

Abstract

A Declarative Extension of PEGs for Recognizing Context-Sensitive Syntax

TETSURO MATSUMURA^{1,a)} KIMIO KURAMITSU¹

Presented: March 10, 2015

Parsing Expression Grammars are a popular foundation for describing syntax. Unfortunately, several patterns of programming language syntax appearing in practice are hard to recognize with pure PEGs. Notorious examples include typedef name in C/C++, indent-based code block in Python, HERE documentation in many scripting languages. To recognize such patterns, we address a declarative extension to PEGs. The “declarative” extension means no programmed semantic actions, which are traditionally used to realize the extended recognizer behaviors. PEG+L is an extended class of PEG notations, including symbol table handlers and conditional parsing. This paper demonstrates the design and semantics of PEG+L with our experience developing language grammars, including C, Java8, JavaScript, Lua, Ruby, bash, etc. In addition, we present that PEG+L still achieves a mostly linear parsing time.

¹ Yokohama National University

^{a)} matsumura.t.lab@gmail.com