ПРИМЕНЕНИЕ СИНТАКСИЧЕСКИ ОРИЕНТИРОВАННОГО УПРАВЛЕНИЯ ОБРАБОТКОЙ ДАННЫХ В ЛОГИЧЕСКИХ СИСТЕМАХ

Л. Н. ФЕДОРЧЕНКО

Санкт-Петербургский институт информатики и автоматизации РАН

СПИИРАН, 14-я линия ВО, д. 39, Санкт-Петербург, 199178

<lnf@iias.spb.su>

УДК 683

Федорченко Л. Н. **Применение синтаксически ориентированного управления обработкой данных в логических системах** // Труды СПИИРАН. Вып. 3, т. 2. — СПб.: Наука, 2006.

Аннотация. *Рассматривается пример применения контекстно-свободных грамматик в регулярной форме для представления логических схем вывода для совместного использования систем компьютерной алгебры и систем компьютерной поддержки доказательства теорем. Представлена функциональная структура программного модуля SynGT.* — Библ. 9 назв.

UDC 683

Fedorchenko L. N. **An Application of Syntax Directed Control of Data Processing in Logic Systems** // SPIIRAS Proceedings. Issue 3, vol. 2. — SPb.: Nauka, 2006.

Abstract. *We consider an application of CF grammar with regular expressions in right hand side of the rules for representation and processing in computer algebra and proof assistants. The program SynGT has been presented*. — Bibl. 9 items.

1. Введение

В последние годы различные математические коллективы направили свои усилия на создание программного обеспечения – систем поддержки аналитических вычислений и возможных доказательств истинности выводов. Такие системы либо встраиваются в уже существующие системы компьютерной алгебры, либо представляют собой автономные программы со своими методами представления данных весьма специальной структуры, языками, позволяющие манипулировать с ними, и разнообразными библиотеками эффективных функций для выполнения базисных операций.

В качестве перспективной формальной базы для совместного использования систем компьютерной алгебры и систем компьютерной поддержки доказательства теорем в СПИИРАН разрабатывается подход, основанный на использовании формальных грамматик и синтаксических методов для представления схем логических выводов и вычислительных схем. Современные системы компьютерной поддержки доказательства теорем используют расширяемые системы правил, состоящих из двух частей. Одна часть содержит некоторое постоянное логическое ядро, другая — состоит из правил, задаваемых пользователем. Поскольку набор и состав этих правил может изменяться, необходимы программные средства, позволяющие быстро настраиваться на вводимые изменения и дополнения в схемы правил. В основе технологии с использованием синтаксических методов лежит принцип раздельного описания данных и методов их обработки. Это придает программным приложениям модульность, гибкость и облегчает их интеграцию в другие программные системы. Формальной моделью используемых синтаксических методов являются регулярные выражения, описывающие входной язык, и конечный автомат – в качестве адекватного средства его обработки. Теоретическим обоснованием этой модели являются

теорема Клини [6,7] и её следствия о том, что класс регулярных множеств является минимальным классом, содержащим все конечные множества, замкнутым относительно операций объединения, конкатенации и замыкания; регулярные множества распознаются конечными автоматами и представляются регулярными выражениями. Использование этой модели в обработке данных возможно лишь в применении к регулярным языкам, например, в лексическом анализаторе компилятора Yacc.

На регулярной модели была сформирована версия методики трансляции языков, примененная в компиляторе языка Алгол, а затем она была обобщена разными коллективами разработчиков трансляторов и применялась в ряде других проектов при построении компиляторов. Естественным обобщением регулярной модели является КС грамматика в регулярной форме (КСР-грамматика), строгое определение которой дано в ряде статей [8,9].

2. Управляющая КСР-грамматика

Регулярные выражения можно использовать в качестве средства для определения бесконечных языков, если правила грамматики языка задавать в виде множества пар вида (A,r), где A — нетерминальный символ грамматики (нетерминал), а r — регулярное выражение, с помощью которого можно вычислять множество слов (цепочек), непосредственно выводимых из нетерминала A. Исторический опыт в создании и реализации различных языков программирования показал практическую целесообразность такого подхода. К примеру, языки программирования типа Паскаль определяются посредством КС грамматики в регулярной форме.

Следуя общепринятой терминологии [7], грамматики, содержащие правила вида (A,r), называются КС грамматиками в регулярной форме. Каждое правило такой грамматики может интерпретироваться как множество правил вида $\{A \to x \mid x \in L(R) : L(R) \in R(V)\}$, где V — объединение алфавитов (словарей) терминальных и нетерминальных символов, $V = N \cup T$.

