

発表概要

Scheme 言語によるリスト構造に特化した XML データ処理の効率化と記述性の向上

坪井 啓明^{†1} 前田 敦司^{†1} 山口 喜教^{†1}

データの記述方式として XML が普及するのにもない、大規模な XML データを扱う機会が増えてきている。木構造である XML データを処理するのに適した言語としては、木構造を表現可能な S 式を用いてプログラムを記述する Scheme 言語があげられる。しかし、Scheme 言語における現在一般的な XML 処理手法には一長一短があり、実行時の処理効率とプログラムの記述性との双方に優れた手法が存在しないのが現状である。このことに対し、Scheme 言語において木構造を扱うようなプログラミングスタイルを維持しながらメモリ消費量や処理速度を低下させないような処理の実現を目指し、データが必要となるまで処理をしない Lazy な XML 処理手法がこれまでに研究された。また、別の研究において Scheme 言語で木構造を扱う際の記述性を向上させるために、正規木表現を用いて木構造を扱うための言語 `trx` が提案された。本発表では、Scheme 言語による XML データ処理の効率化と記述性の向上の両立を目指し、先行研究によって確立された Lazy な XML データ処理による高い処理効率を活かしたまま、正規木表現 `trx` による記述方式で XML に対する正規木表現マッチングを行う手法を提案する。提案手法では特に、大規模な XML に多くの場合出現すると考えられるリスト構造に着目し、リスト処理に特化した手法とすることで、より高い記述性を有したプログラミングスタイルを目指している。この提案手法を実際に実装し、動作の確認と性能評価を行った。

A New XML Data Processing Method in Scheme That Can Pursue both the Run Time Efficiency and Descriptive Capability

HIROAKI TSUBOI,^{†1} ATSUSHI MAEDA^{†1}
and YOSHINORI YAMAGUCHI^{†1}

It is said that opportunities to treat large XML data are increasing with the

popularity of XML as description format. Scheme language is one of the language that is suitable to handle XML data because it is to describe a program with S-expression which can express tree structure easily. However, there seems not to be superior method which has both of run time efficiency and descriptive capability of the program. To solve the above issues, there are some researches to realize the method that doesn't decrease memory consumption and transaction speed while maintaining programming style handling tree structure in Scheme. This is the XML processing based on the Lazy technique. On the other study, there is another research which aims at improvement of the descriptive capability to handle tree structure in Scheme, `trx` that is language to handle tree structure using regular-tree expressions was suggested. In this presentation, we suggest new method that is using regular-tree expression matching with `trx` while maintaining superior efficiency on Lazy method, in order to satisfy both of processing efficiency and descriptive capability on handling XML. In particular, we paid attention to list structure that is often appeared to large XML, and implemented this method and evaluated the performance of it.

(平成 22 年 10 月 29 日発表)

^{†1} 筑波大学大学院システム情報工学研究科

Graduate School of Systems and Information Engineering, University of Tsukuba