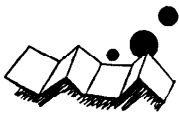


解説



データベース研究の動向†

上 林 弥 彦††

1. ま え が き

データベースシステムは、多数の利用者がデータを共有し利用する目的で構成される。情報工学におけるデータベース研究は、データそのものよりもこのデータベースシステムにかかわる諸問題を扱うものである。

データベースシステムは比較的歴史のある計算機応用の分野ではあるが、最近まで情報工学の中心分野としては認められていなかった。しかし、1975年以降はデータベースを専門に扱う国際会議の組織化や学術誌の刊行がいくつか行われ、現在では、計算機基礎理論(アルゴリズム解析等)に次いで論文の発表の盛んな分野となってきている。また、計算機基礎理論、人工知能、グラフィックス、計算機アーキテクチャ、CAD等の分野の雑誌や国際会議にもデータベースに関連した論文が掲載されるようになり、その他の各種応用分野も含めて、その重要性が認識され始めている。北米の大学では、この分野の求人も多く学生に人気のある分野の一つとなってきており、教育面でも急速に充実しつつある。このようなデータベースの人気は、計算機システムの最近の進歩によって高機能のデータベースが経済的に実現できるようになり、従来潜在的であったデータベースの需要を顕在化させたことが最大の原因となっている。このことが、データベース研究の学術的な側面の発展をもうながしている。情報工学の諸分野で、実際面の経験の蓄積がある程度でき、理論化の可能性や理論的手法によって解決すべき問題があらかくなってから理論化が行われ、その成果がさらにその分野の高度化に役立ったという歴史がある。この点で、現在のデータベースはこのような理論化の波に洗われていると言える。データベースよりも歴史が

古く、性格的にも良く似た情報検索の分野も体系化へ進むきざしを見せており、これらデータ科学の理論化とその発展は、80年代の情報工学研究の一つの目となるであろう。

本稿では、まずデータベース研究の問題点をまとめ、研究室で収集した約3,900件のデータベース関係の論文の分析を中心に最近の研究上の話題をまとめる。

2. データベース研究の問題点

データベース、情報検索、ファイルシステムは、人間の頭脳における記憶の機能を実現するためのものであると言える。しかしながら、記憶の機能にも種々のものがあるため、それぞれ異なった立場より研究がなされている。データベースシステムの目標のうち主なものは次の通りである。

(1) データの共有: データを共有することにより冗長度を減らし、コストを低減させると共に、データ全体を一律に管理することができる。

(2) 視像 (view) の実現: 上記のデータの共有による利点を生かしながら、各利用者は個別ファイルと同じような自由度でシステムを使えるようにするため、各利用者の扱うデータの範囲を規定する概念である。

(3) 使い易く機能の高い言語の実現: データを蓄積・検索するための言語は、データの物理的構造を知らなくても簡単に使え(非手続き性、内部構造との独立性)かつ高い機能を有する(例えば関係完備)必要がある。

(4) 各種独立性の実現: システムのファイル部分、ハードウェア部分、システム内部でのデータの物理的構造、論理的構造、データベースの利用形態等を分けて定義でき、それぞれの部分の変更の波及は最小限におさえられ、システム全体の再構成がほとんど起らないようにすること。

(5) データベースのモデルの利用: 上記の言語の

† An Overview of Database Research by Yahiko KAM-BAYASHI (Department of Information Science, Kyoto University).

†† 京都大学工学部情報工学科

開発や独立性を達成するために、各レベルの定義のために利用できるモデルが必要である。利用者の利用形態、システム内部での論理的モデルに対しては、実体とそれらの間の関連を表現したものが用いられる。ファイル編成やデータ構造に対してもそれに適した形式的記述が重要である。また、これらのモデルによってデータの持つ意味や各レベルの定義の形式化やそれら間のインタフェースの定義が可能となる。

(6) **オンライン更新**: 銀行システムや座席予約システムのように、データの変更は要求のあった時点ですぐ行わなければ問題を起すことが多い。情報検索システムでは、更新をまとめて処理するのがふつうである。

これらのデータベースの諸目的は、現状では完全に達成されていない部分が多い。そのことがデータベース理論研究のための動機となっている。例えばデータの共有は次のような問題を引き起している。

(1) データを共有しているため、各利用者のデータ操作に限界を設ける必要があり、視像上の操作も個別ファイルより制限が強いので、これをゆるめる方法も研究する必要がある。

