データベース産業の課題

阿部耕一郎*

Assignment to database industry

ABE Kohichiroh

[著者抄録] 各種のデータベースは、これから求められる大型システムの必須部品であると考えられ、またデータベースのみによる情報提供というものも、これから、その対象とするマーケットが成熟して行くと考えられる。本稿では、そのような状況が、これまでどのように形成され、これからどう展開して行くかを論じた。

[著者付与キーワード] データベース, 情報提供, 情報マーケット, 知的情報資源, 事実情報資源, 学術文献, 数値データ, オンライン

[Author Abstract] Various kinds of databases are considered to be essential part in future large sized systems. Information provision only by databases is also considered to be growing as the market becomes mature. This paper discusses how such circumstances have been built and will be developed from now on.

[Keywords by Author] database, information service, intellectual information resource, fact information resource, technical literature, numerical data, online

* 広島経済大学 (〒731-01 広島市安佐南区祇園5-37-1) Tel. 082(874)3041
Hiroshima University of Economics (37-1, Gion 5-chome, Asaminami-ku, Hiroshima-shi 731-01)

1. はじめに

最近、コンピュータを中心としたシステムを構築する話題のなかには、必ずと言ってよいほど、データベースの作成という事柄が含まれている。

これは、コンピュータが、単に数値演算のみを人間に代わって処理するというものではなく、何らかの判断を行うことを求められるようになってからの特有現象である。しかも、これからの大型のシステムに対しては、いくつかの知的判断を必要とする処理の複合体であることが、求められており、そのことは、その知的判断の根拠となるものとして、各種のデータベースを内包していることが前提とされている。

このことは、個人ベースでの情報処理を前提とするパーソナルコンピュータについても、具有すべきソフトウェアパッケージとして、Multiplanをはじめとする表作成プログラム、日本語ワープロプログラム、DBMS プログラム（データベース作成用プログラム）の3種が必ず挙げられることから、同様の社会環境の存在することが理解できる。つまり、関心は、「判断」という極めて人間らしい現象に集中しているというわけである。

考えてみれば、人類がこれまでに経て来た社会制度である、農業社会においても工業社会においても、人間の創造物の本質はいずれも手や足の代替であったとすることができよう。それに引き換え、現在進行中の変革においては、いずれも人間の脳の代替物としての創造であることが理解できる。

もちろんそれが、人間の脳の持つ能力のすべてであるとは毛頭考えていない。本来の人間の脳の能力と比較すれば、それらは、ごく限られた部分である。しかし、人間の願望は際限もなく膨らむものであり、第五世代コンピュータの開発、あるいは AI（人工知能）の組み立てが昨今話題としてあがって来ている。

つまり、情報化社会の進展度を表示する指標としては、人間の脳による作業の代替として、どのレベルまで可能になったかを測定することが基本

になりそうである。しかし、読むこと、書くこと、計算することの3種の作業を単一行うこと、あるいは一つの能力のみを代替するものを指して、現代的で高度化されたものと呼んでいるわけではなさそうである。

現在、高度情報化社会にふさわしいと言われる事象をよく見てみると、そこには、これら3種の機能が複雑に組み合わされている複合物、つまりシステムと呼ばれるものが多い。しかも、これらシステムと呼ばれるものの中には、必ずと言ってよいほど、データベースと呼ばれる部分が含まれているのが特徴でもある。

2. データベースとは何か

ではいったいデータベースとは何であろうか。そこで、具体的なシステムに即して眺めてみることにする。

私達が新幹線を利用して旅行する時、まず JR 緑の窓口へ行って座席を予約する。窓口担当者に座席に関する希望条件を告げると、即座に希望に該当するものが入手できる。この一連の作業を支えているシステムの中には、運行される全列車の座席に関する情報がすべて、希望を構成する要件のどの観点から質問されても対応できるように、正確に編集されたファイルとして存在している。

したがって、窓口担当者である人間の記憶とは何かかわりもなく、2重販売その他のトラブル要因を生み出すことなく、新幹線運行の全体が進行しているわけである。このような情報ファイルをデータベースと呼んでいる。

もともと、データベースという言葉の意味は、文字記号で表示されるものも、数値記号その他の記号で表示されるものも、記号と認定できるもののほとんどを対象とする、広い意味での情報を必要に応じて、取り出しやすい形で集積したもの、つまり情報の基地という意味である。

