

機関番号：21201
 研究種目：基盤研究 (C)
 研究期間：2008～2010
 課題番号：20500095
 研究課題名 (和文) 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
 研究課題名 (英文) Research on Efficient Data Usage by Database Virtualization Technology in Ubiquitous Environment
 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

研究成果の概要 (和文) : マルチデータベース仮想化技術の研究とデータマイニングエージェント技術の研究を行った。スキーマ変換機能とクエリ変換機能を異種データベース (RDB、XMLDB) を対象に構築し、利用評価を実施し有効性を確認した。さらに、データマイニングエージェントがマルチデータベースから動的にユーザにとって関心の高いデータを検索して組み合わせる方式として利用者の検索履歴と検索頻度を利用したコンテンツのグループ化手法とそれに基づく効率の良いコンテンツ検索法の開発を行った。

研究成果の概要 (英文) : Our research objective is to develop a database virtualization technique and data mining agent for data analysts or other data mining users to reduce their workloads such as data collection from databases and data cleansing works. We developed schema and query conversion function for heterogeneous data models (RDB, XMLDB) using a single XML schema as user's view and confirmed the efficiency for data acquisition function. We designed and prototyped mechanism to automatically select and combine databases to achieve data mining activity and we developed an efficient content access method based on the grouping technique of the contents using the content access history and frequency.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,600,000	480,000	2,080,000
2009年度	1,500,000	450,000	1,950,000
2010年度	500,000	150,000	650,000
総計	3,600,000	1,080,000	4,680,000

研究分野：総合領域

科研費の分科・細目：情報学 メディア情報学・データベース

キーワード：マルチデータベース、仮想化、データマイニング、マルチエージェント、P2P ネットワーク

1. 研究開始当初の背景

(1) ユビキタスセンサーネットワーク等から大量のデータが収集されるユビキタス時代では、これらのデータの中に隠された知識や傾向をデータマイニング技術を用いて発見し意志決定などに役だてることが重要となっている。こういった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)、演繹DB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センサーDB、XML DB、Web DB、地図DBなど、多種多様なデータベース (すなわち、マルチデータベース) が存在し、分散配置されている。

(2) こうしたマルチデータベースは、テキストデータや数値データだけではなくマルチメディアのコンテンツが格納されている大規模かつ大容量DBであることが多く、データマイニング技術の革新が期待されている。特に、このマルチデータベースに対してデータマイニングを行うデータ分析者（データマイニング・クライアント）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データベース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、これがデータ利用上の大きな負担となっている。

2. 研究の目的

本研究は、上記のデータ分析者の負担を軽減するために、先ずユビキタスコンピューティング環境上に分散配置されているマルチデータベースが、あたかも一つのDBMSで管理されているかのように見えるDB仮想化技術を研究する。そして、データマイニング・クライアントがデータマイニング・サーバ経由でこのDB仮想化技術を用いてアクセスする時、マルチデータベースから適切なDBを選択し、データ収集、ルール発見、ルール可視化などのデータマイニング作業を代行するデータマイニングエージェント技術も併せて研究する。

3. 研究の方法

平成20年度は、マルチデータベース仮想化技術の研究とデータマイニングエージェント技術の研究を開始した。その目標とする具体的な研究課題は以下の通りである。

- ・マルチデータベース仮想化技術の研究：マルチデータベース仮想化の方式を設計する。
- ・データマイニングエージェント技術の研究：同時に、データマイニングエージェントがデータマイニングに必要なデータベースを動的に選択して、そのデータベースを組み合わせる方式を設計した。

平成21年度は、マルチデータベース仮想化技術の研究として、

- ・仮想化DBMSのプロトタイプ作成
- ・マルチデータベース（リレーショナルDB、XMLDB、DWH）の構築を実施した。一方、データマイニングエージェント技術の研究として、

- ・マルチエージェントの方式設計
 - ・内包する知識ベースの方式設計を行った。
- 続いて、平成22年度は、マルチデータベース仮想化技術とデータマイニングエージェント技術を融合化するための研究を実施した。

