

グループ討論：データベース研究で何を行なうべきか

データベース研究の在り方

上林弥彦

京都大学工学部

情報工学分野の大きな特色は新しい分野がどんどん拡大してゆく点にある。データベース分野でも新しい問題が常に生み出されてきているが、最近では細かくならず論文を書くための研究が増えすぎているのではないかと批判が起きている。昨年あたりからSIGMODの顧問委員会、VLDBの理事会(Endowment)の委員の間でこの件に関するe-mailが飛び交い、現実問題に関係した論文を別のトラックで採用する動きがある。これはデータベース分野を代表する2つの国際会議の動向が研究課題に大きな影響があるという観点があるためである。データ工学国際会議ではもともとデータベースの現実的応用を主な課題にしているためこの問題は非常に深刻といえ、2つの基調講演やパネルでこの問題が取り上げられた。ここでは、この問題に関する意見のほかに、SIGMOD顧問委員会でのe-mailによる議論(92年秋)およびデータ工学国際会議のパネル(93年春)の内容についてまとめる。

1. データベース研究の在り方

研究テーマを大きく分けると次のようになる。

1. 新しい概念、新しいシステムに関するもの
2. すでに提案された概念およびシステムであるが将来方向として重要なもの
3. すでに提案された概念の拡張、整理
4. 知られている概念の合成
5. 新規性がないか練習問題程度のもの

1に関しては関係データベースではSystem RとINGRES、分散データベースではSDD-1が新しい概念を生み出すものとなった。しかしSDD-1が失敗したように概念の生成とプロジェクトの成功は必ずしも同じではない。2に関してはオブジェクト指向と従来型の共存や実用システムの開発が重要といえる。よく批判されるのは3や4の立場のものである。論文のための論文なのか本当に重要性があるのかのみきわめが重要といえる。メソポタミアに対数表があったように人間は必要となると発明できるものであるから、100年後に役立つかも知れない研究というのは脆井でしかない。日本人は概念としてまとめるのが下手といわれるが、これは博士課程への進学が少ないことも関係する。アメリカでも外国出身の学者が増えて、1を狙うよりもある程度の品質の論文を多数出すという傾向が現れている。

2. SIGMOD顧問委員会での議論の要約

Peter Bunemanが問題提起をし、Won Kimが現状の問題点として、1つの問題を深く追究したジャーナルスタイルの論文が採用されるために、一般の参加者が理解して学ぶことができないので、よりインダストリ向きな研究開発方向を示す論文を採用することが必要であると述べている。彼自身のRDBとOODBの統合という論文は実用的で方向を示すものだが、SIGMODやVLDBに受理されない性格のものであった。

Mike Stonebrakerは、SIGMODの参加者、特にインダストリからの参加者が減っていることに懸念を示している。対策として

- 1) 論文数を増やす。
- 2) 参加者にプロシーディングスを先渡しする。

- 3) 今までの性格の論文は1/3として発表以外に発表者によるポストセッションのような討議の場を設定する。

- 4) 2/3は招待講演、インダストリ論文とする。

- 5) この方針を何年間に継続する。

これに対してWon Kimは、展示会等で人集めをする会議の論文の質は良くないが、そういう所からも運営法等学ぶことがあると言っており、Mike Brodieは、目的が大学でテニユアを得るための論文となっているのが問題で、SIGMODの目的を変えるように提案した。この点に関してはJeff Ullmanが、科学を発展させるような論文が良い論文として認められるべきで、テニユアはその結果であると反論している。Brodieらがアカデミックとインダストリを区別しようとするのに対しUllmanは区別はないと言ってやる。Brodieは、Hartmanisらの記事を参照して、データベース研究の大きな成果は全て実際から出た(Bachman, Codd, System R, INGRES)ことを強調した。

3. データ工学国際会議のパネル

基調講演：GMDのTsichritzisは応用のあるテーマについて研究せよということ(ある人が数えたら)17回も繰り返していた。INRIAのValdoviezはアーキテクチャの人で、トランザクション処理の重要性を説いていた。

