

第2回データ工学国際会議報告

石川 博	横田治夫	遠山元道
富士通研究所	ICOT	慶応大学

筆者らは1986年2月5日～7日にかけて米国のロサンゼルスにおいて開かれたIEEE主催の第2回データ工学国際会議に出席したので、会議の印象と概要について述べる。本会議の主旨は人工知能、データベース工学、ソフト工学等の複数の学問領域を統合して現実世界の問題を解決することにある。会議には米国、欧州、日本等から約310名の参加者があり、約80件の論文発表があった。分野は分散データベース、ハッシング、問い合わせ最適化等の伝統的なものから、異種データベース、ファジーシステム、知識ベース、時間的データベース、マルチメディアデータベース等の新しいものまでカバーする。この報告では11件の論文についてその要約を述べる。

"A Report on the Second International Conference on Data Engineering"
(in Japanese)

by Hiroshi ISHIKAWA (FUJITSU LABORATORIES, LTD., 1015 Kamikodanaka Nakahara-ku, Kawasaki, 211, Japan), Haruo YOKOTA² (ICOT) and Motomichi TOYAMA (Keio University)

The authors took part in the IEEE Second International Conference on Data Engineering, which was held in Los Angeles, USA on Feb 5-7, 1986. We make a brief report on this. There were around 310 participants from America, Europe, and Asia. About 80 papers were presented. They include new fields such as heterogeneous databases, fuzzy systems, knowledge bases, temporal databases, and multimedia databases.

1 現住所：富士通研究所

2 Current address: FUJITSU LABORATORIES, LTD.

1. 会議の概要

IEEE主催の第2回データ工学国際会議はロサンゼルスにおいて1986年2月5日から7日まで3日間にわたって開催された。会議に先立つ2日間はチュートリアルにあてられた。ロサンゼルスは2月とはいえとても暖かく、日本の初夏といった天候であった。会議はロサンゼルスのダウンタウンにあるボナベンチャー・ホテルでおこなわれた。このホテルは日本のパッケージ・ツアー等の利用客も多い、高級だが日本人にはホピュラーなホテルである。ホテルは4つの円筒を束ねたような構造をしており、概観は未来都市といった雰囲気だが、各円筒の間で自由な行き来ができない点は参加者の不評を買っていた。

オープニング・セッションのあと同時に3つのセッションを並行しておこない、3日間で合計25個のセッションで80件近い論文と4つのパネル・セッションがおこなわれた。会議には米国を中心にして英、仏、独、日本等から約310名の参加があった。

データ工学国際会議はその第一回が1984年に開催されたばかりの若い国際会議である。データ工学とは何かという質問には主催者側の言葉を借りて答えることにする。即ち、データ工学とは「知識工学、人工知能、データベース工学、ソフト工学及びコンピュータ・アーキテクチャ等の幅広いディシプリンを統合して現実の問題を解決すること」(Ramamoorthy, Honorary Chairperson)であり、「情報システムの設計、開発、管理、利用におけるデータと知識の役割に関するものである。従ってその範囲はデータベースの伝統的な側面を越えて知識ベースや文書・画像処理、記号処理に及ぶ。データ工学の研究開発にはモデル、論理設計・物理設計、分散システム、ツール、ソフトウェア、ハードウェア、アーキテクチャ、セキュリティ、インテグリティ、システムモデリング、システム設計、知的システム等が含まれる。」(P. B. Berra, General Chairperson)

では実際の会議はというと分散データベース、アクセスメソッド、統計データベース、ハッシング、物理的更新、問い合わせ最適化、並列実行制御、性能、データモデル、高級インタフェース、理論といった伝統的な分野から関係モデルの拡張、異種データベース、ファジー・システム、知識ベース、時間的データベース、エンジニアリングCAD/CAM、マルチメディアデータベース、Ada、仕様等の新しい分野までカバーしている。参加者の顔触れをみるとやはり従来 of 会議でおなじみのデータベース研究者が多いが、伝統的な分野であってもなんらかのかたちで知識というものに言及している論文が目に着くことが、この会議の特徴の1つであろう。

以下に基調演説と筆者らが独自に選んだ11件の論文についてその梗概を述べる。

2. 基調演説と論文の要約

Keynote Speech (Opening Plenary Session)

E. F. Codd

(The Relational Institute)

まず我々は、あの関係データベースの発案者であるCoddがDateと共にIBMをやめて新しい会社を作ったと聞いて驚いた。新しい会社は、大学で講演すること等を主とするコンサルタント会社のようなものである。彼は、キーノートスピーチで現在までデータベースの分野において行われてきた研究の概要とそれに対する彼の意見を述べた。その中でデータベース分野におけるより学術的な研究が必要であることを力説していた。特に、IBMの代表的関係データベースシステムであるDB2を例に挙げ、実際の利用形態のみに着目しているDB2のようなシステムは本当の関係データベースシステムではないと批判していた。さらに、ANSI-SQL標準についても、拡張性を無視した標準化は好ましくないと否定的であった。最後に、データベースシステム的设计者は、研究者の要求に対してもっと柔軟になるべきであると呼びかけていた。