4. 研究成果

マルチデータベース仮想化技術の研究とデータマイニングエージェント技術の研究を実施した。その研究目標に対する具体的な研究成果は以下の通りである。

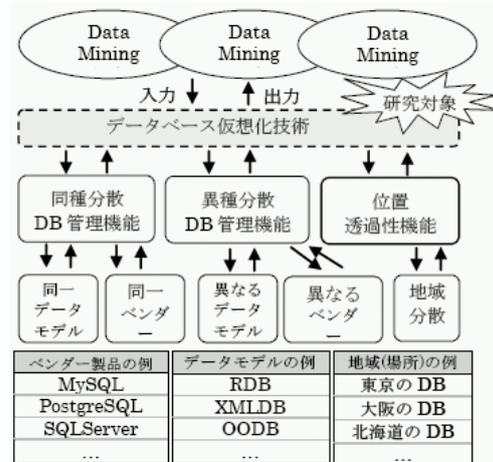


図1. マルチデータベース仮想化技術

(1) マルチデータベース仮想化技術の研究
図1に示すように、DB仮想化技術は同種分散DB管理機能や異種分散DB管理機能、位置透過性機能といった機能を備える。

本研究では、異種データモデルDBの仮想化に焦点を当て、異種DB仮想化システムの開発を行った。これにより、ユーザは実DBのモデルやベンダー製品の差異を認識することなくDB操作を行うことができる。

①異種データベース仮想化システム

異種DB仮想化システムの概要を図2に示す。システムの流れとしては、まずユーザは仮想化DBのスキーマである共通スキーマの情報を取得し、仮想化DBの構造を認識する。次にその構造情報を基に仮想化DBに対して一つのクエリを発行する。仮想化DBシステムは、実DBにクエリを分散問い合わせし、同一形式の結果をユーザに返却する。

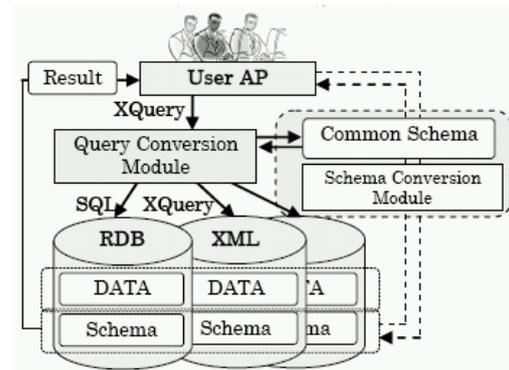


図2. 異種DB仮想化システムの概要

②仮想化データベースの利用者ビュー

本研究では異種DBの仮想化として、異種DBのデータモデルをRDBとXMLDBに焦点を絞って開発を行った。仮想化DBの利用者ビューは4

つの提案がある。提案 1 は全ての DB を一つの表で表現したもの、提案 2 は複数の表で表現したもの、提案 3 は全ての DB を一つのツリーで表現したもの、提案 4 は複数のツリーで表現したものである。本研究では仮想化 DB の利用者ビューをツリー形式かつ DB 毎に複数のツリーとして開発を行った。

③仮想化 DB のスキーマとクエリ

仮想化 DB の利用者ビューを DB 毎の複数のツリーと定めた。そこで本研究では、仮想化 DB の共通スキーマを XML スキーマ、問い合わせ言語を XQuery として仮想化 DB システムの開発を行った。ユーザは共通スキーマである XML スキーマによって仮想化 DB の構造を知り、1 つの XQuery によって問い合わせを行う。DB 仮想化システムはそれぞれの実 DB にクエリを分散問い合わせする。その結果を統一されたデータ形式でユーザに返却する。

表 1. 仮想化 DB の関数ライブラリ

関数	概要
ConnectVDB (String dbname1, String dbname2, ...)	仮想化 DB との接続
CloseVDB ()	仮想化 DB との切断
Execute (String xquery)	クエリの実行
next (elementName)	検索結果の繰り返し処理
getString (String elementName)	テキストノード値の取得
getAttr (String elementName, String attributeName)	属性値の取得

④仮想化 DB の使用

仮想化 DB を利用する際には、本研究で開発したコマンドラインツールと関数ライブラリを使用する。このコマンドラインツールは、アプリケーション構築の前準備として、ユーザに仮想化 DB の情報を提供する支援ツールである。仮想化 DB の関数ライブラリは仮想化 DB とのアクセスを行うための関数インターフェースである。仮想化 DB の関数ライブラリを表 1 に示す。

