

## 分散 RDF データベースに関する研究

的野晃整 谷村勇輔 小島功

産業技術総合研究所 情報技術研究部門

{a.matono, yusuke.tanimura, isao.kojima}@aist.go.jp

### A Research on Distributed RDF Databases

Akiyoshi MATONO, Yusuke TANIMURA, Isao KOJIMA

Information Technology Research Institute,

National Institute of Advanced Industrial Science and Technology (AIST)

#### 1. はじめに

情報爆発時代の到来によって、必要な情報の検索や管理が困難となってきている。我々は、この困難を克服できる解決策の一つとしてメタデータが重要な役割を担うと考えている。現在、メタデータ記述の標準フレームワークとして、RDF (Resource Description Framework) が広く認知されている。RDF は、簡潔なモデルでありながら高度な情報を表現できるという特徴から、近年広い分野での利用が始まっている。たとえば、ユビキタスコンピューティング環境において、ucode と呼ばれる識別子に基づくメタデータ体系には RDF を採用している。ucode は個々のモノや場所に割り振られるため、そのメタデータも膨大となることが予想できる。このように、RDF に基づいて記述されたデータ (RDF データ) は分散した状態で、データの大規模化が進んでいる。そのため、RDF データを効率的に管理できる基盤として、分散 RDF データベースの実現が不可欠となっている。

分散 RDF データベースは、ボトムアップ式とトップダウン式の 2 種類におおまかに分類できる。ボトムアップ式は各組織が生成する RDF データをその組織で管理する方式で、データのアクセス制御を各組織で直接行うため、安全性が高く、またデータの鮮度が保障されるが、格納方法を統一できないために、検索性能は劣る。一方トップダウン式は統一された配置法に基づいて RDF データの格納方法を決定する方式で、検索性能は期待できるが、データの収集や再配置のコストが高く、細かなアクセス制御の実現が困難である。このように、それぞれに利点があり、必ずしも一方が常に優れているとは断言できない。

我々は、これまで問合せ処理演算時に発生する転送量を減少させるために、ブルームフィルタを用いた索引手法を提案し、トップダウン式の一つである P2P 環境とボトムアップ式環境へ適用してきた [1][2]。また現在はトップダウン式としての分散ストレージ上において、RDF 問合せ処理を効率的に実現するために、格納方法やアルゴリズムの研究を進めている [3]。

#### 2. ブルームフィルタを用いた索引手法

分散環境における問合せ処理では、大量の通信が発生してしまう。例えば、二つのノードで選択した結果をいずれかのノード上で結合する場合、選択結果全体を他方のノード

に転送しなければならない。我々はこの問題を解決するため、データの転送前に不要な解をフィルタリングすることで転送量を減少できると考えた。

我々はまず RDF モデルに合わせてブルームフィルタを 3 次元に拡張し、次に RDF データの演算に対応するように、3 次元ブルームフィルタの演算を提案した。これによって、与えられた問合せから生成される実行プラン木は、本来 RDF データに対するプランであるが、それをそのまま 3 次元ブルームフィルタに対して適用できるようになる。3 次元ブルームフィルタは、RDF データのおおまかな存在情報であるため、それらの演算結果から解に含まれる RDF データをおおまかに特定できる。このブルームフィルタ演算結果をフィルタとして利用することで、他ノードへの転送前に不要な解を除外できる。

ブルームフィルタを用いた索引手法の性能を実験によって評価した。その結果、複雑な問合せであるほど転送量を減少する傾向があることが分かった。また、提案手法の効果が無い場合であっても転送量をわずかな増加に抑えられ、効果がある場合には転送量を大きく減少できることも確認できた。

#### 3. おわりに

分散 RDF データベースの実現を目標に、ボトムアップ式とトップダウン式のそれぞれの環境での性能向上を目指して、研究を進めている。今後は、高い検索性能を期待できるトップダウン式分散環境と、安全と鮮度が高いボトムアップ式分散環境のそれぞれの利点を生かしたハイブリッドな分散 RDF データベースを実現したいと考えている。

#### 参考文献

- 1) 的野、サイド、小島：P2P 環境における三次元ハッシュ索引を用いた分散 RDF データベース問合せ処理、情報処理学会論文誌：データベース、第 47 巻、No. SIG8 (TOD 30), pp.121-133, 2006 年 6 月。
- 2) 的野、サイド、小島：ブルームフィルタを用いた分散 RDF 問合せ処理のための索引手法、電子情報通信学会 第 19 回データ工学ワークショップ論文集、C4-1。
- 3) 谷村、的野、小島、田中、関口：MapReduce によるデータベース処理の実装とデータ分散格納方法の提案、並列/分散/協調処理に関するサマー・ワークショップ (SWoPP2008) 発表予定。