

82年度ACM SIGMOD国際会議報告

上林 弥彦
(京都大学工学部)

1. まえがき

82年6月に、次のような日程で北米をまわった。

- 6月1日(火) IBM サンホセ研究所
- 6月2日(水)~4日(金) ACM SIGMOD 国際会議出席
- 6月7日(月) ハーバード大. CCA
- 6月8日(火) マギル大
- 6月9日(水) ウォータールー大

主目的は、SIGMODでの発表で、その内容を詳しくしたもの、IBMサンホセとウォータールー大学でセミナー講演をした。SIGMODの日本からの出席者は、私と共著者の吉川正俊氏および日立の福岡慎一氏のみであったので、会議の内容を中心に、データベース研究の話題についてまとめてみた。

2. データベース関係の国際会議、特にVLDB問題について

データベース関係の国際会議として、特に有名なのはACM SIGMODとVLDBで、SIGMODの論文採択率は1/4弱(81年 25/93, 82年 33/123)、VLDBの論文採択率は1/5程度と、同種の他の会議に比べるとかなり競争率が高くなっている。これは、データベース関係の研究者(特に博士課程の学生)が75年以来大幅に増加したにもかかわらず会議数が増えていないことや、会議や雑誌のランク付けがはつきりしてきたということが大きな理由となっている。

今年は、あらたに理論関係の国際会議(ACM Principles of Database Systems, PODS)も独立し、3本立となった。この独立の1つの理由は、上記の会議がバランスをとるために、投稿数の特に増加した理論の論文の採択数を限っていたためである。前身の「プリンストン出身者会議」のメンバーが中心となって独立を決めたが、SIGMODでは独立しても論文は集まらないので、同じ週に2つの会議をやるという妥協案も出したという。PODSは本年度は37/104の採択率で、来年3月にオ2回を行なうことになっている。この影響で、今年のSIGMODでは理論関係の出席者が以前より減少していた。

VLDB問題については、文献(2)にD. HsiaoとACMの両者の言い分が掲載されており、SIGMODでも委員長のM. Stonebrakerより特に現状説明があった。82年のVLDBをVLDB Endowmentが主催することになり、ACMとIEEEに共催を求めたことから問題が起きた。ACMの言い分は、VLDB Endowmentは、Hsiaoの記念団体とも言える私的な団体で、そういう団体にすでに有名になったVLDBを私物化させることはできないという点にある。学会は功績に応じて賞を出しており、これで個人の業績をたたえるべきで、いくら業績があっても国際会議の私物化という形を許してはならない、もし許すとACMのかわりに多くの私有化された会議ができてしまうというのがACMの言い分である。Hsiaoの言い分は、オ2回はACMと無関係であったし、今までACMに主催をたのんだのはTODSに論文を集めるという目的もあったためで、現状では何の利益も受けていない、論文集の配布もVLDB Endowmentがやりたい

と言うもので、VLDBの利益をACMが吸い上げている(ACMは他の利益のない会議も多く援助しているため)ことに対する不満も(公式には言われていないが)あると考えられる。ACMが共催をしないことになってから、VLDBがACMのメーリングリストを使用したことがわかり、両者の関係はかなりこじれ、一時は別の本拠でVLDBをACMが主催する形でCall for Papersまで用意された。データベース研究者の中にせっかく良い会議であるVLDBをつぶしたくないという意見が多いこともあり、結局、これは中止となった。ACMの代表は、HarrisonとDenningで、この2人は任期切れになったことや、HsiaoもVLDB Endowmentの代表をしりぞいたことから、新しい代表で今後この問題が話し合われることになっている。なお、本年度のVLDBの会議録は今までのようにACMやIEEEを通じて入手することはできない。

3. SIGMOD 国際会議について

データベース研究の応用分野がひろがりをみせている中で、ACM SIGMODの会員数はここ2年ほどほとんど増加しておらず、この分野の研究も一つの転期をむかえつつあると考えられている。PODSの独立でSIGMODが従来の、実際家から理論家までのmeeting placeとしての意義を失ないつつあり、VLDB問題もからんでSIGMODの今後の方角をどのようにするかというのが今年のSIGMODの一つの大きな関心事であった。

