

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 5月31日現在

機関番号：14401

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22500057

研究課題名（和文）

データ独立性の概念を実現する分散情報共有技術に関する研究

研究課題名（英文）

A Study on Distributed Information Sharing Mechanisms Realizing the Notion of Data Independence

研究代表者

春本 要 (HARUMOTO KANAME)

大阪大学・工学研究科・准教授

研究者番号：50263215

研究成果の概要（和文）：本研究課題では、データベース工学の分野での重要な概念であるデータ独立性を分散情報共有に適用する技術の確立を目的としている。主な研究成果として、P2P型エージェントプラットフォーム上でのアプリケーション非依存な分散情報共有を可能にするデータ管理エージェントの設計、および、それに基づくサービスエージェントとして条件合致データ数を提供するサービスエージェント、ルールに基づくデータ処理を行うサービスエージェントの設計と実装を行った。

研究成果の概要（英文）：The objective of this research is to establish a distributed information sharing mechanisms employing the notion of data independence, which is a fundamental idea in the database engineering field. One of the main achievements of this research is the design of a data management agent on a P2P agent platform which enables application-independent distributed information sharing. We also designed and implemented two kinds of service agents based on the implemented data management agents, i. e., an agent to get the number of data which meets a given condition and an agent to perform data processing based on rules.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	1,300,000	390,000	1,690,000
2011年度	1,000,000	300,000	1,300,000
2012年度	700,000	210,000	910,000
年度			
年度			
総計	3,000,000	900,000	3,900,000

研究分野：データ工学

科研費の分科・細目：情報学・計算機システム・ネットワーク

キーワード：データ独立性、分散情報共有、Rete アルゴリズム

## 1. 研究開始当初の背景

1975年、ANSI/X3/SPARC委員会によって内部スキーマ、概念スキーマ、外部スキーマからなる3層スキーマアーキテクチャが提唱され、いわゆるデータ独立性の概念が普及した。現在のデータベース管理システムにおい

ても、3層スキーマアーキテクチャが目指したデータ独立性の概念、つまり、データ管理とアプリケーションあるいはサービスを独立させるという概念は、データの柔軟な活用・共有において重要な概念として捉えられており、実際に3層スキーマアーキテクチャ

に基づいて設計されたデータベース管理システムは情報システムの中核として成功を収めている。

一方、インターネットの普及および高速化により、発生するデジタル情報の量は爆発的に増加している。また、データの利用ユーザは旧来のデータベースシステムでは特定多数であったものが、現在ではインターネットを介した不特定多数へと変化している。このような状況の変化から、情報提供サービスにおける情報管理・提供形態もまた変化している。つまり、情報発生源（人間、各種センサーなど）の情報をサーバに集約し、その情報を直接ユーザに提供するのではなく「サービス」として何らかの形に加工した上で提供するという形態である。このような情報管理・提供形態は、利用ユーザに広く画一的なサービスを提供するには適しているが、個々のユーザから見た場合、情報提供形態がサービスによって決定されてしまうため、柔軟な情報活用ができない。

サービス主導型の情報管理・提供形態では、本来より有効に活用できる情報をサービスが制限してしまっているといえる。データ独立性の概念と対比するならば、データとサービスの間の独立性がなく、サービスの背後にデータが隠ぺいされ、データレベルでは組織を越えたデータ共有ができていない状態といえる。

## 2. 研究の目的

インターネット環境において真に柔軟な情報管理形態を考えた場合、データとサービスが独立していない状態は望ましくない。そこで本研究課題では、インターネット上で真に柔軟な情報管理形態の確立、および、その実現技術の確立を目的とする。具体的には、インターネット上での情報共有においてもデータ独立性の概念を導入し、サービスレベルではなくデータレベルでの安全かつ柔軟な情報共有を実現する手法を考案する。

その形態としては、情報発生源からのデータをサービス提供者に集約し提供する形態ではなく、サービスから情報を切り離し、いわゆる P2P (Peer to Peer) 型の形態、つまり、情報発生源が直接情報を管理し提供する形態が、個々の情報発生源でのアクセス制限を含む情報管理を実現しやすい点で望ましいと考えられる。したがって、P2P 型アーキテクチャを基盤とした情報管理・提供形態を基本として考える。

研究代表者はこれまでに P2P 型アーキテクチャによって地理的範囲を条件とした検索要求に対応する P2P ネットワーク LL-Net を提案し、さらにその上で位置情報を有する情報を効率的に管理・検索する手法を提案している。また、LL-Net の機構を含む P2P 型アプ