パネル：カリフォルニア大のStonebrakerが主宰した"Are We Polishing a Round Ball?"というパネルは最終日の午後であったにもかかわらず多くの出席者を集めた。パネリストはNeuhold(GMD)、Reuter(スツツゲルト大)、Agrawal(IBMアルマデン研)の3人でStonebrakerを含む4名が研究すべきことすべきでないことの項目を列挙し、最後にそれぞれの項目について支持する人数を調べた。以下で()はパネリストの中での支持数である。

[研究すべきこと]

協調・利用者インタフェース(3)
超巨大データベース(3)蓄積型(3)
マルチメディア(3)データ移動(3)移動分散(2)
メソッド最適化(2)マルチデータベース設計(1)
意味制約(1)オブジェクト(1)
システムに統合されたDBMS(1)

[研究すべきでないこと]

関係DB等既存のDBMS(3)再帰性(3)
マルチデータベーストランザクション(3)
新しいトランザクションモデル(3)
データの重複(3)新オブジェクト指向モデル(2)
不完全データ(2)マルチデータベース接続(2)
データ統合(1)関係モデルの一般化(1)
同期理論(1)

マルチデータベースに関しては問題を必要以上に複雑化しているという考えが基本にあり研究者の反論があった。トランザクションモデルではテキサス大のある教授が似たような論文を大量に出していることに対する反感があるようである。エスプリプロジェクトについては、産学共同の良い例だとNeuholdが自賛していたが、会社から資金を吸い上げるための機構に過ぎないとの意見もあった。

グループ別討論：

データベース研究で何を行うべきか

昨年7月の研究会に引き続き本年7月の研究会では、21世紀を展望した次世代データベースを目指した研究開発において、何に重点を置くべきかというテーマでグループ別討論を行った。参加者数は約50名であった。まず、参加者にアンケート用紙を配布し、各自が最も重要と考えている課題を挙げてもらった。その結果をもとに、事務局で3つの分野（グループ）を設定した。上林弥彦氏（京大）の問題提起を受けた後、参加者は各自希望するグループに分れて約2時間の討論を行い、最後に再度全員が集まって各グループの発表と討論を行った。次ページ以降は、各グループの討論結果をまとめたものである。なお、グループの構成は以下の通りである。

Aグループ： マルチメディア・データベースシステム研究（コーディネータ：打浪清一（九州工大）；20名）

Bグループ： オブジェクト指向とその関連研究（コーディネータ：吉川正俊（奈良先端大）；24名）

Cグループ： データベースの高機能化・高性能化研究（コーディネータ：井上潮（NTT）；9名）

（本ページ文責：井上 潮（NTT））

Group Discussion:

What are Research Issues in the Database Field?

We held a group discussion in this July's workshop on what are research issues for next generation databases aiming at the 21st century. This is a succession of the group discussion held last year. About 50 members participated this year. First, each participant proposed his/her most important issues and three groups are organized. Next, Prof. Yahiko Kambayashi made a presentation on future directions. Then, the three groups discussed on their issues separately. Finally, all participants got together, and each group presented its results and answered several questions. The summary of discussions of each group is shown in following pages. The group organization is as follows:

Group A: Multimedia Database System Issues (Coordinator: Seiichi Uchinami (Kyushu Institute of Technology), 20 members)

Group B: Object-oriented and Related Issues (Coordinator: Masatoshi Yoshikawa (Nara Institute of Science and Technology), 24 members)

Group C: High Functionality and Performance Issues (Coordinator: Ushio Inoue (NTT), 9 members)

(Ushio Inoue (NTT))