通常の人間生活に置き換えて言うならば、非常に機能化された情報図書館を持つことに等しい。このような情報図書館が、システムの中に内包さ

れているということである。

もちろん、情報化社会以前から、このような情報を手に入れるための道具、つまりデータベースに相当するものは、数多く作られている。それは、事典であり、辞典であり、年鑑類・名鑑類のすべてが、明らかに現在私達が、データベースと呼んでいるものの原形であると言えることができる。

しかし、このようなこれまでの知的資産である辞典やデータブック類と、現在問題にしているデータベースとのいちばん大きな相違点は、編集の観点の自由度の差にあると言えることができる。

辞典を例に挙げるならば、ある未知の言葉の持つ意味は、その言葉の50音順の見出しによって説明にたどりつき、それがどのような事象を表示するのかを理解することができるが、その辞典を利用して逆に、ある特定の事象がどのような言葉で表現されているかを調べることは、至難のことと言える。それは、編集の観点がほとんど個定化されているからである。

つまり編集の観点が固定化されている従来の辞典、データブック類は、きめられた観点からのみ情報をとり出せるが、現在のデータベースは、観点のとり方に非常に大きい自由度があって、極言すれば、どのような観点からでも取り出せるというのが、きわだった特徴と言えることができる。その意味では、やわらかい辞書、やわらかいデータブックと言うこともできよう。この特徴が、複雑なシステムを支える図書館の機能としては、重要な要件でもある。

3. データベース化技術の誕生

いま情報産業の進展度の最も高い国は、やはり米国である。しかし、現在のようなオンライン情報提供の時代に至るには、特殊な事情があった。それは、宇宙開発競争において、ソビエトに一時期、先をこされたということである。

このことは上下院を含めて国会レベルでの検討にまで及び、別の観点からみれば、産軍複合体のいろいろな意味での典型と見られているNASA

の誕生にまで至っている。このNASAは、宇宙開発において、人類の知的資源を総動員してでもその目的を達するという使命から、莫大な量の情報を集積したとも言えることができる。

しかし、情報は、ただそこにあるということだけでは、何の効用も果たすことはできない。必要が起こったとき、即座に、かつ的確にその情報を取り出して利用できることが、最も重要なことである。

このような事情を打開するため、大量の文献情報を管理し、利用するための情報検索システムの構築を、航空機産業の大手であるロッキード社に発注した。そうして出来たものが、「DIALOG」と呼ばれている情報検索システムである。

1985年ロンドンで開催された国際オンライン情報会議においても、各国の情報機器およびネットワークの説明展示ブースが並ぶ中で、最大のスペースを占有していたのは、このDIALOGであり、今では、社会主義国家群を除くあらゆる国に対してネットワークを形成している、世界最大のデータベース・ディストリビュータとなっているが、これももとはと言えば、ロッキード社が、NASAに納入した後、商用システム用に改善したものを出発点としている。

ここでディストリビュータというのは、データベースの内容自体を作成することはしないが、内容を構成するデータレコードから検索観点の柔軟性をもったファイル化と、通信回線を通じた検索需要者との接点部分をシステム化することによって情報を提供する業者を指している。

現在DIALOGは、200種類近くのデータベースを保有しており、それらの文献データベースは、自然科学から社会科学、人文科学にまで及んでいる。余談であるが、昨年、米国の新聞大手のナイト・リッターが、DIALOGをロッキード社から買収している。

4. データベースの編成

ここで視点を变えて、データベース編成作業の

流れを簡単にたどって見る。

もともと、人類の知的資産をこれからの人間生活の改善や福祉に、可能なかぎり有効に利用しようとする考えが、学術上の研究成果をデータベース化しようという動きに結びつくのは、当然と言えるほど直結しやすいルートではある。しかし、何もない所からいきなり学術文献データベースを作ろうとする動きには、非常に大きい難関が待ちうけていることも確かである。

現在、比較的ポピュラーになっている著名文献データベースはほとんど、データベース化の前段階として、学術文献抄録誌の刊行という、冊子形態による二次情報の刊行を行っていたものが、大部分を占めている。

これは、学術論文を読みかつ理解して概要を作成し、データベース化のための必須レコードを可能な限り客観的な基準に基づき抽出するという作業が、まだ機械化できていないため、そのほとんどを人力によって処理するという大変な部分を残しているからである。

