

発表概要

CPEG：正規表現のキャプチャ付き解析表現文法による 型付きのツリー構築

山口 大輔^{1,a)} 倉光 君郎^{1,b)}

2018年10月31日発表

CPEG は regex 風のキャプチャアノテーションを加えた解析表現文法 (parsing expression grammar, PEG) の拡張である。追加されたアノテーションは 2 つ (それぞれ, *capture*, *fold-capture* と呼ばれる) であり, 柔軟な構文木の構築が可能である。さらに, CPEG は任意の入力文字列の解析結果の構文木すべてについて成り立つ構造的な特徴が静的に導出されるように設計されている。これによって, CPEG による構文解析の結果の構文木が意図したとおりの構造であるかを静的に検査することが可能となる。本研究では, 構造的な制約を記述する型として正規表現型に着目した。正規表現型は, 木オートマトンの別表現であり XML スキーマの関連分野でさかんに研究がなされてきた。CPEG の型推論は与えられた文法から正規表現型を導出する。本発表では, 型推論規則について形式的な定義を与え, 健全性, 完全性, 型推論の一意性について証明する。型推論規則は CPEG を構文仕様としてのみならず, 構文木のスキーマに関する仕様としても利用することを可能とする。

Presentation Abstract

CPEG: A Typed Tree Construction from Parsing Expression Grammars with Regex-like Captures

DAISUKE YAMAGUCHI^{1,a)} KIMIO KURAMITSU^{1,b)}

Presented: October 31, 2018

CPEG is an extended parsing expression grammar with regex-like capture annotation. Two annotations (capture and fold-capture) allow a flexible construction of syntax trees from arbitrary parsing patterns. More importantly, CPEG is designed to guarantee structural constraints of syntax trees for any input strings. This reduces an amount of user code to check whether indented elements exist. To represent the structural constraints, we focus on regular expression types, a variant formalism of tree automata, which have been intensively studied in the context of XML schemas. Regular expression type is inferred from a given CPEG by the type inference which is formally developed in this presentation. The soundness, the completeness, and the uniqueness of the type inference enable a CPEG to serve both as a syntactic specification of the input and a schematic specification for the output.

This is the abstract of an unrefereed presentation, and it should not preclude subsequent publication.

¹ 横浜国立大学大学院
Yokohama National University, Yokohama, Kanagawa 240-8501, Japan

^{a)} yamaguchi-daisuke-bf@ynu.jp

^{b)} kuramitsuk@fc.jwu.ac.jp