

有向グラフに対する Shape Expression Schema の
妥当性検証及び修正手法の提案
Proposal of Shape Expression Schema Validation and
Correction Methods for Directed Graphs

学籍番号：201821627

氏名：藤永 健

Fujinaga Takeshi

グラフデータの構造はスキーマによって表現できる。スキーマを定義することで、問い合わせ式の記述に役立てることができるなどの利点がある。しかし、グラフデータがスキーマに対して妥当でなければ、その恩恵を受けることはできない。妥当性検証を行い、グラフデータの妥当性を保証することで、スキーマを利用した処理を確実に行うことができる。

グラフデータに対するスキーマとして Shape Expression Schema (以下, ShEx) が提案されている。既存の RDF Schema と比較して, ShEx は妥当性に関して形式的なセマンティクスが定義されており, より厳密にデータ構造を定義できるため, 妥当性検証を行うことに適している。ShEx は, 型の集合として記述されるスキーマであり, グラフデータの各ノードに対してその型の定義を満たしている型を割り当てることができる時, そのグラフデータは ShEx に妥当であるといえる。

先行研究において, 一定の制約を設けることで, グラフデータに対する ShEx の妥当性検証が効率よく行えることが証明されている。しかし, この手法はグラフデータ全体をメモリに置くことを前提としたものである。近年大規模なグラフデータが増加しており, そのようなグラフデータに対応するために, よりメモリ消費の少ない手法が必要である。また, 近年のグラフデータは複雑化している。そのため, グラフデータが妥当でない場合, 単に妥当でないという結果だけでなく, どのようにノードを修正すればよいか提示できればより有用である。

本論文は, グラフデータに対する ShEx の妥当性検証アルゴリズムおよび修正手法を提案する。提案アルゴリズムは, サイズの大きいグラフデータも検証できるように, グラフデータ全体をメモリ上に保持しない点に特徴がある。また, 効率性の向上及びメモリの消費量を更に抑えるため, ShEx の型の依存関係に着目し, 階層の概念を導入した妥当性検証を行うアルゴリズムも提案する。さらに, 修正手法は, グラフデータが妥当でない場合, 妥当でないノードに最適な型とその修正方法を提示する。これは, 文字列に対する正規表現の修正方法をグラフと型に適用させたものである。

評価実験の結果, 提案アルゴリズムの実行時間がほぼ線形であり, グラフデータに対してメモリ消費量が比較的小さく抑えられているという結果が得られた。また階層化を用いたアルゴリズムの方がやや省メモリであることも確認できた。また, 妥当でないノードの修正に関しては, 誤りが少数であれば, 概ね所望の修正が可能であ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：永森 光晴