Adaptive Techniques for Distributed Query Processing

C. Yu, L. Lilien, K. Guh, M. Templeton, D. Brill, A. Chen

(イリノイ大学シカゴ校) (System Development Corporation)

この論文は分散データベースにおける問い合わせの最適化に対する新しい方式を提案している。この技術は2グループに分類される：(1)問い合わせの実行効率を直接向上させる技術と(2)問い合わせの実行戦略におけるコストの見積りを改善する技術。(1)にはセミジョインをするフィールド間に集合的包含関係が成り立つ場合にそのセミジョインを省略することができるという技術や、またジョインするフィールド間に集合的包含関係が成り立つ場合にそれを省略できるか、問い合わせをより単純な問い合わせに分解してデータのアクセス量を減らすという技術が含まれる。(2)にはジョインやセミジョインの結果のテーブルの大きさ、データ転送率、及び局所的な処理コストに関する見積りを実際値と比較して改善していく技術が含まれる。さらにエキスパート・ユーザから最適化の方略を学習するメカニズムを提供している。即ちユーザに対してシステムが選んだ最適化の方略を示し、ユーザがもしよりよいと思う方略があればそれを入力してもらう。システムは両者を比較し、その結果を記憶していく。そのためにグラフィクス・インタフェースを提供する。

A Knowledge-Based Approach to Design a Portable Natural Language Interface to Database Systems

H. Ishikawa, Y. Izumida, T. Yoshino, T. Hoshiai, A. Makinouchi

(富士通研究所)

この論文はK I D (Knowledge-based Interface to Database systems)というポータブルな自然言語インタフェースの設計に関する知識にもとづくアプローチについて述べてい

る。K I Dはデータベースのユーザが問い合わせ言語やデータベースの定義について知らなくても、自然言語（特に日本語）でデータベースにアクセスできるようにした実験システムである。そのためにK I Dは世界モデルという知識ベースを持っている。世界モデルは対象分野のモデル、言語のモデル及びデータベース写像の3つの知識を統合したものである。K I Dを新しい分野に移行するには、ユーザがその分野に対する世界モデルを作りさえすればよい。世界モデルはいわゆるオブジェクト指向のモデルであり、理解しやすく、作りやすい。K I Dはユーザの自然言語の問い合わせを意味解釈する際に言語のモデルと対象分野のモデルを用い、データベースの問い合わせコマンドを作成する際にデータベース写像を用いる。K I Dの世界モデルの参照の仕方は一般的であり、理解しやすく、知識ベースのデバッグがしやすい。例えば、意味解釈では特殊化原理という規則にもとづいて世界モデルを参照する。また意味構造からデータベースの問い合わせコマンドへの変換にはルールベースを用いているため、これを書き直すことによって新しいDBMSへも移行できる。対象分野に依存する知識はすべて世界モデルに閉じ込めることによってK I Dは分野の移行性を得ている。

Inclusion of New Types in Relational Database Systems

M. Stonebraker

（カリフォルニア大学バークレー校）

ビジネス領域以外の新しい分野では新しいデータ型を要求している。例えば、地理的応用では点や線分、多角形などを、また科学的応用では複素数や時系列データを扱う必要がある。さらにそれらのデータ型に伴うアクセスメソッドも定義できなくてはならない。つまりデータベースシステムは(1)ユーザによる新しいデータ型とそれに対する操作の定義を許し、(2)新しいデータ型に対するアクセスメソッドの定義を許し、さらに(3)新しいデータ型を含む問い合わせの最適化も可能にしておく必要がある。この論文はこれらの要件に対するアプローチについて述べている。新しいデータ型の定義とその操作の記述、及びアクセスメソッドの記述はシステムで用意したテーブルに格納される。アクセスメソッドの導入に伴い、それをトランザクションにどう組み入れるかという問題がでる。そのうちリカバリに対してはイベントからなる論理ログをとるようにする。さらにUNDOやREDOといった操作もシステムが提供する。並列実行制御に関してもシステム定義のルーチン群を提供する。問い合わせの最適化に関してはアクセスメソッドを記述したシステムテーブルの中の情報（例えば、見積りタプル数、見積りページ数等）を用いてルールベースで行う。

