講座 ◇ これからの情報担当者 ◇ 第3回

情報の対象・ニーズの変化Ⅱ ―一次情報(フルテキストデータベース)

細江 孝雄*/HOSOE Takao

*Change in information objects and needs Ⅱ —
the primary information(Full text database)*

Series ◇ Information specialist for a coming age ◇ 3

[著者抄録] 最近、一次情報としてのフルテキストデータベースへの関心や要求が高まりつつある。またフルテキストデータベースの提供数も増加しつつある。しかし、利用の実態は、かならずしも多いとは言えない。これらの理由、背景を示すとともに、フルテキストの提供面、利用面の双方からみた現状、技術的課題・問題点、二次情報系データベースを超える有用性、波及効果等について述べる。

[著者付与キーワード] 一次情報、一次資料、フルテキストデータベース、図表、データ圧縮、情報検索システム、電子出版、文書処理、オンラインシステム、技術開発

[Author Abstract] Full text databases of primary sources have become integral parts of the online searching. The number of full text databases online is also increasing. Users, however, have not adopted these files for use at the same rate. This paper describes several reasons for the slow adoption of full text, the present services and uses, technical problems, usefulness and influence of full text databases compared with secondary databases.

[Keywords by Author] primary information, primary source, full text database, graph and table, data compression, information retrieval system, electronic publishing, document processing, online system, technology development

* JICST 技術管理室 (〒100 千代田区永田町2-5-2) Tel. 03(581)6411
JICST Technical Coordination and Development Office
(5-2, Nagatacho 2-chome, Chiyoda-ku, 100)

1. はじめに

現在の情報担当者は、単なる情報の利用者としてではなく、本誌の30巻講座「企業における社内データベースの構築」でも特集されたように、その規模はともかくとして、社内外で発生する様々なデータ（一次情報）を活用して、今や、自らがデータベース作成に携わる時代に入っている。

したがって、ここでは、90年代のホットなテーマである一次情報としてのフルテキストデータベースについて、利用面のみからではなく作成・提供の面からも概観してみることにする。

最近の情報の利用動向を見ると、これまで主に提供されてきた二次情報のみでなく、原著論文等の詳細かつ具体的内容への直接アクセスを行うために、原著論文の全文（フルテキスト）等の一次情報への関心や要求が高まりつつある。

しかしながら、これらのフルテキストデータベ

ースへの要望が高まりつつあるとは言われながらも、表1および表2に示すように、80年代に入ってフルテキストデータベースの数あるいは一次情報としてのファクト系データベースの数は増えたが、その利用は必ずしも同程度には増加していないというのが現状のようである^{1,2)}。

以下、これらの背景、理由について述べるとともに、最初に、利用の必然性・効用を二次情報との関連で整理し、次に現状の紹介、フルテキストデータベースがもたらす様々な波及効果、技術的問題・課題等について述べる。

2. 一次情報としてのフルテキストデータベースへの期待が高まっている理由、必然性

フルテキストデータベースをめぐる内外の状況を紹介する前に、後述する多くの技術的・体制的な問題、課題を抱えてはいるものの、フルテキス

表1 データベースの形態別にみたデータベース分布

タイプ	データの形態	57年度	58年度	59年度	60年度	61年度	62年度
リファレンス	文書情報	339	501	654	906	933	1,012
	書誌	92	133	173	202	194	210
	抄録	10	11	47	112	211	236
	書誌／抄録	188	286	353	421	380	404
	その他	49	71	81	171	148	162
ファクト	全文情報	52	87	198	338	616	786
	全文	35	66	180	313	503	593
	全文／その他文書組み合わせ	17	21	18	25	113	193
	数値情報	130	229	207	239	389	616
	画像情報	2	2	4	6	2	1
	文書／数値／画像組み合わせ	81	97	179	213	19	25
合 計		604	916	1,242	1,702	1,959	2,440
リファレンス DB 比率		56.1 %	54.7 %	52.7 %	53.2 %	47.6 %	41.5 %
ファクト DB 比率		43.9 %	45.3 %	47.3 %	46.8 %	52.4 %	58.5 %