SIGMODの現在の方針としては、応用面を含めて、製品展示も行なう形で行なうというもので、今年は、少數の製品展示が行われ、応用面の論文も多く採録されている。来年度は、製品展示を行なうとともに、SIGDAやSIGBDBと共催して、応用面の論文も集める予定になっている。

本年度は、INGRESをはじめとして数社からの展示があった。すこし変っていたのは、マギル大学の大学院生が自分で開発したDateの教科書の言語を実行する教育用のシステムの販売を行っていたことで、あとでマギル大に寄った時に聞くと1件売れたということであった。我々のところの大学院生であったレ・グエト・チュン氏も、HDR社のデータベースシステム開発の中心となっており彼のシステムも電話回線を使ってデモンストレーションをした。データベースマシンIDM500とヌー8000を組合せたシステムで、QUEL, SQL および各種の応用プログラムをサポートしており、ペンタゴンをはじめかなりの引合いがあるということであった。SIGMOD後、吉川正俊氏はHDR社に2ヶ月滞在した。

会議では、統計データベース、画像・グラフィックス・プレイ・文字データベースの統合、1990年のDBMS、個人用情報管理の道具の四つのパネルと33件の論文発表があった。次頁以後の論文番号で分野をまとめると次のようになる。

- | | | | |
|----------------|--|------------|--------|
| ・データベース理論 | 1, 2, 5, 7 | ・データ圧縮 | 33 |
| ・利用者言語とインタフェース | 4, 6, 10, 11, 12, 13 | | |
| ・質問処理 | 16, 17, 18, 28, 29, 30 | | |
| ・ファイル | 14, 15, 29 | ・データベースマシン | 26, 27 |
| ・復旧 | 19, 20, 21 | ・トランザクション | 31 |
| ・応用 | {図形(3, 23, 25), 各種拡張(24), 統計DB(8), 系図(10)
計算用DB(9, 22), ドキュメント(32)} | | |

• SIGMOD 論文の要約

1 普遍関係スキーマデータベースにおける結合の表現 (D.Maier, D.S.Warren)

関係データベースの質問処理における一つの問題は、関係名を指定しないような質問表現で、利用者の考えている解が得られるかどうかという点にある。普遍関係を求めて、質問の対象としている属性集合に射影する方法は、コストがかかる上、必要なデータを落としてしまう可能性もある。本論文では、更新の単位としての属性集合を「連係」として定義すると共に、検索の単位としての属性集合を「対象」として定義している。質問の指定した属性集合を含む対象集合のうち極小なものに対応する答の和集合が質問の答であると定義されている。すべての連係が対象となっており、対象が空でない共通集合演算で閉じている場合に、質問の答は自然なものとなる。本論文では、質問は冗論理表現（比較と論理積）によるもののみを扱っている。さらに、汎化、集約、更新の容易さ、関数従属性との関係についても述べられている。

2 関係データベース理論に対する組織的考察 (T.Imielinski, W.Lipski, Jr.)

関係データベース設計論の主流である、はつきり判らない普遍関係より、はつきりした関係集合を得るという方法は通常的设计と逆方向であることを指摘し、基本関係より関係表現によりデータベーススキーマを得るという考え方を提案している。基本関係は他の関係によって表現できないもので、概念スキーマは基本関係の集合として定義される。基本関係では関数従属性は存在しうるが、結合従属性は存在しない。データベーススキーマでは、これらのほかに普遍インスタンス制約といったものも制約となりうる。データベース状態の与えられた場合、それがどのような意味（概念スキーマの状態）に対応するかという解釈として、閉世界仮説によるもの、開世界仮説によるもの、その中間のものという3つの解釈を与え、開世界仮説の場合のデータベース状態の同値性について論じている。

3 作物状況査定用データベース (W.G.Driggers)

アメリカ農務省の開発した、リモートセンシングデータを用いた、農業、気象データベースシステムについて述べている。システムは、4台のミニコンより構成されており、内3台 (PPP 11/70 × 3) はヒューストンに、1台 (11/34) はワシントン (D.C.) に置かれており、DECNETで結合されている。それぞれ役割を分担しており、情景処理 (LANDSATやMETSATの情報より必要なデジタル画像を抽出)、分析用端末 (画像表示、演算等の機能を持ち、デジタル画像の分析を補助)、利用者インタフェースシステム (この中にデータベースがあり、文字や図形が表示される)、端末インタフェースシステム (ワシントンより気象情報を送る) に分かれている。データベースはDBMS-IIである。