(2) データの安全性の問題があり、利用者識別、利用権付与、利用権制御、暗号の利用等の種々の手法を確立する必要がある。

(3) 各利用者はデータの一部しか見ることができない。全体のデータが矛盾しないようにするため一貫性の検査機構が必要である。

(4) 複数の利用者がデータベースを同時に使うため、並行処理の問題が生じる。効率が良く、意味的に正しく、デッドロックのない処理を行う手法を研究する必要がある。

(5) 共用によってデータ量が増加する。大量のデータに対する処理では、アルゴリズムの効率によって処理時間が大幅に変わるので、良いアルゴリズムの研究がより重要となる。

代表的なデータベースシステムのモデルとして、階層型、ネットワーク型、関係型が良く知られている。この順はシステム効率の高い順で、自由度や機能の低い順となっている。このため、効率と機能のかね合いでこれらの三種とも今後も残ってゆくと考えられる。

理論的な扱いを考える場合には、多種独立性の実現、自由度の高い言語の実現、制約の少ない視像の実現等前記の目標を達成することが容易であることから、関係モデルが最適であると言える。さらに、関係

モデルに基づく問題の定式化は他のモデルに基づくものより単純になることが多く、かつその成果を他のモデルに適用することも容易であることも、関係モデルが理論面で使われる大きな理由となっている。データベースの諸問題を関係モデルで定式化することは、アルゴリズム研究において構造の単純なチューリング機械が用いられるのと類似しているとも考えられる。チューリング機械と異なるのは、チューリング機械は単なる計算モデルのままでは止まっているのに対し、関係モデルは将来のデータベースに幅広く用いられることが期待されている点であろう。(例えば、ADABAS や Model 204 (CCA) のような準関係データベース言語に対する利用者の人気が高いというサーベイがある)。

3. データベース関係の研究分野

筆者らは、データベースおよびその関連分野の文献を約3,900件収集している。これらの文献は、1970年代については代表的な出版物のほとんどを調べ上げて集めたもので、1960年代およびそれ以前さらに1980年に入ってから文献も含まれている。これらの文献について約3カ月かけて分類作業を行った。この作業は、まず大分類を行い、つぎつぎと細分して、細分野どうして関係のあるものを統合するという形で行ったもので、研究室で開発した文献情報処理用のプログラム群を利用した。その結果、データベース関係の研究テーマ等を示す一つの分類が得られたので、その分類に含まれる文献数と共に以下に示す。文献数は、その分野の研究の盛んなことを示してはいるが、階層型データベースのように実用的なものに対してはあまり文献がないことから判るように、むしろ実用段階に入る前の方が文献数が多い。

(A) 一般的文献

A 1	本および論文集	52
A 2	サーベイと文献表	23
A 3	その他	23

(B) データベースモデル

B 1	階層モデル	5
B 2	ネットワークモデル	35
B 3	関係モデル	46
B 4	実体関連モデル	40
B 5	他のデータモデル	87
B 6	モデルの等価性と比較	33

