

## 平成27年度 情報工学コース卒業研究報告要旨

関 研究室	氏 名	川 本 将 也
卒業研究題目	XPath式からボトムアップ先読み付き決定性選択 トップダウン木オートマトンへの変換	
<p>近年、クラウドサービスなどの普及によって大量のデータを効率よく保存することが益々必要となっている。データの保存は様々な形式によって行われているが、その中の1つにXMLがある。XMLは代表的なマークアップ言語であり、タグを用いてデータ構造を容易に表現することができる。また、人が見ても構造がわかりやすいという利点があり、汎用性が高いため多くの分野で用いられている。XML文書はタグによる階層構造をもっているため、1つのXML文書を1つの木と考えることができる。</p> <p>XML文書に対する代表的な問合せ言語としてXPathが知られている。XPathは、例えば子頂点へ移動せよといった経路方向(XPathでは軸と呼ぶ)と頂点のラベルとを組み合わせたパス指定と呼ばれる言語構文をもち、これによりXML文書の要素の位置を指定する。XPathによる問い合わせは直感的で記述しやすいのが特徴である。XPathの問い合わせ評価方法の1つとして、木オートマトンとよばれる、木を入力として動作する計算モデルがある。XPath式からそれと等価な評価を行う非決定性木オートマトンや交代木オートマトンに変換する理論的研究が行われている。しかし、それらの変換法は極めて煩雑であり、実装については報告者の知る限り報告されていない。</p> <p>本研究ではXPathの部分クラスXPath/0を定義し、そのクラスのXPath式をボトムアップ先読み付き決定性選択トップダウン木オートマトンへ変換する手法を提案した。さらに、提案法に基づく変換ツールの実装と評価実験を行った。</p> <p>提案する変換法の基本方針は次の通りである。変換の対象としているXPath/0では、指定頂点集合を限定する記述(XPathでは述語とよぶ)を根からのパス指定の最後にしか許さないため、述語部分を決定性ボトムアップ木オートマトン(DBTAと略記) <math>A</math> へ変換し、パス指定部分については <math>A</math> を先読み部としてもつ決定性選択トップダウン木オートマトン(DSTA<sup>B</sup>と略記)に変換する。また、提案法で生成される木オートマトンは、既存研究と同様に、XML文書の木表現をFirst-Child Next-Sibling符号化(fcns符号化)して得られた2分木を入力とする。そのため、述語部分の評価を行うDBTA <math>A</math> では、各頂点を走査するとき、その子頂点に加えて兄弟頂点の情報を保持しながらボトムアップに親頂点へと情報を伝えていかなければならない。提案法では、それらの情報を保持するのに必要な数のビット列で状態を表現し、適切な情報の伝達を表す遷移規則を構成する。</p> <p>次に、提案した変換法の正当性の証明を行った。証明は、与えられたXPath/0式の述語部分を真にする入力木 <math>t</math> の要素の位置の集合と、式の述語部分を提案法により変換したDBTAが指定する、<math>t</math> を fcns 符号化して得られた木 <math>t'</math> における要素の位置の集合が等しいことを示した。その結果を用いて、XPath/0式全体が指定する <math>t</math> の要素の位置の集合と、その式を変換して得られた DSTA<sup>B</sup> が指定する <math>t'</math> における要素の位置の集合が等価であることを示した。</p> <p>また、提案法に基づく変換ツールを作成した。ツールによって変換した木オートマトンが指定した要素の位置を正しく返すか、また、手作業で記述した木オートマトンと比べ評価時間はどうかを実験で調べたところ、自動変換で得られた木オートマトンは指定した要素の位置を正しく返し、また、その評価時間も手作業のそれに劣らないことが分かった。XPath式を手作業で木オートマトンに書き直す作業では頻繁に誤りが混入するため、提案法はXPathで記述された問合せ処理に有用であることが分かった。今後の課題として、提案法で得られた木オートマトンを最簡形に変換すること等が残されている。</p>		