

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 26 日現在

機関番号：21201

研究種目：基盤研究(C) (一般)

研究期間：2012～2014

課題番号：24500122

研究課題名(和文) ユビキタスデータベースにおける知識融合技術に関する研究

研究課題名(英文) Research on Knowledge Integration Technology for Ubiquitous Databases

研究代表者

澤本 潤 (Sawamoto, Jun)

岩手県立大学・ソフトウェア情報学部・教授

研究者番号：50438082

交付決定額(研究期間全体)：(直接経費) 4,000,000円

研究成果の概要(和文)：分散データファイルの統合化の研究、複数のDB視点からの知識融合技術の研究、負荷分散や並列処理技術に関する研究を行った。具体的な実施内容は以下の通りである。(1) MySQL, MongoDB, Linked Open Data に対して統一クエリを介して情報の取得をする異種情報融合フレームワークを開発・評価した、(2) 知識融合技術の方式設計と作成・評価を行い、分散型データマイニングへの適用・評価を実施した、(3) シミュレータにより仮想化処理方式負荷解析を行ない、性能上の問題点を明確にし、仮想化処理の実行効率低下への影響度合いを確認した。

研究成果の概要(英文)：We conducted the research work concerning integration of distributed data files, knowledge fusion technology from multiple DB viewpoints and acceleration of access operation by the load sharing and parallel processing technology, etc. We output some of the research results such as development of a framework for integrating multiple heterogeneous databases, data mining application technology, efficiency evaluation of database virtualization in parallel distributed processing environment.

研究分野：ユビキタスコンピューティング

キーワード：マルチデータベース 仮想化DB 分散データファイル データマイニング 知識融合技術 並列処理技術

1. 研究開始当初の背景

ユビキタスコンピューティング社会では大量のデータが収集利用されるが、こうしたデータの中に隠された知識や傾向を発見し意思決定などに役立てることが重要となっている。このユビキタスコンピューティング環境で利用されるデータベース(DB)を考えると、リレーショナルDB(RDB)、階層DB、オブジェクト指向DB(OODB)などの従来型のデータベース管理システム(DBMS)だけではなく、新しい概念を実現したXML DB、Web DB、NoSQL DBなど、多種多様なDBが存在し、分散配置されている。こういった大規模かつ大容量DBにより、データベースマーケティングや教材推薦サービス等へのデータマイニング技術の利用拡大が期待されているが、こうしたサービスを行う情報分析者は、本来的にはルール発見やそのルール解釈の作業に集中したいにも拘わらず、データマイニングの準備過程であるDB選択やデータ収集などの作業に多大な時間を割かねばならず、こうした作業負担を軽減する技術が必要となってきた。

一方、こうした背景の下、申請者は平成20年度から平成22年度にかけて、科学研究費補助金基盤研究(C)「ユビキタスデータベース仮想化技術によるデータ利用の効率化に関する研究」(課題番号20500095)を実施し、その成果としてRDB、XMLDBなどを仮想化するユビキタスデータベース仮想化技術を新たに提案し、実装した。しかしながら、まだ研究課題は数多く残っており、主な課題は以下の通りであった。

<課題1> 現状のユビキタスデータベース仮想化技術では、現在インターネットの商用サイトで大容量データを扱う為に広く使われている分散データファイル(例えばNoSQLなど)を仮想データベースへ統合化しなければならない。

<課題2> 現状のユビキタスデータベース仮想化技術では、一つのDBの見え方を提供するため、逆に種々のDB視点から複眼的にデータマイニングしたいという要求には応えられない。

<課題3> 現状のユビキタスデータベース仮想化技術では、負荷分散技術や並列処理技術などの利用により、高速化が期待できる。しかし、現時点ではこの高速化を定量的に評価する段階には至っていない。

