

正規木文法の差分抽出に関する研究

Extracting Differences Between Regular Tree Grammars

学籍番号：201121736

氏名：堀江 和磨

Kazuma HORIE

XMLはWeb上の標準的なデータ記述フォーマットとして広く普及している。XMLデータをデータベース等で継続的に蓄積・管理する場合、格納すべきデータの構造をスキーマで定義しておき、それに沿った構造のデータを作成・格納することが一般的である。また、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し、それに応じてスキーマ定義が更新されることも多い。このような状況では、スキーマの更新履歴の管理、スキーマの更新に応じたXMLデータの修正等が必要となるため、スキーマの更新内容を適切に把握しておく必要がある。特に、管理者が複数で更新内容の共有が必要な場合や、スキーマが複雑で更新内容が多岐にわたる場合等は、スキーマの更新内容を把握することがより重要となる。スキーマの更新内容を把握す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の差分抽出問題について考察し、以下の結果を示した。

1. 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、同問題が計算困難であることを示した
2. 正規木文法の差分抽出が効率よく行えるための十分条件を求めた
3. 上記十分条件の下で、正規木文法の差分抽出を行う効率の良いアルゴリズムを構成した。更に、このアルゴリズムを実装し評価実験を行った

提案アルゴリズムをRubyを用いて実装し、評価実験を行った。その結果、本アルゴリズムを用いた場合、既存のXML差分抽出ツールを用いた場合と比較して、より適切にスキーマ間の更新内容が把握できるという結果が得られた。

研究指導教員：鈴木 伸崇

副研究指導教員：森嶋 厚行