Entity-Relationship Modeling of Fuzzy Databases

A. Zvieli, P. Chen

（ルイジアナ州立大学）

この論文はファジー (fuzzy) 情報に対する実体関係モデル (Entity-Relationship model) に基づくフォーマルなアプローチについて述べている。まずモデルがファジーであることを3つのレベルに分類している。第一のレベルは概念モデルがファジーであること、即ち実体セット、関係セット及び属性セットがファジーである場合を指す。第二のレベルは実体と関係の特定のオカランスがファジーである場合を指す。第三のレベルは特定のオカランスの属性と関係がファジーである場合を指す。モデル上ではファジーな概念に対してファジーであることを示す記号とメンバーシップ関数の値を付加して表現する。第一レベルのファジーはデータベースの設計や変更の段階にのみ現れる。第二レベルのファジーはあるセットにあるオカランスが含まれるかどうかを指定するのに使用される。第三レベルのファジーはあるオカランスに属性値があるかどうかを指定するときに使用される。さらにファジー実体関係代数の基本演算について述べている。例えば、2つのファジー・セットに関して集合和、集合積、集合差等の演算を提供している。これらの演算は同時にメンバーシップ関数の値も計算する。また実体の属性に関するメンバーシップ関数値を条件にして実体を選択する演算がある。

The Bounded Disorder Access Method

W. Litwin, D. B. Lomet

(INRIA) (IBM)

現在、データベースのアクセスメソッドとしては、ハッシングとB氏木等の木構造索引が主流となっている。しかし、両方式ともそれぞれに問題点を抱えている。ハッシングは索引ファイルの保守をしなくてよい代わりに、順アクセスやレンジ検索には不向きである。一方、木構造索引の性能には索引ファイルの大きさとデータファイルの大きさの比率が重要となる。近頃の大型システムでは、主メモリの容量と二次記憶容量の比は1:1000程度であり、これはB氏木で1回の索引操作でディスクページを検索するのに適すると思われる比率の40倍にもなっている。この比率を改善するために、マルチバケット方式等が提案されているが、単にノードを複数ページにしたのではオーバフロー等により全体としてキーの順番が崩れてしまう。そこで、この論文ではハッシングと木構造索引を組み合わせた限定不順序 (Bounded Disorder) アクセスメソッドという新しい方法を提案している。この方式は、索引を用いてノードアドレスを捜した後でハッシングを行うことにより、大局的にキーの順序を守り、レンジ検索にも有効な効率の良いアクセスメソッドを提供するものである。

Deductive Database System based on Unit Resolution

H. Yokota, K. Sakai, H. Itoh

(ICOT)

この論文は演繹データベースシステムでの汎用的な再帰呼出しの取り扱いについて述べている。知識ベースシステム構築の一アプローチとして、またデータベースへの高級問い合わせ言語の実現方法として、関係データベースと一階述語論理を結合して演繹データベースシステムを構成する方法が注目されている。しかし、一階述語論理で記述された各種再帰呼出しの停止条件を判定する方法について、統一的な扱いで停止性を証明したような研究発表は今までなかった。この論文では、単位演繹(Unit Resolution)をもとにしてセッティング評価法(setting evaluation)という方法を導入し、最小浮動点処理を拡張することにより、ホーン節で記述されたどのような複雑な再帰呼出しであっても、停止し全解を求めることができることを証明した。

セッティング評価法では、与えられた確定節とゴールからセッティングと呼ばれる演繹中に現われる負リテラルの集合を求め、その各要素間の関係を関係代数を用いて記述する。この関係代数演算の繰返し実行によって各要素の補インスタンスとなる正リテラルを求め、セッティングの全要素に対して補インスタンスが一つも見つからなくなった時点で停止する。ゴールもセッティングの一要素に含めることにより、ゴールに対する全解を求めることができる。データベース管理システム側で繰返し処理をサポートすることにより、効率のよい演繹データベースシステムが実現できる。

Some Performance Results on Recursive Query Processing in Relational Database Systems

J. Han, H. Lu

(ウイスクンシン大学)

一階述語論理の再帰呼出しを関係代数演算の繰返しに変換して実行する場合のいくつかのアルゴリズムについて、解析的および実験的評価を行っている。この論文では、典型的な再帰呼出しを設定して、その展開アルゴリズムを、制限演算と結合演算の組み合わせ方と繰返し処理における中間結果の利用方法から、4種類に分類している。そして、各アルゴリズムについてI/OコストおよびCPUコストを見積って解析するとともに、ウイスクンシン大学でVAX 11/750上に作成した関係データベースシステム上で実際に実験して評価している。

評価の結果、毎回全演算をしないおすアルゴリズムより繰返しの前段階の結果を利用するアルゴリズムのほうが優れているのは当然であるが、毎回2回ずつ中間結果との結合演算を行った後で制限演算を行うアルゴリズムより、中間結果に制限演算を行った後3回ずつ結合演算を行うアルゴリズムの方がコストが低いことが分った。これは、単純に結合演算の回数だけでアルゴリズムを比較したのでは不十分で、中間結果としてできるリレーションの大きさも考慮する必要があるということを示している。

