

## 解説



## フルテキスト・データベースの応用動向†

根岸 正 光††

## 1. はじめに

いわゆる商用データベース・サービス，すなわちオンライン情報検索システムの上に各種のデータベースを乗せて，一般からの検索利用に供する事業は，1972年の米国のDialogとOrbitがその初めとされる。これらは当初科学技術系の文献抄録データベースを対象にするものであったが，翌1973年には米国の判例の全文をおさめたLEXISもサービスを始めているから，フルテキスト（全文）データベースのサービスも古い歴史をもつといえる。その後出版界では電算写植（CTS）による印刷が普及し，これにともなって，大規模な抄録誌だけでなく，一般の図書，雑誌の類についても，その写植用ファイルから全文データベースを作成して，オンライン検索サービスを行う例が増加している。現在ではたとえばデータベース・サービス最大手のDialogでは，その380種のデータベース中の85件ほどが全文データベースとなり，これらに含まれる雑誌，新聞などは1,300種に至っている<sup>1)</sup>。本稿では，昨今成長の著しい全文データベースについて，関係機関の実地訪問調査による知見を軸として，実用事例を整理し，今後の動向と問題を検討する。

## 2. データベース・サービスにおける

## フルテキスト・データベース

商用データベース・サービスの対象になるデータベースには各種のものがあるが，一般には，そのデータ内容によって2次情報データベースと1次情報データベースに大別される（米欧の商用データベースの名鑑であるCuadra/Elsevier，

“Directory of online databases”ではおのおのReference databases, Source databases といっている<sup>2)</sup>。これは，データベース・サービスが，2次情報である文献目録や文献抄録を中心として発展してきたという経緯によると思われる，それ以外のものをことさら1次情報データベースと称している。文献抄録データベースでは，検索の結果をみた上で，図書館などを介して現物文献を取り寄せる作業がさらに必要になる。1次情報データベースは，必要なデータそのものが収容されていて，検索の結果としてこれを直接入手できるという点特徴で，統計や測定値などの数値データあるいは画像データ，またプログラムを収容したデータベースはこれに該当する。文献データベースの場合も，その要旨だけでなく，本文の全文をも含むデータベースは1次情報データベースとされる。

オンライン・データベース・サービスにおける全文データベースの成長はここ数年めざましく，前記Cuadra/Elsevierの名鑑で集計すると，1980年版では，収録された500件のデータベースのうち25件(5%)が全文データベースであったが，その後，1984年版では2,453件中の422件(18%)，1987年版では3,369件中の842件(25%)，1989年版では4,062件中の1,381件(34%)というように，絶対数だけでなく比率も増加している<sup>3)</sup>。この名鑑の1990年版は4,615件のデータベースを掲載しているから，そのうちの全文データベースは1,600件を超えているものと思われる<sup>2)</sup>。全文データベースは複数の雑誌，新聞をまとめて収めたものが多く，こうした個々の雑誌，新聞別に整理した全文データベース専門の名鑑も出版されており，この1990年版には3,100誌紙が掲載されている<sup>4)</sup>。

なお，データベースに関しては，昨今は戦略的情報システムとも関連して議論の盛んな，機関内で形成・運用される大規模なデータベース，すな

† Recent Development in Full Text Database Applications by Masamitsu NEGISHI (National Center for Science Information Systems).

†† 学術情報センター

わちインハウス・データベースや、事務文書のデータベース、予約・発注機能に連動するトランザクション・サービスのためのデータベースなどにも全文型のものがありうる。ただし、本稿ではもっぱら公開商用のオンライン全文データベースと、ポータブル・データベースとかパッケージ・データベースとも称される CD-ROM 版の全文データベースについて検討する。