リケーションプラットフォームとして PIAX (P2P Interactive Agent eXtensions) と呼ぶシステムを開発してきた。このように構築してきたシステムを基盤として安全かつ柔軟にインターネット上でのデータ共有を実現するアーキテクチャを提案し、いくつかのサービス実装を通じてその実現可能性を示すことが本研究の目的である。

## 3. 研究の方法

(1) これまでに開発してきた P2P エージェントプラットフォーム PIAX を基盤とし、まず P2P 型アーキテクチャによる細粒度のデータ共有を実現するための手法を考案し、それに基づいたデータ管理エージェントを設計・実装する。

(2) サービスエージェントとして、指定した条件に合致するデータ数を提供するサービスエージェントの構成法を検討し、実装・評価する。

(3) サービスエージェントとして、ルールに合致するデータをユーザに提供するサービスエージェントの構成法を検討し、実装・評価する。

## 4. 研究成果

(1) P2P エージェントプラットフォームを基盤としたデータ管理エージェントの設計・実装

まず、データ独立の概念を実現するためのデータ表現形式の検討を行った。細粒度のデータ共有のためのデータ表現形式として、データベースにおいて広く利用されている関係データモデルに基づくデータ表現形式と、インターネット上で用いられている XML や RDF について検討を行った結果、データ表現能力、扱いやすさの両面で RDF が優れており、また、アプリケーションからも効果的に活用できることから、RDF を基礎とするデータ表現を採用することとした。また、RDF データを大規模分散環境で共有する際に必要となるデータ格納形式について、関係データモデルに変換し関係データベースを用いる手法や、RDF をネイティブにサポートする RDF データベースの調査検討を行った。

次に、RDF によるデータ表現を基盤としたデータ管理エージェントの設計を行った。特に、センサから出力されるデータを管理するセンサエージェント、および、センサエージェントから登録されるデータを管理するデータ管理エージェントについて、センサデータのコンテキストとしての有効活用を考慮した管理手法の検討を進めた。具体的には、センサが設置された位置に基づく情報共有だけでなく、センサの有効範囲も考慮し、さらにセンサデータから得られる高レベルのコンテキスト情報も効果的に共有できるよ

うな、R-tree を基礎とした分散型センサデータ管理手法の設計を行った。本手法はセンサデータを単純に共有する手法よりもスケラビリティに優れた手法である。さらに設計したセンサデータ管理手法を PIAX エージェントとして実装し、その動作検証を行った。

また、データ独立性を実現するために必要となる、データ生成者によるアクセス管理機構について検討を行った。アクセス権の設定については、個々の RDF データにアクセス権を設定するのは煩雑であるため、ある程度まとまった単位でアクセス権を設定できるように、RDF データのクラスや条件を用いたアクセス権指定法を考案した。

### (2) 条件合致データ数を提供するサービスエージェントの設計・実装

サービスエージェントの一例として、単に条件に合致するデータを提供するサービスではなく、P2P ネットワーク上の各ノードが保持するデータについて、指定された条件に合致するデータ数を効率的に推定し提供するサービスエージェントの設計と実装を行った。特に、条件に合致するデータをすべて検索しカウントする方法は大きなコストがかかるため、データ管理エージェントを P2P ネットワークの一種である Chord のリンク構造に従って組織化し、Bloom Filter で表現された各エージェントの保持データを Chord のリンク構造に従って算術和を用いて集約した Spectral Bloom Filter から、指定条件に合致するノード数を効率的に推定する手法を考案した。提案手法の評価結果から、よい精度で合致ピア数を推定できることを示した。この成果は、大規模分散環境で発生するデータ群について、効率的に概数を把握できる点で、大きな成果である。

### (3) ルールに基づくデータ処理を行うサービスエージェントの設計・実装

時々刻々と生成されるセンシングデータなどの情報を効果的に活用するためには、それらをアプリケーション毎に定義されるロジックに基づいてデータストリームとして処理することが求められる。そのような機能をサービスエージェントとして提供できれば、データ独立性を保ちつつ、容易に多様なアプリケーションを開発することが可能になる。

そこで、分散配置された様々なセンサから発生するセンシングデータを「環境コンテキスト」として抽象化し、環境コンテキストに応じてルール処理により適切なデータ（コンテンツ）を提供するサービスエージェントの設計と実装を行った。

