

Shape Expression Schema の更新操作とそれに伴うグラフデータの修正手法

松原 尚利

近年、グラフデータが急速に普及している。グラフデータの特徴は、データ間の関連づけが容易なことである。また、データにはその構造を定義するスキーマと呼ばれるものが存在する。スキーマは、データの利用状況の変化などに応じて更新する必要がある場合がある。スキーマが更新された際、その下にあるデータは通常妥当ではなくなる。そのため、新しいスキーマの下で妥当になるようにデータを修正する必要がある。しかし、データを適切に修正するのは容易ではない。そこで、スキーマに施された修正によってどのようにデータを修正するかという手法を一般化することが必要になる。

本研究では、Shape Expression Schema (以下 ShEx) と呼ばれるグラフデータ (RDF データ) の構造を定義するスキーマ言語に着目する。ShEx は、Regular Bag Expression という規則に基づいて各ノードに割り当てる型を記述する。ShEx は W3C において使用策定中のスキーマであるため、その研究についてはまだ途上の段階であり、スキーマの利用に関する研究の進展が望まれている。本研究では、ShEx を更新した際にその下のグラフデータをどのように修正すれば更新後のスキーマの下で妥当になるのかについて考察する。より具体的には、ShEx の更新操作を形式的に定義すること、および、ShEx が更新された際にグラフデータの妥当性を維持するように修正する手法を確立することの 2 点を本研究の目的とする。なお、本研究では、修正の複雑さを軽減するために、各ノードが一つだけ型を持つ `single-type` のグラフに限定して研究を進める。

先行研究では、DTD を更新した際に XML データを更新する手法について研究がなされている。しかし、XML データは木構造であるため、これらの研究をグラフデータに適用することは困難である。グラフデータのスキーマ更新に関する研究は、著者の知る限り行われていない。

本研究ではまず、ShEx に施す更新操作を提案する。ここでは、スキーマの更新に必要であると考えられる操作を網羅的に列挙し、ShEx においてはどのような操作を施すかを定義した。次に ShEx に施された更新操作に応じてグラフデータを更新後の ShEx の下で妥当になるように修正する手順を考案した。最後に、ShEx の更新に応じて、グラフデータを自動で修正するシステムを作成した。

実装したシステムを用いて評価実験を行なった。その結果、更新操作によって処理時間は異なるが、データサイズに関して概ね線形に近い時間でグラフデータの修正が行えることがわかった。

(指導教員 鈴木伸崇)