ちなみに、日本国内で最大の国産文献データベースを作成している日本科学技術情報センターにおいても、正職員以外に、抄録作成協力者として、大学の教員および国公立の試験研究機関の研究員を中心とした外部協力者を約 6000 人組織化して、その事業の運営を支えている。

しかもこれらの事業全般を運営するために年間約 120 億円を費消しており、そのうちの過半の部分は、原論文の掲載されている学術雑誌を、世界的規模で収集することから始まり、抄録作成と機械可読のための必須レコードの抽出という、いわゆる機械化できていない部分によって費消されていることから、学術文献データベースを作成するためにいかに多くの経費を要するかが理解できるであろう。

このことは、世界における著名な学術文献データベースの作成が、ほとんどと言ってよいくらい、国立機関かまたは、国の援助資金を多額に得られる機関によって運営されていることから、理解できるであろう。

(出典：通産省，データベース台帳より)

情報資源のうち、知的資産の性格の強い部分については、ビジネス分野および自然科学・技術分野を中心として展開されている現況が良く理解できる。

図1 分野別データベースの分布 (I)

またこれだけの経費をかけて作成したデータベースの利用に基づく収入は、全経費の 50 % に達していないのが実情である。

しかし、このことは、逆に見てみると、機械入力までの段階を国の社会資本充実の出資対象として処理すれば、残りの部分は、営利事業として成立するかもしれない、という考え方が出てくる理由の根元となっているとも言えることができる。

5. 知的資源と事実資源

データベース化によって、より高度の価値を生成できる対象情報というのは、必ずしも学術諸分野における文献情報のみに限られるわけではない。学術成果として生成する数値データもまた、貴重な対象情報となり得るし、すでに現在、先進的情報処理機関においては、物質の物性に関する数値データあるいは、スペクトルデータなどのファクトデータベースを編成しており、すでに提供

図2 分野別データベースの分布(Ⅱ)

サービスを開始したものもいくつか存在する。

ただ、数値によるファクトデータベースの社会的効用は、社会科学分野における利用度の方が大きいように思われる。例えば、アンケート調査や社会事象を表示する数値データは、すでに冊子体でも刊行されており、社会的意志の決定や行政施策のシミュレーション、あるいは諸種の調査や予測にとっては必須のデータであり、その利用度は、第一段階としては行政側からの需要が大きく高まって、成長して行くことが考えられる。

自然科学分野におけるファクトデータベースの利用は、現在着実に進展を見せている。今後、いわゆる材料革命と言われる分野の主流動向として発展して行くことになるのではなかろうか。

つまり、これまでは人間が、地球上において自然な形で存在し得た物質の性格を認識し、それらを人間の利用の条件の要素と比較して、都合のよいもののみを材料として採択するという流れ（採択されることによって当該物質は、あらためて資源として認知されるわけであるが）であった。

しかし、人間の利用に要求される物性要件を満たした新物質（自然の形では地球上に存在し得ない物質）を、人間の手による分子設計によって創造するという時代が、すでに非常に近づいてきている。そのための基盤として、既存の物質の物性

その他に関するデータは、十分にユーティリティの高い形でデータベース化されて置くことが、前提となるからである。

このような時期は、材料開発という分野が地球上の制約を離れる、いわゆる離陸期を迎えはじめていることを表しており、その意味では、社会活動の中の経済的効果を持ったイノベーションは、ここ当分とだえることなく続きそうである。

ファクトデータベース、つまりあらゆる現象・事象の事実に関する情報というのは、ビジネス世界においては、もっと有力な位置にあることも見逃すわけにはいかない。

冒頭にあげた JR 緑の窓口のシステム事例において、あのシステムを支えているデータベースは、大量にある販売可能な座席のどれが売れており、どれが残っているかという事実に関するもので構成されている。

また社会生活においてもはや否定することができない程、確固たる位置を占めている銀行オンラインシステムにおいても、これを支えているデータベースは、個々人の持っている預金金額に関する残高という事実、言い換えれば、預金通帳という“もの”がすべてデータベース化されていて、システム全体を支えているわけである。

