

Shape Expression Schema の下でのパターン問合せ修正手法の提案

船木 龍

近年, RDF/グラフデータ (以下, グラフデータ) が様々な分野で利用されている. グラフデータのスキーマとして, これまでにグラフスキーマや RDF Schema が提案されているが, グラフデータの構造を定義するには不十分である. そこで, 新たに提案されたのが Shape Expression Schema (以下, ShEx)である. ShEx はノードに対する型を Regular Bag Expression (RBE)という規則に基づいて定義する.

パターン問合せは, グラフデータに対する問合せ式の 1 つである. パターン問合せを q , ShEx スキーマを S とした時, S に妥当でかつ q を含むグラフデータが存在しない場合, q は S の下で充足不能であるという. 一般に, グラフデータは膨大な量のデータを有する. そのため, 充足不能な問合せを実行した場合, 膨大な量のデータを検索した後に充足不能という結果が返されることになり, この検索は明らかに無駄である. 一方, 充足可能性はグラフデータを検索することなく, スキーマのみで判定可能である. そこで, 充足不能な問合せ式を実行前に効率よく検出することが出来れば有用である.

しかし, 充足不能性の判定を行うだけでは不十分である. 特に, パターン問合せや ShEx スキーマが複雑な場合, およびユーザが ShEx の構造を十分に把握していない場合, ユーザがどのようにパターン問合せを修正すればよいかを発見するのは容易でない. そのため, 充足不能という結果を返すだけでなく, パターン問合せの修正方法を提示することができればユーザにとってより有用であると考えられる.

そこで本研究では, ShEx の下でのパターン問合せの修正手法を提案する. パターン問合せの結果が充足不能であった場合, その問合せは誤りを含んだ式であり解は得られない. そこで, 充足不能な式を修正して充足可能なものにする方法を提案する. 提案手法は, エッジラベルの誤りとエッジの過不足を許容したマッチングを行なうことで, 充足不能なパターン問合せに対する修正方法を発見する.

関連研究として, DTD を参照して XPath 式を修正する研究がなされている. しかし, XML データは木構造であること, XPath 式とパターン問合せでは問合せの構造が異なることなどから, この研究を本研究に適用することは困難である. 著者が知る限り, ShEx の下でのパターン問合せの修正手法に関する研究はこれまで行われていない.

提案アルゴリズムを実装し, 充足可能なパターン問合せを元に作成した充足不能なパターン問合せを用いて評価実験を行った. その結果, 提案アルゴリズムを用いることで, 充足不能なパターン問合せを充足可能なものに修正可能であること, および, 修正が効率よく行えることを確認した.

(指導教員 鈴木伸崇)