### 3. フルテキスト・データベースの類型

これまで、全文データベースとして商用サービスが行われているものは、図書、論文、判例、新聞記事、特許などである。以下これらを事例に即して概観する。

#### (1) 判例データベース

全文データベース・サービスとして最も歴史の古いものは、米国における判例データベースである。これは判例法を主体とし、また弁護士社会ともいえる米国の固有の事情によるところが大きく、Mead Data Central 社の LEXIS (1973 年～) と、West Publishing 社の Westlaw の 2 大サービスが展開されており、これらは米国の弁護士事務所のいわば必需品となっている<sup>5)</sup>。内容的には、判決全文そのままをデータベースに流し込み (Westlaw では、独自の「Shepherd 索引」と称する判例引用索引を付加している)、その全単語から機械的に索引を抽出し、後述の BRS などと同様に、語間距離を指定するいわゆる隣接演算などを用いて所望の判例を検索する。データベースの更新処理面では、全米各地の裁判所で出される判決を一刻も早く弁護士に提供する必要があり、新聞記事データベースと同様の状況にある。各地に設置された支所で、その地域の裁判所の判決文を取り寄せ、これを OCR などを入力・送信して、判決の数日後までにはデータベース化する。

我が国は実定法国で、判例への需要がさほど高くなく、特定の分野についての判決の要旨を入力したデータベースのサービスが行われる程度であったが<sup>6)</sup>、最近判決全文を入力したサービスも行われるようになった。

#### (2) 新聞記事データベース

新聞の全文データベースは、新聞の入稿、編集、版下出力という一連の工程の電算化、CTS 化にともなって集積される記事ファイルを、別途

データベースとして整備し、独立した商品に仕立てたものである。前記 Mead 社の NEXIS と称するデータベース群の中で提供される New York Times が有名であるが、我が国でも各全国紙はすでにオンライン・サービスされている。データベース化に際して、あらためて検索用のキーワード、分類などを付与する場合もあるが、ともあれ検索手法的には従来からの文献抄録型データベースと同様に考えてよい。これは、「全文」とはいふものの、一本々々の新聞記事は、分量的にもまた内容的にも、もともと抄録的であるからである。全文であることの効用は、たとえば個人名を指定してその人物に関するすべての記事を網羅的検索するというように、特定性の強い語で検索する場合に著しい。

なお、通信社が常時配信するニュース、外電などを収容したものも新聞記事データベースの一種である。この種データベースでは、オンライン的無停止更新という運用方式が特徴的で、電話回線を通じて配信されてくる記事を自動的にデータベースに追加して、24 時間無停止サービスを行う。このためには専用のソフトウェアが必要で、米国では、ミニコン上にこの種システムを搭載して、全世界にサービスする NewsNet の例がある<sup>7)</sup>。

#### (3) 完結テキスト・データベース

主として文学系統で、古典作品のテキスト全文をコンピュータ・ファイル化したものを、とくに「テキスト・データベース」と称している。我が国では、樋口教授(九州大)によるトーマス・マン全集全巻を入力したものを初めとして、欧文、和文の数 10 点が研究者自身により作成されている。この種のデータベースでは更新のない完結性が特徴の一つである。また利用上でも、語法の統計的研究など、要するにテキスト全文を走査しながら、特定の文字パターンを検出するといった体裁のものであるので、一般の文献データベースとはおよそ様子が異なる。この種の処理のためのソフトウェアとしては、オックスフォード大学で開発された OCP (Oxford Concordance Program) が有名で、昨今これがパソコン上でも動くようになり、欧文系の研究に活用されている<sup>8)</sup>。また、九州大の SIGMA システムもこの分野でよく使われており、文学や哲学などにおけるまったく新たな研究方法を提供している<sup>9)</sup>。

(4) 新聞・雑誌版面イメージ・データベースすでに出版された雑誌や新聞などをデータベースに仕立て直す最も簡便な方法は、それらの版面、ページ・イメージをそのまま画像データとして電算ファイルにすることである。この際、見出し、日付など、最小限のデータを別途入力して索引ファイルを作り、これを手掛かりにして、所望の文書を探し、表示させる。OA 機器の一種として開発された光ディスク・ファイリング装置はこの手法によるものであるが、これは省スペース効果を主眼にしており、検索機能は限定されるので、データベース的ではない。