このようなデータは、個々単一に評価すれば価

値自体もあやしくなりそうな事実というのも、網羅的にデータベース化されることによって大変な効用を生みだす、情報事例であろう。

したがって、情報資源というのは、一方の極点を学術的資源として、またもう一方の極点を単純な事実資源として持っている。そしてその中間に位置する、事実資源と知的資源との微妙なバランスを取る情報資源は、無数に存在すると考えなければならぬ。このことは、これからデータベース化すべき対象資源が、際限もなく存在し続けることを暗示するものでもある。

またここでもう一つ見逃すことのできない点は、これまで述べたように、知的資源には、機械可読の形にして入力する以前に、莫大な先行投資部分が存在し、産業として成立するための大きなブレーキとなっていることに対して、緑の窓口の場合も銀行オンラインの場合も、データベース化すべき情報は、その利用者が無料で提供してくれるということである。銀行オンラインの場合は、キャッシュディスペンサーの設置してある場所で利用者自らキーボードをたたくという、言うならば入力費用まで利用者側の無料サービスで支えられているわけである。つまり最もシステム効率の良い対象情報であると言うことができよう。

このことは、これからデータベース化の対象情報となるもののうち、事実資源の性格の割合が強ければ強いほど、入力以前の事前処理のコストが無限に0に近づくという性質を持っていることを、例証していると言うことができる。つまり、企業として成立させやすいものと言い換えることもできよう。対象情報の持っている性格の差異が、その利用システムを設計しようとするとき、際立った性格差をもたらす点は、以上のようなことである。

6. 対象情報の表示媒体

一方、検索するためのシステムを考える場合、これらの対象情報が画像によって表示されるか、記号によって表示されるか、あるいは音によって

表示されるのかによって、システムの作り方は大きく変わってくる。

表示方式というのも、画像→記号→音→画像というように、一つの表示パターンサイクルを形成しており、同時にこの3種だけでなく、それぞれ隣接する表示方式の中間帯というものが考えられる。現在までのところ、記号（文字および数値）中心の情報に対応する部分が、いちばん開発が進んでいるということが出来る。画像中心とする部分については、まだまだ未開発と言っても良いくらいであり、人間本来のコミュニケーション媒体として最も有力な部分である所については、これからの可能性を含めて、いちばん楽しい部分が残っていると言えよう。

このように対象情報の表示媒体の差は、検索システムのデザイン自体に大きく影響を与える。それは、検索の手がかりとなるものをどのように決定するかということにかかっている。

したがって、データベース化すべき情報資源の沃野というのは、縦欄を、知的資源から事実資源にいたる軸とし、横欄を表示パターンとする表が構成でき、その交差する枠の数だけの種類が予測可能であるし、まだまだ開発すべき部分が数多く残っていることがわかる。

7. 利用の多様化、高度化

データベースのオンラインによる提供サービス開始の初期、短期間ではあったが、出力端末専用機の時期というのがあった。その当時出回りはじめたばかりのパーソナルコンピュータのように、何らかのインテリジェンシーを一切持ち合わせないで、ただ、検索用データの入出力だけを受け持つという機器である。しかも、インテリジェンシーを備えたパーソナルコンピュータよりも、はるかに高い価格であった。したがって当時のオンラインデータベースディストリビュータは、その利用者層を開拓するために、創草期の難渋度を幾重にも経験しているのである。

もちろん、パーソナルコンピュータに比較して

その機能と価格のバランスが著しく破格の商品である端末専用機が消えて行くのに、それほど時間はかかっていない。そのライフサイクルは、せいぜい5年であったと記憶している。

その頃、8ビットCPUにOSとBASICを複合したソフトとして初めからロードしたパーソナルコンピュータは、以後目覚ましい機能拡充を遂げ、たちまちのうちに16ビットCPUの時代も過ぎて、現在では32ビットの時代に突入してしまっている。

しかも、これらの個人使用には過剰とも思える能力の豊かさは、いわゆるパソコン通信時代なるものも造成してしまった。

このような社会環境の変化は、データベースディストリビュータから見れば、サービス対象となる個々の潜在需要者自らのコスト負担で、整除されたマーケットに変化して行くことに、つながってくる。ディストリビュータ側から見れば、まさに“天の恵み”とさえ言える程の環境変化であると言うことができよう。

