

004.434

Синтаксический анализ для динамически расширяемых грамматик на основе PEG

Авторы: Кореньков Ю. Д., Логинов И.П., Лаздин А.В., Университет ИТМО, Санкт-Петербург
Научный руководитель: Лаздин А.В. Университет ИТМО, Санкт-Петербург

На сегодняшний день существуют различные классы программного обеспечения, для которых синтаксический анализ является неотъемлемой частью. К таким программам можно отнести в первую очередь компиляторы и интерпретаторы языков программирования, а так же среды разработки. Кроме этого, существует целая парадигма разработки, для которой разработка ориентированных для конкретной задачи языков является одним из этапов создания решения – программного продукта, решающего конкретную прикладную задачу. Эта парадигма – языково-ориентированное программирование, или расходящаяся разработка.

Ключевым элементом языково-ориентированной разработки является использование предметно-ориентированных языков для моделирования предметных областей, затрагиваемых конструируемым прикладным решением. С помощью таких языков может записываться как само решение, его части, так и данные, которыми это решение может оперировать. Так как каждая прикладная задача чем-то отличается от других, часто это может требовать разработки новых предметно-ориентированных языков, заметно сокращающих объём программного кода и время, требуемое для создания конечного продукта.

Таким образом, минимизация времени итерации разработки предметно-ориентированного языка, времени поддержки изменений в синтаксисе при адаптации к изменяющимся требованиям прикладного решения, времени введения в него новых синтаксических конструкций становится одним из ощутимых требований к процессу разработки предметно-ориентированных языков, а значит и к инструментарию, задействованному в этом процессе.

В рамках данной работы рассматривается построение синтаксического анализатора для текста на основе PEG-грамматик (грамматик типа RB – разбирающее выражение), для которого одним из основных требований является способность интерактивно изменять своё поведение в соответствии с изменениями, вносимыми пользователем, – разработчиком предметно-ориентированного языка, – в грамматику. Грамматика предметно-ориентированного языка при этом рассматривается как модель синтаксической структуры текста не в формальном смысле, а в модельном. Формально, в соответствии с теорией формальных языков и грамматик, грамматики должны соответствовать ряду ограничений в зависимости от классов используемых анализаторов, что приводит пользователя к проблемам конфликтов и неоднозначностей, требующих разрешения посредством модификаций грамматики. В соответствии же с модельным подходом грамматика рассматривается как метамодель, спецификация структуры желаемого синтаксического дерева. При этом синтаксическое дерево для конкретного текста будет являться моделью, устроенной соответственно структуре моделируемой предметной области, а не формальным ограничениям теории.

Разработанный анализатор поддерживает гранулярное обновление и расширение грамматик с частичной верификацией затрагиваемых обновлением элементов внутренней архитектуры. В отличие от генераторов компиляторов, при этом ни на каком из этапов работы не требуется повторного полного анализа всей грамматики.

Для решения типичных для классических PEG-анализаторов проблем типа цикла бесконечной левой рекурсии, в разработанном анализаторе методы PEG-анализа сочетаются с методами TDOP – нисходящего разбора с учётом приоритетов операторов, обобщённого для случая произвольных расширяемых правил.