版面イメージ方式のオンライン全文データベース・サービスは、版面の配信にファックスを用いるため、ファックスの普及が著しい我が国独自のサービス形態といえる。エレクトロニック・ライブラリ社による ELNET は、昔からの新聞記事切抜きサービスをオンライン化したものとみることが出来る。1988年1月以降、全国紙、地方紙、専門紙の30数紙(ほかに雑誌140誌ほど)から毎日主要記事を切り抜いて、画像データとしてファイル化すると同時に、それぞれに検索用のキーワードを付与し、これを索引ファイルに入力する。利用者は索引ファイルをオンラインで検索して必要な記事を選定し、記事本体をファックスで受信する。

学術情報センタでは、この手法による全文データベースのサービスを、「学会予稿集電子ファイル」という呼称で1991年から開始する。学会の年次大会などの予稿集のページ・イメージを光ディスクに焼き込む一方、各論文について著者名、標題、その他書誌事項をタイプ入力して索引ファイルとする。なお、これは、著者に抄録作成を依頼し、これを入力するという通常型の文献抄録データベースである「学会発表データベース」の簡易版という位置付けにある。

#### (5) BRS の学術論文全文データベース

米国 BRS Information Technologies 社では、1981年から米国化学会との共同で学術論文の全文(図表含まず)データベース・システムを開発し、1983年からこれを一般に提供している。その後米国化学会は独自にサービスを行うようになり、BRSは、現在では CCML (Comprehensive Core Medical Library) と称する医学雑誌百数十誌の

論文全文データベースのサービスで有名である。

BRS では、雑誌出版社から電算写植用磁気テープの供給を受け、これから写植機能コード類を除去すると同時に、著者、標題などのデータ項目種別を判別して、データベース向きに再編集を施すプログラムを、欧米で普及している写植機、数機種に対応して開発した。これにより、写植ファイルからデータベースへの変換がほぼ自動的にできるようになっているとのことである。

検索システムは、基本的に隣接演算によるものであって、「単語 A と単語 B が何語以内に連続して現れる文献を検索せよ」といった指令を組み合わせる検索を行う。こうして検索した結果の表示の際、該当(ヒット)語の出現頻度を表示でき、また該当語の前後のみを抽出して表示できる。こうした機能により、少ない出力文字数で検索結果の当否の判断が行われるようにしている<sup>10)</sup>。

また、BRS では従来の文献抄録データベースに、全文データベースを連結して検索できるシステムを開発している。すなわち、医学文献抄録データベース MEDLINE の検索において、該当文献の抄録を表示させたとき、CCML にその文献の全文が収録されている場合にはその旨表示され、コマンドによりその全文を呼び出すことができる。つまりここでは2次情報から1次情報に連続的にアクセスするという形態が実現されている<sup>11)</sup>。

なお、CCML の普及には、米国における医療過誤訴訟に対する医師側の防衛措置的な側面が強い。すなわち、医師は CCML を利用して、日頃から治療法の進歩をウォッチしており、この面で過失はないと主張しようというわけである。判例データベースと同様に、ここにも米国の社会的な特殊事情が反映されている。こうしたエンド・ユーザ(医師)による直接的検索需要に対応して、BRS は1985年から COLLEAGUE と称するメニュー型の検索方式も用意しているが、機能的にはコマンド方式と同一のものである<sup>12)</sup>。

#### (6) CAS における雑誌と全文データベースの一貫生産

米国化学会傘下の CAS (Chemical Abstracts Service) では、同学会発行の論文誌19誌を中心に (CJACS: Chemical Journals of the American Chemical Society と称する)、さらに英国王立化

学会発行誌などを加えた約 35 誌を対象として、全文データベース（図表含まず）のオンライン・サービスを 1985 年から提供している(CJO:Chemical Journals Online)<sup>13),14)</sup>。

CAS は雑誌の版元であるから、全文データベースは、雑誌の編集・印刷の CTS 化と一体化して運用されている<sup>15)</sup>。すなわち、(1)受け付けられた原稿は電算入力され、(2)棒組み状態で著者校正に回される（著者は仕上がりページ・イメージを確認できない）、(3)図表枠を空けた形で CTS 出力を得る、(4)図表は別組みのものを貼り込んで、(5)印刷する、という工程である。