しかし、このようなパソコン通信によるネットワーク化社会の急速な成熟は、データベースディストリビュータの経営基盤を確立するために好都合な変化であることも確かであるが、これらに附随して、新たな問題提起をする契機にもなっている。それは利用者の側からあらためて「ダウンローディング」、「アップローディング」、「ゲートウェイ機能」などを求める度合が急速に強まって来たことである。このことは、利用者側のネットワーク化の成熟に伴って、データベースへの要求が極端に多様化し始めることの徴向であるとも言えることができる。

ここで「ダウンローディング」と言うのは、利用者が自分の端末であるパーソナルコンピュータを通じてデータベースにアクセスした時、その検索結果の出力を、紙に印字させず、機械可読の状態のまま自分の記憶用磁気ディスクに書き込んでしまうことである。

このような出力をして置くと、後で再度CRTディスプレイに再生することも、また紙に印字す

ることも可能であり、さらに自分個人用のデータベースを構築する時には、それらを構成レコードとして再使用することも可能である。それだけではなく、当該出力データを他の解析用ソフトを利用して、より高度のデータを得るために解析・加工するといった処理も容易にできるというわけである。

このような要求は、知的資源を増殖させる。つまり研究活動の活性化の観点からは、大いに推進させたい項目である。しかし、一方で、それでもなくとも大きい負荷を負わされているデータベース製作者側の権益と、どこでバランスさせるかという新しい問題を提起している。

つぎに、アップローディングと言われるものの実態は、パソコンネットを構成する小グループの共有資産として、グループ共有のデータベースを、ディストリビュータ側の高機能な検索システムを利用するために、ディストリビュータ側のホストメモリーに間借りさせろという要求である。利用者側は、これによって自分達だけで作成する検索システムよりは格段の高機能を獲得でき、はるかに便利となることは確かである。

また、ゲートウェイ機能というのは、あるオンラインシステムに入って検索した結果が不満足であった場合、そこから別系列のオンラインシステムに自由に入出力できる機能を持つという要求である。

このように、これから起こってくる多様化ニーズというのは、いずれも、データベース製作者、ディストリビュータ、利用者のそれぞれの間で利害の対立するものばかりが増加してくるになり、これらの整除には、別種の社会的エネルギーを必要とすることになるだろう。

以上のようなさまざまな問題群とはやや異質な問題にも目を向けなければならない。その一つは、学術文献情報のデータベース利用が高度化すればするほど、一次情報の社会的地位よりも、二次情報の社会的地位の方が、上位に来るようになることである。

このことは、ダウンローディングの要求に見ら

表1 国産・外国製データベースの分野別オンライン化状況

(オンライン化の数/データベース総数)

	国 産		海 外 製		合 計	
1. 一般	34/56	(60.7) [%]	212/212	(100.0) [%]	246/268	(91.8) [%]
2. 自然科学・技術	55/67	(82.1)	566/568	(99.6)	621/635	(97.8)
3. 社会科学・人文科学	0/3	(0.0)	136/136	(100.0)	136/139	(97.8)
4. ビジネス	86/154	(55.8)	471/476	(98.9)	557/630	(88.4)
5. その他	4/21	(19.0)	9/9	(100.0)	13/30	(43.3)
合 計	179/301	(59.5)	1,394/1,401	(99.5)	1,573/1,702	(92.4)

いずれも、これらのデータベースにはファクトデータベースが含まれており、ビジネス分野においては、その数もかなり多いはずである。なお、社会科学・人文科学分野については、国産データベースが極端に少なく、しかもオンライン化数が0である点は、情報サイクルの国際化を考えるうえで、大変な欠陥である。

(出典：通産省、データベース台帳より)

れるように、情報生産者である研究者はデータベース化されていない情報を、次第に無視して行くようになるからである。

そうやって来ると、著名データベース編成者の意向によって入力対象論文となるかどうか、生まれても死滅を待つばかりの情報になるか、多くの人に読まれて、次の再生産サイクルの中に入る情報となるかを決めかねない事態を生み、情報の生産現場に歪みを生じさせる要因となる恐れが、十分にある。

このように情報生産者の社会的位置と、データベース編成者の社会的位置の相互関係に変化が起これば、少なくとも学術文献情報をデータベース化する事前処理の部分は、次第に生産者側に押しつけられるようになることは確かであろう。