(C) データベース言語

C 1	データ定義言語	31
-----	---------	----

C 2	関係言語		F 23	論理的アクセスパス の自動生成	10
C 21	関係論理	11	F 3	他の質問処理アルゴリズム	23
C 22	関係代数	9	F 4	ビュー処理	
C 23	SQUARE/SQL	15	F 41	ビュー上での更新	12
C 24	QBE	19	F 42	その他	20
C 25	言語の完全性	6	(G)	一貫性, 安全性, 並行処理, 回復	
C 26	その他	50	G 1	一貫性	67
C 3	その他のデータ操作言語	42	G 2	安全性	
C 4	ホスト言語		G 21	統計データベース	19
C 41	APL	14	G 22	利用者の同定	5
C 42	その他	43	G 23	利用権付与	15
C 5	データベース言語—その他	15	G 24	利用権制御	13
(D)	データベースシステム		G 25	プライバシー	13
D 1	データベースシステム	90	G 26	安全性—その他	68
D 2	階層型システム	25	G 3**	関連した安全性問題	26
D 3	ネットワーク型システム	16	G 4	データベースの並行処理	55
D 4	関係型システム		G 5**	関連した並行処理問題	25
D 41	System R	16	G 6	回復	27
D 42	INGRES	15	(H)	データベースの保守	
D 43	PRTV	8	H 1	データディクショナリ/ディレクトリ とデータ通信	17
D 44	その他	61	H 2	データベースの変換	65
D 5	ミニコン, マイクロコン DB	16	H 3	再構成	17
D 6	個人用データベースシステム	8	H 4	データベース管理	12
(E)	データベースの設計と解析		H 5	その他	13
E 1	データベース設計	151	(I)	分散型データベース	
E 2	ANSI/SPARC モデル	16	I 1	分散型データベース一般	49
E 3	概念スキーマ	25	I 2	分散型データベースシステム	
E 4	要求分析	20	I 21	SDD-1	12
E 5	関係データベース設計		I 22	不均一型分散 DB	12
E 51	サーベイ	5	I 23	その他	28
E 52	従属性	54	I 3	分散型データベースの設計と評価	
E 53	設計基準/正規化	32	I 31	資源の配置	21
E 54	設計アルゴリズム	32	I 32	システム設計	47
E 55	その他	10	I 33	モデルと評価	22
E 6	パフォーマンス解析		I 4	分散型質問処理	28
E 61	モデル	16	I 5	分散型並行処理	71
E 62	パフォーマンスモニタ	5	I 6	信頼性と一貫性	16
E 63	パフォーマンスの評価	23	(J)	高水準データベース	
E 64	DBMS の選択	4	J 1	自然言語型のデータベース言語	43
(F)	質問処理		J 2	推論付データベース/知識ベース	62
F 1	ブール式の処理	12	J 3	自己組織型データベース	10
F 2	関係型言語の質問処理		J 4	発展形データベース	4
F 21	関係表現の処理	51			
F 22	結合操作	16			

J 5	ファジィデータベース	16	N13*	他の文献検索の話題	34
(K)	データベースシステムのハードウェア構成		N14*	モデル	23
K 1	データベースハードウェア一般	5	N15*	評価	17
K 2	記憶装置	20	N16*	情報検索システム	24
K 3	連想記憶	14	N 2**	機械可読データベース	6
K 4	特殊プロセッサ		N 3**	クラスタリング	24
	K41 連想プロセッサ	19	(O)	データ構造とアルゴリズム	
	K42 整列・併合プロセッサ	11	O 1**	データ構造とアルゴリズム一般	23
	K43 その他	15	O 2*	データ構造	53
K 5	データベースマシン	76	O21*	平衡木	30
	K51 RAP	16	O22*	二進木	38
	K52 CASSM	14	O23*	その他の木	25
	K53 後置型データベース	11	O 3*	探索・整列・併合	
	K54 テキストプロセッサ	6	O31*	探索	42
(L)	物理的データ構成		O32*	整列	43
L 1	物理的データ構成一般	26	O33*	併合	14
L 2	ファイル編成		O 4*	系列処理	46
	L21 転置ファイル	40	O 5**	その他	50
	L22 平衡ファイル	18	(P)	情報理論	
	L23 部分マッチハッシュ	16	P 1*	データ圧縮	63
	L24 併合型インデックス	7	P 2*	暗号	59
	L25 他の多属性用ファイル	25	P 3**	誤り訂正	10
	L26 一連検索ファイル	33	(Q)	その他の関係する計算機科学の分野	
	L27 自己組織ファイル	12	Q 1**	オペレーティングシステム	12
	L28 B 木, VSAM, ISAM	33	Q 2**	プログラミング言語とその処理	16
	L29 関係モデル向きファイル	25	Q 3**	データ型, 抽象データ型	36
L 3	データベースに用いられるハッシュ	27			
L 4	関連するハッシュの話題	81			
L 5	ファイル編成のパフォーマンスのモデルとその解析	85			
L 6	記憶階層	40			
L 7	ファイル割当て	23			
L 8	その他	50			
(M)	データベース応用				
M 1*	図形データベース, グラフィックス	88			
M 2**	CAD	21			
M 3**	医療情報処理	34			
M 4**	事務処理	28			
M 5	計算機科学への応用	11			
M 6**	その他	86			
(N)	情報検索				
N 1*	情報検索一般	88			
	N11* 用語分析と索引付け	41			
	N12* 文献分析	12			

上記の分類記号のうち印の付いていないものはデータベースの分野に属すると考えられるもので、主要な論文はほぼ集められている。*印の付いた分類記号に関しては、データベースの論文の現われる文献のものや関連の深いものはほぼ集められていることを示している。**印は、教科書や最近の文献のみといったものを示している。文献の分類はできるだけ一カ所になるように試みたが、数カ所に現われているものもある。3,900 件の中には上記以外のものもあるが分野として不完全であるため省略した。論文数の非常に多い分野の一部には筆者の勉強不足で細分できなかったものが含まれている。