ここでは、そもそも(1)の工程で、著者、標題などのタグ付け（データ項目の区別がコンピュータで容易にできるように設計された特殊な記号、データ項目識別子）が行われるので、最初から全文データベースになっているという体裁である。

このようなデータベースと雑誌の一貫製作は、経費的にみると、雑誌の編集・印刷に通常かかる程度の費用で、雑誌と同時に全文データベースも作成されるという点で有利であるとのことである。BRS や後述の学術情報センタでの事例のように、雑誌印刷工程の中間生成物である写植ファイルを、全文データベース用に改めて加工する方式では、どうしてもデータベース化のための追加費用を必要とする。この点で、今後は CAS のような当初からデータベースの作成を折り込んだ、データベースと雑誌の一貫作成システムが望まれる。CAS のシステムはこの種システムとして先駆的なものであるが、その後こうしたシステムへの認識が高まり、のちにみるように SGML という全文データベース向きの国際規格が制定されるに至った。CAS はこの規格制定に積極的に関与しており、早晚 SGML 対応のシステムに移行することになろう。

CJO の検索システムは、BRS とおおむね同等の機能のものであるが、各単語について文字回転索引（“INFORMATION”, “NFORMATION/I”, “FORMATION/IN”, …）を生成して、単語の中間部分でも検索できるようにしているのが特徴である。

(7) 学術情報センタの学術論文全文データベース

学術情報センタでは、1987 年から学会誌の全文

データベース化について検討を開始し、1989 年 9 月から化学系学会誌のサービスを開始した。これを「学術論文データベース」と呼称し、1991 年度当初には化学系と電子系の 5 学会 6 誌を収録している。

データベースの作成方式は BRS と基本的に同様で、学会側から雑誌印刷用の CTS 磁気テープの供給を受け、これから写植機能コード類を削除すると同時に、データ項目を区分け、整理してデータベースとする。図表の類および複雑な数式、化合物式は画像として原誌から光ディスクに焼き込み、ファックスで配信する。検索は、文献抄録向きソフトウェア製品に手を加えて全文検索に対処している。図表の説明文はデータベースのほうにも収容されるので、利用者はこれにより図の要否を判別し、図番と電話番号により出力要求を出せば、その図表がファックスで送られてくる。

写植ファイルの変換は、我が国では和洋を含めて写植機の機種が多く、BRS のように機械的変換では対応しきれず、人手による編集を併用している。ところで、一般に雑誌の最終校は印刷期限に迫られるため、写植ファイル本体で修正せずに、印画紙上での貼込みで処置されることが多く、ファイルと雑誌上の文面とは不一致が生じる。この点は、印刷所と協議して、最終校正分のファイルへの反映を別途後追的に行ってもらおうにしている<sup>16)</sup>。

雑誌はおのおの独自の論文スタイルをもっているが、データベース化の際に、これらを統一的に再編集する必要がある。これには、(1)著者姓名の逆転などの正規化、(2)著者と所属機関の対応、(3)特殊記号・記法の読み下し、(4)数式、化学構造式などの図番起こしと挿入位置指示、(5)脚注の付番などのように、機械的に難しいだけでなく、人手による場合にも、作業者に一定の知識水準が要求される作業が含まれる。

#### (8) 全文データベースの種類

上記の事例を踏まえて、全文データベースの作成経路を整理すると図-1 のようになる。そして、本文や図表の保持方法から類型化すると、次のようになる。一般のオンライン・データベース・サービスでは、「全文」データベースが増加しているが、図表、写真は収容されず、本文のテキストをコード化データとして収容するのみである。