A Word-Parallel, Bit-Serial Signature Processor for Superimposed Coding

D. L. Lee

(オハイオ州立大学)

テキストの検索には、シグナチャファイル(Signature File)と呼ばれるテキストの特徴部分を抽出したファイルを作成しておいて、テキストの全走査を行わずに高速に検索処理を行う方法がある。この論文は、重ね合わせコーディング(Superimposed Coding)をもとにするシグナチャファイルを用いたテキスト検索用の新しいハードウェア構成法を提案している。重ね合わせコーディングでは、データベースは幾つかの固定長ブロックシグナチャに分解される。各ブロックシグナチャは、もとのテキストに対応するビット列の重ね合わせによって得られる。検索は、問い合わせに対応するシグナチャのビット列とシグナチャファイルのビット列の間のパターンマッチによって行われる。従来、シグナチャファイルを幾つかに分解して、ファイル単位の並列処理を行うが、1つのファイルに対するビットの検査は順次行うWSBP(Word-serial Bit-parallel)法は提案されていた。しかし、二次記憶からのローディングを考えた場合には、あまりファイル単位の並列性は期待できない。そこで、ビット単位の比較器をシグナチャの長さ分持つことにより、I/Oの転送に合わせた処理が可能なWPBS(Word-parallel Bit-serial)法を提案している。バブルメモリを用いる実現法についても言及している。

SNAP: A Graphics-based Schema Manager

D. Brice, R. Hull

(南カリフォルニア大学)

IPOモデルは、発表者らによって開発中の意味論的データモデルの1つである。古くはBRモデルからはじまり、データの意味を現実世界の実体の集合と、抽象化を含む実体間の様々な関連として捉え、これらの間の位相をグラフで表し、視覚にうったえるよう図形で表現することが行われてきた。本論文ではビットマップディスプレイ上にIPOスキーマを表示し、これに対話的に操作を加えて、スキーマ設計・スキーマブラウジング・問い合わせ・結果の表示の4つの機能を果たすスキーママネージャSNAPシステムの概要が報告されている。最初の2つの機能については、Symbolics3600上でオブジェクト指向グラフィックスパッケージBOOGIEを用いて実現され、すでに稼働中である。このようなシステムの実現上で、(1)データモデルで扱われるあらゆるタイプの関連を、抽象化も含めすべて図形的に表現すること、(2)スキーマの一部をモジュールとして扱えるよう幾何学的なクラスタリングを行うこと、(3)抽象化のいくつかの段階においてスキーマの表示が可能なこと、(4)表示形式の柔軟な変更が可能であり、かつ、(5)画面の急激な変化によってユーザがとまどわないよう変化に連続性をもたせること、の重要性が指摘された。

Parameterized View Definition and Recursions

M. Toyama

(慶応大学)

関係データベース言語SQLのビュー定義機能を拡張し、パラメータを持たせることにより、(1)複数の同種のビューインスタンスを1つのビュー定義から生成できる、(2)ビュー定義のなかで定義されるビュー自体を再帰的に参照することにより、2項関係の推移閉包等の定義ができることを示した。(2)については、関係DBのフロントエンドとしてProlog等を利用する研究が盛んである。再帰的ビュー定義によれば、データベースのジョイン等の集合演算を直接利用できること、基となる2項関係が非巡回的な場合に定義を工夫することによって推移閉包の効率のよい導出が可能であることなどの利点を主張している。

Towards an Implementation of Database Management Systems with Temporal Support

I. Ahn

(ノース・カロライナ大学)

時制を扱うデータベースシステムの実現手法を考案している。まず、「時制」という用語の使い方を整理し、STATIC, ROLLBACK, HISTORICAL, TEMPORALの4つの類型を定義する。次に、履歴データをタプル単位で残すか属性値単位で残すかにより、タプルバージョンングと属性バージョンングの2つの論理的構造を示す。前者では、値の更新が起きる度に新しいタプルを時刻とともに追加し、後者は変更された値を時刻とともにタプル内のフィールドに追加する。両者の表現する情報は等価で、いわゆる非第一正規形と第一正規形の相互変換と類似の操作で変換できる。

物理的構造は、自由に読み書きできる一次記憶と、より安価で読み書きにコストのかかる履歴記憶の2階層におくことを前提としている。基本的な物理構造として、(1)逆チェイニング、(2)アクセスンリスト、(3)クラスタリング(固定ブロック)、(4)履歴スタック、(5)セルラチェイニングを挙げ、どれがいずれの論理的構造の実現に向いているか、また履歴記憶に追記型光ディスクの利用が可能か否かを検討している。選択の助けとなるような詳細な性能評価などは行っていない。

参考文献

Proc. International Conference on Data Engineering, Feb. 1986.