出典：データベース白書 1989 年

表2 我が国で利用可能なデータベース

(1)データベースの総数および科学技術分野の割合

区 分	総 計	国内作成データベース			海外作成データベース		
		科学技術 分野	その他の 分野	計	科学技術 分野	その他の 分野	計
文献データベース	566	39	62	101	215	250	465
全文データベース	501	7	56	63	104	334	438
ファクトデータベース	589	20	199	219	64	306	370
計	1,656	66	317	383	383	890	1,273

(2)科学技術分野のデータベース数の国別内訳

作成国	日	米	英	仏	西独	その他	計
区分							
文献データベース	39	121	49	14	15	16	254
全文データベース	7	94	3	3	1	3	111
ファクトデータベース	20	42	12	1	7	2	84
計	66	257	64	18	23	21	449

(3)科学技術分野の文献データベースにおける収録記事件数の国別内訳

作 成 国	日	米	英	仏	西独	その他	計
延べ収録記事件数 (百万件/年)	2.3	5.9	1.3	0.6	0.4	1.2	11.7
全体比(%)	20	50	11	5	3	10	99

注) ①「データベース台帳総覧昭和62年度版」より JICST 調べ

②我が国でデータベースサービスを行っているプロデューサ、ディストリビュータ、代行検索企業、代理店が扱っている商用のデータベースを対象とする。

③「文献データベース」には、文献の二次情報の他、研究課題情報等の案内型データベースやシソーラスファイル等の用語辞書を含む。また、「全文データベース」には、一次雑誌・新聞等の文章データベースを含む。

④(1)において、「科学技術分野」のデータベースは、データベース台帳総覧において「自然科学・技術」に分類されたものがほとんどであるが、一部「一般(辞書、補助ファイル)」に分類されたものを含む。

⑤(3)における収録記事件数は、収録記事件数が不明のデータベースがあるため若干の誤差を含む。

トデータベースへの期待、ニーズが高まっている主な理由、必然性を以下に挙げてみる。

(1)文献検索の精度向上

- ・多様な検索が可能（フリータム、否定・ネガティブ情報、質・量的情報等）
- ・誤りの減少（抄録、索引等で人手を介さない）

(2)原文献を得るまでの時間遅れが無い

(3)文献の分析

- ・精密な分析が可能
- ・必要な結果、データ、方法、材料等を効率よく知る（比較、新規性、新製品開発の指標）
- ・編集者や査読者のためのツール

(4)図や表などの情報の利用

- ・マルチメディア情報

(5)シソーラスの作成

- ・動的な変化を把握
- ・索引用語の自動選定

(6)情報量の増加と省力化

- ・情報量増加への迅速な対応
- ・二次情報加工には時間、コストを要する
- ・情報選別、抄録、索引等の専門家の省力化

(7)研究動向サーベイツール（二次情報でもある程度可能）

- ・情報量の増加
- ・文献集の作成（「参考文献」欄の活用等）
- ・重要課題、課題の変遷、研究の流れ
- ・雑誌や研究者のランク付け
- ・研究課題と機関の関係
- ・研究協力関係
- ・専門家を検索、講演依頼等

(8)論文作成ツール（二次情報でもある程度可能）

- ・論文リスト
- ・「データ」のみからの論文作成

(9)教育への活用

3. フルテキストデータベースをめぐる状況

「フルテキストデータベース」と一口にいても、次のように、データベースの規模（データ量、対象データの分野等）、利用形態（提供形態）、実

現内容、実現機能等により様々なレベルがある。

特に最近では、オフィスレベルあるいはパーソナルレベルでもデータベースの構築や文書の蓄積・検索等が可能な環境が整いつつあるのだなおさらである。

例えば、後者の文書の蓄積・検索について言えば、国内でも昨年あたりからオリジナルデータ（一次情報）をそのまま入力し検索可能とするような新製品、あるいは関連技術の発表等が相次ぐという注目すべき状況にある。

また、大規模な商用データベースについて言えば、図、表、写真等のフルテキストにおける重要な一次情報についての本格的な検索は、内外ともに未だ実現されていない。日本語を対象とすると、さらに日本語特有の技術的課題がこれに加わる。