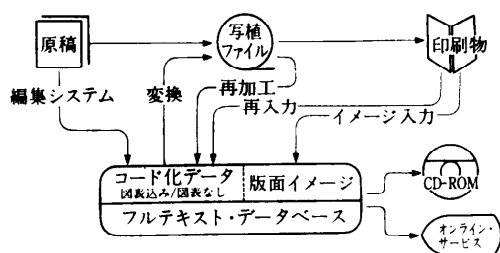


図-1 フルテキスト・データベースの作成と応用の経路

一方、既存印刷物の版面をそのままイメージ・データとして蓄積する方法のものもあるが、これは本文に対する検索ができない。つまり、全文データベースには、(1)図表なし全文データベース、(2)図表込み全文データベース、(3)版面イメージ・データベース、の3種があることになる。こうした状況を念頭に、以下、全文データベースの作成・検索方式について検討する。

#### 4. フルテキスト・データベースの作成方式

##### (1) 全文データベースの作成

全文データベースを作成するには、これに対応する印刷物との関係で、(1)既成の印刷物を改めてタイプ入力、OCR 入力する方法、(2)印刷物の版面を画像としてスキャナ入力する方法、(3)印刷に用いた写植ファイルをデータベースへ変換する方法、(4)まず全文データベースを作って、これから印刷物も作成する、すなわちデータベースと印刷物の一貫作成の方法がある。

まず、(1)はすでにみたように、判例など既定の印刷物で、データベース作成者がその印刷工程に介入できない場合や、過去の印刷物をさかのぼってデータベース化の際に採られる方法である。(2)は版面イメージ・データベースを作成する場合である。このとき検索に必要な項目は別途タイプ入力する必要がある。(3)は BRS や学術情報センタの例にみるとおり、全文データベース作成にあたって、現在多く採用されている方法である。

(4)は今後発展が期待される方式であるが、現状では、CAS や新聞各紙など、おのおの独自のシステムによっている。データベースと印刷物の一貫作成は、文献抄録データベースでは普通のことである。元来、大部の抄録誌の編集・印刷を効率化するために電算機を導入したことから、自然

にデータベースも作成されるようになった(現今では抄録誌よりもデータベースのほうに販売収入の重点が移っている)。新聞や抄録誌のように、記事の入稿から編集にわたって独自の電算化システムを開発するなら、印刷物とデータベースの同時作成が可能になるのは当然である。編集作業の効率化を主たる動機として電算システムが導入される例は辞書類にもみられ、その CD-ROM 版全文データベースが出版されている。

一方、一般の図書・雑誌についてみると、その編集段階での電算化要求は概して低く、単に末端の印刷工程が電算写植に替わってきているに過ぎない。これら一般の図書雑誌における全文データベース化の進展には、電算化編集システムが安く導入でき、かつこれにより編集経費が低下すること、印刷物以外の媒体での出版からも相当の収入が期待されるという、費用と収入両面からの動機付けを必要とする。前者のためには、標準規格が浸透し、これに対応する廉価なシステムが出回る必要がある。後者に関しては、オンライン全文データベースよりも CD-ROM での出版が当面関心を集めている。

##### (2) SGML

全文データベースに関する標準規格として SGML (Standard Generalized Markup Language, ISO 8879) の普及が期待される。SGML は、著者、標題、章節、段落等々のデータ区分とその包含関係(これを文書の論理構造と称している)が計算機プログラムで容易に判別できるように、項目ごとにタグを挿入しながら、文書を作成するという方法に関わる ISO 規格である<sup>17)</sup>。SGML は、こうした論理構造に着目する点で、もっぱら印刷様式を関心の対象とする電子出版とは一線を画している。

SGML の先駆的実用事例としては、米国マサチューセッツ医学会がある。同学会は、1989年の時点ですでに SGML 方式による複数メディア出版を始めた。“The AIDS Knowledge Base”と称するエイズ関係の教科書がこれである。この研究分野は現在急速な進展がなされているので、教科書も頻繁な改訂を要する。そこで、数十名の共著者に簡略なタグを折り込んだ原稿をワープロで作ってもらい、そのフロッピー・ディスクを学会で集める。学会では、これをコンピュータに