2. 研究の目的

本研究は、先述した3点の課題を解決するために、以下のように研究目的を設定する。

<課題1> 実装方式の異なるNoSQL等の分散データファイルをいくつか取り上げ、仮想化DBに統合化する方式を開発する。

<課題2> 仮想化DBに対して利用者レベルの種々のDB視点を利用者に提供する機能を研究し、それぞれのDBからデータマイニングした結果から得られる複数のルールを調

停・統合する「知識融合技術」を研究する。
<課題3> 負荷分散技術や並列処理技術を利用した仮想化DB環境上で実行効率を測定し、対象データベースの接続数や個々のデータベースの規模が大きくなったときを想定して定量的に検証する。

3. 研究の方法

ユビキタスデータベース仮想化技術により構築された仮想化DBに対して、分散データファイルの統合化を研究するチーム、種々のDB視点から知識融合技術を研究するチーム、および負荷分散や並列処理技術等の利用による高速化の研究を行うチームの構成で研究を進めた。平成24年度は、チームの確立、調査検討・方式設計に注力し技術課題の抽出・整理、その解決法に目途をつけた。平成25年度は、プロトタイプ制作・評価を中心に行い、実現に目途をつけた。最終年度の平成26年度は実用に近い応用の構築・評価により性能なども含めた総合的評価を行った。

4. 研究成果

3つの課題に対してそれぞれ具体的な達成度評価は以下の通りである。

(1)分散データファイルの統合化の研究：幾つかの分散データファイル方式(特に、NoSQLとLinked Open Data)に対応した仮想化DBの統合プロトタイプの作成・評価を実施した。ほぼ、予定通りの達成度であり、プロトタイプによりその有効性と性能に関するユーザ評価を行った。

異種情報融合フレームワーク

本研究では、MySQL、MongoDB及び、Linked Open Dataからの情報取得クエリであるSPARQLに対して、統一クエリを介して情報の取得を可能とすることで、あたかも一つの仮想データベースから情報を抽出することができる異種情報融合フレームワーク(以下本フレームワーク)の開発と、実験として、本フレームワークを用いた場合と用いない場合での情報取得速度の計測と比較並びに、統一クエリの優位性をアンケートにより評価を行った。統一クエリは学習者の多いSQL文に準拠して作成した。本フレームワークの構成を図1に示す。

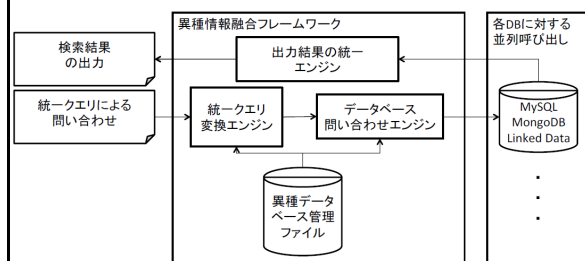


図1. 異種情報融合フレームワーク

利用者は、アプリケーション上やターミナル上で統一クエリを発行する。フレームワーク内では、融合データベース情報の管理ファイ

ルから登録されている DB の種類を抽出し、種類に応じて各クエリへの変換を行う。クエリ変換後、融合データベース情報の管理ファイルから各データベースへのアクセス情報を取得し、変換したクエリを用いて各データベースに並列にアクセスを行う。各データベースから得られた出力結果は、本フレームワークを介して JSON 形式に統一して利用者へ出力を行う。

評価アプリケーション

本フレームワークの応用例として、Wikipedia の情報を SPARQL により取得可能な DBpedia から、本フレームワークを介して札幌市、盛岡市、並びに郡山市の地域コードを取得し、総務省統計局消費者物価指数 (CPI) の都市階級の年平均データを MongoDB に挿入し、本フレームワークを介して、それぞれの市の食料の CPI 値を取得した後に、Google Chart の Geo Map を用いて日本地図上に表示した。アプリケーションの構成を図 2 に記載する。

