

## 平成25年度 情報工学コース卒業研究報告要旨

関 研究室	氏 名	後 藤 健 志
卒業研究題目	木文法に基づく XML 圧縮文書に対する直接問合せ法の評価	
<p>近年、クラウドサービスなどの普及により、大量のデータを効率よく保存することが必要不可欠となり、データの保存方法が高い関心を集めている。データは様々な形式で保存されているが、その中の1つにXMLがある。XMLはマークアップ言語の1つで、タグを自由に定義することができ、データ構造を容易に表現することができる。人が見ても構造が分かりやすいという利点があり、汎用性が高く多くの分野で用いられている。しかしながら、XML文書はその内部構造を指定するためのタグを多数含むため、テキスト形式もしくはタグ構造を反映した木構造で記憶すると、サイズが大きくなるという欠点がある。ゆえにXML文書は多くの場合、圧縮されて保存されている。圧縮されたXML文書に対して問合せをする際に、解凍してから行うのは非効率的であり、プライバシーの観点からも情報漏洩の危険がある。そこで本研究では、圧縮されたXML文書に対して、圧縮した状態のまま問合せを行う方法を提案することを目的とした。</p> <p>上述のようにXMLの文書構造はタグによって指定される階層構造となるので、1つのXML文書を1つの木として表現することができる。木構造を圧縮する方法は種々提案されており、最も単純な方法に共通部分木をまとめることによるDAG表現がある。またそれを拡張した方法として、与えられた1つのXML文書木のみを生成する木文法に変換することにより、圧縮することも可能である。圧縮方法として様々なアルゴリズムが提案されている中に、SLCF (Straight-Line Context Free) 木文法を用いた圧縮法がある。また、与えられたXML文書木のみを生成するSLCF木文法を効率よく求めるTreeRepairアルゴリズムが知られている。一方、SLCF木文法によって圧縮された木に対して問合せが行えることが理論的には示されているものの、実際の間合せ効率は不明であった。</p> <p>本研究では、圧縮されたXML文書に対して、圧縮した状態のまま問合せを行うアルゴリズムを提案し、それを実行するツールを作成した。具体的に、遷移可能な全てのノードを走査する方法と、その改良として同一の部分木は重複走査を回避する方法を提案する。作成したツールは、TreeRepairによって変換したSLCF木文法と、問合せを表す決定性セレクトティングトップダウン木オートマトンを入力として受け取り、問合せ結果となるノードのラベル及びポジション、及び実行時間などの付加情報を出力する。また、非圧縮のXML文書に対する問合せを行うツールも作成し、双方の実行時間から圧縮文書に対する直接問合せ法の評価を行った。</p> <p>作成したツールを用いて様々なXML文書に対し問合せを行った結果、圧縮されたXML文書に対する問合せでは、重複走査を回避することにより、多くの場合でそれを行わない場合と比較して問合せ効率が改善されることが分かった。また、非圧縮のXML文書に対する問合せと比較すると、圧縮率が極めて高い場合では、圧縮された文書に対する問合せのほうが効率がよい場合もあることが分かった。</p>		