4 フォーム応用開発システム (L.A.Rowe, K.A.Shoems)

カルフォルニア大学バークレー分校で開発中のFADS (Form Application Development System) は、データベースの応用を、フォームに基づいて会話的に設計できる機能を持っている。システムはVAX-IIを用いて、現在15000行のCによるプログラムで構成されており、データ管理にはINGRES (言語EQUELによる) を用いている。各種応用プログラムをフォームを中心に定義できるようにしたのが特色である。1つの定形的な応用は、いくつかの手続きより構成され、各手続きに1つのフレームが対応する。各フレームは、データを示すフ

フォームと操作を示す操作リストより構成されている。フォームは、レイアウトの指定がなされているもので、関係を示すものと1つの組を示すものとがあり、内容はデータベース内のデータによって変化する。これらのフレーム、フォーム、関係等もデータベースで管理されているため、新しいアプリケーションの開発時にも利用できる。デバッグも会話的に行なえ、中断点の指定も可能である。

5 限定作用子付き関係論理式表明の効率の良い検査法 (P.A. Bernstein, B.T. Blaustein)

データベースDに更新操作をほどこした場合Duに変形されるとすると、Duが与えられた表明Aを満足しない場合更新操作は無効となる。Du上でのAの成立を調べるかわりに、D上でのA'の成立を調べることで済ませることができると、実際に更新をしないので処理の効率化がはかれる。AよりA'を求める方法についてはすでに研究されているので、本論文ではA'の成立の判定が効率良くできるようなクラスとして *complacent* という概念を提唱し、その判定法を示している。A'がこのクラスにない場合はこの判定法はA'の成立の十分条件となる。[なお、本論文の基本定理に誤りのあることが勝野裕文氏によって指摘されている。]

6 論理的データ独立を目的とした関係名なしの質問言語 (D. Maier, D. Rozenshtein, S. Salveter, J. Stein, D.S. Warren)

①で提案された、連係・対象データモデル(AODM)のもとでの質問言語とその処理について議論したもので、関係名を用いない質問言語(QUEL型)を提案している。問題となるのは組変数の対象となる属性集合を決定することで、自明なものは属性名をそのまま用いている。ANDで結ばれている場合は、同じ組変数でも対応する属性集合の異なることもある。この言語は、関係名が不要だけでなく、組変数も極力使わないようにできるのが特色である。

7 データベーススキーマ間の状態変換の手法 (T. Imielinski, W. Lipski, Jr.)

②で提案されたモデルのもとでの問題を論じたものである。ある概念スキーマ上で定義された2つのデータベーススキーマのある場合、1つのデータベーススキーマの状態に対応する他のデータベーススキーマの状態を求める、テーブルを用いた方法を示している。テーブルは、Ahoらのタブローと似た概念で、著者らによって不完全情報の表現手段として先に提案されたものである。さらに、開世界仮説のもとでのデータベーススキーマの同値性についても議論している。

8 統計解析のためのデータ管理研究の基礎 (計算機科学者のための統計データ管理問題の基礎) (D. Bates, H. Borat, D. De Witt)

現在の統計計算パッケージやDBMSは、どちらもデータの統計解析に完全に適合しているとは言えないため、本論文では問題点をまとめ、統計解析のためのDBMSについて考察している。このようなデータの性質としては、大量で更新が少ない、使用期間が長いといったことが考えられ、利用形態の性質としては、定形的、一部のデータのみを対象とすることがふつうである等が考えられる。提案の主要点は、計算結果を何度も利用できるように、それらを管理し、データの更新による補正も行なえるようにした点である。現在、Wisconsin Storage System(ファイルシステムパッケージ)による実現を考えているが、将来はデータベースマシンも検討する予定である。

9 科学的シミュレーションプログラムのデータモデル (J. Bell)

シミュレーションプログラムには、大量のデータを使うものがあり、これらのプログラムの途中で使われるデータの管理や科学者の利用法に適したデータモデ

ルを構成することが重要である。統計解析と似ているのは、大量データを扱い、多くの属性が1つの組を一意に決める性質を持たない（人間の作ったデータではないため）点であるが、データ品質に関するデータのない点やデータの更新が非常に多いという点では全く異なっている。また、電子の位置は計算時点ごとに变化するが、同じ電子であるという情報は不要であるように、集約的な性質が重要な点、組単位の管理をする関係モデルと異なった点もある。本論文では、シミュレーションプログラムSLASHを例にとり、实体、属性、定義域の定義、冗長データ、集約データ、時間の扱い、アクセスパスの問題について検討している。