より本格的な SGML 形式に変換する。これがすなわち、すべての元となる全文データベースである。

このデータベースは、第1に、必要な変換を施して既述の BRS に送られ、そのオンライン・データベースとしてサービスされる。第2には、これから CD-ROM 版が作成され頒布される。第3は通常の図書形態の出版であるが、これも SGML 形式のデータベースに、データ項目別の活字指定などを別途与えることで容易に電算写植ファイルに変換され、実際の印刷が行われる。オンライン・データベースは月次に更新される。CD-ROM は四半期刊の雑誌類似の購読形態で、購読者には3カ月ごとに改版したものが届けられる(前版は不要になる)。図書は本体を年次に改版し、これに対して毎月追補分が配布される。全文データベースを前提とすれば、このように媒体を複線化した出版活動が、媒体ごとに付加的な経費をそれほどかけずに可能となるはずである。

### 5. フルテキスト・データベースの提供方式

全文データベースをどのような形態で利用者に提供するかについては、マサチューセッツ医学会での事例のとおり、現状では、(1)印刷物、(2)CD-ROM、(3)オンライン・データベース・サービスの3方式が考えられる。ここで電子媒体である後二者については、検索方式が利用者にとって大きな関心である。実際、現行の検索システムでは全文データベースの効果的な検索は難しい<sup>18), 19)</sup>。

まず、オンラインでの全文データベース・サービスをみると、全文向きの検索・表示機能としては、BRS や CJO にみられるような「隣接演算」と「ハイライト表示」がある。隣接演算は、文献抄録データベースでも従来から用いられていた機能である。ハイライト表示は、該当文献表示の際、文中の検索対象となった語を星印で囲むなどして明示する機能で、その語がどのような文脈で使われているかを、すばやく点検するのに有効である。

現行の隣接演算には、文書における章→節→段落→文という「構造」に即した検索機能はない。われわれが、印刷文書を通覧しながら、必要個所を探索する場合、キーワードを念頭にして本文中

を平板にながめてゆくのではなく、同時に章・節の見出しとか、図表の説明などにも着目しながら探索している。こうした臨機応変の連続的な過程をそのまま全文データベースの検索にも実現させるのは無理としても、文書の構造を指定した検索指示ができれば、事態はよほど改善されるであろう。たとえば、章題中に「検索」と「データベース」があって、その章の本文中には「全文」が現れるような個所を検索して、その前後数行を表示せよといった指示は、現状のシステムでは基本的に不可能である。この際、文書の構造を図式的に表現するような指示法で構造的検索が可能なシステムが望まれ、今後の開発に期待したい<sup>20)</sup>。

CD-ROM 版の全文データベースでは当面ハイパーテキストの応用があり、たとえば、本文中に挿入されている引用文献番号を指示すると、別のウィンドウに文末の引用文献リスト中の該当文献が表示されるといった体裁である。文献内での相互参照個所がコンピュータ的に連結され、その間を自由に飛び回って表示できることは、ひとつの新しい読書法を提案しているともいえる。しかし、多数の文献の集合体としての全文データベースを考えると、CD-ROM の場合でも上述のような構造的な検索機能が望まれよう。

### 6. フルテキスト・データベース普及・発展への問題点

#### (1) 技術的問題

全文データベースの作成面に関しては、SGML の普及が鍵であると思われる。すでに米国防省は、CALIS (Computer-Aided Acquisition and Logistic Support) 計画の中で SGML 実施規格を制定し、国防省への提出文書は 1990 年以降これに準拠するよう義務付けられた。また、通常の出版物については、米国出版者協会 (AAP) が 1988 年に実施規格を制定しており、これらを受けて SGML 対応のソフト類がずいぶん出回りつつある。欧州でも、EC 出版局における官報への適用や英国政府出版局 (HMSO) での法令データベースへの適用事例がある。

一方、我が国では、SGML の JIS 化は 1991 年には原案がまとまる見込みであるが、その上で実施規格の制定が必要であり、これにはさらに時間を要するとみられる<sup>21)</sup>。この状況と関連して、

SGML 回りのソフトウェアの開発はようやくその緒についたところといえる。ソフト的道具類には、著者個人向き、編集者向き、印刷所向きなどがありうるが、現時点では編集者向きの和文エディタ製品が開発されたという段階である。

