

学 位 論 文 題 名

New Frameworks for Integrating and Visualizing Relational Databases and Web Resources

（関係データベースおよびウェブ上の資源に対する
新たな統合・可視化フレームワークの提案）

学位論文内容の要旨

本論文では関係データベース上のデータを対象として、これをインタラクティブにウェブ上のデータやサービスと組み合わせて利用し、さらにこれを可視化する対話的な作業環境を実現するフレームワークを提案する。

近年の科学研究では実験および観測技術の高度化と自動化に伴い、莫大な研究データが日々収集され蓄積されている。このような研究現場において、対話的な情報可視化技術はデータの単なる表示を超え、データを様々な観点から観察し新たな知見へと結びつける有用な技術の一つである。

また各分野で研究プロジェクトの国際化と大規模化が進みつつある中、多数の研究者による円滑な共同作業を実現し、さらに研究成果の速やかな社会への還元を目的として、実験データや解析サービスなど研究資源の共有化と公開の重要性が高まりつつある。

結果として、現在は各研究分野において様々な研究資源をウェブを介して利用可能な状況にある。

この状況下において、各研究者は一方に実験データなど自身の研究資源、他方にウェブ上で広く公開された研究資源を持ち、これらを組み合わせることでデータの解析を行う事が日常的な要請となっている。さらに解析結果を評価するために可視化技術と組み合わせる必要がある。

今日このような組み合わせ作業を実現するにはプログラムの技術を必要とするのが一般的である。

しかし個々の研究者が持つ研究課題やアイデアは多様である。

その個別のニーズに対応するには、プログラミング技術を持たない研究者であっても、自分自身で手早く様々な情報・サービスを組み合わせ可視化を行うことが可能な環境の実現が求められる。

この問題に対して、本論文では関係データベース上の表データ、ウェブ文書として公開されたデータやウェブアプリケーションといったウェブ上の資源、可視化表示機能の三者を、エンドユーザ自身が自由に組み合わせ可視化可能な環境を実現する基盤フレームワークを提案する。

提案するフレームワークは、データベース可視化フレームワークである VERD フレームワーク、Web 上の資源の統合モデルである Web View Relation モデル、Web 文書から表形式で情報を切り出すダイナミック・プレビュー・インターフェイスの三つの技術の組み合わせとして実現されている。提案手法の主な特徴は次の三点である。

一点目は、関係データベース中のデータ、ウェブ上の情報やサービス、可視化表示の三者の連携を、関係データベースモデルにおける関係演算を用いて定義する点である。

具体的には、関係データベース上の表データのみならず、ウェブ上の資源や可視化表示についてもこれに関係データベースモデルにおけるビュー表を対応付け、これらの表データ間で関係演算を適用することで三者の連携を実現する。

可視化表示をビューに対応づける枠組みとして、VERD フレームワークは Visualization view モデルを提供する。また Web View Relation モデルはウェブ上の情報やサービスをビュー表として扱う枠組みである。

本論文ではこれらの提案モデルについて詳述し、実際に様々なデータやサービスの連携・可視化が関係演算を用いて定義される事を幾つかの具体例と共に示す。

二点目は、可視化表示やウェブ上の情報・サービスにビュー表を対応づけ、表データ間で関係演算を適用する新たなユーザインターフェイス技術を提案している点である。

これにより、プログラミング技術を持たない研究者であっても容易にデータやサービスの連携・可視化を行うことが実現される。

可視化表示とビュー表を対応づける方法として、VERD フレームワークは Query Flow モデルを提供する。これは可視化表示と対応したビュー表の定義をフロー図として視覚的に表現を行い構築する枠組みである。

またウェブ資源とビュー表を手早く対応づけるため、ダイナミック・プレビュー・インターフェイスを開発した。これはウェブ文書から表形式で情報を切り出すラッパーを手早く構築するインターフェイスである。ラッパー構築が一般的に試行錯誤の反復作業であることに着目し、この反復作業を積極的に改善する仕組みを実現した点は本研究のみが持つ新規性である。

三点目は前述の二点に基づく可視化環境をユーザ自身がアプリケーションを構築可能な部品化ソフトウェア・アーキテクチャーを用いて実現した点である。

本論文では提案手法に基づいて実装した可視化システムについて報告する。

この可視化システムにおいて、個々のアプリケーションは分解不可能な一つのソフトウェアではなく、組み合わせを変更可能な機能部品の集まりとして実現されている。これにより、研究者は自在に部品の組み替えや追加を行うことで、個々の研究目的に対応したデータやサービスの連携・可視化アプリケーションを構築することが可能である。

本論文で提案するフレームワークでは、まずウェブ上の資源も含めたデータやサービス・可視化機能の連携を関係データベースモデルに基づいて取り扱うモデルを導入し、次にこのモデルとユーザの間をインターフェイス技術や部品化ソフトウェア技術を用いて結びつける事で、ユーザ自身によるデータやサービスの連携・可視化を実現している。

学位論文審査の要旨

主 査 教 授 田 中 譲

副 査 教 授 原 口 誠

副 査 教 授 栗 原 正 仁

学 位 論 文 題 名

New Frameworks for Integrating and Visualizing Relational Databases and Web Resources

