

## 発表概要

# OPGを使ったアドホックな大規模な文字列データ解析のための並列処理系

リュウ ケイコウ<sup>1,a)</sup> 田浦 健次朗<sup>1</sup>

2019年7月26日発表

アドホックなデータ向けのパーサジェネレータは、逐次に動くパーサを生成するものが多く、プロセッサの計算資源を十分に利用できず、大規模なデータを効率良く解析できない。OPGは、local parsabilityと呼ばれる、入力の途中から構文解析を始めることができるという性質を持ち、文法中にある演算子に着目して考案された形式文法である。OPGを使った既存のパーサジェネレータの実装であるPAPAGENOは、空白文字や改行文字を使って入力を分割するが、そういった文字が十分に含まれないデータを効率的に処理できない。本発表では、OPGに基づいた並列処理系をタスク並列の手法を用いて実装し、今後並列字句解析の手法を取り入れることを検討する。さらに、実装した処理系の性能を、Wikidataの一部のデータをを用いて評価する。

## Presentation Abstract

### A Parallel Ad-hoc Data Processing System for Large Text Data Based on Operator Precedence Grammar

QIHENG LIU<sup>1,a)</sup> KENJIRO TAURA<sup>1</sup>

Presented: July 26, 2019

Most parser generators for ad-hoc data can only generate sequential parsers, which cannot fully utilize the computational resource, and are not suitable for parsing large data. OPGs have a feature named local parsability, which allows us to parse only a part of the whole input string. The existing PAPAGENO implementation, which is a parser generator based on OPGs, uses spaces and newline characters to split the input file, and thus cannot efficiently process the data which does not have such characters sufficiently. In this presentation, we introduce a parallel data processing system based on OPGs which uses the task parallel method, and consider to add parallel lexical analysis to that system. Besides, we evaluate our parser generator using part of the Wikidata dump file.

---

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

<sup>1</sup> 東京大学大学院情報理工学系研究科  
Graduate School of Information Science and Technology,  
The University of Tokyo, Bunkyo, Tokyo 113-8656, Japan

<sup>a)</sup> kko@eidos.ic.i.u-tokyo.ac.jp