

特集「学術情報データベースの構成と利用」の編集にあたって

杉本重雄† 宮崎収兄†† 伊藤 潔†††

汎用計算機、ワークステーション、パソコンを問わずさまざまな計算機の上にデータベースが構築され、現在の我々の営みにとって不可欠なものとなっている。データベースの規模はさまざまであり、数十年をかけて構築されてきた巨大なデータベースから、社内規模の大規模データベース、個人利用目的の小規模データベースまである。学術分野においても多様なデータベースが構築され利用されている。それらには学術論文や図書などの文献情報（書誌情報）を集積したもの、実験や調査から得られた事実情報や統計情報を集積したものなどがある。

本特集では、データベースを研究開発し提供する立場とデータベースを利用する立場の両方の観点からデータベースを論じることを目的にして、データベースに直接携わっておられる研究者や開発者の方々に執筆をお願いした。「作る立場」と「使う立場」の両方からの解説を含めることで、データベース開発者のみならずエンドユーザからのデータベースシステムに対する観点をも理解できるものとなることをめざした。作る立場からは、人文・社会・自然科学の各分野におけるデータベースを具体的に示すため、日本史データベース、統計データベース、ゲノムデータベースおよび高速増殖炉データベースについて、それぞれの分野の特性や実際に利用されているデータベースの構成などについて解説している。また、学術情報と流通には図書館が大きな役割を担っている。そのため、全国の大学図書館等を結び学術情報の集積と利用のセンターとして大きな役割を果たしている学術情報センターのデータベースについて解説する。学術データベースの将来を考える上で電子化した文書そのものを提供するデータベース、いわば「電子図書館」を考えることが重要である。そのため、欧米での「電子図書館」研究開

発プロジェクトについても作る立場の解説として位置づけた。使う立場については、データベース利用の専門家であるデータベースサーチャのデータベースの利用方法と、データベースとユーザとのインタフェースを考える上で重要な要因であるデータベース検索におけるユーザの特性について解説している。

本特集は、データモデルやデータベース管理システムといったデータの入れ物主体ではなく、データそのもの、いわば中味主体にデータベースを見ることを目的として構成した。そのため、解説の中に現れる応用分野に特徴的な話題、あるいは専門用語などに戸惑いを感じられることもあるかもしれないが、それぞれの分野での問題や特徴を理解し、それらを見比べていただくことができればありがたいと考えている。また、データベースの分野は異なっても、次のような共通の観点があると思われる。これらが本特集の解説を理解する一助となれば幸いである。

- ・データベースに格納するデータをどのように収集するか
- ・データベースに格納するために原データをどのように加工・入力・蓄積するか
- ・データベースから情報を得るためにどのような機能を設けるか
- ・データベースを検索することで得たデータを利用者がいかに利用するか、たとえば、利用者はデータベースを利用することで新たな事実を発見できるか

以下に各解説の簡単な紹介を示す。

1. 日本史データベース

日本史データベースに蓄積されるデータは、古文書あるいは古文書から取り出したデータである。ここでは、古文書に現れる文字の扱い方、資料の分類や配列の方法、漢語や古語の読みなど、歴史資料をデータベースに入力し、検索する過程に生じるさまざまな問題について述べている。また、テキスト情報だけではなく古地図のような

† 図書館情報大学

†† 沖電気工業(株)総合システム研究所

††† 上智大学理工学部

画像情報をも利用すること、データベースを利用することで新たな事柄を発見すること、すなわちデータベースの発見的利用が可能であることといったデータベース化に関する新しい話題もある。

2. 統計データベース

統計データベースに蓄積される統計情報は事実情報を集計したデータに基づくものであるため、文献情報や個別の事実情報と異なった性格をもっている。統計データベースの解説では、統計情報がどのように作成され、作成者によってどのような特性があるか、どのように利用されるか、実際に利用されている統計データベースにはどのようなものがあり、それらがどのように提供されるかを述べている。また、著者等が開発した長期経済統計のデータベースについて述べている。

3. ゲノムデータベース

ヒトゲノムを解読しようという研究が全世界的に取り組まれており、その中で計算機科学からのアプローチも重要な役割をもっている。ゲノムデータは分子生物学の解析手法に基づいて生成される多種多様かつ大量のデータであり、計算機を利用してゲノム情報を体系化することが求められている。ゲノムデータベースの解説では、ゲノム情報とそれをデータベース化する上での問題点、研究開発の動向を述べている。

4. 高速増殖炉開発データベース

原子力発電は巨大科学の代表であり、さまざまな最新技術の上に成り立っている。巨大科学を支えるために技術情報の利用性を高めることが重要であり、国際間での技術情報の交流が進められ、データベースの相互利用も進められている。高速増殖炉開発データベースの解説では、動力炉核燃料開発事業団の大洗工学センターで開発され利用されている燃料材料データベース、構造材料データベース、信頼性データベースについて述べている。

5. 学術情報センターのデータベース

文部省大学共同利用機関である学術情報センターでは大学等の研究者向けにオンライン学術情

報検索サービスを提供している。学術情報センターは大学図書館等を結んで構築されている学術情報ネットワークの中心である。この解説では学術情報センターが提供するデータベースとデータベースサービスの状況、および学術情報センター独自のデータベースの作成と将来のデータベース像について述べている。

6. 欧米における「電子図書館」プロジェクト

図書館の電子化が進み、書誌情報データベースを中核とする図書館情報システムが一般に広く利用されている。将来の図書館情報システムは文書資料そのもの、すなわち一次情報を提供する「電子図書館」を指向したものになってゆくと考えられる。ここでは欧米における一次情報を指向したシステムとしてコーネル大学の CORE プロジェクトとカーネギーメロン大学のマーカープロジェクトを紹介する。

7. データベースサーチャの視点

データベースサーチャは情報検索の依頼を受けてデータベースを利用する検索の専門家である。情報検索の対象となる分野は幅広く、かつ利用するデータベースも多様であるため、分野に応じたデータベース検索を行わねばならない。ここではサーチャの業務を紹介し、いかに検索の戦略をたて、いかにデータベースを選択し利用するかについて解説する。

8. データベース検索におけるユーザの特性

ユーザがデータベースをより効率的・効果的に利用するためには、データベースユーザの特性を理解することが重要である。ここではデータベース検索におけるユーザの特性に関して、情報利用者の要求、エンドユーザと検索仲介者の間の相互作用、検索スタイルと検索戦術、検索者の背景的要因といった分野の研究について解説する。

本特集の構成にあたっては執筆者の方々をはじめとして多くの方々の協力と助言をいただいた。末筆ながら感謝の意を表したい。

(平成4年8月24日)