コンテキストウェアサービスにおいては、ユーザ（人）の状況とフィードバックす

べきコンテンツの状況とをマッチングし、状況に適合するコンテンツをユーザにフィードバックすることが必要である。例えば飲食店を推薦するようなコンテキストウェアサービスでは、ユーザの環境コンテキストとユーザ周辺の飲食店の周辺の環境コンテキストとをマッチングすることが必要である。図1は、ユーザuに対し、周辺のコンテンツ（飲食店）のうちで屋内型の飲食店か、もしくは屋外型飲食店で周辺の雨量センサの値が0であるような飲食店の情報をフィードバックするようなルール例である。このように、一般にセンシングデータから得られる環境コンテキストとユーザのマッチングは、ルールを用いて表現することができる。

```

for user u, content c, sensordata s
if u.requestArea contains c.loc and
(c.status = 'INDOOR' or
(c.status = 'OUTDOOR' and s.type = 'RAINFALL'
and s.observedArea contains c.loc and s.value = 0))
then
deliver content c to user u

```

図1 ユーザとコンテンツのマッチングルールの例

このような膨大なユーザ、コンテンツ、センサのマッチング処理をリアルタイムに実現するサービスエージェントの設計においては、何らかの最適化が必要である。ルールに基づくマッチング処理を効率的に実現するアルゴリズムとして、Rete アルゴリズムがある。Rete アルゴリズムは、ルールの集合から Rete ネットワークと呼ばれるネットワーク構造を生成し、そのネットワーク構造上にユーザ、コンテンツ、センサなどのデータを投入することによって効率的にマッチング処理を行うアルゴリズムである。図2は図1のルールに対する Rete ネットワークの例である。Rete ネットワークの各ノードはフィルタ処理を行う  $\alpha$  ノード、複数のデータのマッチング処理を行う  $\beta$  ノードなどから構成され、それぞれ処理結果を蓄積するワーキングメモリ（それぞれ  $\alpha$  メモリ、 $\beta$  メモリと呼ばれる）を有する。このワーキングメモリを活用することにより、Rete アルゴリズムはデータのマッチング処理を効率化している。

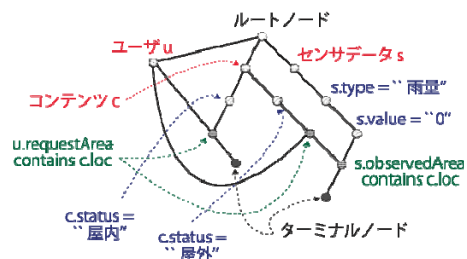


図2 Rete ネットワークの例

しかし、Rete アルゴリズムは元来プロダクションシステムにおけるルール処理を効率化する目的で開発された技術であり、地理的な要素を含むマッチング処理はそのままでは効率的に処理できない。また、膨大なユーザ数やセンサ数を扱うためには、分散型の処理によって負荷分散を実現する必要がある。そこで、特にセンシングデータとコンテンツの地理的な位置関係に基づくマッチング処理を効率化するため、Rete ネットワークの各ノードの処理を地理的なブロックに分割して処理するとともに、分割したブロックの処理を P2P オーバレイネットワーク上の多数のピアで分散処理する手法を提案した。

図 3 はこの手法の概要を表している。Rete ネットワークの各ノードの処理を地理的なブロックによって分割し、それぞれのブロックを担当するピアはそのブロックに関するデータのみを上流ノードから受け取る。これにより、地理的な近接性や包含性を検査するようなマッチング条件の処理を局所化し、マッチング処理を効率化する。また、各ノードの各ブロックの処理を担当するピアを、マッチング条件とブロック ID から決定することにより、P2P オーバレイネットワーク上の多数のピアでの分散処理を実現する。

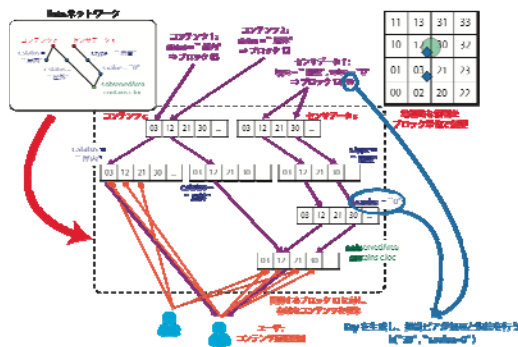


図 3 提案するサービスエージェントの処理の概要

提案手法の有効性を評価するため、シミュレーションによる評価を行った。このシミュレーションでは、32km 四方の領域を想定し、センサとコンテンツがそれぞれ領域内に 512 個ずつランダムに配置された環境において、ユーザに対して周囲 4km 内で図 1 のルール例に基づきコンテンツを配信するようなアプリケーションを想定した。また、センサの観測範囲は周囲 1.5km とした。

