

発表概要

OPG を利用したアドホックな並列データ処理系

リュウ ケイコウ^{1,a)} 井原 央翔¹ 田浦 健次朗¹

2018年8月1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 と呼ばれる形式文法はその性質から並列パーサに応用できるが、並列スキャナを必要とする点では、実装が複雑になりがちである。本研究では、通常の OPG よりスキャナを取り除き、マルチコアの特徴を活かせるパーサを生成できるパーサジェネレータの実装を行う。さらに、JSON の文法をスキャナレスな OPG にする試みを行い、その実用性について検討する。OPG は、local parsability と呼ばれる、入力途中から構文解析を始めることができるという性質をもち、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。この性質を利用すれば、最良の場合プロセッサの数に従いスループットが線形に大きくなるようなパーサを作成できる。

Presentation Abstract

Parallel Ad-hoc Data Processing System Using Operator Precedence Grammar

QIHENG LIU^{1,a)} HIROKA IHARA¹ KENJIRO TAURA¹

Presented: August 1, 2018

Most parser generators for ad-hoc data can only generate sequential parsers, which cannot fully use the computational resource, and are not suitable for parsing large data. Though operator precedence grammars (OPGs) can be used by parallel parsers, we still need parallel scanner, which may result in a complex implementation. In this research, we remove the scanner from common OPGs, and implement a parser generator which can generate parsers work on multi-core computers. We also try to represent the grammar of JSON to scannerless OPG, and consider the practical use of the system. OPGs have a feature named local parsability, which allows us to reduce only a part of the whole input string. Thanks to this feature, we can make parsers such that the throughput grows linearly along with the processor counts.

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

¹ 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

^{a)} kko@eidoss.ic.i.u-tokyo.ac.jp