Следующим шагом на пути обобщения грамматик является преобразование КСР-грамматики в *управляющую КСР-грамматику,* которая по сравнению с обычной имеет дополнительные словари — словари контекстных символов (семантик) и предикатов, а правые части ее правил – регулярные выражения относительно символов всех ее словарей. С контекстными символами связываются некоторые преобразования вычислительной среды, а предикаты проверяют выполнение требуемых условий при текущем состоянии вычислений. На практике семантики служат для вставки необходимых обрабатывающих кодов в компилируемый по данной грамматике текст. Эти преобразования и предикаты вместе с данными, на которых они определены, задаются описанием вычислительной среды, которое подобно описанию библиотеки (пакета) процедур без параметров и логических функций, ассоциированных с предикатами.

Например, с помощью семантик можно проверять контекстные условия идентификации индикаторов операций и типов, определяемых пользователем (если это разрешено в языке). Индикаторы определяются в описаниях операций, приоритетов и в описаниях типов. Различаются определяющие и использующие вхождения индикаторов. Определяющее вхождение индикатора — это его вхождение в соответствующее описание типа или операции, и именно оно определяет априорный тип индикатора во всех его использующих вхождениях.

Труды СПИИРАН. 2006. Вып. 3. Т. 2. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 2. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

С другой стороны, если в языке есть возможность определения приоритетов операций, то приоритет индикатора бинарной операции описывается в конструкции описание приоритета. Вхождение индикатора операции в конструкцию описание приоритета называется определяющим приоритет вхождением индикатора. В этом случае контекстное условие идентификации индикаторов операций состоит в том, что любое использующее вхождение индикатора операции и определяющее его вхождение должны идентифицировать одно и тоже определяющее приоритет вхождение такого же индикатора. Наличие нескольких определяющих вхождений для одного индикатора (или идентификатора) в одном блоке или их отсутствие вовсе является нарушением контекстных условий идентификации и рассматривается как контекстная синтаксическая ошибка.

Такого рода семантические проверки могут быть внесены заранее в управляющую грамматику в виде семантик.

Таким образом, можно рассматривать управляющую КСР грамматику, представляющую собой совокупность множеств $G_R = (N, T.\Sigma, \Pi, P, S)$, где N – множество нетерминалов, T — множество терминалов, Σ — множество семантик, Π — предикатов, P — множество КСР-правил, $P = \{A : R_A \mid A \in N, R_A$ — регулярное выражение над символами из множества $N \cup T \cup \Sigma\}$, S — начальный нетерминал грамматики.

КСР-правила записываются в виде:

<нетерминал>:<регулярное выражение>.

Не умаляя общности, считаем, что для каждого нетерминала имеется единственное определяющее его правило. Для контроля правильности записи во время набора текста грамматики в системе SynGT (Syntax Graph Transformations) реализована подсветка различными цветами метасимволов (то есть знаков операций регулярных выражений, скобок и ограничителей), терминалов, нетерминалов, семантик, предикатов и комментариев, семантики специфицируются специальным символом.

Управляющая грамматика порождает синтаксическое управление — множество управляющих цепочек, состоящих из терминальных, контекстных и предикатных символов. Каждая управляющая цепочка (слово) определяет предложение входного языка, представленное ее терминальными символами, со всеми контекстными свойствами и семантикой, выраженной через контекстные символы и их интерпретацию в вычислительной среде. Механизм порождения в КСР-грамматиках отличается от механизма вывода в КС грамматиках. Регулярные выражения, составляющие правые части правил, трактуются как формулы над множествами слов в алфавите терминальных и контекстных символов. Точное определение этого дается в статье [8 стр.2].

Управляющая КСР-грамматика может трактоваться двояко: как анализирующая или как порождающая. В зависимости от этого по-разному определяется интерпретация управляющих цепочек.

В процессе построения анализирующего процессора управляющая КСР-грамматика трансформируется в синтаксическую граф-схему, промежуточную форму представления грамматики и эквивалентную с точки зрения порождения языка.

Синтаксическая граф-схема — это ориентированный псевдограф, состоящий из нескольких компонент связности, по одной на каждое правило грамматики.

Каждая компонента имеет начальную вершину, в которую не входит ни одна дуга, помеченную символом вида «begin-*A*», конечную вершину, из которой

Труды СПИИРАН. 2006. Вып. 3. Т. 2. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 2. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

ни одна дуга не выходит, помеченную символом вида «end-*A*», где *A* — определяемый нетерминал, внутренние вершины, помеченные символами терминалов и нетерминалов — операндов регулярного выражения правой части правила, и дуги, помеченные контекстными символами — операндами того же регулярного выражения (контекстные символы не могут быть операндами итераций).