10 経路表現のできる質問言語の開発 (S.M. Dintelman, A.T. Maness)

本論文では、系図情報システムGENISYSのための質問言語について述べている。GENISYSは、遺伝に関係した病気等の研究に用いられるもので、各種の検査データと統合されている。質問は多くの関係にわたることが多いので、結合のかわりにあらかじめ設定されたリンクによって効率をあげている。言語GQLにおける、リンクの設定法、複数関係にまたがる質問の表現、1:nリンクの場合に対応する組を制限するための制約の表現、システムの実現法等について検討している。

11 CQLF-CODASYL型データベースに対する質問言語 (F. Manola, A. Pirotte)

CQLF (CODASYL Query Language, Flat) は、CODASYL DDLの81年ANSI dbANS版に対応するもので、ネットワーク型のほかに関係型の処理を可能としたものである。CODASYLモデルのすべての特色を生かし、また、関係型処理を行なうための設計上の問題点について検討している。関係型の処理として、結合、集合演算、質問結果の質問への利用、スキーマ変更機能等が追加されている。質問は、SELECT, USING, WHEREの形で書かれ、WHERE文の中に、E HAS MEMBER EP IN E-EPといった表現ができる。

12 ネットワーク質問言語NOAH (G. Schlageter, M. Rieskamp, U. Prädell, R. Unland)

NOAHは、n:m関連も扱える非手続き型のネットワーク言語で、16ビットミニコン上のCODASYLデータベースシステム上で開発されている。質問は、経路表現、条件表現、出力表現の3つの部分よりなる。経路表現は、PATH R₀, S₁R₁, S₂R₂, ... という形で表現される。ここで、R_iはレコード型、S_iはセット型で2つのレコード型を1:nで結合している。条件は、セット型に対するSETCONDとレコード型に対するWHEREの2つで表現され、出力表現はSELECTである。質問は、PATH, SETCOND, WHERE, SELECTの順で定義される。

13 ERQ: DBMS質問言語に対する導出制御と教育機能 (J. Longstaff)

ERQ (ER Query Interface) は、ERモデルを利用者インタフェースとしたもので、初心者にはERモデル、関係モデル、関係言語を教育する機能を有している。また、導出制御機能として、システムが質問の一部を自動的に生成したり、質問の意味を自然言語で示すことができる。[質問の意味を自然言語で示すという提案は、HDR社のC. Le Vietも独立に行なっている。]

14 近似順序従属性のインデックス容量圧縮への応用 (J. Dong, R. Hull)

順序従属性ODは、関数従属性 $X \rightarrow Y$ に、Yの値がXの値の増大に伴って、単調非増大または単調非減少するという条件を追加したものである。本論文では、

近似的にODの成立する場合として、2種の近似的ODを定義し、それぞれの場合に適した、XとYの両方に対するインデックス構成法を示している。一部の例外を除いてODの成立する場合は、例外を密インデックスで、その他を疎インデックスで扱う方法が示されている。Yの値が $\bar{x}(x)$ の値を中心とした範囲にある場合には、Xの値で分割されたブロックが複数個1つのYの値に対応することになる。これらの方法の解析や一般化についても検討している。

15 データベース設計における水平方向への分解 (S.Ceri, M.Negri, G.Pelagatti)

本論文では、データの水平方向への分解について、アクセス回数をコストとした考察を行っている。単純述語 π_i は、 π_i を満足するかしないかでファイルの内容を2つに分けるもので、それぞれのサブファイルについて、その中の任意のデータに対するアクセス回数が同じであれば完全であるという。最小項述語は、 π_i および $\bar{\pi}_i$ の論理積で示される。すべての π_i について、 π_i と $\bar{\pi}_i$ がどれかの最小項述語に現れるような π_i の集合は最小である。ファイルが最小項述語に対応して分割されている場合、任意の述語を満足するデータを求めるには、述語を最小項述語の論理和で表すと良い。モデルとして、ファイルFと要求集合Rおよびその間の対応を与えたものを考えており、いくつかの仮定のもとで検討を行っている。また、分散データベースでのファイル分散、主記憶と2次記憶間でのファイルの水平分解、与えられた負荷分散を実現するためのファイルの各装置への分散という3つの応用分野についても考察を加えている。

