

論文の内容の要旨

論文題目 Regular Expression Types for XML

(XML のための正規表現型)

氏名 細谷 晴夫

XML は木構造データのための単純でかつ柔軟性のある汎用データ形式である。また XML は、そのサブセットをスキーマを用いて定義することができ、それを個別のアプリケーションのデータ交換形式として用いることができる。近年 XML が次世代 Web アプリケーションへ急速に採用されてきているに従い、XML を用いたソフトウェア開発に助けとなるようなよりよいプログラミング言語機能のサポートが求められている。特に、強い要望のある機能は、(1) 出力データが、与えられたスキーマに必ず従うことを保証する静的プログラム解析、(2) 木構造を操作するために便利な言語機能である。

本論文では、XML 処理言語の核となる機能を2つ、すなわち「正規表現型」と「正規表現パターンマッチ」を提案する。正規表現型は、スキーマ言語によく見られる「繰り返し (*)」や「選択 (|)」などの正規表現の記法を型システムに採り入れる。さらに、「意味論」的に定義された部分型をサポートする。本研究では、この部分型がもつ柔軟性が、XML を用いたソフトウェアをスムーズに進化させるのに必要であることを論じる。正規表現パターンマッチは、従来の (関数型言語にあるような) パターンマッチ機構に、上記の繰り返しや選択の演算子を追加したものである。これら演算子を使うことによって、任意の長さを持つ部分木「列」をマッチすることができるようになるので、列の中の任意の位置に一挙に飛んで、そこから欲しいデータを抽出するようなパターンが簡潔に記述できる。本研究ではさらに、これら2つの機能を使った言語の例として XDuce 言語を設計し、その核言語を定式化し、型健全性を証明する。

さらに本研究では、上記の機能を実現する上で発生するアルゴリズム上の問題にも取り組む。部分型の判定問題は、正規ツリーオートマトンの包含判定問題に簡単に帰着できるが、この問題自体が指数時間完全という高い計算量をもつことが分かっている。この問題点に対し本研究では実際の入力に対しては効率的に動作するアルゴリズムを開発する。まず出発点として Aiken と Murphy の集合制約の解消アルゴリズムを用いる。本研究での追加点は (1) 完全性の証明、(2) XML 処理で発生する典型例に対応した実装技法、(3) 予備実験である。予備的実験では、ある程度現実的な例をいくつか試した結果、実用的な効率で動作することを確認した。

正規表現パターンマッチを型付きプログラミング言語に用いる際、パターンマッチに出現する変数の型を自動的に計算し、過度な型注釈をユーザに要求しないようにすることが重要である。そのために本研究では、パターンマッチを取り巻く文脈からパターンの変数へ、型情報を伝播するような型推論アルゴリズムを開発する。このアルゴリズムは、まず型とパターンをツリーオートマトンへ変換し、ツリーオートマトンの閉包演算 (和・積・差) を使って型情報の伝播を行なう。技術的に困難な点は、繰り返し演算子と、選択演算子の「第一マッチ則 (先に出現する選択肢が優先的にマッチする性質)」が相互作用し、これによって解析の停止性と精度を同時に保証することが難しくなるということである。本研究ではこの問題点に対し、閉包演算を明示的に含んだような型の表現を導入することによって解決する。