学術情報センターでは、1990年度に凸版印刷(株)との共同研究「学術雑誌全文データベースのためのSGML方式による投稿・編集システムの研究」において、先の「学術論文データベース」中の論文(和文と英文)をSGML形式の全文データベースに変換し、これから実験誌を印刷し、またCD-ROMを作成するという実験を行った。この過程での知見は別途まとめて公表する予定であるが、ともあれ、SGML対応のソフト的道具類の開発が急務であることは明らかになっている。

全文データベースの検索利用の面では、先にも述べたような文書の構造に即した検索・表示機能をもったシステムの開発が必要と思われる。これは諸外国においても現在研究中の段階であるから、我が国の貢献が期待される場所である<sup>22)</sup>。

## (2) 社会的問題

全文データベースの普及には、執筆者側での全文データベースに対する意識の高揚が望まれる。しかし現状では、依然として印刷物を唯一の最終成果物とみるのが常識であり、したがって副産物である全文データベースについては出版者側で考えればよいという認識になる。全文データベースに基づいて出版形態が多様化し、それだけ流通性が向上することは、むしろ筆者も望むところであろうが、現状ではそうした効用が実感される段階にないので、著者側の意識変化は容易でない。

マサチューセッツ医学会の例に照らしても、こうした状況の打開は出版者に求めることになると思われる。もっとも、ここで問題は、出版社と印刷所の関係である。米国との対比でみると、米国では大手出版社は印刷部門を抱えているし、マサチューセッツ医学会でもページ・ポジティブまで内作するというように、出版者自身、印刷を含む技術的問題を自己の関心事としている。SGMLの実施規格が出版者の団体であるAAPで制定されたということも、こうした業態の反映であり、さらに、AAPはこの規格の普及促進のためにEPSIG (Electronic Publishing Special Interest Group) を開設している。その発行にかかる

“EPSIG News”は、米国の状況を知るのに便利である<sup>23)</sup>。

我が国の場合、大手の印刷会社が存在しており、印刷やCD-ROMに関する技術開発はこれらに負うところが大きい。通産省の指導のもとに1989年に日本規格協会内に発足した「SGML懇談会」でも印刷会社や端末機器関連会社が大きな役割を担っているようであるから、我が国ではこの方面から展開が図られることになるのであろう。もっとも、SGMLによる全文データベース化、出版形態の多様化という新たな状況は単なる印刷技術の問題にとどまらないから、出版者側の積極的な取り組みに期待するところは大きい。

## 7. おわりに

全文データベースは文献型データベースの発展の方向として今や確定的である。既出 Dialog社の Summit 氏も、今後10年以内に、おもな新聞・雑誌はすべて全文データベースに移行するであろうとの見通しをのべている<sup>24)</sup>。我が国においても、ワープロや電算写植の普及をみれば、基本的にこうした方向にあることは明らかである。もっとも、印刷物と全文データベースの関係は出版者にとって微妙であって、既述のBRSでも、雑誌の売上げが低下するのではないかという出版社側の危惧を反映して、全文データベースのほうは、雑誌よりも2~3カ月遅れで提供されるものが多いとのことである。しかし、一方においてNewsNetに収容されている専門誌の類では、印刷物よりオンライン販売のほうが主力になっているという状況もあるので、全文データベース化は時代の趨勢といつてよかろう。我が国における出版文化の長い歴史を考えると、外国の事例が単純に我が国にもあてはまるとは思われないが、社会全般をおおう情報化の中で、全文データベースについても、我が国独自の発展の仕方が模索されつつあるというのが現状であろう。

## 参 考 文 献

- 1) *Database Catalog 1991*, p. 118, Dialog Information Services, Palo Alto, CA (1991).
- 2) *Directory of Online Databases*, Vol. 11, No. 3, p. 826, Cuadra/Elsevier, New York (1990).
- 3) Tenopir, C. and Jung Soon Ro: *Full Text Databases*, p. 251, Greenwood Press, New York (1990)

