

研究課題別研究評価

1. 研究課題名： 次世代応用指向データモデルの開発

2. 研究者名： 遠山 元道

3. 研究のねらい：

データベースシステムは、企業などにおける大規模データ管理の必須技術として、広く利用されている。従来の応用では、レコードベースの入出力でデータベースの外部インタフェースは大体用が足りていたが、近年 WWW や PDA との連携などにおいて、これらのアプリケーションドメインにおける構造化データと関係データモデルの構造上のミスマッチが大きなストレスを生み、多くのプログラマがインタフェースプログラミングに従事している。本研究では、検索言語 SQL にデータの構造化と各種メディアへの変換機能を持たせた SuperSQL の提案を中心に据え、これに基づいて種々の応用データを、データ、構造、メディア抽象の三要素に直交分解してモデル化する Trinity データモデルを提唱する。

4. 研究結果及び自己評価：

研究結果

データモデルを提案するということは、システム設計、インプリメント、プログラミングなどのためのフレームワークを与えることに相当し、その直接・間接の効果を端的に評価することは容易ではない。本研究では、従来混在していた概念を3つの直交要素に分離し、独立に論じる提案のメリットを、様々な切り口からの検証を積み重ねるアプローチをとった。

1) 新しい質問処理最適化手法： SuperSQL の質問文1つの実行によって、データベースのコンテンツから多段階のメニューナビゲートを含む Web サイトを生成することができる。しかし、全体を一つの質問で処理しようとする、データベースにわずかな更新があった場合でも全体の再生成が必要になる。ここでは、SuperSQL 質問言語に、別質問の動的呼び出し関数 invoke の機能を与える拡張を施し、これを用いて質問を複数の等価な質問に分解する手法を与えた。これにより、再生成を局所化することができる。複雑な質問では等価な分解が複数存在するので、コスト評価に基づき、最適な分解を選択することができる。このような最適化は、質問文が HTML 生成を行うという予備知識を利用することで初めて可能になる。Trinity モデルの三要素のうち、構造とメディア抽象から情報を受け、データ (SQL 質問) の部分の飛躍的な効率化を実現するという意味で、三元分解のメリットが大きく寄与している。

2) 高水準メディア抽象： ここでは三要素のうち、メディア抽象のみを考察対象としている。データベースという前提を一旦離れて一般のプレゼンテーションを観察してみる。たとえばルーブル美術館は多数の絵画を展示しているプレゼンテーションである。単に絵画の内容を利用者に伝えるだけならば、画集で事が足りるが、美術館での展示はむしろ美術館という建物自体のメディアとしての比重が極めて大きい。カーネギーホールでの演奏と、壁一枚外のセントラルパークでの演奏では同じコンテンツでも異なったプレゼンテーションとなる。データベースに絵画の画像情報が蓄積されるとして、SuperSQL によってこれらのプレゼンテーションを行う際、HTML、VRML というレベルでメディアを定義していたのでは、美術館の3D 仮想現実はあまりにも複雑すぎる。このとき「VRML で定義したルーブル美術館」をメディアとして捕らえ、フロア、部屋、壁面などをコンストラ

クタとして抽象化することにより、データベースの検索結果とメディアの任意の組み合わせによるプレゼンテーション（展覧会）の現実的な生成が可能となる。このように、VRML を言語レベルでなく、特定のデザインに限定して高水準のメディアと捕らえる考えかたにより、Trinity モデルの現実的な多くの応用への適用に道が開かれる。

3) 木構造スキーマの情報容量：この視点は、三要素のうち構造にのみ着目したものである。関係データベースから得られた情報に対し、プレゼンテーションにおいてグルーピングを施す。グルーピングの仕方によっては、情報の損失が発生する。損失は必ずしも悪いことではなく、関係代数の射影演算のように、情報のサマライズの目的で積極的に利用される。同一の属性集合を対象とした SuperSQL の質問文の集合において、得られるプレゼンテーションの情報量に基づく同値類を定義し、これらが完全束を成すことを示した。

4) 応用 SuperSQL の総合的な応用として、関係データベースから XML ドキュメントへの変換など、重要な応用に関する成果について、適宜報告を行った。

自己評価

データモデルの提案は、データベースという研究分野において、考えられる最大のテーマである。多くの研究は、特定のデータモデルに基づく言語の提案、論理的・物理的な設計手法の提案、データ構造やアルゴリズムの提案などである。しかも本提案は、従来データベースの範疇では無視されていた、データベース周辺領域をモデルの対象として取り込むことを目的としている。この提案が本当に実を結ぶには、アカデミズムや産業界において、その優位性が認められ、これに基づく研究開発が広く行われたり、国際規格が定められたりするようにならなければならない。そこまでの成功は、正直に言って遥かに先のことである。E.F. Codd が 1970 年に発表した関係データモデルにしても、製品が供されるのに 10 年、SQL の ISO 規格ができるまで 20 年近くがかかっている。

前項で述べた成果はいわば土台づくりの段階で、先は長い、私自身にとっては 3 年間の試練に耐え、ますますこのモデルの有用性、必然性に確信を深めることとなった。残念ながら 3 年の範囲に完結しなかったが、慶應義塾大学理工学部のデータベースの授業において、本年から SuperSQL を演習に取り入れる試みを始めた。いままで研究室の内部で、いわば作り手が自ら味見をしてうまいうまいと言っていたようなものだが、71 名の第三者に実際に試用し、評価をして貰った。作品集（非公開）は <http://www.db.ics.keio.ac.jp/ssql/submit2001J.html> にあり、多くの利用者から SuperSQL の有用性についてポジティブなコメントが寄せられた。これを励みに、ソフトウェアを広く世に問い、さきがけ研究の成果が長期的に、かつ世界的に広がる日を、かなりの確信と共に夢見ている。

5. 領域総括の見解：

20 世紀後半にデータベースとして、関係データモデルが提唱されて以来、確固たる成功を収めているが、今日 WWW のホームページなど日常的に目にする多くの情報は平坦な表とは異なる多彩なレイアウトに基づく表示になっている。遠山元道の研究は、関係データモデルの成功を踏まえて 21 世紀においてこれに代わる技術の基礎理論の創造と応用開発を行った。具体的には、関係データベースに対する質問言語 SQL の拡張として SuperSQL を開発し、これを道具として、データベースの対象を各種応用データの管理に拡大し、これを包括するデータモデルとして Trinity データモデルを開発、実装して応用につなげた成果は高く評価できる。

6. 主な論文等：

- *赤堀，有澤，遠山， SuperSQL による関係データベースと XML データの統合利用，情報処理学会論文誌 Vol.42 No.SIG 8 (TOD10), 2001 年 7 月 pp66-95
- Y. Maeda, M. Toyama, ACTIVIEW: Adaptive Data Presentation Using SuperSQL, 27th International Conference on Very Large Data Bases (VLDB2001), 2001 年 9 月
- 有澤，遠山 SuperSQL 処理系におけるグルーピング操作の効率的な実装，電子情報通信学会データ工学ワークショップ 2001，2001 年 3 月
- 高畑，石橋，藤沼，遠山 Magic Mirror Mailing: 個人情報データベースを利用する柔軟なメール配送シテム，情報処理学会 DBWS2001，2001 年 7 月