

## 正規経路式から XPath 式への変換手法の提案

丹野 由宇志

XML(Extensible Markup Language)はデータを柔軟に記述出来るデータのフォーマットとして、様々なシステムで用いられている。大量の XML データを蓄積・管理する場合、DTD(Document Type Definition)などのスキーマ言語を用いてあらかじめ格納するデータの構造を定義しておき、スキーマの定義に妥当なデータを生成・蓄積するのが一般的である。XML データへの問合せ言語として XPath(XML Path Language)がよく使われている。

他の問い合わせ式としては、正規経路式(regular path query)がある。正規経路式は正規表現を用いて記述される問い合わせ式である。XML など様々なデータに適用可能で、これまでに多くの研究がなされている。正規経路式は XPath と比べ記述力が高いが、一般の XML 処理系 (Apache Xalan, REXML など) ではサポートされていない。

そこで、記述力の高く有用な正規経路式を XML の問い合わせ式として扱うために、本研究では、正規経路式から XPath 式への変換を行うアルゴリズムを提案する。ただし、DTD なしに正規経路式から XPath 式に変換することは困難なので、DTD があることを前提とする。また本アルゴリズムで生成する XPath 式は軸を child 軸と descendant-or-self 軸に限定し、ノードテストは要素のみかつ述語は使用しないとする。本アルゴリズムでは、入力された正規経路式をまず DTD に妥当な child 軸のみを用いた式に変換し、次に簡略化出来る部分を descendant-or-self 軸に置き換えた簡潔な XPath 式へ変換する。

DTD がある要素を始点とした閉路を 2 つ以上含む要素を持ち、また複数の要素を含む閉包のある正規経路式が与えられた場合、等価な XPath 式での表現が不可能であるため、正確な変換が困難である。この場合、正規経路式の意味を包含する近似的な XPath 式を生成することで対応する。

変換の方法としては、まず与えられた DTD のグラフ化と与えられた正規経路式と同じ正則集合を持つ NFA の作成を行い、グラフと NFA の正則集合の共通部分を受理する積オートマトンを作成する。次に、DTD と積オートマトンを比較することで簡略化出来る要素を判定し簡略化リストを作成する。最後に、簡略化リストを参照して積オートマトンの受理言語を簡略化し、形式を整えることで簡潔な XPath 式として出力する。

本アルゴリズムを Ruby で実装し、評価実験を行った。その結果、本アルゴリズムにより正規経路式を正確に XPath 式へ変換が出来るという見込みが得られた。

(指導教員 鈴木伸崇)