⑤仮想化 DB のデータマイニング活用評価

仮想化 DB をデータマイニングに活用した場合の有効性を検証する。我々は、データマイニング結果の可視化支援ツールを提案し、データ分析作業の効率化について評価を行った。被分析データとデータマイニング結果を合わせて可視化をすることにより、分析者がデータマイニング結果を理解し、より有効な情報を選択することを目的としたデータマイニング支援ツールの開発を行った。

また、事例としてコンビニエンスストアのケースを考える。ここでは、コンビニ A では顧客の購買データは RDB に保存されており、コンビニ B では、XMLDB に保存されているとする。通常、データ分析者は、両方のデータベースからのマイニング結果を上述の可視化

ツールを活用し比較しながら分析を進めることになる。

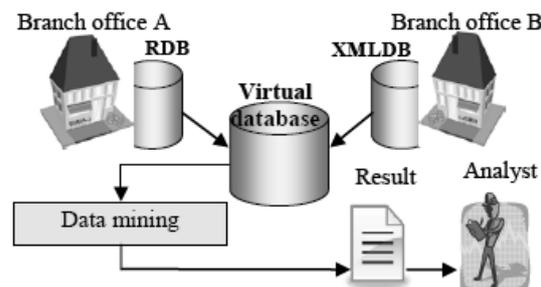


図 3. 仮想化 DB を用いたデータマイニング

一方、同じ状況で図 3 に示されているように、我々が提案する仮想化 DB を利用した場合を考える。データ分析者から見た場合、データマイニング結果は一つであり、複数データベースからのマイニング結果を比較、統合する必要もなくなる。結果として、複数の異種データベースに渡るデータ分析作業の効率化が期待できる。

⑥評価実験

この章では実際にアプリケーションを実装し、従来方式と仮想化 DB を利用した場合とでプログラム工程数の比較を行った。検証を行うためのサンプルデータとして、気象庁のサンプルデータから RDB の表形式のエリア情報データと XML 形式の天候情報データをそれぞれ取り扱った。このデータの中から、例題としてエリア情報からは「地域名」と「気象台」、天候情報からは「降水量」と「気温」をそれぞれ同時に出力するアプリケーションを構築した。複数種類の DB を複合的に利用しデータを取得する際に、従来方式ではその DBMS に合わせたクエリを発行し、得た違う形式の結果を統合する必要があった。今回開発した仮想化 DB を用いることによってユーザは DB の差異を意識することなく、あたかも一つの DB を扱うかのように情報を取得することができる。その結果、ユーザは本来的な作業に集中することができた。

(2) データマイニングエージェント技術の研究

同時に、データマイニングエージェントがマルチデータベースの中からデータマイニングに必要なデータベースを動的に選択し、そのデータベースを組合せる方式の開発を行った。具体的には、P2P オーバレイネットワーク上での分散ネットワーク環境において動的にユーザにとって関心の高いデータを検索して組み合わせ提示する方式のプロトタイプ開発および基本的な評価を行った。ユーザがシステムを利用し検索ワードを投入した際に、システムはユーザのコンテキストを持った検索エージェントを生成する。ユ

ユーザのコンテキストとは、検索をかけた時点のユーザに関連する状況を表すものであるとする。空間的情報、時間的情報に加え、過去の検索履歴もコンテキストに加えることによって、過去の検索履歴も容易に現在の検索に反映することが出来るようにする。

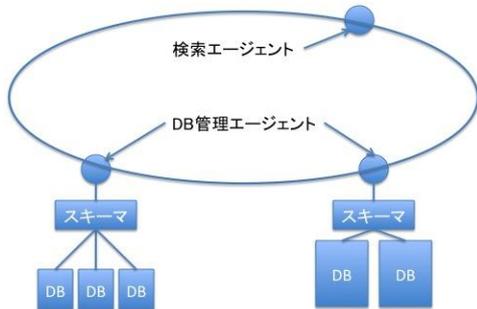


図4. 検索エージェントとDB管理エージェント

図4に示すように、検索エージェントが検索を行う場合はデータベース管理エージェントへクエリを投げる。クエリを受け取ったデータベース管理エージェントは仮想データベーススキーマに対してクエリを実行することによって、異種データベースに対して検索を行う。