テーマ マルチメディアデータベースシステム

参加者 着席順, 敬称略

脇山 俊一郎 (仙台電波高専),	植村 俊亮 (奈良先端大),	北川 文夫 (広島大),
松田 正樹 (PFU),	藤城 一成 (お茶の水大),	津田 宏 (ICOT),
伊藤 左智子 (SRL),	清光 英成 (図書館情報大),	池田 秀人 (広島大),
山崎 達也 (筑波大),	金森 吉成 (群馬大),	辻村 敏 (松下電器),
白田 由香利 (リコー),	平賀 瑠美 (日本IBM),	有沢 誠 (横浜国大),
横田 一正 (ICOT),	清木 康 (筑波大),	阿江 忠 (広島大),
増永 良文 (図書館情報大),	打浪 清一 (九工大),	

オーガナイザ: 打浪 清一, 書記: 池田 秀人

報告書執筆担当者 打浪 清一 (九州工業大学 情報工学部)

参加者の構成

最初自己紹介と研究テーマの紹介を行った。マルチメディアデータベース, ユーザインタフェース, GUI, 並列分散DBMS, マルチエージェント, サイエントフィックビジュアライゼーション, 辞書・自然言語, 知的生産システム等に関して研究或いは議論したい人が20人集まった。マルチメディアデータベースには, オブジェクト指向DBMSも向いていると言われているが, オブジェクト指向のグループが別に有った為, オブジェクト指向を研究している研究者が少なかった。その為マルチメディアデータベースシステムにオブジェクト指向の適否に付いての議論は余りなされなかった。

1. 討議内容

今はマルチメディアの時代と言われているが, その割には, マルチメディアの大きな適用分野, 商品が見えない。本当にマルチメディアデータベースシステムに関して研究する事が有るのだろうかという事から議論が始まった。

マルチメディア, マルチメディアデータベースシステムと言っても, そのイメージは各人によって異なるのでは話がしにくいという事で, 先ずこれらの定義について議論を行なった。

2. 現状認識

2.1 マルチメディアの定義

ビットマップにしてハードディスクに記憶してしまえば, 全部統一的に記述でき, シングルのメディアで十分で, マルチメディアというものは無いのではないかとの意見が出た。しかし, 収集し, 再生するメディアを考えれば, 動画像, 静止画, 音声, 文字は夫々, 収録

再生の仕方が異なり、これらの記述媒体を考えるとマルチメディアという区別は有るということが了解された。しかしマルチメディアとは言っても、その中の一つ、例えば図面だけではシングルメディアでは無いか、マルチと言うからには、2つ以上のメディアを統合する必要が有る。静止画のデータベースシステムは、本当はシングルメディアなのだけれど、マルチメディアの重要な要素なので、マルチメディアのサブセットとして、やはりマルチメディアと言って良いのではないか、という事になった。但し、文字と画像は別々に扱われていて、生き別れの状態にあり、これを統合するのは今後の課題である。

2. 2 マルチメディアデータベースシステムの定義

マルチメディアデータベースシステムのデータモデル像がはっきりしていないが、これをはっきりさせるべきだとの意見が出た。

究極のマルチメディアデータベースシステムでは、対象物は知識・実体として管理されるべきであり、各メディアでの表現はビューの様なもので、ビュー更新は、本体の更新を引き起こさなければならず、例えば、音声で更新した後、動画像でそれを再生したら、更新された結果が動画像で見えなければならない。

このような考え方に対して、そのようなものは実現不可能で、まだ、各メディアで十分な処理が出来ないので、各メディアのデータベースシステムを詳しく研究して充実させなければならず、より低レベルで地道に研究を行なうべきだとの意見が出た。

- (1) 現状認識としては、マルチメディアと言っても、文字と画像の生き別れであって統合されていない。
- (2) マルチメディアでは範囲が広くて、一般論は考えにくい。より対象を限定して考えるべき。
- (3) マルチメディアデータベースシステムの構造は次のように考える。
 - (a) 何層かのレイヤー構造を考える。最上位レイヤーは、各オブジェクト（個体）に関する情報は全てメディア的に統合されたレベルである。下位レイヤーはデジタル化されビットマップで表現されたレイヤーである。夫々のメディアは統合されて無く、各メディア毎の体系を持つ。
 - (b) 各レイヤーでのデータモデルを確立する必要がある。
 - (c) 各レイヤー間の写像（リンク付け）の方法を確立する。
 - (d) ビュー（表現）の取扱い法を確立する。
 - (e) データ型変換（例：文字→ビットマップ→音声）が自由に出来るようにする。
 - (f) 複数層を統合的に管理することは重要だが、どちらに重点を置くかで立場が異なる。2つの研究の立場が有る。下位層を主に研究、上位層を主に研究（これは全体の研究となる）する立場である。
 - (g) シングルメディアのより深い研究が必要である。