フルテキストデータベースでは、まずこれらの点について整理しておく必要がある。

①当該データベースの目的

- ・パブリックデータベース（商用等）であるかプライベートデータベース（オフィスでの文書管理等）であるか

②データベースの規模

- ・データ量はどの程度であるか（対象分野、記憶媒体、検索速度等に関係）

③オンライン型であるかスタンドアロン型（CD-ROM等）であるか

④原文（テキスト）の詳細な検索が可能であるかどうか、それとも原文複写代替機能に主体が置かれているか

⑤二次情報（人手による索引等）の活用度合い

⑥一次情報（原データ）の入力手法

- ・印刷体か電子媒体か
- ・計算機可読化の範囲
- ・原データの文献・文書構造の尊重度合い

⑦検索機能のレベル

- ・図、表、写真等の検索（本格的な検索は、現段階では技術的に不可能）
- ・検索速度
- ・検索精度

⑧出力表示、伝送機能のレベル

- ・図、表、写真等の出力表示、伝送の方式
(現状は、省略または図表等の説明文のみの表示または光ディスク、CD-ROM、ファックスを利用して出力、伝送しているのが大半)

⑨コスト

3.1 オンラインによる商用フルテキストデータベース

大規模な商用データベースのプロデューサ、ディストリビュータにおいては、対象とする一次情報のデータ量(データ圧縮技術)、図、表、写真等の扱いを中心に共通する多くの未解決の技術的課題を抱えている。日本語を対象とする場合には、さらに日本語特有の技術的課題がこれに加わる。

このことが、総じてコストや提供内容にそのまま反映し、フルテキストデータベースの利用が期待される割りには今ひとつ伸びていない大きな要因となっている。したがって、当然ながら内外のデータベース関係機関およびコンピュータ関連のメーカーにとって、これらの技術的課題が最大の関心事となっている。

3.1.1 ビジネス分野と科学技術分野での利用状況

技術的課題を抱えながらも、表1からもわかるように、最近ではフルテキストデータベースの増加率が著しく高くなっている。

開発、利用の進んでいる米国をみると、代表的な Mead Data Central の LEXIS (法令・判例) や NEXIS (新聞・雑誌記事) をはじめ、企業財務、有価証券、市場調査、産業調査等いわばビジネス分野での利用が中心を占めている。科学技術系分野では、最終的にはフルテキストの検索に移行していくことと思われるが、現状では、利用は必ずしも多くないようである。

フルテキストデータベースといっても、検索対象は文字部分であり、記憶容量、オンライン出力時の回線速度・時間、プリンタやディスプレイ装置の性能・解像度等の理由により、図、表、写真

等のグラフィック情報はテキストから省かれているのが実情である。

このためにグラフィック情報を別ファイルとして文字情報ファイルとリンクさせたり、CD-ROM や光ディスク等のオフライン系の大容量記憶媒体を利用しての出力や、ファックスを用いての伝送などの方法が試みられている(3.2項参照)。

多くの利用者は、オンラインで提供されるフルテキストデータベースを印刷体出版物の完全な代替品として利用するまでには至っていないというのが現実であろう。目的に応じてオンライン系の二次情報(書誌系)データベースやフルテキストデータベースおよびオフライン系の CD-ROM 等や印刷体を使い分けているというのが現状である。

図、表等の検索が実現されないと、図、表データの多い科学技術分野における利用は、ニーズのある特定の分野を除いて低いままとなると思われる。

CAS でも、「利用者をひきつけるためには、ある特定分野または小分野においてフルテキストの対象データをかなりの量、集めることが重要である。ACS Chemical Journals のみでは不十分であるが、CAS や JICST の文献データベース(60万件以上/年 二次情報系)ほどの網羅性は必要ないと思われる。真に網羅的にフルテキストデータ(資料)を集めた物が提供できるか、または経済的に可能になるまでには、年月がかかるであろう」との見解を示している。

現在の商用サービスでのフルテキストデータベースは、1記事のデータ量が少ない新聞(Mead Data Central の NEXIS や日経テレコン等)や分野が限定もしくは特定されている法律(Mead Data Central の LEXIS 等)、特許(同 LEXPAT 等)などが比較的大規模データベースとなっており、利用も多いようである。

一方、JICST ファイルのように科学技術全分野を対象とし、雑誌数も1万種を超えるような大規模データベースでのフルテキスト化は、現在、国内外ともに実現されていない。この世界では草分

け的なBRSのCCML (Comprehensive Core Medical Library 医学雑誌等70余誌)あるいはSTNのCJACS (米国化学会発行の化学系19雑誌)に代表されるように、対象分野と収録雑誌を絞ってデータベース化しているのが現状である。