この分野名からデータベース研究の性格をつかむことができる。例えば、モデル論や安全性は人間社会の扱いを考えた文科系的な思考も必要であるし、システムに対しては、ハードウェアとソフトウェアがからみ合った扱いが必要である。データベース設計、質問処理、統計データベース、分散データベース、並行処

---1973---	---1974---	---1975---	---1976---	---1977---	---1978---	---1979>---
CHAMB7312	CHANH7403 CHU-07411	COMB 7509 THOM 7507	ALSB 7605A ALSB+7603 ALSB87605 LEGOL7608 MANMR7603 STEAL76	BERNG7711 BERNF7711 BERNS77 BERNS7712C GOLD 7706 LEE-S7711 ROSES7705	BADAF7808 BERNF7805 BERNS7806 BERNS7806 BERNS7808 GOLD 7808 ISLON7811 LEE 7805 MENAM7808 HRNAP7805 ROSES7806 SCHL 7806 STON 7808 STUCC7808 THOM 7802	BADA 7911 BADA 8010 BAYE8010 BERNC7912 BERNG7906 BERNG8010 BERNS8003C BERNS8003S BERIK7910 CHENB8010 GARC 7908 HERMV7910 KAME 7910 KAWAN7910 LEE 8005 LIN 7908 LOME 7906 MENAM7905 MING 7908 PAPA 7910 RANIF7910 RIES 7908 ROHA 8005 SCHL 7909 SEGU 7910 STON 7903C THOM 7906 YANAM7911

図-1 分散型並行処理の論文の年度別頻度

理、ファイル編成等の分野では、計算機基礎論の手法が有効である。その他高水準データベースでは人工知能の手法が用いられ、データベース応用では、グラフィックス、CAD、医療、事務処理等、種々の分野に関連を持っている。暗号の研究も安全性と深くかかわっている。このようにデータベース研究は、まさに多様性を持った分野であると言える。

さらに、各分野の論文の年度別頻度をみると、研究分野の推移を知ることができる。例えば、言語関係の論文数は最近減ってきているが、分散処理の論文は非常に増えている等の傾向がある。図-1には、分散型データベースにおける質問処理の研究論文の年度別頻度が示されている。

このほか、文献標題に含まれる共通キーワードを抽出し、各フレーズを含む文献数を求めれば、分野別とは若干異なる情報を得ることができる。

4. データベース研究の話題

前章のデータベースの各分野の中から選択して、データベース研究の最近の話題についてまとめてみる。

モデル論: 階層、ネットワーク、関係モデル以外のものでは、実体関連モデルの論文が多い。モデル論の研究者は日本では、小林、石田らがいる。Ullmanの教科書では、実体関連モデルより出発して、関係モデル、ネットワークモデル、階層モデルの順で、データベースの基本設計法を示している。関係モデルの欠点である意味的な扱いに対する研究が多く、このモデル

の提唱者である Codd も、ACMTODS の 79 年 12 月号に意味論的拡張を考察した論文を発表している。この論文を含めて不完全情報の扱いに関する論文は現在 7 件あり、そのうち 6 件は 79~80 年に発表のものである。各モデルに対する比較論の論文は総数としてはかなりあるが最近のものは少ない。ネットワークデータベース上での関係モデルの実現等の質問変換にも関係するものが 8 件ある。

データベース言語: データベース言語の分野では関係言語に関する研究が多い。関係言語には、関係代数、関係論理、写像型、定義域指向型、グラフィックス向き等の種類があるが、特に、System R で使われている SQL (SEQUEL) や、市販されている IBM の QBE の論文が目立つ。自然言語インタフェース (43 件) の中にも関係言語を基礎としたものが多くみられる。データベース言語の完全性は、Codd により関係論理に基づく言語と同じ表現力を持つことと定義されている。この定義に疑問をなげかけたのは、Bancilhon (7809) であり、79 年には、Aho, Ullman (POPL 7901) や Chandra, Harel (STOC 7904) が新しい完全性の定義を行った。関係代数の表現力に関しては Paredaens (IPL 7802) があり、その一般化に相沢 (情報 8007) の研究がある。プログラミング言語とデータベースとの関連に関する研究も盛んで、とくに APL (14 件) に関するものが目立っている。高級言語を拡張してデータモデルを 3 種共扱えるようにする試みとしては、Data の UDL (Unified Database Language)