3. 今後の研究課題

時間が足らなくなり十分な議論が出来なかったが、以下の研究が必要だとの合意を見た。

3. 1 データベースシステム分野でユニークな研究テーマ

最近のデータベースシステム分野の研究では何も新しいものは出ていなくて、オブジェクト指向は言語屋の分野の研究の真似だし、メディア変換は、パターン認識屋の分野の研究なのでは無いかとの意見も出た。オブジェクト指向の考え方は、言語、OS、DB、AI等の分野から自然発生多発的に起きたもので言語屋のものでは無い。また、パターン認識屋は各画像の認識で、そのファイル化、統合化、データベース化はやっていない。

マルチメディアのデータを記述処理する為にはデータ構造としてどれだけのものを準備し、その上での演算(処理)としてどれだけのものを準備すればよいのかという体系化を行なうのはデータベースシステム研究の分野である。

ここでは、マルチメディアDBMSのプリミティブを明確にしないと発散する。時間が入り、位置+時間で考えるのはマルチメディアデータベースシステムの特徴である。

サイエンティフィック・ビジュアルイゼーションに於いては、PixelではなくてBoxelが単位となる。Volume Dataの可視化がVRでは問題になっている。これをマルチメディアにどう適用するのかが問題となる。

3. 2 マルチメディアシステムマニフェストの必要性

上述のマルチメディアデータベースシステムに対する認識を更に深く、普遍的なものとするためにマルチメディアシステムマニフェストをまとめ提言すべきである。こういうグループ討論を繰り返し収束させて行ったら良いとの意見も出た。

3. 3 応用からのニーズを重視すべきである。

利用者の望む情報を、使いやすい形で、提供できるシステムを考え設計・試作すべきである。マルチメディアデータベースシステムの真の応用が有るのかどうか、確認すべき。例として、画像検索(類似剖検写真検索、辞典、知的生産システムに於ける画像検索等)の実世界での必要性が挙げられた。

3. 4 マンマシンインタフェースの改善の必要性

現在のデータベースシステムのマンマシンインタフェースは旧来の文字型の検索(SQL等)から脱していない。マウスやGUIも十分でない。マウスに代わるマルチメディア用の入力デバイスの開発が必要である。

例えば、画像をキーとして画像を検索する(医療、辞典)、音声で検索する等、頭の切り換えが必要である。

この様な観点から、VR、ペンコンピュータ、メタファ等が候補に挙がった。

謝辞

熱心にご討論抱いたグループのメンバー、後の全体討論でご意見を頂いた諸氏に謝意を表す。なお本報告をまとめるに当っては、池田 秀人氏、増永 良文氏に議事メモを始めとしてお世話になった。ここに記して深謝します。

Bグループ： 「オブジェクト指向とその関連研究」

参加者： 天野浩文 (九大), 市川哲彦 (お茶の水女子大), 大芦雅弘 (松下電器産業(株)), 大津浩二 (群馬大), 大蒔和仁 (電総研), 掛下哲郎 (佐賀大), 片山紀生 (東京大学大学院), 上林弥彦 (京大), 北川博之 (筑波大), 鈴木健司 (NTT 情報研), 高橋千恵 (JIPDEC), 高山 毅 (筑波大), 鶴岡邦敏 (NEC C&C 研), 遠山元道 (慶大), 永田元康 (大阪教育大), 西岡利博 (三菱総研), 福田紀彦 (群馬大), 古川哲也 (九大), 牧之内顕文 (九大), 村上 敦 (協同システム開発), 村田美恵 (高度通信システム研究所 AIC), 山北隆典 (学習情報通信システム研究所 SRL), 吉川正俊 (奈良先端大), 龍 忠光 (富士通研) 計 24 名

