

論説

データベース・システム

酒井 博 敬†

データベース工学の研究は、この2年間に発表された研究会報告だけを見ても、めざましい転換を遂げた。

その特長としてまずあげられるのは、CAD、オフィス、医療、研究、建設、行政などの分野での高度応用志向の顕在化である。歴史的“もの”資料、歯科矯正学における履歴データ、化学グラフ・データ、プロジェクト管理データ、音声データなどのデータベースが開発され、実用化されたものも少なくない。これら多様な応用分野では、文字、数値、図形、画像、音声などのマルチメディア・データの扱いに対するデータベース工学的なアプローチとともに、知識工学的手法がみられるようになった。新しい応用分野はデータベース工学研究テーマの宝庫ともいえる。

次にこれら応用の要請から、データ・モデリングに関する新しい提案が相次いでなされたことがあげられる。マルチメディア・データ、履歴データ、統計データ、工業データなど、伝統的なデータ・モデルの枠に適合し得ないデータを扱うために、その多様な側面をモデル化し、データ設計に反映させなければならない。新たな視点からデータの多次元性が考察され、形式化の試みがなされ、システム設計の方法論に浸透してきている。

第三に高度応用を支援するための高機能データベース・システムあるいはデータ言語が現れ始めたことがあげられる。CAD、オフィス、研究開発など、それぞれ専用ではあるが、マルチメディア DBMS ともいえるべきものが相次いで発表された。各システムには日本語を含む高度な言語とともに、文書、グラフ、図形、画像などを関連づけるマン・マシン・インタフェースが具備され、また質問応答のための知能インタフェースも提供されている。

高度応用の実現は基礎技術の開発に支えられている。第四にあげられるのは、ハードウェアおよび計算機アーキテクチャのデータベース工学的な展開である。計算機がデータベース・センタとしての色彩をもつよ

うになったこと、およびローカル・エリア・ネットワークやワーク・ステーションが普及したことによって、データベース・マシンと分散データベース・システムは実用化、商用化の段階に入った。この分野ではとくに大規模関係データベース向けデータベース・マシン、大容量主記憶向けデータベース編成法、分散データベース・システムにおけるデータ資源管理と質問処理の最適化など多彩な研究成果が報告されている。

いくつかの応用分野で扱われている新しいデータの性格をあげてみる。

医療におけるマルチメディア・データベース：関係、図形、画像の各データベースから構成され、文字・数値と図形・画像が関連づけられている。とくに時間区間と事象を属性にもつ病歴データを対象として、履歴検索空間の概念を取り入れた問合わせ言語が用意されている。

工業データ向けデータベース：関係データベースの中に引数付き実現値を導入し、類似図形の圧縮表現を図っている。また概形から細部へのデータ階層が維持され、階層をわたる図形検索機能が実現されている。

プロジェクト管理向けデータベース：ルール・ベース中のルールを満たすように構成された構造物、作業、プロジェクト・チーム、およびプロジェクト管理の各データベースからなる。それぞれの階層構造と階層間関係の維持とともに、作業進捗にともなうスキーマと実現値の経時変化をモデル化している。

これらの応用におけるデータの扱いに共通しているのは、データ・モデルにおける構造表現、操作、および完全性の各局面が多次元的にとらえられ、設計手法に反映されていることである。とくに重視される次元として、内包・外延(intension-extension)、時間あるいは時制、および抽象レベルの階層性があげられる。

内包・外延は、メタ・データ対データ、スキーマ対実現値の関係である。ANSI/SPARC/DBSSGのDAFTG(Database Architecture Framework Task Group)が提案する参照モデルでは、データ・モデ

† 京都産業大学経営学部

ル・スキーマ、データ辞書スキーマ、データ辞書データ、AP データの4段階が設けられている。この次元は、たとえばCADにおける設計変更やモデル変更、オフィス・システムにおけるルール変更など、状況の進展とともに起こるスキーマの変化、ルール・ベースにしたがったスキーマの自動生成、スキーマ変更にもなうデータの再編成、同じ実体に関する実現値の過去から現在に至る必ずしも構造的に一律でない履歴の重ね合せなどの問題を提起している。

