

ИССЛЕДОВАНИЕ И РЕАЛИЗАЦИЯ АЛГОРИТМА СИНТАКСИЧЕСКОГО АНАЛИЗА

Павлова Е.В. (Университет ИТМО, Санкт-Петербург)

Научный руководитель – Кореньков Ю.Д. (Университет ИТМО, Санкт-Петербург)

Большинство алгоритмов разбора контекстно-свободных грамматик, использующихся в парсерах языков программирования, имеют свои ограничения на принимаемые грамматики. Как правило, они исключают некоторые пограничные случаи для улучшения производительности и упрощения алгоритма обработки текста программ. Это не является проблемой, пока не стоит задача обработки заранее неизвестных грамматик без возможности гарантировать отсутствие во входных грамматиках неопределенности. Данный подход сложно реализовать для задачи обработки в большей степени обобщенных КС-грамматик, например, при конструировании парсеров для предметно-ориентированных языков (DSL). Тем не менее, для данной области применения существуют алгоритмы, которые могут обрабатывать общие случаи КС-языков, с неоднозначными грамматиками и токенами переменной длины, при этом теряя в производительности.

Цель работы – исследование одного из алгоритмов, применимых для обработки наибольшего множества контекстно свободных грамматик - Marpa, а также создание его реализации в контексте возможного последующего применения в качестве алгоритма разбора для среды разработки предметно-ориентированных языков. Данный алгоритм интересен тем, что в каждый момент времени хранит полный набор состояний разбора. Это позволяет применять его для обработки изменений входного текста «на лету», во время редактирования. Хранение полного набора состояний разбора позволяет обращаться к отдельным частям дерева разбора и перестраивать их там, где произошли изменения, без необходимости полного повторного разбора всего текста, что может значительно увеличить скорость обработки изменений исходной строки текста. Также выбранный алгоритм позволяет эффективно реализовать обработку ошибок, не только не останавливая дальнейший разбор входного текста, но и предлагая возможный вариант исправления ошибки на основе списка доступных токенов. Это может быть использовано для автоматического исправления опечаток.

В результате получена реализация парсера КС-языков на основе алгоритма Marpa, который может быть использован как модуль для обработки текста на основе заданной грамматики или встроен в инструмент создания парсеров для предметно-ориентированных языков.