- 4) Orenstein, R.M. ed.: *Fulltext Sources Online*, Vol. 2, Issue 2, p. 232, BiblioData, Needham Heights, MA (1990).
- 5) *West's Law Finder—A Legal Research Manual*, West Publishing Co., St. Paul, MN, p. 76 (1988).
- 6) 梅本吉彦: 法律情報検索の現段階, *ジュリスト増刊・ネットワーク社会と法*, pp. 130-135 (1988).
- 7) 米国におけるデータベースの現状と展望, p. 150, (財)データベース振興センター, 東京 (1989).
- 8) 長瀬真理, 西村弘之: コンピュータによる文章解析入門—OCP への招待, p. 210, オーム社, 東京 (1986).
- 9) 有川節夫他: テキストデータベース管理システム SIGMA とその利用, *情報処理学会研究報告*, Vol. 89, No. 66 (89-FI-14), pp. 7.1-7.8 (1989).
- 10) *BRS Full Text Syllabus*, p. 65+99, BRS Information Technologies, Latham, NY.
- 11) Nesbit, K.: *BRS/LINKS to the Future: Online Hypertext Is Born, Online*, Vol. 14, No. 3, pp. 34-36 (1990).
- 12) *BRS Colleague User's Manual*, p. 87, BRS Information Technologies, Latham, NY (1988).
- 13) *The CJACS File Database Description*, p. 66, American Chemical Society, Washington, D.C. (1987).
- 14) 廣田勇二: STN で利用できるフルテキストデータベースについて, *情報処理学会研究報告*, Vol. 89, No. 66 (89-FI-14), pp. 4.1-4.8 (1989).
- 15) Martinsen, D. P., Richard, A. L. and Lorrin, R. G.: *Multiple Use of Primary Full Text Information—A Publisher's Perspective, Online Review*, Vol. 13, No. 2 (Special issue on full text databases), pp. 121-133 (1989).
- 16) Negishi, M.: *Problems in Compiling and Servicing Full Text Databases and Numerical Databases: Experiences and Plans at NACISIS, CODATA Bulletin*, ICSU/CODATA, Paris, Vol. 22, No. 3, pp. 113-122 (1990).
- 17) Bryan, M.: *SGML: An Author's Guide to the Standard Generalized Markup Language*, p. 364, Addison Wesley, Wokingham, U.K. (1988).
- 18) Basch, R.: *The Seven Deadly Sins on Full-text Searching, Database*, Vol. 12, No. 4, pp. 15-23 (1989).
- 19) Pagell, R. A.: *Primary FTDs for the End Users: New Roles for the Information Professional, Online Review*, Vol. 13, No. 2, pp. 143-154 (1989).
- 20) 影浦 峽, 大山敬三, 宮澤 彰, 根岸正光, 鳥居俊一, 絹川博之: 文献の論理構造を考慮した全文検索システム, *学術情報センター紀要*, No. 3, pp. 49-58 (1990).
- 21) 内田富雄: SGML をとりまく環境と JIS 化への動き, *CEPS Japan '91 講演要録*, pp. N. 1.1-N. 1.5 (1991).
- 22) 芝野耕司: SGML と全文データベース, *情報処理学会研究報告*, Vol. 89, No. 66 (89-FI-14), pp. 2.1-2.8 (1989).
- 23) *EPSIG News*, Electronic Publishing Special Interest Group, Association of American Publishers, Dublin, OH, Vol. 1, No. 1 (1987).
- 24) Summit, R. K.: *Outlook for Electronic Information Services*, データベース国際セミナー, (財)データベース振興センター, pp. 21-44 (1989).

(平成3年4月23日受付)



根岸 正光 (正会員)

1945年生。1968年東京大学経済学部卒業。1976年同大学院経済学研究科博士課程修了。東京大学助手。

講師, 助教授 (情報図書館学研究センター, 文献情報センター) を経て, 1986年より学術情報センター教授 (データベース研究部門)。オンライン情報検索システム, 学術雑誌総目録データベースの開発に従事。最近, 論文引用度の分析による研究水準の国際比較, 全文データベースの編集・検索システムなどを研究。ASIS 会員。