16 CODASYLデータベースシステムに対する質問の最適化 (U.Dayal, N.Goodman)

不均一データベースシステムMULTIBASEではDAPLEXという言語が用意されており、この言語による質問は各サイトで処理できる部分質問に分けて処理される。各サイトでは、アクセスパスの最適化とそのサイトの言語への翻訳が行われる。本論文では、アクセスパスの最適化について考察しており、木質問や存在記号のない質問とある質問に対する処理、アクセスコストの計算法、最適な方式の選定法等についてまとめている。

17 一般化導結合を用いた分散データベース質問処理 (上林, 吉川, 矢島)

分散データベースの質問処理において、導結合を用いると処理コスト(通信コスト)を低減できることが知られている。質問が木質問の場合には導結合による効率の良い処理方法が知られているが、一般の巡回質問の場合を念めた方法は従来知られていなかった。本論文では、後者の場合を扱うため、あらたに一般化導結合という概念を導入し、それを用いた質問処理方式について検討している。質問グラフの全域木に対応して質問が処理されるので、この方式は属性付加による木質問への変換とみなすこともできる。また、この処理に適したデータ圧縮手法も提案している。

18 Abe 統計質問機構におけるアクセスパス (A. Klug)

関係データベースシステムにおいて、COUNT, SUM, AVERAGE等の統計(集約)演算の利用が増えつつあるが、現在のシステムではその機能が不十分であったり使うのがむづかしいという欠点がある。Abeは、定義域関係論理に基づいているが、論理積型質問に制限されている反面、判りやすく、統計演算の表現しやすい言語となっている。統計演算を結合演算と同時に実現しているが特色である。System Rの研究で、結合方式としてソートマージ法とネストスキヤン法の2つが優れていることが知られており、統計演算の実現も同等にできるため、

の2つの方法を採用し条件によって選定する。従来、System R ではこの2つの方式の選択は質問の形式で決められ、INGRES では常に前者が使われていた。

19) 分散データベースにおける全順序的な検査点 (H. Kuss)

分散データベース内のあるサイトで、記憶装置の事故が起って回復データが失われた場合、分散データベース全体を以前の無矛盾な状態にもどす必要が生じる(コールドリスタート)。質問処理の途中で故障がおこったときによく使われる2相コミットを用いた場合に適したコールドリスタートの処理方法を提案している。この方法は、通信オーバーヘッドが少なく、2相コミットの実行中にコールドリスタートの準備ができてしまうという利点がある。

20) 分散コミットプロトコールの分析 (E. C. Cooper)

分散データベースで質問処理の途中で故障がおこった場合、その処理をなかったことにして無効とするか、または最後まで実行する場合は問題がないが、そのどちらでもない状態(ブロックと呼ぶ)に入ると困ることになる。ブロックを避けるようなプロトコールが存在しないことが示されているので、本論文では、いくつかのプロトコールに対してブロックを起こす頻度について解析している。対象としたものは、2相コミット、その拡張した共同回復スキーム、3相コミット、ネストされた2相コミット、SDD-1 で用いられたバックアップ付基本コミットである。共同回復のものは2相コミットやバックアップ付コミットに比べて悪くならないことが示され、ある条件のもとでは3相コミットが最も良いこと等が示されている。

21) 並行処理と復旧スキームの統合した実現 (A. Chan, S. Fox, W. K. Lin, A. Novi, D. R. Ries)

CCA では高いパフォーマンスの分散DBの開発を行っている。言語はADAPLEX (Ada と DAPLEX をまとめたもの)で、集中型のLocal Database Manager (LDM) と、それを複数回用いたDistributed Database Manager (DDM)を開発する予定である。本論文は、LDMのためのマルチバージョンデータを用いた並行処理と復旧スキームの統合システムの実現法を示したものである。更新トランザクションは2相ロックに従い、最新のデータを利用して更新する。更新がおこると古いデータは、version pool に古いものは以後になるようにリンクを付けて入れられる。読出し専用トランザクションではロックを使わず、旧バージョンのリンクで最初に出会った、自分の開始前に終了した更新トランザクションの作ったデータを用いる。Version pool の廢領域回収をうまく行なうことが重要である。デッドロック、強制中止、ソフトクラッシュ、ディスクのエラー等の原因で必要となる回復についても、原因の種別に分けて検討している。