3. 1. 2 通信コストとフルテキストデータベース

ところで、現在は、情報を得るためのコストのかなりの部分を通信コストが占めている。フルテキストの利用が普及するためには、コスト、特に通信コスト低減化が必要となろう。この点からも、前述したように、データ圧縮をして通信コストを下げる等の工夫が必要となる。ちなみに、IBMの過去30年の調査によると、CPUコストは年25%下がっているが、通信コストは10年ぐらいで年11%ぐらいしか下がっていないとのことである。

通信コストがかからないという点からみて、データ量や使い勝手等はともかくとして、パソコン、ワークステーションレベルやCD-ROMでのフルテキスト的なものの製品化が進めば、個人あるいはオフィスレベルでの利用が普及すると思われる。これについては、3. 2項で後述する。

3. 1. 3 代表的商用システムの検索機能

次に、フルテキストデータベースの先導的役割を果たしてきたBRSをはじめ、STNやDIALOG等の代表的なシステムの検索システムの全般的概要をながめてみる。検索例を図1～図5に示す。

これらのシステムでは、従来の検索機能に加えてフルテキストデータベース用の検索機能が用意されており、ページ単位またはパラグラフ単位(段落単位)で入力キーワードの出現頻度を知ることができたり、検索結果を文単位、パラグラフ単位、ページ単位で出力可能にしたり、出力中の当該キーワードを画面上で知り得るハイライティング機能などを付加している。

また、Mead Data CentralのNEXISのように、変化形(単数・複数形、所有格等)や米語・英語のつづり、省略形・頭字語とフルスペル等の自動検索(同義語検索)機能を持っているものもある。これらに共通していることは、記事をパラグラフ

単位(段落単位)に分割して、検索範囲を狭めることにより、検索の高速化や効率化を図っていることである。

3. 1. 4 代表的なサービス機関とフルテキストデータベース

内外の代表的なサービス機関で現在提供されているフルテキストデータベースの内容の詳細については、「データベース台帳総覧」(通産省)、「Directory of Online Databases」(Curra/Elsevier)、「Computer Readable Databases」(Martha E. Williams)、「Encyclopedia of Information Systems and Services」(Gale Research Company)、「IMI (Information Market Indicators)」(IMI, Inc.)等により知ることができる。但し、これらのツールは、発行されるまでにタイムラグがあるので、最新の情報を得るには、データベースサービス機関に直接問い合わせて確認することが必要である。

各機関で提供している代表的なフルテキストデータベースについては、昨年、原田³⁾がドキュメントデリバリーの観点からレビューしている。また、昨年3月に刊行された「データベースの先端技術」(科学技術庁)⁴⁾や「全文データベースのシステムに関する現状調査報告書」⁵⁾(学術情報センター昭和62年7月)などによっても知ることができる。

内外の各データベースの詳細な紹介については、紙面の都合で別の機会に譲り、本稿では、日本国内の最近の動向について述べておくことにする。

```

1__ : migraine or cluster adj headache$1
RESULT 33

2__ : 1 with treat$5
RESULT 13

3__ : 1 with 322$1
RESULT 9

4__ : 2 or 3
RESULT 16

5__ : ..p 4 gn,hits/doc=1-2
1
GN  GENERIC DRUG NAME: INDOMETHACIN.
TX
(本文)(50 OF 86) 段落番号
INDOMETHACIN HAS ALSO BEEN USED FOR
SYMPTOMATIC TREATMENT OF POSTOPERATIVE PAIN+,
BILIARY PAIN+, CHRONIC ERYTHEMA NODOSUM+, AND
CERTAIN TYPES OF HEADACHE+ (E.G. CLUSTER
HEADACHE, EXERTIONAL HEADACHE).

2
GN  GENERIC DRUG NAME: LITHIUM CARBONATE. LITHIUM
CITRATE.
TX
(40 OF 88)
SUBHEADING: OTHER USES
(3213) IN SEVERAL UNCONTROLLED TRIALS OF LITHIUM
IN THE TREATMENT OF PATIENTS WITH CHRONIC
CLUSTER+ OR MIGRAINE+ HEADACHE, THE DRUG HAS
REDUCED THE INCIDENCE AND SEVERITY OF HEADACHE.
FURTHER STUDIES ARE REQUIRED BEFORE USE OF THE
DRUG CAN BE RECOMMENDED IN THESE CONDITIONS.