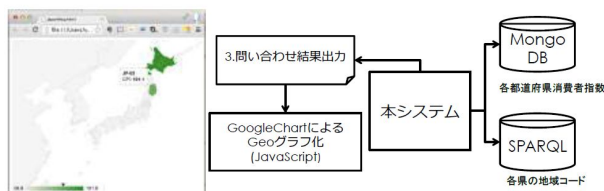


図 2. 評価アプリケーション構成

実験・評価

本フレームワークの有用性を評価するために、データベースに対する基本的な知識を持ち、実務経験の無いデータベース初心者7名に対して、MySQL、MongoDB、SPARQL 並びに本フレームワークを実際に利用し、問題形式で情報の取得を行い取得するまでに経過した時間を測定した。結果のグラフを図 3 に記載する。

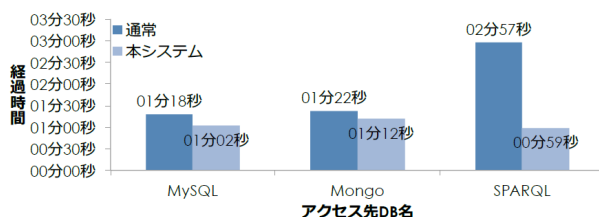


図 3. 情報取得にかかる平均時間の評価結果

この結果に対して t 検定 ($p < 0.05$) を実施した結果、本フレームワークを介した場合 MySQL と SPARQL に対して有意差が認められた。Mongo クエリに対しては、別途実施したアンケート結果より、取得は早いですがクエリの理解しやすさは統一クエリに劣る結果となり、本フレームワークが、データベース初心者に対して有用であることが分かった。

(2) 複数の DB 視点からの知識融合技術の研究：仮想化 DB 上での複数 DB 視点構成手法のプロトタイプ作成・評価については、Linked Data を利用した情報統合について試作・評価を行った。知識融合技術のプロトタイプ作成・評価については、Web 教材データベースからの教材推薦、SNS サイトにおけるレコメンド機能に関する研究、ゲームにおけるレコメンドの研究開発などの各種の分散型のデータマイニング応用プログラムへの適用・評価を実施し、その研究成果をまとめ学会発表を行った。

(3) 負荷分散や並列処理技術等の利用による高速化の研究：シミュレータの実装を行い仮想化処理方式の性能上の問題点を明確にしていた。具体的には、仮想化処理基準を決定し、データベースの処理時間の算出方法などを明確にし、プロセス間の相互通信方式を試作・評価しその研究成果をまとめ学会発表を行った。

シミュレータ実装

実行効率の測定を行うためのシミュレータを作成する。本シミュレータでは図 4 のような分散環境におけるデータベース仮想化を想定して実装した。実際の仮想化の変換処理やデータベースのクエリ実行処理などはシミュレーション実行時には行わず、各処理に要する処理時間の算出・換算をする。また通信に関しても実際の通信時間の計測はせず、通信時間を得るための計算を行う。

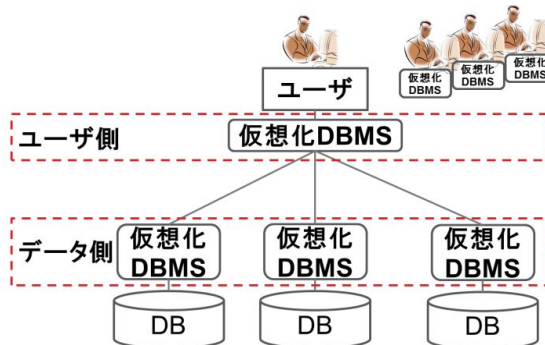


図 4. 分散環境におけるデータベース仮想化

シミュレーション実行

ユーザ数やデータベース接続数等の設定を変えて複数回実行した結果、ユーザ数が少ない場合、データベースの接続数が多い場合、結果件数が多い場合に仮想化処理の割合が大きくなるという知見が得られた。また、全体の処理時間に対する仮想化処理時間の割合の検証範囲内での最大値は、9.569%程度であり、この値は十分に小さいと考えられる。

5. 主な発表論文等

〔雑誌論文〕(計 3 件)

Yuji Wada, Yuta Watanabe, Keisuke Syoubu, Hiroshi Miida, Jun Sawamoto, Virtual Database Technology for