22) 数値データベースシステムのモデル (O. A. Daini)

数値データ処理に必要な行列を、その性質に応じた形で2次記憶に記憶するためのデータ管理について述べている。行列としては、稠密(直接記憶する)、疎(間接記憶、非要素素を行または列優先順に並べる。別に位置も記憶)、対称、非対称、帯、三角等が考えられている。

23) データベースに対する効率の良い空間的アクセス (M. Tamminen)

地図データベース等で必要な、「点(x,y)はどの領域に含まれるか?」という質問処理について扱ったものである。順序保存型拡張可能ハッシュと図形圧縮のための4分木の考え方をを用いたEXCELL法によって、上記の問題は $O(1)$ の

アクセスコストで処理できる。

24 関係データベースシステムの機能拡張について (R. L. Hashin, R. A. Lorie)

関係データベースシステム (System R を考えている) を、工学的設計、オフィスオートメーション、グラフィックス等に用いるための機能拡張として、(1) 構造のあまりない長いデータの処理、(2) 複合したオブジェクトの扱い、(3) 会話的利用の効率化という3つの問題について扱っている。(1) の目的のため、データ型として LONG を許し、CURSOR によってデータの一部をとってこれるように SQL を拡張している。複合したオブジェクト (部品名を表わす関係の1つの組が部品の構成要素を示す関係では複数の組に対応する) を陽に扱うために、IDENTIFIER (オブジェクトの識別)、COMP-OF (親子関係)、REF (子というしのリンク) という属性型を導入している。システムよりの結果を見てデータを変更しながら会話的に処理するという方式の効率化のためには、ロックや回復法について考えなおさなければならない。このため、オブジェクト単位のロックを許し、コピーを利用者領域に作るための \$CHECKOUT および戻すための \$CHECKIN という命令が追加されている。また、種々のバージョンのデータも保持できるように拡張されている。

25 土地データベース管理システム (F. Antonacci, P. Dell'Orco, A. Turtur)

土地データベース管理システムは、地図、航空写真、サテライト情報、統計等の種々のデータを統計でき、計画決定に利用しやすい機能を持つことが要求される。本論文では、Codd の拡張モデルを用い、AQL を拡張してこれらの要請に答えるシステムの提案を行っている。[論文の遅れのため、Proceedings には含まれていない]。

26 関係データベースマシンのための定義域に基づいた内部スキーマ (M. Missikoff)

本論文では、データベースマシンに適した内部スキーマとして、定義域の重複する属性をまとめて逆ファイルを作る方式が提案されている。[これは、ファイル構成法として、Header が提案したもの (ACM TODS, 1978) と全く同じものである。我々のシステムではこれを拡張した構成を用いている (Lecture Notes 75, 1978; COMPSAC 79)。]

27 マルチプロセッサシステムにおける準結合 (P. Valduriez)

準結合をバス結合のマルチプロセッサシステムで並列処理するための方法について述べている。システムは、キヤッシュ付のサブシステムに2つのバスがつながれた構成となっており、1つのバスにはデータベースとフィルタプロセッサからなるサブシステムが複数つながっており、他のバスには結合プロセッサが複数つながっている。等結合をする場合、一つの関係の結合属性値を結合プロセッサに分散させ、ハッシュ値によるビットベクトルを作り、これを用いて準結合を実現する。もう一つの方法は、1つの関係の結合属性値集合を用いてフィルタプロセッサを働かせるものである。この方式は最大値・最小値をまず計算することにより不等号結合にも使える。論文では準結合を使わない結合も含めてこれらの方式の比較を行っている。

28 データベース応用における共通表現の解析 (S. Finkelstein)

データベースの利用において、共通の表現を持つ質問が繰り返し表われることが多い。これらを考えた複数質問の最適化が本論文の主題である。質問を質問グラフにより表現して、他の質問処理の結果 (システムに蓄えられている) が利用

できるかどうかを検査する。このような処理のため、Pascal で COMMON というプログラムも開発している。

[29] 質問最適化のための構造 (A. Rosental, D. Reiner)

