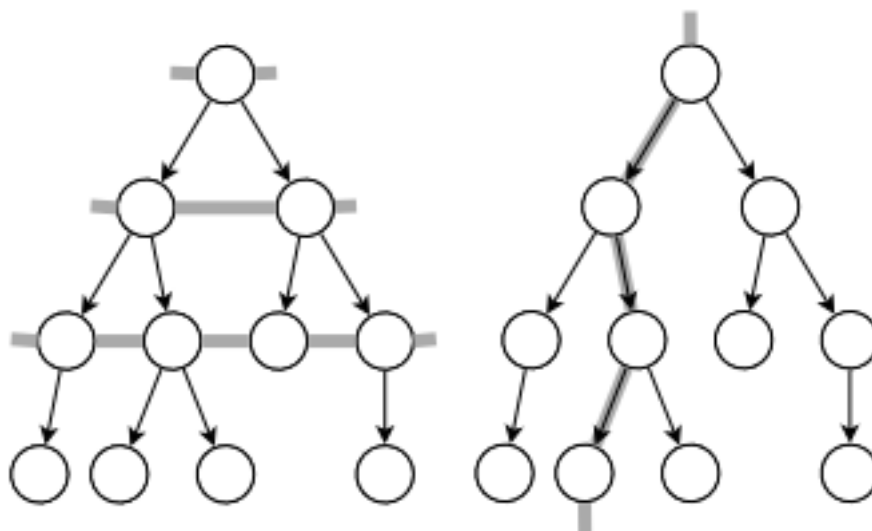
$$\begin{aligned} G &= (\{S, A, B, C\}, \{a, b, c\}, P, S) \\ P &= \{S \rightarrow ABC, \\ &\quad A \rightarrow aA, A \rightarrow a, \\ &\quad B \rightarrow bB, B \rightarrow b, \\ &\quad C \rightarrow cC, C \rightarrow c\} \\ R &= \{S, ABC, aAbBcC\} \end{aligned}$$

Tato gramatika generuje jazyk $L(H) = \{a^n b^n c^n : n \geq 1\}$, který zřejmě není bezkontextový.

2.2 VERTIKÁLNÍ KONTROLA DERIVAČNÍHO STROMU

Vertikální kontrola derivačního stromu spočívá v kontrole cest v daném derivačním stromu, tj. v kontrole řetězců vzniklých zřetěžením symbolů ležících *na cestě* mezi počátečním neterminálem (kořenovým uzlem) a jedním z terminálů. Možnosti jsou: kontrolovat, jestli alespoň jedna cesta v daném derivačním stromu má požadovanou vlastnost, jestli více cest (dvě, tři, ..., ale ne všechny) má tuto vlastnost, případně jestli všechny cesty mají tuto vlastnost. Další možnost spočívá v konkatenaci jednotlivých řetězců pro všechny či jen některé cesty v daném derivačním stromu a kontrola tohoto jediného řetězce.

Příkladem řízené gramatiky s vertikální kontrolou derivačního stromu je *stromová gramatika s řízenou cestou*, jež byla popsána v práci *A New-Old Class of Linguistically Motivated Regulated Grammars* [3]. Jedná se o gramatiku s kontrolou jediné cesty v derivačním stromu.



Obrázek 1: Principy kontroly derivačního stromu: horizontální (vlevo) a vertikální (vpravo)

Gramatiky s vertikální kontrolou více cest v derivačním stromu nebyly (pokud vím) dosud dostatečně vědecky zpracovány.

Gramatika s řízenou cestou (*path-controlled grammar*) H je dvojice $H = (G, G')$, kde $G = (N, T, P, S)$ a $G' = (N', N \cup T, P', S')$ jsou bezkontextové gramatiky.

Tato gramatika generuje řetězec x právě tehdy, když je tento řetězec generován gramatikou G a zároveň existuje derivační strom gramatiky G pro řetězec x , který obsahuje alespoň jednu cestu popsanou řetězcem generovaným gramatikou G' .

3 ZÁVĚR

Gramatiky řízené derivačním stromem jsou velmi perspektivní skupinou řízených gramatik. Existuje velké množství možností, jak provádět řízení bezkontextové gramatiky prostřednictvím jejího derivačního stromu. Ne všechny tyto možnosti byly dosud dostatečně prozkoumány.

REFERENCE

- [1] Meduna, A.: *Automata and Languages*, Springer-Verlag, London, 2000, ISBN 1-85233-074-0
- [2] Culik, K. a Maurer, H. A.: *Tree Controlled Grammars*, Computing, Vol. 19, pp. 129-139, 1977
- [3] Marcus, S., Martín-Vide, C., Mitrana, V., Păun, Gh.: *A New-Old Class of Linguistically Motivated Regular Grammars*, Language and Computers, Computational Linguistics in the Netherlands, pp. 111-125, 2000
- [4] Navrátil, P.: *Syntaktická analýza založená na řízených gramatikách*, Diplomová práce, Vysoké učení technické v Brně, Fakulta informačních technologií, Brno, 2004