Distributed Database in Ubiquitous Computing Environment, American Journal of Database Theory and Application 2012, 1(2): 13-25, DOI: 10.5923/j.database.20120102.02
Y. Wada, T. Segawa, J. Sawamoto, and H. Sato, Improvement of Attribute Correlation Method and Proposal of Collaborative Attribute Method in Text Recommender Systems for E-Learners, International Journal of Informatics Society (IJIS) 査読有 VOL.6, NO.2 (2014) pp.57-65 (November 2014).
T. Kashiwagi, J. Sawamoto, H. Sato, N. Segawa, E. Sugino, and Y. Wada, A Proposal of P2P Content Delivery System for Supporting Streaming Applications, International Journal of Informatics Society (IJIS) 査読有 VOL.6, NO.2 (2014) pp.89-96 (November 2014).

[学会発表](計 20 件)

Yuji Wada, Takuya Segawa, Jun Sawamoto, Hiroyuki Sato, Research on Recommender Functions for Learning Support Systems, 2nd. International Conference on Applied and Theoretical Information Systems Research, 27th to 29th December 2012, Taipei, Taiwan.
Takuya Sasaki, Jun Sawamoto, Hiroyuki Sato, Norihisa Segawa, Eiji Sugino, Yuji Wada, A Proposal of Access-based Contents Recommendation in Virtual Database Environment on P2P Network, 12th WSEAS Int. Conf. on APPLIED INFORMATICS AND COMMUNICATIONS (AIC'12), pp. 142-147, August 21-23, 2012, Istanbul Turkey.
鈴木吉輝, 澤本 潤, 杉野栄二, 瀬川典久, 広告重要度と配信距離に基づくユーザー適応型地域情報配信システムの提案, 情報処理学会第 75 回全国大会, 3-191-192, 3/6-8 2013. 東北大学
杉山暢彦, 澤本潤, 杉野栄二, 瀬川典久, 異種データベースの仮想化技術(NoSQL データベース"Cassandra"の仮想化), WISS2012 予稿集, WISS2012 デモ・ポスター発表 (2012) 青森県三沢市古間木山 56 古牧温泉 青森屋 12 月 6 日 8 日 2012
杉山暢彦, 澤本 潤, 杉野栄二, 瀬川典久, 佐藤裕幸, 和田雄次, NoSQL データベース Cassandra の XML 仮想化システム, 信学技報, ライフインテリジェンスとオフィス情報システム研究会 (LOIS), vol. 112, no. 466, LOIS2012-72, pp. 17-22, 2013 年 3 月 7 日 8 日 沖縄県宮古島市.
鈴木吉輝, 澤本潤, 商店街を対象としたデジタルサイネージ作成・配信システム

の提案及び実装,平成 24 年度電気関係学会東方支部連合大会発表, 8 月 30 日(木)・31 日(金) 2012 年、秋田大学 栗原隆平、Web 教材データベースからの教材推薦サービスに関する研究、情報処理学会研究報告 コンピュータと教育研究会 2013 年 2 月

蔣 淵舒, 和田雄次, 信頼度を考慮した協調フィルタリングに基づく書籍の推薦方式の提案、第 12 回情報科学技術フォーラム FIT2013、D-029、2013 年 9 月、鳥取大学。

Yuji Wada, Takuya Segawa, Jun Sawamoto and Hiroyuki Sato, Proposal and Evaluation of Collaborative Attribute Method in Text Recommender Systems for E-Learners, International Workshop on Informatics 2013 (IWIN2013), pp.41-48, September 2013, Stockholm, Sweden, 2013 年 9 月 1 日-4 日

Takanori Kashiwagi, Jun Sawamoto, Hiroyuki Sato, Yuji Wada, Norihisa Segawa, and Eiji Sugino, A Proposal of P2P Content Delivery System for Supporting Streaming Applications, International Workshop on Informatics 2013 (IWIN2013), pp.3-8, September 2013, Stockholm, Sweden. 2013 年 9 月 1 日-4 日