データベースへの時間概念の導入は多くの面から議論されている。発生から消滅までの状態変化をとらえ、動的完全性を分析する手段としての実体履歴のモデル化、およびこの観点からのデータ抽象化、同一実体の継続的な変化を対象とする履歴データ・モデル、事象と時間区間に関する条件が指定される問合わせの処理、統計データベースにおける分類基準の経時変化など、データの時間次元の扱いは最近の重要課題である。マルチメディア・データにおける時間の問題は、構造要素間の関連性と完全性の維持をともなうため、一層複雑なものとなっている。またデータ・モデルの形式化については、履歴データの数学モデル、スキーマ記述における時制論理の導入などの研究が進められている。

抽象レベルの階層は、大局から小局へ、一般から特殊へ、粗から密へといった意味合いをもつ。ユーザ・インタフェースの面からデータベース全体像の理解把握を容易にし、会話型問合わせの便宜を図ることが一つの目的である。たとえばマルチメディア・データ間の関係、属性間の時間的關係、属性のもつ意味などを把握するために概要空間の表示から始めてデータ細部のズームアップへと導く機能、あいまいさを含む問合わせの処理、知能的問合わせ応答方式による目標探索機能などがあげられる。一方複雑な構造を管理する意味でも、対象を階層分解し、必要に応じた詳細レベルのデータを検索、更新する機能が要求される。これはCADにおける構造体のモジュール分解、プロジェクト管理における作業手順の分解、地図情報の地域分解などに見られる。個々のモジュールの構造は一般に不均質である。また問合わせ操作は垂直的および水平的な階層間関係にわたる。

さて本記念論文には、過去2年間の発表論文および新たな応募論文の中から、履歴データベースの基礎的考察に関するものが選ばれた。

履歴データベースでは、実体変化の経緯、すなわち

同一実体の実現値である組(tuple)の系列が考察の対象となる。従業員履歴や銀行小切口座履歴、CAD/CAMにおける種々のバージョンをもつ設計情報、患者の検査履歴や病歴などの医療情報、対象別・時系列で整理した新聞記事情報などがその例である。これらの応用では、個々のデータよりも、その履歴そのものが意味をもつ。

履歴データベースに関する主な研究課題としては、次のものがあげられる。

- (a) 時間情報や履歴情報の表現、操作、管理のための適切なモデルは何か。
- (b) 履歴データベースのよい設計法とは何か。
- (c) 時間情報や履歴情報に関する完全性制約とは何か。またそれをどのように形式化すべきか。
- (d) 履歴データベースの質問言語はどうあるべきか。またその実現方法と利用者インタフェースをどうするか。

本論文のモデルでは、履歴(組系列)および履歴集合(組系列の有限集合)に関する完全性制約をまず一般的な写像ととらえ、それに現実的と考えられる種々の性質を付加することによって、それらを適当なクラスに限定するというアプローチをとっている。その結果として以下のような制約を形式化している。

- (1) 一様な制約: 組系列がこれを満たせば、その任意の部分系列もまたこれを満たすという制約(例:「論文の連続発表は同じトピックでなければならない」)。
- (2) 初期化制約: 組系列の最初の部分が常に満たさねばならない制約(例:「すべてのセミナーは一月中に始まる」)。
- (3) 履歴集合に関する制約: いくつかの組系列(すなわち複数個の履歴)が共存するための制約(例:「同一人物は同じ日に複数のセミナーでは発表できない」)。

さらにある履歴に対して、ある条件を満たす部分履歴を答として返すような質問を区間質問と名付け、その性質を明らかにしている。

総じて本論文で扱われている事項やそのアプローチは独創性の高いものであり、履歴データベースの研究に今後必要となるであろう一般性のある数学的枠組を与えている。またここに導入された諸概念は、実際の履歴データベースや質問言語の設計にも適用できるものと考えられる。

(昭和60年7月1日受付)