①コンテンツのグループ化概要

P2P 環境でのマルチデータベース検索において、利用者の検索履歴と検索頻度を利用したコンテンツのグループ化手法とそのグループ化に基づく効率の良いコンテンツ検索法の開発を行った。コンテンツがグループ化されていない場合、利用者が欲しいコンテンツの情報毎にコンテンツの検索を行う必要がある。一方コンテンツがグループ化されている場合、あるコンテンツの検索を利用者が行った際に、そのコンテンツが属しているグループのコンテンツ情報を同時に取得することが可能となるため一回の検索で複数のコンテンツ情報を取得することができる。グループ化されているコンテンツ群は、多数の利用者の連続したコンテンツの検索によって行われているものである為、同じ趣向を持った利用者にとっても有用なコンテンツグループである可能性が高いと考えられる。コンテンツの主な例として料理の材料であったり、ファミリーレストランのメニューなどといった消費材など複数回検索されることが考えられるコンテンツなどを想定した。

②グループ化に関連するパラメータ

パラメータを表2に示す。ノード数、検索傾向度、検索回数期待値はシナリオファイルを利用して変更を行う。検索履歴リスト保持数、グループリスト保持数、関連性強度減少率、関連性強度基準値、推薦有無のパラメータに関してはオーバーレイネットワークを変更す

る。ノード数は、起動ノード数とその中でコンテンツを保持しているノード数とを示している。本実験では、起動している全ノードがコンテンツを保持しているとは限らないとしている。現実でも情報取得者が必ず情報提供をしているとは限らない為である。その為、コンテンツ保持ノード数は1000ノードと固定し検索ノード数の変更を行い両者の割合を変更させ実験を行っている。

表2. グループ化に関連するパラメータ

パラメータ名	説明
起動ノード数	実験時に起動する全ノード数.
コンテンツ保持ノード数	コンテンツを保持するノード数.
検索履歴リスト最大数	検索を実行する利用者ノードが保持する検索履歴リストの最大エンタリー数.
グループリスト最大数	自ノードのコンテンツと関連性の高いノード情報を保持するグループリストの最大エンタリー数.
関連性強度減少率	1日経過した際に各ノードとの関連性強度が減少する割合. 減少率がX%, 現在の関連性強度をSとした場合, 減少後の関連性強度は $(S \times X / 100)$ となる.
関連性強度基準値	グループ化が行われているかを判断するための関連性強度の閾値.
検索傾向度	各利用者が1日に検索するコンテンツの傾向度. この値により検索を行う際に予め設定した仮グループ内のコンテンツの検索, 取得を行う傾向を設定する.
検索回数期待値	各利用者が1日に検索を実施する回数の期待値.
推薦機能の有無	検索した際に, 検索対象ノードのグループリストを利用者へ推薦し, そのグループリストを利用し検索を行うかどうかの有無.

③実験結果と考察

本実験では、各パラメータを変更し実験を行った。特徴的な評価結果と考察について述べる。グループ化の評価は以下の精度と再現率を用いて行った。

精度：グループ化されているノードの中で、初期設定の際に設定したコンテンツの仮グループ通りにグループ化が行われているノードの割合

再現率：コンテンツを保持しているノードのうち初期設定において関連性のあるコンテンツの仮グループに設定したとおりのグループ化が行われているノードの割合

<検索傾向度 80%の場合>

検索傾向度が80%の場合は初期設定で予め設定していた仮のグループ内の検索傾向度が80%となっていることから、各ユーザが1サイクルの検索において似た性質のコンテンツを頻繁に検索していることになる。そのため、推薦機能が無いケースにおいても非常に高い精度、再現率となっている。

< 検索傾向度 50%の場合 >

検索傾向度 50%の場合は推薦機能が無い場合は検索傾向度 80%に比べて精度、再現率共に低下しているのに対し、推薦機能がついている場合は、検索傾向度 80%と比較しても精度、再現率共に大きく低下していないのが特徴である。この実験の結果から、ユーザが1回の検索サイクルにおいてリピートの度合いが半分程度の検索をしていたとしても、その中から必要なコンテンツを選びグループ化を行うことが出来る事が分かる。