書記： 鶴岡邦敏 (NEC C&C 研)

コーディネータ： 吉川正俊 (奈良先端大)

報告書執筆担当： 吉川正俊 (奈良先端大)

1 OODB について

1.1 OODBMS の使用経験

現在 OODBMS を用いた応用開発を行なっている参加者に、OODBMS を使用する理由やその他の意見を求めたところ、以下のような回答を得た：

- 機械系 CAD/CAM
階層、継承 (RDB より OODB の方がすぐれている)、柔軟性。
- フルテキスト
データのカプセル化。論理構造は RDB でも良いかも知れない。OODB は reference や関連の扱いにまだ不満があり研究の余地があるかも知れない。
- マルチメディア (Objectivity 使用)
C++ のクラス階層をエンドユーザには見せないことにより OO を知らない人にも使える環境を構築。多種多様なデータ型、階層があり、たとえば同じバイナリデータをどう扱うかを応用側ではなく、データ側で管理してくれるため便利。
- 医療 (ONTOS 使用)
Coad/Yordon 法から ONTOS への変換は大変。
- 学習教材
マルチメディア、複合オブジェクトの扱い。
- ネットワーク環境 (ONTOS 使用)

1.2 データモデルと応用との関係について

- OODB はあくまで手段であり、どのような応用システムが作りたいかという応用指向アプローチが重要ではないか？

- requirement needs に基づかないデータモデルの研究はやめるべきである。
- DBS を解体して应用到 DB 機能を入れる / DB に応用の機能を入れることが重要である。
- 応用から来る要請とデータモデルから来る要請を分けるべきか？

1.3 現在の OODBMS の問題点および今後の課題

- OODB の形式化、理論化について
 - OO モデルの理論化 (例: typed λ calculus) が必要。
 - データの入れ物の理屈はあるが、入れ方の理屈が必要。
 - RDB を超えるためには代数系が必要である。
 - Kifer は、F-logic は OODB の基礎を与えたと言っているが、確かに関係モデルにおける関係論理に相当するような基礎を与えたのではないか？
- クラスについて
 - C++, OODB はクラスライブラリを提供すべきである。(Smalltalk はライブラリがついてくるためすぐれている。)
 - 上流工程のクラスと下流工程のクラスは異なる。上流工程のクラスは semantic object であり応用のドメインに依存する。OODB は下流工程のクラスを扱っている。
 - クラスを再利用、改造するためにはその内容を知る必要があり、また他人が使っていると改造できない。従ってクラスの再利用、改造は困難である。
 - ライブラリの共有は難しい。(副作用)
- SQL のような簡易言語がないためエンドユーザのビュー構築が困難。
- オブジェクト指向の design methodology はない。
- OODB の標準が必要である。
- 大規模な分散化 OODB のバックアップは問題。
- OODB の並列化はどの側面で有効か？

2 次世代データベースの位置付け

次世代 DB を RDB と同じ枠組で議論するのは無理ではないか？という問題提起がなされた。その背景には、DBMS と応用、プログラミング言語、OS などとの境界が RDB と次世代データベースでは異なるという認識がある。(関連して、OO は広がりがあることが良い (PL, OS, CASE)、という意見が出された。) 特にプログラミング言語との融合との関連で、OODB におけるデータ独立性については、以下のような多様な意見が出された。

- OODB プログラミングと Smalltalk プログラミングの本質的な違いは何か？(ないのでは?)
- OODB にはデータ独立性がない。そのためデータ工学がソフトウェア工学と近いものとなり、OODB の設計が難しくなる。
- RDB でもデータ独立性は限定的にしかない。
- RDB ではデータ独立性があることが理想である。
- OODB では抽象データ型を扱うためデータ独立性が高い。

