

平成22年度図書館情報メディア研究科プロジェクト研究 研究成果報告書

種 目	萌芽研究		研究代表者 氏 名	鈴木伸崇
研究課題	正規木文法の差分抽出アルゴリズムに関する研究			
研究組織（研究代表者及び研究分担者）				
氏 名	所属研究機関・部 局・職	現在の専門	役割分担	
鈴木伸崇	図書館情報メディア研究科・准教授	情報科学	研究代表者	
研究目的				
<p>XML は Web における事実上標準のデータ形式であり、大量の XML データが生成・蓄積されている。XML データを蓄積・管理する場合、通常、時間の経過と共に格納すべきデータの構造や種類が変化する。このため、時間の経過と共にスキーマ定義も更新する必要があるが、スキーマの更新履歴を適切に管理することが重要である。スキーマの更新履歴を得るには更新前後のスキーマ間で差分抽出を行う必要があるが、これを適切に行える手法はこれまでほとんど提案されていない。本研究は、スキーマ定義言語としては最も表現力の高い正規木文法を対象とし、正規木文法に関する差分抽出アルゴリズムの開発を行う。</p>				
研究成果				
<p>正規木文法は、終端記号、非終端記号、開始記号、および生成規則から構成される。正規木文法の差分抽出問題とは、二つの正規木文法 <math>G</math> と <math>G'</math> が与えられた時に、<math>G</math> を <math>G'</math> へ更新するために必要な最短の編集操作列を求めることをいう。ここで、編集操作列とは、「生成規則の追加、削除」「生成規則の左辺の非終端記号の変更」「生成規則の右辺の終端記号の変更」等の編集操作の系列である。本研究では、まず、このような編集操作を形式的に定義し、正規木文法の差分抽出問題の計算複雑さについて考察した。その結果、正規木文法の差分抽出問題は文法の意味を考慮した場合は EXPTIME 困難、構文的な等価性のみ考慮した場合でも NP 困難であることを示した。次に、正規木文法の差分抽出問題が多項式時間可解となる十分条件を求め、その条件の下で動作する多項式時間アルゴリズムを開発した。最後に、得られたアルゴリズムを Ruby で計算機上に実装し、評価実験を行った。その結果、スキーマの更新が行われた場合、従来の手法と比較して提案アルゴリズムを用いた方がより容易に更新内容を把握できることが示された。</p>				
代表的な研究発表・特許等の成果一覧、特記事項等				
<p>堀江和磨, 鈴木伸崇, "正規木文法のための差分抽出アルゴリズム", 情報処理学会第 73 回全国大会, 4N-1, 2p., 2011.</p>				