< 検索傾向度 20%の場合 >

検索傾向度 20%の場合は各ユーザの検索が極めてランダムの場合となる為、推薦を行う際に誤った推薦を行ってしまう。その為、効率的なグループ化を行うことが出来ず、精度、再現率共に低い結果になったと考えられる。

今回の評価は、ノード数を固定し本稿で提案したP2Pネットワークにおけるコンテンツのグループ化の基本的な機能が有効であるかを検証した。検索傾向度 50%のケースでも、簡単な推薦機能を付加することによりグループ化が有効に働くことが確認できた。今回示したのは、関連性強度減少率が5%、関連性強度基準値 3、の場合である。特に、減少率5%はもっとも緩やかな設定値であるが、15日ではほぼ関連性強度が半減するペースであり今回のパラメータ設定値として妥当であったといえる。本グループ化機能は、コンテンツが利用者からコンテンツ要求を受けた際に動的にコンテンツのグループ化を利用者の検索状況に応じて行っている。そのためP2Pネットワークを対象とした動的なノードの変化(参加や離脱)に対応した機能であるといえる。また検索要求を行った際にそのコンテンツが属しているコンテンツのグループを提供することによって、結果的には検索数を減少させ効率的にコンテンツの検索・取得が行えるようになった。

5. 主な発表論文等

[雑誌論文] (計 2 件)

- ① 佐々木 拓也, 澤本 潤, 和田 雄次, 加藤 貴司, P2P ネットワークにおけるコンテンツのグループ化に関する研究、情報処理学会論文誌、査読有、Vol. 52, No. 2, pp. 359-367 (Feb. 2011).
- ② 赤崎 英之, 加藤 貴司, ベッド バハドゥール ビスタ, 高田 豊雄, 澤本 潤, Skip Graph における範囲を持つ属性をキーとしたコンテンツ共有手法の提案, 電子情報通信学会技術研究報告 (NS2009-153), 査読無, pp. 65-70 (Jan. 2010).

[学会発表] (計 17 件)

- ① 三井田 浩, 和田 雄次, データマイニン

グ結果の可視化、比較・分析支援ツールの開発、情報処理学会第 71 回全国大会、立命館大学 びわこ・くさつキャンパス (2009 年 3 月 11 日).

- ② 赤崎 英之, 加藤 貴司, ベッド バハドゥール ビスタ, 高田 豊雄, 澤本 潤, Skip Graph における範囲を持つ属性をキーとしたコンテンツ共有について, 2009 年ソサイエティ大会講演論文集 (通信講演論文集 2), p. 32, 新潟大学 (2009 年 9 月 16 日).
- ③ 渡辺 裕太, 菖蒲 佳右, 和田 雄次, 澤本 潤, 加藤 貴司, 異種データベースの仮想化技術, 第 8 回情報科学技術フォーラム (FIT2009) 講演論文集, pp. 211-214, 東北工業大学 八木山キャンパス (2009 年 9 月 2 日).
- ④ 菖蒲 佳右, 渡辺 裕太, 和田 雄次, 澤本 潤, 加藤 貴司, 仮想 DB のバックアップ・リカバリ方式, 第 8 回情報科学技術フォーラム (FIT2009) 講演論文集, pp. 215-218, 東北工業大学 八木山キャンパス (2009 年 9 月 2 日).
- ⑤ 佐々木 拓也, 澤本 潤, 瀬川 典久, 杉野 栄二, 加藤 貴司, 和田 雄次, データマイニングによるデータハッシュテーブルの階層分類構造自動構築機能を利用した知的検索システムの提案, 第 8 回情報科学技術フォーラム (FIT2009) 講演論文集, pp. 423-428, 東北工業大学 八木山キャンパス (2009 年 9 月 2 日).
- ⑥ Takuya Sasaki, Hideyuki Akasaki, Jun Sawamoto, Takashi Katoh, Yuji Wada, Norihisa Segawa, Eiji Sugino, A Proposal of an Adaptive P2P Database Retrieval System Using a Hierarchically Categorized DHT Structure, International Workshop on Informatics 2009 (IWIN2009), Hawaii, USA (2009 年 9 月 11 日).
- ⑦ Yuji Wada, Yuta Watanabe, Jun Sawamoto, Takashi Katoh, Database Virtualization Technology in Ubiquitous Computing, 6th International Conference on Innovations in Information Technology (Innovations '09), pp. 170-174, Al Ain, UAE (2009 年 12 月 15 日).
- ⑧ Yuji Wada, Yuta Watanabe, Keisuke Syoubu, Jun Sawamoto, Takashi Katoh, Virtualization Technology for Ubiquitous Databases, 2010 International Conference on Complex, Intelligent and Software Intensive Systems, pp. 555-560, Krakow, Poland (2010 年 2 月 15 日).
- ⑨ 佐々木 拓也, 澤本 潤, 和田 雄次, 加