関係モデルおよびネットワークモデルのための質問最適化の方式について述べている。質問を置き換え可能な結合を示すグラフ表現にし、これを場所やソート情報を考慮した物理グラフに変形する。このグラフに対して最適化操作をほどこし最終の処理方式を決定している。

[30] PASCAL/R 関係データベース管理システムにおける質問処理方式 (M. Jarke, J. W. Schmidt)

本論文では、関係データベース機能を持つ Pascal 型言語 Pascal/R において、限定作用子のある質問の処理方式について述べている。一階述語で表わされた質問は prenex 標準形に変換される。質問は結合、論理積、論理和の順で実行される。全称記号のない場合は、すべての積項は独立に処理できるが、そうでなければ、まとめた処理の方が効率が良くなることもある。手法としては、同じ関係を用いる部分表現の並列処理、ネストされた部分表現の一括処理、限定作用子のある関係数の減少、その他の技法が述べられている。

[31] トランザクションのモデル化 (C. Rolland, C. Richard)

本論文ではデータベースのモデル化と同じような手法でトランザクションのモデル化を行なう。概念レベルのモデル化では、C-Object (時間的な変化が同じ性質のものをまとめる)、C-Operation (実世界の操作のクラス)、C-Event (実世界のイベントのクラス、基本的な状態変化で、C-Operation をトリガーする) という概念の関係をグラフで表現している。次に、論理レベルのモデル化と、概念レベルよりの変換をまとめており、最後に DBMS でトランザクションを扱うために、トランザクション管理モニター、トランザクションベースを統合したシステムを提案している。

[32] データベースにおけるドキュメントの概念 (I. Kowavski, M. Lopez)

本論文では、データベースによるドキュメントの扱いとして、Text をデータ型に追加し、Document をテキストを記述する概念として導入している。ドキュメントの内部的側面として、レイアウト、論理構造 (論文は標題、本体、結論、文献よりなり、本体は節に分かれている、という木構造を持つ)、固定部分と変化部分 (フォームと内容等) をあげている。外部的側面としては、検索のキーとなるものがあげられ、論文の場合は、日付、著者等である。機能的側面は、変更、コピー、分配等の行ないやすいことをあげることができ、ドキュメントは、計算機内では電子ドキュメントとなって扱われる。Document は、特性 (主に外部的側面に対応) とテキスト (主に内部的側面に対応) より構成されている。テキストは、内容を示す文字列と木構造を示す構造、変数表より構成される。実体とドキュメントの対応付けをモデル化した、実体ドキュメントモデルを提案し、データベースにおける操作についてまとめている。このシステムは、決めた書式でのテキスト出力、郵便、情報検索に利用される。

[33] データ圧縮のための多グループ化手法 (K. A. Hazboun, M. A. Bassiouni)

データの性質により、文字の出現頻度が異なるいくつかのグループのあるとき、これらを一律に1つのハフマン符号にしないで、各グループに対応した符号化を行うとデータ圧縮率を向上させることができる。この場合、グループが異なる

時の遷移記号も符号化しなければならない、あるグループに属していない記号の
くる確率を用いて、これらの記号をまとめた節点をハフマン木内に含める。つぎ
に、グループ間の遷移確率を用いて、遷移情報を示す記号列もハフマン木で符号
化する。実験では、10~40%ほどハフマン符号より圧縮でき、また、個々のグ
ループの木がちいさくなることから復号時間も6%良くなったと報告されている。

謝辞 日頃御指導いただく本学矢島脩三教授に深謝します。また、本稿をまと
めるにあたり御助力いただいた、麗々公社勝野裕文氏、愛知教育大中津植男氏、
本学大学院生武田、小島、近藤の各氏に感謝します。

文献

- (1) ACM SIGMOD Proceedings 1982
- (2) ACM SIGMOD Record, April 1982

来年度のSIGMODは、5月23日から26日までサンホセで開かれる。論文の期
限は12月1日で、分野に応じて宛先が異なる。

{ SIGMOD: Prof. D. DeWitt, Univ. of Wisconsin
オフィス関係への応用: Dr. E. Carlson, IBM San Jose Lab.
エンジニアリングへの応用: Dr. R. Lorie, IBM San Jose Lab.

なお、来年度のPODSは3月にアトランタで、VLDBはシンガポールでの
開催が予定されている。

361-23 (日程が変更される)