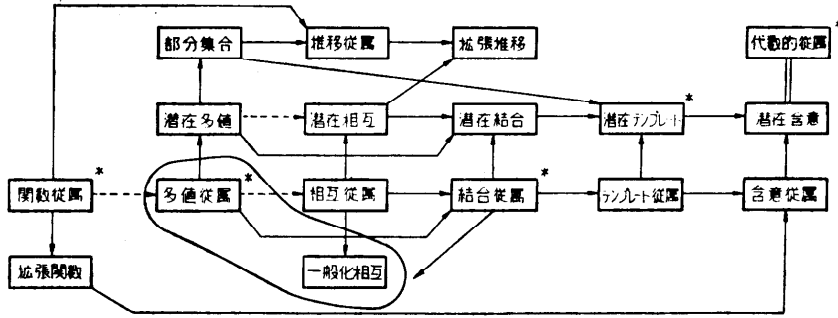


図-2 各種従属性間の関係

がある。

データベースシステム：データベースシステムの論文の中でも、関係モデルに関する研究が多い。商用のものとしては、Tymshare 社の MAGNUM や、IBM の QBE 言語によるものがあり、研究中のものとしては、IBM の System R、カリフォルニア大学の INGRES、英国 IBM の PRTV 等が有名である。これらの中では、MAGNUM 以外はビュー処理の機能を有しており、System R は、利用権制御、一貫性制御、並行処理等のできる最も進んだシステムである。そのほか、Tandem の ENFORM、Honeywell の MULTICS のもとで動く MRDS や日電の INQ (竹内ら)、富士通の RDB/V1 (林、牧之内ら) 等も知られている。最近の傾向としては、効率があまり問題とならない中小規模のデータベースシステムを関係モデルで実現するものがあり、その例として、IBM のシステム 38 や Britton-Lee Inc. のデータベース計算機 IDM 500 を掲げることができる。マイクロコンピュータのオフィスへの利用の拡がりとともに、COBOL やデータベースがマイクロコンピュータでも使えるようになってきている。Micro Data Base System 社では、BASIC や PASCAL 等の言語から呼び出して使えるネットワーク型データベースシステムおよび階層型データベースシステムを市販しており、Z-80 の代表的なオペレーティングシステム CP/M のもとで利用できる。簡略化したデータベースシステムやコンピュータノートブック、会計システム、図書管理システム等のソフトウェアはすでにいくつかの会社から供給されている。

データベースの設計と解析：データベース設計に関しては、種々の手法が使われており、この中でままとまりの良いのは、ANSI/SPARC の多層構成とそれに関係した概念スキーマ設計および関係データベースの設計である。特に、関係データベースの設計論は、デ

ータベースの数学的理論の中では、かなり確立してきたものの一つとなっている。研究者の中には、Ullman, Aho, Ginsburg, Chandra 等、計算機基礎理論で良く知られている人達が含まれている。日本でも、嵩、谷口、増永、上林らは同様の背景を持った研究者である。他に小林、国藤、竹島、勝野、田中克己らがいる。データベースのデータの中には、なんらかの性質が常に保たれていることが多く、この性質の中には、システムの記憶するデータ量を減少させるのに役立つものがあり、それを用いて冗長性の持つ諸問題を解決できることがある。このような性質の中に意味論とも結びついた従属性がある。従属性によって設計された冗長度の少ない記憶形式を正規形という。関数従属性による正規形には、第一、第二、第三正規形、Boyce-Codd 正規形のほか、関係表の横方向の分解を考えた (3,3) 正規形がある (Smith)。多値従属性による正規形には第四正規形があり、さらに進んだ正規形には、射影結合正規形 (SIGMOD 7905) や定義域・キー正規形がある (Fagin)。

関数従属性の一般化として多値従属性のほかに図-2 に示すような種々の従属性が知られている。これは、関数従属性 (FD) や多値従属性 (MVD) の研究で次のような点が問題となったため、それらを解決するために導入されたものである。(a) MVD では1つの関係表を2つに分解することを考えていたが、3つ以上の表への分解の中には MVD の連続では表わせないものがある。(b) FD や MVD では、与えられた集合から導かれる従属性をすべて求めるための公理系が存在したが、MVD を一般化した潜在多値従属性では有限の公理系が存在しないことが判っている：