図 4 と図 5 は、分割するブロックのサイズが提案手法での分散処理における通信負荷に与える影響を、ピア数を 32 と 128 とした場合で評価した結果である。この評価結果から、ブロック分割しない場合は特定の 7 ピアに通信負荷が集中するのに対し、ブロック分

割を行った場合は通信負荷が分散されていることがわかる。特に、ブロックサイズを 8km とした場合が全体として効果的な負荷分散を実現していることがわかる。ブロック分割を細かくしすぎると多くのブロックがセンサの有効範囲に含まれることになるため、全体として通信負荷が増大することから、適切なブロックサイズは、センサがもつ有効範囲と密接に関連していることがわかる。また、この評価は通信負荷だけを表しているが、特に複数のデータのマッチング処理を行う  $\beta$  ノードの処理はブロック分割によって地理的に局所化されており、ブロック分割を行わない場合と比較して大幅に処理を効率化している。

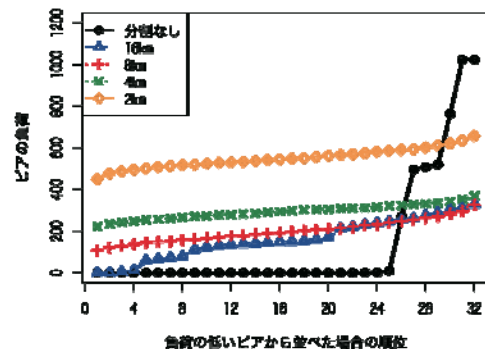


図 4 32 ピアの場合の負荷分散効果

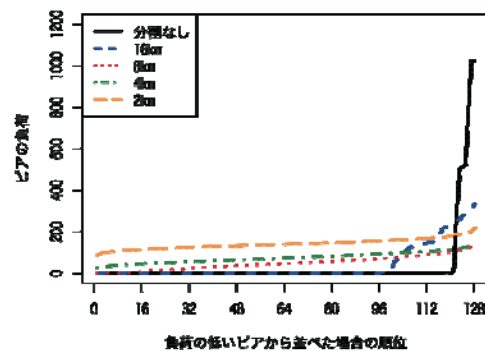


図 5 128 ピアの場合の負荷分散効果

## 5. 主な発表論文等

〔雑誌論文〕(計 3 件)

① 川上朋也, 寺西裕一, 春本 要, 下條真司, 大規模コンテキストウェアサービスのための Rete アルゴリズムを用いた地理的コンテンツ配信手法, 電気学会論文誌 C (査読有), Vol. 133, No. 4, pp. 785-793, 2013 年 4 月.

② 川上朋也, 寺西裕一, 春本 要, 下條真司, センサデータの意味的解釈に基づく分散センサ情報管理システム, 情報処理学会論文誌 (査読有), Vol. 54, No. 2, pp.

730-739, 2013年2月.

③ 川上朋也, 竹内 亨, 寺西裕一, 春本 要, 西尾章治郎, 下條真司, 地理的オーバーレイネットワークを用いた位置依存コンテンツ分散管理手法, 情報処理学会論文誌(査読有), Vol. 52, No. 12, pp. 3289-3299, 2011年12月.

[学会発表](計3件)

① 川上朋也, 寺西裕一, 春本 要, 下條真司, 大規模コンテキストウェアサービスのための Rete アルゴリズムを用いた地理的コンテンツ配信手法について, 情報処理学会研究報告, 2012-DPS-153, No. 4, 2012年11月, 滋賀県彦根市.

② 白木 徹, 寺西裕一, 春本 要, 竹内 亨, 西尾章治郎, P2P ネットワークにおける属性条件合致ピアの概数を推定可能なピア検索手法の検討, 第3回データ工学と情報マネジメントに関するフォーラム(DEIM2011)論文集, C6-3, 2011年3月, 静岡県伊豆市.

③ 川上朋也, 竹内 亨, 寺西裕一, 春本 要, 西尾章治郎, 地理的オーバーレイネットワークの使い分けが可能な位置依存コンテンツ分散管理手法の提案, 情報処理学会第18回マルチメディア通信と分散処理ワークショップ論文集, pp. 81-87, 2010年10月, 宮崎県宮崎市.

## 6. 研究組織

### (1) 研究代表者

春本 要 (HARUMOTO KANAME)

大阪大学・大学院工学研究科・准教授

研究者番号: 50263215