Ориентированные дуги соединяют вершины таким образом, что множество следов всех маршрутов от начальной к конечной вершине образуют регулярное множество цепочек, представленное регулярным выражением. Любое регулярное выражение представимо в этой форме.

В системе SynGT синтаксическая граф-схема отображается таким образом, что никакие дуги не пересекаются, а текстовое регулярное выражение вынесено вниз, на линейку Desktop.

Приведем пример грамматического представления схемы логического правила (в секвенциальном исчислении).

Рассмотрим правило:

$$\frac{\Gamma \mapsto A; B\Gamma' \mapsto C}{\Gamma(A \Rightarrow B)\Gamma' \mapsto C}$$ (введена импликация слева)

С грамматической точки зрения терминалами здесь являются символы ; , $\mapsto, \Rightarrow$, (,) , Γ, Γ' — нетерминалы для списка формул, A, B, C — нетерминалы для формул. Сами по себе схемы правил являются регулярными выражениями над терминалами и нетерминалами, причем, на этапе вывода до терминальных выражений используется «почти КС»-грамматика (используются параллельные продукции). Конкретное применение правила получается применением грамматических продукций к нетерминалам, в результате чего нетерминалы Γ, Γ' превращаются в конкретные списки формул (параметрические формулы), A, B, C — в конкретные формулы.

Построение схемы вывода требует решения грамматических уравнений (обычно несложных) для проверки возможности отождествления заключения одних правил с посылками других. Это позволяет использовать уже существующие системы грамматического анализа, такие как система SynGT, разработанная в СПИИРАН.

3. Информационная структура SynGT

Ниже приведена структура основных информационных компонент системы и возможных операций с ними.

3.1. Основные компоненты системы

В качестве основных информационных компонент SynGT выделяются:

- SynGT-документ;
- рабочее поле (desktop) с областями;
- синтаксическая граф-схема;
- подграф;
- дуга (семантика);
- нетерминальная вершина;
- терминальная вершина;

Труды СПИИРАН. 2006. Вып. 3. Т. 2. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 2. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

- правило грамматики.

SynGT-документ — объект, содержащий информацию о синтаксической граф-схеме, состоянии рабочего поля, о соответствующей синтаксической граф-схеме КСР-грамматики и комментарии к текущему SynGT-документу.

3.2. Операции с SynGT-документом

SynGT-документ можно:

- создать;
- загрузить в систему с носителя (МД);
- удалить из системы;
- переименовать;
- распечатать в специальном формате GRW;
- записать на носитель (МД).

3.3. Рабочее поле

Рабочее поле **(desktop)** может занимать часть или полный экран дисплея и предназначено для непосредственной работы с синтаксической граф-схемой. Рабочее поле состоит из кнопок:

- устранения левой рекурсии;
- устранения обеих рекурсий;
- устранения правой рекурсии;
- подстановки вместо нетерминала его порождения;
- выделения правила;
- анализа грамматики;
- минимизации грамматики.

и областей для:

- редактирования изображения графа (нетерминальных вершин);
- словаря нетерминалов;
- отображения полного правила грамматики;
- словаря терминальных символов;
- словаря дуг с семантиками;
- инструментальной линейки;
- комментария.

Синтаксическая граф-схема в SynGT — это совокупность ориентированных графов (компонент), не имеющих общих вершин и обладающих следующими свойствами:

- каждая вершина может иметь символьную метку;
- каждая дуга может иметь символьное выражение (семантику),
- существует только одна вершина графа, из которой дуги только выходят (входная вершина);
- существует только одна вершина графа, в которую дуги только входят (выходная вершина);
- две любые вершины графа могут быть соединены между собой однонаправленной дугой.

Для *эквивалентной трансформации* графа предусмотрены следующие операции:

Труды СПИИРАН. 2006. Вып. 3. Т. 2. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 2. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

- выделение подграфа и его свертка в нетерминальную вершину (редукция);
- вставка вместо одной нетерминальной вершины ее порождения — подстановка;
- устранение левой рекурсии;
- устранение обеих рекурсий;
- устранение правой рекурсии;
- неявная подстановка нетерминала;
- анализ грамматики на предмет крайних рекурсий и самовложенных нетерминалов;
- минимизация представления грамматики;
- копирование графа в буфер обмена;
- вставка графа из буфера обмена;
- изменение положения вершин в окне;
- распечатка графа в виде картинки в точечном формате;
- изменение масштаба изображения.

