

УДК 004.4'413

ИТЕРАТИВНОСТЬ И АДАПТИВНЫЙ СИНТАКСИЧЕСКИЙ АНАЛИЗ

Кузенкова Е.В. (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»),

Научный руководитель – ассистент Кореньков Ю.Д.

(федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

В данной работе представлен подход к обработке рекурсивных и циклических конструкций в грамматиках при синтаксическом анализе предметно-ориентированных языков. Рассматриваются особенности реализации этапов подготовки и работы синтаксических анализаторов, производных от РВ (PEG) и Эрли (Earley) для контекстно-свободных грамматик.

В рамках различных подходов к задаче автоматизированного синтаксического анализа существует хорошо известная проблема обработки рекурсивных правил в составе грамматики анализируемого текста. Рекурсивность правил грамматики возникает, когда цепочка в правой части некоторого правила начинается или заканчивается нетерминалом, из которого выводима сама данная цепочка. В рамках контекстно-свободных грамматик данная проблема может быть классифицирована на различные сценарии, такие как: левая и правая рекурсия (по положению рекурсивного нетерминала), прямая и косвенная рекурсия (входит ли рекурсивный нетерминал в левую часть того же правила, что и цепочка правой части), скрытая рекурсия (когда фрагмент между рекурсивным терминалом и границей цепочки выводим в пустую строку). Широко используемые на практике генераторы синтаксических анализаторов, такие как ANTLR, зачастую хотя и включают обработку для некоторых специализированных случаев, таких как, например, арифметические выражения и некоторые из выше названных, однако не поддерживают обработку рекурсивных конструкций в общем случае. Это приводит к ограничению выразительной способности грамматик, поддерживаемых синтаксическими анализаторами и усложняет разработку предметно-ориентированных языков.

Предлагается подход к обработке общего случая рекурсии в грамматиках предметно-ориентированных языков, применимый к синтаксическим анализаторам, производным от алгоритмов РВ-анализаторов и анализаторов на базе алгоритма Эрли. На этапе подготовки анализатора подход включает оценку набора выводимых из рекурсивного правила цепочек с целью сбора соответствующей метаинформации о структуре грамматики. Начинается подготовка с обнаружения и классификации соответствующих паттернов в структуре грамматики для случаев левой и правой рекурсии. Для случаев правой рекурсии в выводимых цепочках размечается выход из праворекурсивного правила. Для случаев левой рекурсии в выводимых цепочках размечается выход из леворекурсивного правила и с ним ассоциируется информация о выводимых на пути рекурсии вариантах нетерминалов. На этапе работы анализатора для случая правой рекурсии при обработке правой рекурсии в момент завершения обработки цепочки выполняется итеративная проверка локального контекста разбора на соответствие нетерминалу покидаемого правила. Пока эта проверка успешна, финализируется очередной узел дерева разбора с переходом вверх, что соответствует обычному выходу из правила при работе РВ-анализатора. Для случая левой рекурсии в момент завершения обработки цепочки выполняется разветвление состояния анализатора пропорционально количеству выделенных альтернативных цепочек вывода. Для каждой из них в текущем месте дерева разбора выполняется построение ряда промежуточных узлов, соответствующих нетерминалам из выделенных на этапе подготовки цепочки, с последовательным выходом из этих узлов. Для обоих рассматриваемых семейств алгоритмов данные действия являются локальной контекстной проверкой над метаинформацией о структуре грамматики. Для анализаторов на основе алгоритма Эрли предлагаемый подход применим двумя способами: на

фазе предсказания с раскрытием всех предполагаемых для рекурсивного вывода цепочек в соответствии с правилами обработки альтернатив разбора (так достигается снижение количества итераций внутреннего цикла анализатора) или на фазе сканирования (среди входных цепочек для рекурсивного правила при этом рассматриваются только не циклические) при обработке выходной цепочки (так достигается снижение общего числа обрабатываемых элементов состояния анализатора за счет того, что влияющая на выбор полезных альтернатив разбора цепочка уже была сканирована). Помимо обработки рекурсивных правил, рассматривается так же обработка кванторов со счетчиками за счет контекстных проверок. В предлагаемом подходе как обработка рекурсивных правил, так и обработка кванторов со счетчиками выполняется в рамках одних и тех же механизмов разбора – локальных проверок состояния анализатора. Различаются только условия данных проверок.

Предложенный подход к обработке рекурсивных и циклических конструкций в грамматиках развивает ранее представленное приближение к построению адаптивных синтаксических анализаторов. Он позволяет сохранить их оперативность и значительно расширить полезную функциональность, а рассматриваемые проверки на этапах подготовки и анализа соответствуют принципу локальности, соблюдение которого необходимо для минимизации требующих обновления элементов анализатора в случае внесения изменений в исходную грамматику.

Кузенкова Е.В. (автор)

Подпись

Кореньков Ю.Д. (научный руководитель)

Подпись