図-2 では次の記号を用いている。A→B: BはAの一般化であり、Bの特殊な形の中にはAと一致するものがある。A→B: BはAの一般化ではあるが、B

の特殊な形として A は表現できない。A=B: A と B とは等価である。(A B): A と B を合せたものを一つの単位として表現する場合、A*: A に対する公理系が知られている。

ある従属性が与えられた場合、あらかじめ決められた従属性集合から導かれるかどうかを判定するには、公理系が存在するか、なんらかの判定法が存在しなければならない。潜在多値従属性が有限の公理系を持たないため、部分集合従属性や潜在テンプレート従属性への拡張が行われた。部分集合従属性はさらに FD を含む推移従属性へと拡張された。残念なことに、この推移従属性や部分集合従属性も有限の公理系を持たないことが分かった (SIGMOD 8005)。潜在テンプレート従属性は公理系を持つことが示されており、FD 以外の有用な従属性を含む、現在知られている中では最小の従属性のクラスである。Sadri, Ullman は、この従属性と FD の干渉問題についても考察している (SIGMOD 8005)。さらに、FD を特殊な場合として含む代数的従属性が公理系を持つことが知られており、この従属性 (および等価な潜在含意従属性) が、有用な従属性をすべて特殊な場合として含む従属性の中で公理系を有する、現在知られている中で最小の従属性である。現在の研究課題の一つは、このような性質を有するより小さい従属性の発見ないしは、FD 以外のものを特殊な場合として実現する従属性と FD との間の相互干渉の公理系を求めることである。

データベース設計法として知られている、分解法、合成法、独立成分法について、それぞれ長所、短所のあることが知られており、有用な設計法の開発も期待されている。とくに、種々の従属性は、冗長性の減少や公理系の確立を目的として定義されているため、意味との対応には無理のあることもあり (上林他 COM-PSAC 7911)、利用者のみるモデル (例えば、ER モデル、酒井 SIGMOD 8005) と異なりうる点に設計上の問題がありうる。

データベースの解析には、シミュレーションを含む種々のモデル化の手法のほか、効率測定モニタや各種評価手法の研究がある。

質問処理: プール式での最適化や、関係度数での操作の順序を変更することによる最適化はかなり前より行われている。Chandra らは、特定の質問に対しては最簡形が一意に決まることを示した (STOC 7705)。Aho らは、その結果と従属性とを結びつけ、Tableau や Chase という概念を確立した。これは、質問の等

価性、従属性集合の変換、無損失結合等の種々の計算に利用できる有用な方法である。結合処理の方法や論理的アクセスパスの自動生成も重要な研究テーマとなっている。視像 (view) 処理については、視像上の更新の処理が大きな問題となっている。

一貫性: 一貫性はデータの上に課せられた制約で、静的な制約のほか、値の変更時に変更前と後の値の関係を規定する動的な制約もある。QBE では、これらの制約の表現が容易である。INGRES では質問の変更という形で一貫性の実現を行っているが、データ変更の途中での検査は行わないようにしなければならない。一貫性の一般化とその実現の手段としては Eswarn の提唱した trigger がある。

安全性: 計算機システムの安全性の確保は、計算機犯罪の防止やプライバシーの保護といった面で、単にデータベースに限らず重要な問題である。利用者に対して許す質問の制約が不相当であると、いくつかの質問を合成することにより許されていないデータの内容を知ることのできる場合がある。これは統計データベースの問題として知られていて、多くの研究がある。また、利用権制御のように OS ですでに知られていたものがさらにデータベース向きに研究されているものもある。公開キー方式等の暗号の利用も重要である。

並行処理: 多数の利用者の質問を同時に処理する場合、デッドロックを避けるだけでなく、結果として誤ったデータ操作にならないように注意しなければならない。このようなことの起らない十分条件として直列可能性という条件が知られており、それを実現する単純な方法として二相ロックという方法が知られている。最近の研究としては、その条件の一般化やロック管理の計算量等がある。