8. 産業的観点からの展望

以上のようにデータベースをめぐる動きを見ると、この分野における企業化の動向が、その立ち上がり期にあるという風に言うことができる。したがって、今後ますます豊かな進展を約束

されている分野であるように思える。

ちなみに、物材の生産部門に眼を向けてみても、人間の豊かな生活を支えるうえで、物材の消費が減少することなど考えられない。しかも一方で、情報を扱う分野での人材要求増大傾向を考えると、より一層の生産性向上が求められる。また、消費側の需要も今後ますます多様化することを考えれば、これまでのような量産による生産性の向上では、それらの要求に耐えられず、衰退の憂き目を見ることになるだろう。

となれば、生産ラインの全分野を総合的な単一システムで包括することが必要となり、そのシステムの多様化需要に対する柔軟性を保持するためにも、中小規模のデータベースを無数にシステム内に内包することが求められる。

ここで求められるデータベース化の対象情報は、どちらかと言えば、単純な事実資源の様相の濃いものであるから、その効率から考えて容易に着手しやすいものと言える。恐らくこの分野での進展は、予想以上に速いものと考えた方が無難である。

一方、データベース化のためには、いくつかの

難点をもっている知的資源の分野を見ても、入力以前の処理つまり、最も労働集約的部分の機械化というの、全く不可能というわけではない。ここでは、これまで集積されて来た自然言語処理に関する知的資産の総動員を前提とすることは当然であるが、これらを可能とする社会環境の整備も、着々と進行していると言うことができよう。

それは、一次情報の冊子体刊行の段階での変化を指している。いま、冊子体作製の工程においては、何らかの形でワードプロセッサ的機能を持ったシステムを利用することによって、その編集および印刷版下作製工程の効率化を図っているはずである。ということは、一次情報が社会的存在となるいちばん最初の段階で、機械可読の状態になっているということである。

これを処理対象素材として、内容分析を行うことは、それほど困難なことではない。しかし現実には、機械可読化のためのソフトウェアの差異などによって、現在の段階では、いくつかの難点があり、内容分析用のソフトウェアについても、いま既に実用的レベルにあるかという点では、多くの疑問が残る。

しかし、これらの難点が克服されるのも、それほど先の話であるとは考えられない。となれば、この分野のデータベース化事業というの、企業として成立するための基盤は、いずれ整備完了という時期に至ると考えられる。

9. お わ り に

最近の話題の中に、出版物のコピーに対する防護措置の一つとして、複写機械の利用度数に応じ

たコピーライト料を徴収しようという動きがある。このこと自体の可否はともかくとして、このような措置を行うことにより、従来、著作権のみが表面に出て、とかく出版権というものの定義があいまいにされていたのが、表面にはっきり出てくるという効果があり、その意味では前進と考えられる。

しかし、出版権のみを明確化し、ある意味で強化されるのであれば、その反対給付として、公的機関に機械可読の形態で納入し、出版権登録に代えるという義務を負わせた方が、はるかに好ましい措置になると思われる。そうなれば、前記の問題の解決はより加速化された変化となって、効果はより大きくなる。

一方、個人レベルでの私的データベース作成という動きも、今後ますます強化されてくるように思われる。この動きは現在の段階でさえも、学術的資産の生産に従事する人達だけでなく、広く情報生産者の持つべき要件となってきており、恐らく、情報生産者以外の一般層にも広がって行くであろう。

ちなみに、パソコンソフトの中で、データベース作成のためのソフトウェア DBMS はかなり大きい需要者を持っており、初期のものと比較すれば、最近は随分レベルの高いものも出はじめている。

このように見てくると、情報の生産、加工（データベース編成者）、ディストリビュータ、情報消費者のそれぞれをめぐる社会環境は、今後、急速に変転を見せるものと考えられる。そろそろ、来たるべき社会を想定したアセスメントに力をそぐべき時期ではないかと考える。

参 照 文 献

- 1) 未来工学研究所，科学技術情報の流通促進のための進展する情報技術の活用方向に関する調査研究報告書—科学技術情報の流通システムの将来動向調査及び総合的分析—，東京，昭和 61 年，232 p.
- 2) 日本科学技術情報センター，科学技術情報ハンドブック，東京，昭和 61 年，428 p.
- 3) データベース振興センター，データベース白書 1987，東京，昭和 62 年，592 p.