Yoshiki Suzuki, Jun Sawamoto, Norihisa Segawa, Eiji Sugino, Hiroyuki Sato, and Yuji Wada, LOCAL ADVERTISEMENT DISTRIBUTION SYSTEM BASED ON THE LEVEL OF USER'S INTEREST AND THE DISTRIBUTION DISTANCE, International Conference on Applied and Theoretical Information Systems Research, November 22-24, 2013 (in CD-ROM), Taipei, Taiwan. 2013 年 11 月 22 日-24 日

狩野大地, 佐藤裕幸, 澤本潤, 和田雄次: 「分散環境におけるデータベース仮想化技術の実行効率測定のためのシミュレータの開発」, 情報ネットワーク研究会 (IN) 電子情報通信学会技術報告, Vol.113 ,No.389 ,IN2013-140 ,pp.133-138 , 2014 年 1 月, 名古屋国際センター .2014 年 1 月 22 日-24 日

Yuji Wada, Ryuhei Kurihara, Jun Sawamoto, Hiroyuki Sato and Hisao Fukuoka, Learning Material Recommendation Service with E-Learning Database, International Workshop on Informatics 2014 (IWIN2014), 査読有, pp. 79-85, September 2014.

Hisao Fukuoka, Akane Kawakami and Yuji Wada, A Proposal of CS Unplugged Utilizing Regional Materials, International Workshop on Informatics

2014, 査読有, pp. 109-114, September 2014.

Daichi Kano, Hiroyuki Sato, Jun Sawamoto and Yuji Wada, A simulator for the execution efficiency measurement of distributed multi-database virtualization, International Workshop on Informatics 2014, 査読有, pp.43-50, September 2014.

狩野大地, 佐藤裕幸, 澤本 潤, 和田雄次, 並列分散環境におけるデータベース仮想化処理負荷の解析, D-4-17, 電子情報通信学会 2015 年総合大会, 2015 年 3 月 13 日。(立命館大学 びわこ・くさつキャンパス)

鈴木吉輝, 澤本 潤, 瀬川典久, 杉野栄二, 佐藤裕幸, 和田雄次, 異種情報融合フレームワークの開発と評価, D-6-19, 電子情報通信学会 2015 年総合大会, 2015 年 3 月 11 日。(立命館大学 びわこ・くさつキャンパス)

鈴木 吉輝, 澤本 潤, 瀬川 典久, 杉野 栄二, 佐藤 裕幸, 和田 雄次, 異種情報融合フレームワークの提案, 情報処理学会第 161 回 マルチメディア通信と分散処理研究会(DPS), 査読無, 2014 年 9 月.

関野裕通, 和田雄次, SNS サイトにおけるレコメンド機能に関する研究 コールドスタート問題の緩和・解決

ISS-P-131 情報・システムソサイエティ特別企画 学生ポスターセッション予稿集, pp.131, 電子情報通信学会 2015 年総合大会, 2015 年 3 月 11 日。(立命館大学 びわこ・くさつキャンパス)

張 亜偉, 和田雄次, ゲームにおけるレコメンダの研究開発 - 重み付け攻撃対策 - ISS-P-129 情報・システムソサイエティ特別企画 学生ポスターセッション予稿集, pp.129, 電子情報通信学会 2015 年総合大会, 2015 年 3 月 11 日。(立命館大学 びわこ・くさつキャンパス)

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

名称:

発明者:

権利者:

種類:

番号:

出願年月日:

国内外の別:

取得状況(計 0 件)

名称:

発明者:

権利者:

種類:

番号:

出願年月日:

取得年月日:

国内外の別:

〔その他〕

ホームページ等

6. 研究組織

(1) 研究代表者

澤本 潤 (SAWAMOTO, Jun)

岩手県立大学・ソフトウェア情報学部・教授

研究者番号: 50438082

(2) 研究分担者

和田 雄次 (WADA, Yuji)

東京電機大学・情報環境学部・教授

研究者番号: 30366398

佐藤 裕幸 (SATO, Hiroyuki)

岩手県立大学・ソフトウェア情報学部・教授

研究者番号: 70443679