データベースの保守: この範囲に入る話題としては、データディクショナリ/ディレクトリ、データ通信、再構成、データベース管理等がある。データディクショナリは本来データベースシステムの管理者に、データベースの設計・開発・実行・管理のための種々の情報を提供するためのものであったが、その機能を拡張して他の目的にも利用しようという試みが盛んになってきている。例えば、(a) データの使われ方から、データのライフサイクルを知り、企業のデータ資源の管理に役立てる、(b) データベースの変更時の処理効率を上げる、(c) 種々の変更の影響分析に用いる、(d) 応用分野の分析に用いる、(e) 報告書作成に用いる。

分散型データベース: 最近分散型データベースに対する論文数が急速に増加している。分散型データベース研究の重要性としては、地理的に離れたデータベースを統合して一つの巨大システムとする必要性や、安全性の確保のため、およびデータの生成や利用に地域性のあるために分散システムを採用する必要性のある点等が考えられる。集中型と異なる問題点は、処理コストより通信コストがはるかに大きいと仮定できる点や同期が困難である点、データの重複を許す場合の多いこと等であり、全く同じ問題に対しても集中型とはかなり異なる処理をする必要がある。質問処理に関して、準結合を利用して効率を向上させる方法が最近注目されている。分散並行処理に関しては、一般の方法のほかに、SDD-1 での方法や多数決法 (Thomas, TODS 7906) 等が知られている。CCA の SDD-1 のプロジェクトは79年で終了し現在不均一分散データベース Multibase の研究が行われている。IBM では System R を結合した R* を研究中である。他にフランスの SIRIUS 等が知られている。日本では、穂鷹、椿、滝沢、大野、田畑らが分散データベースの研究で成果を得ている。

高水準データベースシステム: すでに挙げた項目の中では、発展形のもの論文 (国井ら, COMPSAC 7811) は少ないが研究課題として注目されている。本分類では応用の範囲に入っている、グラフィックスを利用するシステムの中にも高水準データベースに対応するものがある。知的データベースに対しては、日本では、大須賀、古川らの研究がある。山本毅雄、弘原海、有川、武谷、打浪、富樫、金森らはデータベース応用の研究を行っている。

ハードウェア: データベース向きのハードウェアとしては、連想記憶、知的ファイル、ソート・マージ用ハードウェア、大容量ファイル等が考えられ、それらを用いたデータベースマシンの研究も盛んである。初期のものは結合演算に向いていなかったが、最近には種々のデータベース演算を考慮したものが出ている (Yao の DIALOG 等)。RAP を開発していた Schuster は、トロント大学から Intel へ、さらに Tandem Computer へ移っており、そこで、1セル当り1メガバイト、セ

ル数が数十のシステムを開発中である。最近では、Hsiao のデータベースコンピュータ DBC の論文発表が盛んである。UNIVAC 等でも実用化研究がなされており、日本でも、田中護、植村 (EDC)、箱崎 (GDS)、前川、有澤らの研究があり、活動が盛んである。商用のデータベース計算機としては、Britton-Lee の IDM 500 (関係データベースマシン、ファイルは最大 32G バイト、IEEE 488 バスでホスト計算機と接続) や ADABAS データベースマシン 41 があり、共に 81 年初めより市販が予定されている。

ファイル編成: B木に対する研究では、効率の研究 (丸山、溝口) の他、索引部のデータ圧縮により、同じデータ量に対して索引の高さを下げる研究がある。拡張可能なハッシュについても、効率や速度がB木以上であるという報告もある。ページングを考えたアルゴリズムの効率化や、アクセスパスのモデル化等の研究も盛んである。いくつかのキーを指定した場合に有効なファイル編成としても、部分マッチハッシュファイルの他に、組合せインデックス、均衡ファイル (山本純恭ら)、K-D 木等の種々の手法が開発されている。

5. むすび

データベース研究には多種多様なものがあり、そのすべてを公平にカバーするのは非常に困難である。本稿は、読者の参考のために筆者の立場からの研究動向をまとめたものである。本稿で示した分野分類も、文献集作成中に大きく変化しており、今後も改訂の必要があると考えている。なお、紙数の都合で参考文献は省略する。なお、文献集は Database-A Bibliography という標題で Computer Science Press より刊行される。

謝辞 データベース研究に対してご指導・ご援助・ご討論をいただき本学矢島脩三教授に深謝します。文献集の作成に対するご助力および内容面での検討をしていただいた、レ・ヴィエト・チュン氏をはじめとする研究室の諸氏 (吉川正俊、小島功、徳田成穂、武田浩一) ならびに、名古屋大学小西修助手、神戸大学田中克己助手に感謝いたします。

(昭和55年9月19日受付)