

Shape Expression Schema におけるパターン問合せの包含性判定

藤本 春奈

近年、グラフデータの普及が進んでおり、こうしたグラフデータに対するスキーマ言語が提案されている。グラフデータは莫大なデータ量を持つ場合が多いため、より効率の良い問合せ処理が必要とされている。包含性判定は問合せ最適化などに有用であるため、莫大なデータ量を持つことの多いグラフデータにとって重要である。

本研究では、RDF データ(グラフデータ)のスキーマ言語である Shape Expression Schema (ShEx), 及びグラフデータに対する問合せのパターン問合せに着目する。ここでパターン問合せとは、データグラフにおいて指定したパターングラフと同型の部分グラフを検索する問合せである。一般に、スキーマが定義されていないグラフデータに対して行う通常のパターン問合せであれば、部分グラフ同型問題で包含性の有無を判定できる。しかし、スキーマの存在を仮定し、射影演算したパターン問合せの包含性を判定する場合には、部分グラフ同型問題を解くだけでは判定できない包含性が存在する。そこで、射影演算した2つのパターン問合せと ShEx スキーマが与えられた際に、包含性の有無を判定する手法を提案することを本研究の目的とする。

先行研究では、XPath や SPARQL などの問合せ言語の包含性判定手法について研究がなされている。しかし、ShEx スキーマの下での射影演算したパターン問合せの包含性判定手法に関する研究は著者の知る限り行われていない。

本研究の提案手法では、まずパターングラフのノード数 3 以上の 2 連結成分について処理を行う。これは、2つの問合せで同じ変数とラベルを使っても、3つ以上のノード同士が繋がっているか否かで、解が異なるために行う。次に、ShEx スキーマの定義に従い、一方のパターングラフに対してエッジの追加処理を行った後に部分グラフであるかを確認することで、包含性を判定する。エッジを追加する際には包含判定のための条件を設定しており、条件を満たすエッジのみ追加する。この条件は、エッジを追加したことにより、元の問合せの解から変化させることのないように設定している。

提案アルゴリズムを Python により実装し評価実験を行った。評価実験では、まずグラフデータの構造を定義した ShEx スキーマを用意した。その後、ShEx スキーマに対して充足可能なパターングラフを複数個作成し、その ShEx スキーマとパターングラフを用いて提案アルゴリズムの実行時間を計測した。その結果、ShEx スキーマの下でのパターン問合せの包含性を効率よく判定できることを確認した。

(指導教員 鈴木伸崇)