(関係データベースおよびウェブ上の資源に対する
新たな統合・可視化フレームワークの提案)

近年の科学研究では実験および観測技術の高度化と自動化に伴い、莫大な研究データが日々収集され蓄積されている。このような研究現場において、対話的な情報可視化技術はデータの単なる表示を超え、データを様々な観点から観察し新たな知見へと結びつける有用な技術の一つと位置づけられつつある。一方、科学技術の種々の分野で研究プロジェクトの国際化と大規模化が進み、多数の研究者による円滑な共同作業の実現と、研究成果の速やかな社会への還元を目的として、実験データや解析サービスなどの研究資源の共有化と公開の重要性が高まりつつある。その結果、あらゆる分野で種々の研究資源がウェブを介して公開されるようになってきた。

研究者は、ローカルデータベースに管理された大規模な実験データなどの研究資源と、ウェブ上に公開されている研究資源を相互に関連付けながら、特定の条件を満たすデータ群を抜き出し、それらの解析を行うという行為を日常的に手作業で行っている。このような研究過程を支援するには、ローカル・データベースとウェブ上のデータを区別することなく、データソースの指定、相互の関連付け、特定条件による検索、検索結果の可視化、条件の変更、可視化への反映、可視化結果におけるレコード選択、解析ツールの適用、といった操作のすべてをユーザがインタラクティブに定義し実行することができる可視化システムの実現が望まれる。従来、データベース可視化システムとウェブ情報可視化システムは可視化目的も機構も別のものとして研究されてきており、両方のデータソースを区別することなく相互に関連付けて可視化する技術は過去にはなかった。可視化のためのソースの指定や検索条件の変更、可視化スキームのインタラクティブな定義と実行をすべて支援できるシステムも過去には存在しなかった。

本論文は、このような問題を解決するために、以下の3つの基盤技術を新たに確立し、ローカル・データベースとウェブ上のデータを区別することなく相互に関連付けて可視化することができ、可視化のためのソースの指定や検索条件の変更、可視化スキームの定義と実行をすべてユーザ自身がインタラクティブに行うことができる可視化システムのフレームワークを提案している。

第1の基盤技術は、直接操作可能な視覚的部品を用いたデータベースの可視化フレームワークである。このフレームワークでは、検索要求式と可視化スキームの定義の両方を関係ビューとして定

式化し、ソースとなる表の指定、種々の関係演算、レコード表示に用いるテンプレート・オブジェクトの指定、可視化空間の座標軸指定など、検索要求式定義と可視化スキーム定義におけるあらゆる操作を関係ビューの書き換え演算として部品化する理論基盤を構築し、これらを直接操作可能な視覚的部品として実現し、ユーザがこれらの部品を組み合わせることによって検索要求と可視化スキームの定義・変更がインタラクティブに行える環境を実現することができる。レコード表示に用いるテンプレートとして、機能を持ったインタラクティブな視覚的オブジェクトを用いることができるので、単なる可視化ではなく、可視化結果としてインタラクティブなアクセス空間を生成することができる点と、検索要求と可視化スキームを区別することなく関係ビューとして統一的に扱い、その部品化に関する理論的根拠を与えた点の2点に、独創性と先駆性がある。

第2の基盤技術は、ウェブ文書中のリストや、ウェブアプリケーションの入出力を表データとして抽出するためのインタラクティブ・ウェブラッパーである。従来のウェブラッパーに関する研究は、抽出データの学習例を与え、機能推論を用いて精度の高い抽出の一般化規則を求めることに研究の主眼があった。これに対し、インタラクティブ・ウェブラッパーでは、抽出結果のフォーマットを表形式に限定し、学習例の提示数も2に限定することにより、抽出規則生成の計算時間を100m秒以内に短縮した。これにより、ユーザが第1学習例を固定した後、マウスドラッグ操作にあわせて第2学習例を次々と変えると、ラッパーは実時間で抽出規則を生成し、その規則で抽出される全項目を実時間でハイライトすることが可能になった。これにより、ユーザは、抽出結果をみながら正しい結果を導く第2学習例をマウスドラッグとマウスクリック操作のみで選択できるようにした。このような実時間インタラクティブ・ラッパーの研究は世界的にも先駆的である。さらに、ウェブアプリケーションにもインタラクティブ・ラッパーを拡張適用し、ウェブアプリケーションの入力をも関係表の属性とみなす拡張を行った。

第3の基盤技術は、ウェブ文書中のリストやウェブアプリケーションからインタラクティブ・ウェブラッパーで抽出された表データを、データベース中の表と区別することなく相互に組み合わせて利用な統合可視化フレームワークである。このフレームワークでは、ラッパーが生成した抽出規則も関係ビューとみなし、データベース上の関係ビューと区別することなく組み合わせて可視化を定義・実行することができる点に、独創性と先駆性が認められる。

これを要するに、著者は、データベースとウェブ上のデータに対するインタラクティブ統合可視化フレームワークに関して、理論的基盤を与えると共に、アプリケーション・フレームワークを確立したものであり、知識ソフトウェア工学に対して貢献するところ大なるものがある。よって著者は、北海道大学博士(工学)の学位を授与される資格あるものと認める。