データベースの高機能化・高性能化研究

井上 潮 (NTT), 目木信太郎 (岡山県立大), 高倉弘喜 (京大),
西澤 格 (東大), 木實新一 (九大), 原田リリアン (富士通研),
佐藤 聡 (筑波大), 白 光一 (九大), 掛下哲郎 (佐賀大)

1. 検討の背景

近年, データベース処理を行う計算機環境のダウンサイジング, マルチベンダ化が急速に進んでいる。また, 計算機間を接続するネットワークもATM(Asynchronous Transfer Mode)等の技術により, 従来のディスク入出力を上回る速度となりつつある。このような環境下で, データベースの高機能化・高性能化を実現するためには, 並列・分散・協調処理に関する研究が重要な役割を果たすことになる。以上の観点に基づき, 今後研究すべき課題について議論した。

2. 研究すべき重要課題

(1) 並行制御(concurrency control)

並行制御とは, 複数のトランザクションが1つのデータベースに並行してアクセスする際に, トランザクションのACID性(Atomicity, Consistency, Isolation, Durability)を保証するものである。従来の並行制御は, 直列化可能性(serializability)の理論に基づくものが大半であったが, 並列・分散・協調処理の観点からは以下の問題があることが指摘されている。

□ 性能向上のボトルネック

トランザクションを直列化することにより, 処理の並列性を阻害する。また, 直列化のための通信オーバーヘッドが特に分散形態のシステム性能の低下を引き起こす。

□ ユーザ間協調作業の阻害

トランザクションの直列化によってデータ資源が特定ユーザに占有されることにより, 複数ユーザによる1つのデータベースに対する協調作業の円滑な進行を困難にする。

従って今後の研究は, 従来の直列可能性に代わる新しい「正しさ」の概念を確立することと, トランザクション性能と一貫性のトレードオフを明確化することが重要である。具体的には, 以下の問題に対する解答が求められている。

□ トランザクションの概念は本当に必要か?

- 復帰(rollback)と補償(compensation)の役割は？
- 並行制御と協調処理の相互関係は？

(2) 回復制御(recovery control)

回復制御とは、システムまたはデバイスの障害に備えてトランザクションの処理結果を冗長化するとともに、障害発生時にはこの冗長情報をもとにデータベースを復旧するものである。従来の回復制御は、データベースの更新時にデータベースとは別のデバイスにログ情報を取るものが一般的であったが、以下の問題が指摘されている。

- 通常処理性能の低下、コスト増大
ログ情報を取るためのオーバーヘッドにより通常処理（本来行うべき処理）の性能が低下する。または、性能低下を防ぐためには余分なプロセッサやメモリが必要になる。
- 回復処理が長時間、可用性の低下
高性能システムの場合、単位時間当たりのログ情報が膨大な量になり、障害発生時にこれを用いてデータベースを復旧するのに長時間かかり、システムをなかなか再開できない。

従って今後の研究は、フラッシュメモリ、不揮発RAM等の最近の半導体テクノロジーを活用すると並行して、RAID等の技術によりディスク自体を高信頼化することが重要である。

3. その他話題になった事項

(1) 最近の半導体メモリへの格納を効率化するデータ構造（ファイル構成）

ディスクへのアクセスと同様に、半導体メモリへのアクセスもシリアルアクセスの方がランダムアクセスよりもはるかに早いというのが最近の半導体メモリの動向である。従って、半導体メモリ向きのデータ格納構造が重要になる。

(2) 異種分散データベース（マルチデータベース）のための統合スキーマ

特にマルチベンダ環境では、異なるデータモデル、スキーマを持つ複数の既存データベースを統合して使いたいという要求が強くなる。この要求に対応するためには、現実的な性能で統合スキーマを実現する手法の研究が重要になる。

(3) エージェント指向等の新しい並列・分散処理のパラダイム

従来、データベースと知識処理は演繹データベースで代表されるような上流域（データモデル）で融合が図られてきた。今後は、並列・分散処理のパラダイムといった下流域でも融合を図っていくことが重要になる。

（報告書執筆担当者：井上 潮（NTT））