

論文内容の要旨

博士論文題目 経路式に基づく RDF データの格納と検索に関する研究

氏名 野 晃 整

(論文内容の要旨)

本論文では、大規模な Resource Description Framework (RDF) データを効率的に格納、検索するための経路式に基づいた手法を提案する。RDF は普及しはじめており、今後 RDF に基づくメタデータは増加することが予想されるので、大量の RDF データを高速に処理することのできる RDF データベースや索引手法が重要である。

本論文では、RDF データを関係データベースへ格納する手法と、経路式の接尾辞配列に基づく RDF データのための索引手法を提案した。この手法では、RDF データを部分グラフへ分割し、それらの特性に適した手法で個別の関係表へ格納する。具体的には、スキーマデータには継承関係の有無を判別するためにナンバリングスキームを採用し、インスタンスデータには結合演算の回数を減少させるために経路式に基づく格納方法を採用した。

RDF データのための索引手法は、RDF データから抽出した部分グラフごとに経路式を生成し、その経路式を文字列に見立てて、接尾辞配列を作成する手法である。検索処理では、接尾辞配列を二分探索によって一致する経路式を検索する。この手法は、従来の RDF データベースの索引として利用することができ、経路式に基づいた問合せ処理の結合演算を減少させることが可能になる。

本論文では、実験を通して提案手法の性能評価を行い、大規模な RDF データを効率的に格納できることを確認した。従来の RDF データベースは、経路式が長くなるにつれ、処理時間が増加したのに対し、提案したデータベースでは、徐々に処理速度が減少した。これは、提案手法が経路式の長さではなく、解集合の大きさに依存していることを示している。また、接尾辞配列に基づく索引手法の実験では、従来の手法での処理時間は経路式の長さとお対象データの大きさの両方に依存していたのに対し、提案手法では、データの大きさだけに依存し、経路式の長さには依存していないことを確認した。

(論文審査結果の要旨)

2004年12月28日に開催した公聴会の結果を参考に、2005年2月18日に本博士論文の審査を行った。次のとおり、本博士論文は、提案者が独立した研究者として、研究活動を続けていくための十分な素養を備えていることを示すものと認める。

申請者は、本博士論文において、大規模な Resource Description Framework (RDF) データを効率的に格納、検索するための経路式に基づいた手法を提案している。RDF は Web セマンティクスを記述する言語として、受け入れられ始めており、今後 RDF データは急速に増加していくことが予想される。したがって、大量の RDF データを高速に処理できる RDF データベースやその索引手法は重要な研究課題である。

本研究で申請者は、RDF データを関係データベースへ格納する手法と、経路式の接尾辞配列に基づく RDF データのための索引手法を提案し、実験して評価している。

格納に関しては、RDF データを部分グラフへ分割し、それらの特性に適した手法で個別の関係表へ格納する。スキーマデータには継承関係の有無を判別するためにナンバリングスキームを採用し、インスタンスデータには結合演算の回数を減少させるために経路式に基づく格納方法を採用した。

索引に関しては、RDF データから抽出した部分グラフごとに経路式を生成し、その経路式を文字列に見立てて、接尾辞配列を作成した。検索処理では、接尾辞配列を二分探索によって一致する経路式を検索する。

本論文では、実験を通して提案手法の性能評価を行い、大規模な RDF データを効率的に格納できることを確認した。従来の RDF データベースは、経路式が長くなるにつれ、処理時間が増加したのに対し、提案したデータベースでは、徐々に処理速度が減少した。

博士論文に述べられている提案は、今後重要になると考えられる未開拓分野に挑んでおり、RDF 情報のデータベース化に有用な基礎を構築している。よって、本論文は博士(工学)の学位論文として価値があるものと認める。