

論文内容の要旨

博士論文題目 Studies on Integration of Heterogeneous Information Sources
(異種情報源の統合に関する研究)

氏 名 宗像浩一

コンピュータネットワークの発達に伴い、様々な組織で独立に管理された種々の情報源から情報を取得できるようになった。このように個別に管理された既存の豊富な情報源を最大限に活用するためには、これらを統合したビューを提供することが望まれる。

本研究では、異種の情報源を統合するための問題点を取り上げ、それらを解決するための手法を提案する。システム構成としては、メディエータ (mediator) によるデータ統合システムを想定する。即ちアプリケーションプログラムからの問合せ要求を受けて、メディエータと呼ぶソフトウェアモジュールが、各情報源に対して副問合せ要求を発行する。それらの結果を取得して統合し、アプリケーションプログラムに問合せ結果として返す。

まず最初に、メディエータが連言 (conjunctive) 問合せを受けて、複数の情報源から得られる最大限のデータを結合によって生成するための手法を述べる。異種情報源から取得したデータは一般的に、その内容とデータ構造が一樣ではなく、データが欠落したりデータインスタンス毎に構造が異なる場合がある。これらを統合して最大限の情報を取得するためには、アウトジョイン (outerjoin) が必須となることを示し、これを用いた問合せ処理手順の作成アルゴリズムを述べる。次に情報源統合のために述語と外部関数が必要になることを述べ、それらを含めた問合せ処理手順の作成アルゴリズムを提案する。

次に、自己記述形式のオブジェクトモデルを用いて異種情報源を統合するための、問合せ処理方式を提案する。このデータモデルにおいて各データは、オブジェクト識別子、自己の値の意味を記述したラベル、値、および値の型、を持つサブオブジェクトの入れ子構造によって表現される。このように構造が単純なため、種々のデータモデルに基づくデータをこのデータモデルに変換することが可能である。データインスタンス毎に構造が異なる半構造データも表現することができる。このような柔軟な構造を許容する反面、このデータモデルに基づいて問合せ処理を行うためには、従来の関係データベースで用いられていた問合せ処理では対処できない種々の問題が発生する。そこでそれらを解決するための手法を示し、これにより異種情報源から柔軟に所望の情報を取得できることを示す。

最後に、一定周期で発生するデータで構成される周期データ列において、周期が互いに異なる場合に、これらを統合するための処理方式を提案する。リアルタイムシステムにおいて、異なる周期を基本に設計されたセンサやシステムを統合するためには、発生間隔の異なる複数の周期データ列を取り扱う必要が生じる。この際に、各データ列から1個ずつのデータを選択してデータ組合せを得るための基準として、鮮度と同期度を提案する。鮮度は、データ組合せ中で最古のデータ発生時刻と現時刻の差とし、同期度は、最新と最古のデータの発生時刻差として定義する。さらに両者を線形結合した評価関数を定義する。まずデータの同期度に着目していくつかの性質を明らかにし、同期度の上限を決定する。そして最適データ組合せを得るための、問合せ処理の枠組みを示す。最後にこれを、道路上の障害物を2種類のセンサで検出するシステムに適用する例を検討し、単純なデータ選択方法に比べて同期度がほぼ半減することを示す。

(論文審査結果の要旨)

近年のコンピュータネットワークの発達に伴い、種々の情報源を統合して、利用者にビューを提供することが重要な課題となっている。ここで対象とすべき情報源としては、関係データベースやオブジェクト指向データベースのみならず、独自仕様のデータベース、シミュレータ、センサなどが考えられる。このような情報源は一般に種々の異種性および自律性を持っており、これが情報源を統合する際の障害となっていた。本論文ではこのような問題点を解決して情報源を統合するための、問合せ処理およびデータ取得方式を示している。本論文の主な成果は次のように要約される。

(1) 複数の関係を結合したビューに対して連言 (disjunctive) 問合せを発行した際に、問合せ条件を満たす最大限の情報を取得すべきことを主張している。そのためには関係の結合にアウトジョインを使用してフルディスジャンクション (full disjunction) を求め、その中で問合せ条件を満たす組 (tuple) を取得するのが適当であることを示している。そしてそれを効率良く実現するアルゴリズムを示している。さらに情報源統合のために述語と関数を利用する必要性を主張し、それを実現するためのアルゴリズムを提案している。

(2) 異種情報源に含まれる種々の構造を持つデータを統合するために、自己記述形式のデータモデルを用いて問合せ処理を実現するための方式を提案している。このデータモデルに基づく問合せ処理においては、関係データベース向けの従来の問合せ処理方式で対処できない種々の問題点が生じる。そこでそれらを解決するための問合せ処理方式を提案し、これにより異種情報源から柔軟に所望の情報を取得できることを示している。上記 (1) および (2) で提案した方式は実システムに適用され、各地の自治体の上水道設備管理に使用されている。

(3) 一定周期でデータを発生する複数の情報源を統合するために、ある瞬間のデータの組合せ (スナップショット) を近似するデータを選択する必要性を主張している。このようなデータの選択基準として、データの鮮度と同期度を挙げ、これらを勘案した評価関数に基づいて最適なデータの組合せを取得するという新たな考えを提案している。またデータの同期度に関する種々の性質を明らかにしている。そして最適なデータの組合せを効率良く選択するため、異なる目的に応じた3種類のアルゴリズムを示している。さらにこのアルゴリズムのひとつは、周期データ列のみならず、ランダムに発生するデータ列にも適用可能であることを示している。以上のデータ取得方式を知的交通システム (ITS) に適用した例を示し、提案した方式の有用性を確認している。

以上のように本論文では異種情報源を統合する際に生じる、種々の問題点を解決する手法を提案している。これらの研究成果は、学会論文誌3件、査読付国際会議4件として公表され、学術面での貢献が認められる。さらに成果の一部は特許出願されるとともに産業分野で実用されており、実用面での貢献が認められる。よって本論文は博士 (工学) の学位論文として十分に価値あるものと判断する。