

## Shape Expression Schema におけるパターン問合せ充足可能性判定

松岡 栞

Shape Expression Schema (ShEx) はグラフデータのスキーマを記述するために新たに提案されたスキーマ言語である。一般に、グラフデータは莫大なデータ量をもつ。このようなデータに対してパターン問合せを行う場合、その莫大なデータの中からマッチするパターンを検索しなければならない。ここで、パターン P と ShEx S に対して、S に妥当かつ P を含むようなグラフデータ (RDF データ) が存在しない場合、P は S の下で充足不能であるという。充足不能なパターンで問合せを行うと、膨大なデータを検索したのちに解が存在しないという結果が返される。しかし、その検索は明らかに無駄である。そのため、充足不能なパターンを ShEx の定義から発見できることが望ましい。

充足不能なパターンは、ShEx が選言による分岐を含まない場合、部分グラフ同型問題を解くことで発見可能である。しかし、ShEx に選言による分岐が存在した場合、部分グラフ同型問題をそのまま適用することでは解くことができない。なぜなら、選言によって分岐しているエッジのうち的一方が選択された場合、他方のエッジは以降のマッチングにおいて利用することができないが、そのことを部分グラフ同型問題では判別できないからである。

そこで本研究では、ShEx を修正 AND/OR グラフで表現し、パターン問合せの充足可能性を判定するアルゴリズムを提案する。ここで、修正 AND/OR グラフとは、選言による分岐に円弧が描かれたグラフである。より具体的には、まず ShEx を修正 AND/OR グラフで表現し、その中で選言による分岐をもつノードを抽出する。また、パターンと修正 AND/OR グラフとのマッチングが行えるように、部分グラフ同型問題の解法の一つである Ullman Algorithm を次のように拡張して用いる。マッチング中のノードが選言による分岐をもつノードであった場合、同時に分岐しているエッジを使用せずにマッチング可能かの判定を行い、マッチング中のノードに接続している元のノードが選言による分岐をもつノードであった場合、ほかの分岐先のエッジを使用済みか否かの判定を行う。パターンのノード全てが修正 AND/OR グラフのノードとマッチすることが可能であれば、充足可能として処理を終了する。一方、途中で修正 AND/OR グラフのいずれのノードにもマッチしないパターングラフのノードが発見された場合、その時点で充足不能と出力し終了する。

提案アルゴリズムを Ruby により実装し評価実験を行った。充足不能なパターン問合せを複数作成し、RDF データに対する問合せ、および、提案手法による充足可能性判定を行った。その結果、前者の方が大幅に実行時間を要することが判明した。特に、階層が深い時点で充足不能なパターンが存在した場合、より顕著な差が生じることがわかった。また、前者の場合、パターンのノードのマッチングを行う順番によって実行時間は大きく変わるが、後者の場合は実行時間はいずれも短時間で済むことがわかった。

(指導教員 鈴木伸崇)