В графе может быть выделен подграф. Если выделенный подграф не является регулярным, то он дополняется до минимально объемлющего регулярного подграфа, и с таким подграфом предусмотрены следующие операции:

- преобразование в нетерминальную вершину и формирование нового правила грамматики;
- удаление;
- отмена удаления;
- перемещение в окне;
- копирование в буфер обмена;
- вставка из буфера обмена;
- отмена вставки из буфера обмена.

Дуга, соединяющая две вершины графа, имеет семантику, комментарий и стиль изображения. Семантика — это строка символов, которую можно создать, редактировать, копировать, выделять, удалять частями или полностью, производить вставку символов в любое место строки. Стиль дуги — это множество чисел, определяющих толщину линии, её вид, цвет и форму стрелки. В системе SynGT возможно накопление нескольких фиксированных стилей.

Нетерминальная вершина — вершина, соответствующая правилу грамматики. Нетерминальную вершину можно создать, изменить стиль изображения, а также применить для неё все операции, допустимые для подграфа.

Терминальная вершина — вершина, соответствующая символу алфавита грамматики. Символ алфавита здесь — это строка символов, над которой возможны все операции, допустимые над строкой символов в текстовом редакторе.

Правило грамматики представляется в виде $A : R.$, где A — нетерминальный символ, R — регулярное выражение над терминальными, нетерминальными символами, семантиками и возможно другими (вспомогательными) символами.

Устранение крайних рекурсий (левой и правой) производится серией замен и эквивалетнтых преобразований, ведущих к исчезновению вхождений рекурсивного нетерминала.

Анализ грамматики позволяет выявить все виды рекурсий для каждого правила и помогает при работе с большим набором правил.

Труды СПИИРАН. 2006. Вып. 3. Т. 2. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 2. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

Минимизация грамматики позволяет добиться представления правил при помощи наименьшего количества символов. Минимизированная грамматика строится посредством преобразования исходного множества правил к автоматному виду, затем — минимизации автомата и, наконец, обратного преобразования.

Подстановки, устранение рекурсий, а также минимизация грамматики позволяет регуляризировать исходную грамматику, то есть, представить ее в виде минимального регулярного выражения.

Литература

1. *Fedorchenko L. N., Soloviev S. V, et al.* Syntax Graph Transformations in the system SynGT and their Applications (English). Rapport IRIT/2003-06-R, UMR 5505 CNRS-INP-UPS. 15 p.
2. *Федорченко Л. Н., Соловьев С. В.* Синтаксические преобразования в системе SynGT и их новые приложения // Тезисы докладов VIII Санкт-Петербургской международной конференции «Региональная информатика-2002». Ч. 2. СПб.: СПОИСУ, 2002. С. 50.
3. *Федорченко Л. Н., Флегонтов А. В., Соловьев С. В.* Теория типов, компьютерная алгебра и поддержка доказательств // Труды международной конференции «Региональная информатика-2002». СПб.: СПОИСУ, 2003. С. 108–117.
4. *Флегонтов А. В., Федорченко Л. Н., Соловьев С. В.* Использование теории типов в аналитических вычислениях // Информатика и связь. 2004. № 1. С.10–18.
5. *Hopcroft J. E., Ullman J. D.* Formal languages and their relation to automata. Addison-Wesley pub. comp., 1969. 242 p.
6. *Kleene S. C.* Representation of events in nerve nets and finite automata. In **C**. E. Shannon and J. Mc-. Carthy, editors, Automata Studies, pages 341.
7. *Ахо А., Ульман Дж.* Теория синтаксического анализа, перевода и компиляции. Т. 1. Синтаксический анализ. М.: Мир,1978. 612 с.
8. *Федорченко Л. Н.* Извлечение крайней рекурсии из КСР-грамматики в системе SynGT // Труды СПИИРАН. 2001. Вып. 1, т. 1. СПб: СПИИРАН, 2001. С. 350–359.
9. *Федорченко Л. Н.* О регуляризации контекстно-свободных грамматик // IX Санкт-Петербургская международная конференция «Региональная Информатика-2004 (РИ-2004)», Санкт-Петербург, 22–24 июня 2004 г.: Труды. СПб., 2005. С. 89–95.

Труды СПИИРАН. 2006. Вып. 3. Т. 2. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2006. Issue 3. V. 2. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru