

О странных деревьях

Semyon Grigorev

25 декабря 2016 г.

Контекстно-свободные грамматики широко распространены как формализм описания языков соответствующего класса. Существует строгое расширение контекстно-свободных — конъюнктивные и булевы грамматики — грамматики, допускающие кроме дизъюнкции (альтернативы, $|$) также конъюнкцию и отрицание. Данный формализм позволяет описывать более широкий класс языков, однако сложен. Давайте рассмотрим язык $L_0 = \{\omega\omega \mid \omega \in \{a,b\}^*\}$, который не является контекстно-свободным. Для него в работе Александра Охотина “Boolean Grammars” предложена булева грамматика (обозначим её G_0):

$$\begin{aligned} S &\rightarrow \neg AB \ \& \ \neg BA \ \& \ C \\ A &\rightarrow XAX \mid a \\ B &\rightarrow XBX \mid b \\ C &\rightarrow XXC \mid \varepsilon \\ X &\rightarrow a \mid b \end{aligned}$$

Очень просто, правда? Задание для самопроверки — убедитесь, что данная грамматика действительно описывает язык L_0 . Теперь давайте рассмотрим язык $L_1 = \{\omega c \omega \mid \omega \in \{a,b\}^*\}$, который получается из языка L_0 простым добавлением разделителя ‘с’, что наводит на мысль, что грамматики этих языков должны быть очень похожи. Давайте попробуем получить грамматику для L_1 из грамматики G_0 . Не уверен, что это очень простая задача, поэтому предъявим грамматику для этого языка, приводимую в работе “”. Любопытный читатель может попробовать построить дерево вывода цепочки $aabsaab$ в данной грамматике, а так же получить из неё грамматику для языка L_0 .

$$\begin{aligned}
S &\rightarrow C \ \& \ D \\
C &\rightarrow aCa \mid aCb \mid bCa \mid bCb \mid c \\
D &\rightarrow aA \ \& \ aD \mid bB \ \& \ bD \mid cE \\
A &\rightarrow aAa \mid aAb \mid bAa \mid bAb \mid cEa \\
B &\rightarrow aBa \mid aBb \mid bBa \mid bBb \mid cEb \\
E &\rightarrow aE \mid bE \mid \varepsilon
\end{aligned}$$

Не самое очевидный результат, не правда ли? Некоторые трудности в работе с булевыми грамматиками отмечает и Andrew Stevenson в своей диссертации в разделе “Managing Grammar Complexity”.

Давайте вместо букв оперировать более сложными объектами (предпосылки к этому есть в гиперграмматиках, где оперируют множествами). Деревья. Созвучны с языками деревьев, однако там всё выражено в терминах функция для построения новых узлов, а здесь в терминах всё тех же грамматик. Строгое расширение КС. Видна лемма о накачке. Структура грамматики более простая.

Для примера приведём грамматику для языка L_0 .

Из неё можно получить грамматику для языка L_1 действуя согласно интуиции (добавив “коробочку” для буквы c в середину).

Примеры для приведённых ранее языков. Ещё для некоторых.

Идея синтаксического анализа: левые и правые буквы.

Проблемы: композиции сложных участков не выразимы непосредственно в предложенном формализме. Это тема для дальнейших исследований. Кажется, есть некоторый “переход”: только парные скобки регулярны, однако полноценные скобочные последовательности уже нет.

Список литературы

- [1] Alexander Okhotin, “Parsing by matrix multiplication generalized to Boolean grammars”, *Theoretical Computer Science*, V. 516, p. 101–120, January 2014