- 藤 貴司, 杉野 栄二, 瀬川 典久, P2P ネットワークにおけるコンテンツのグループ化に関する研究, 情報処理学会東北支部研究会, 東北学院大学(2010年2月12日).
- ⑩ 渡辺 裕太, 菖蒲 佳右, 和田 雄次, 澤本 潤, 加藤 貴司, 異種データベースの仮想化技術—スキーマ変換方式—, 情報処理学会創立50周年記念(第72回)全国大会(第1分冊), pp.815-816, 東京大学本郷キャンパス(2010年3月8日).
- ⑪ 菖蒲 佳右, 渡辺 裕太, 和田 雄次, 澤本 潤, 加藤 貴司, 異種データベースの仮想化技術—クエリ変換方式—, 情報処理学会創立50周年記念(第72回)全国大会(第1分冊), pp.813-814, 東京大学本郷キャンパス(2010年3月8日).
- ⑫ Yuji Wada, Yuta Watanabe, Keisuke Syoubu, Jun Sawamoto, Takashi Katoh, Virtual Database Technology for Distributed Database, IEEE 24th International Conference on Advanced Information Networking and Applications Workshops (FINA2010), Perth, Australia, pp.214-219 (2010年4月20日).
- ⑬ 佐々木 拓也, 澤本 潤, 和田 雄次, 加藤 貴司, P2P ネットワークにおけるコンテンツのグループ化に関する研究, マルチメディア, 分散, 協調とモバイル (DICOMO 2010) シンポジウム論文集, 岐阜県下呂温泉 水明館, pp.652-658 (2010年7月7日).
- ⑭ 渡辺 裕太, 菖蒲 佳右, 三井田 浩, 和田 雄次, 澤本 潤, 加藤 貴司, ユビキタス環境におけるマルチデータベースの仮想化技術, マルチメディア, 分散, 協調とモバイル (DICOMO 2010) シンポジウム論文集, 岐阜県下呂温泉 水明館, pp.1262-1267 (2010年7月7日).
- ⑮ Takuya Sasaki, Jun Sawamoto, Takashi Katoh, Yuji Wada, Norihisa Segawa, Eiji Sugino, A Proposal of P2P Content Retrieval System Using Access-Based Grouping Technique, R. Setchi et al. (Eds.): Knowledge-Based and Intelligent Information and Engineering Systems (14th International Conference on Knowledge-based and Intelligent Information and Engineering Systems (KES 2010) Proceedings, Part III), LNAI 6278, Springer, Cardiff, UK, pp.455-463 (2010年9月8日).
- ⑯ Yuta Watanabe, Keisuke Syoubu, Hiroshi Miida, Yuji Wada, Jun Sawamoto, Takashi Katoh, Technology for Multi-database Virtualization in a Ubiquitous Computing Environment, International Workshop on Informatics 2010 (IWIN2010), Edinburgh, UK, pp.89-96 (2010年9月13日).
- ⑰ Takuya Sasaki, Jun Sawamoto, Takashi Katoh, Yuji Wada, Norihisa Segawa, Eiji Sugino, Access-Based Contents Grouping on P2P Network, International Workshop on Informatics 2010 (IWIN2010), Edinburgh, UK, pp.131-138 (2010年9月13日).

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
加藤 貴司 (KATOH TAKASHI)
岩手県立大学・ソフトウェア情報学部・講師
研究者番号: 20323115
瀬川 典久 (SEGAWA NORIHISA)
岩手県立大学・ソフトウェア情報学部・講師
研究者番号: 20305311
杉野 栄二 (SUGINO EIJI)
岩手県立大学・ソフトウェア情報学部・講師
研究者番号: 10293391
(H20→H21: 連携研究者)