```

図1 BRS の DIFT ファイル検索例

```

=> S SOLAR (W) ENERGY (5P) WINTER#
669 SOLAR
20522 ENERGY
335 WINTER#
L6 3 SOLAR (W) ENERGY (5P) WINTER#

=> D KWIC ← ヒットタームの前後 20 語のものを出力表示

L6 ANSWER 1 of 3
TX TX(2) of 35. Introduction. . . . total energy as well as
(本文) ← 段落番号 ヒット(ハイライティング)
wavelength distribution. The purpose of this study was to observe
how the natural variation in **solar** **energy** influx in an Arctic
area (69.degree.30'N) would influence the formation of photoproducts
with respect to oxygen content, water solubility, and. . .

TX(5) of 35. Experimental . . . stirring bar and a thermostat
system which ensured that the water temperature did not drop below 4
.degree.C during the **winter** experiments (Table I). .Table. The
water surface was then completely covered with 16.0 g of a 3:1
mixture of crude. . .

```

図3 STN の検索例 (本文中の段落の検索)

出力例 1

AN ACCESSION NUMBER: 384056. 8502.
 OC PARAGRAPH SENTENCE NS-WORD
 AN (1)
 SO SOURCE: AMERICAN HOSPITAL FORMULARY SERVICE.
 TI MONOGRAPH TITLE: Nicotine Polacrilex.
 GN GENERIC DRUG NAME: Nicotine Polacrilex. Nicotine.
 TN TRADE NAME MFR: Nicorette (Merrell Dow).
 SY SYNONYMS: Nicotine Resin Complex.
 RN REGISTRY NUMBER: 54-11-5.
 DI DRUG INFORMATION:
 DI CHEMICAL NAME: S-3-(1-Methyl-2-pyrrolidinyl)pyridine.
 MOLECULAR FORMULA: C₁₀H₁₄N₂.
 TC THERAPEUTIC CLASSIFICATION: MISCELLANEOUS AUTONOMIC
 DRUGS (12-92).
 DE DESCRIPTORS: (3113) Chemical description. (3115) Chemical
 description. (3123) Constants. (3133) Physical description.
 (3143) Solubility. (3303) Stability. (3345) Storage.
 (3400) Preparations. (3410) Other chemical ion ingredient.
 (3415) Other chemical ion ingredient. (3810) Absorption.
 (3203) Pharmacology. (3205) Pharmacology. (3500) Dose.
 (3635) Dependence addiction. (3685) Treatment toxicity.
 (3200) Pharmacology. (3813) Absorption. (3815) Absorption.
 (3823) Distribution. (3825) Distribution. (3833) Elimination. . . .
 SH SECTION HEADINGS/SUBHEADINGS (PRIMARY TEXT):
 CHEMISTRY AND STABILITY.
 CHEMISTRY.
 STABILITY.
 PHARMACOLOGY.
 PERIPHERAL NERVOUS SYSTEM.
 PHARMACOKINETICS.
 ABSORPTION.
 DISTRIBUTION.
 ELIMINATION.
 USES.
 SMOKING CESSATION.
 CAUTIONS.
 ADVERSE EFFECTS.
 LOCAL EFFECTS.
 :
 :
 SH SECTION HEADINGS/SUBHEADINGS (ADDITIONAL TEXT):
 CHEMISTRY AND STABILITY.
 CHEMISTRY.
 STABILITY.
 PHARMACOLOGY.
 NERVOUS SYSTEM EFFECTS.
 PERIPHERAL NERVOUS SYSTEM.
 CENTRAL NERVOUS SYSTEM.
 CARDIOVASCULAR EFFECTS.
 GI EFFECTS.
 OTHER EFFECTS.
 :
 :
 AB PRIMARY TEXT.
 (1 OF 35)
 (抄録) SECTION HEADING: CHEMISTRY AND STABILITY
 SUBHEADING: CHEMISTRY
 (3115) Nicotine is a naturally occurring autonomic drug.(4).
 AB (2 OF 35)
 SECTION HEADING: CHEMISTRY AND STABILITY
 SUBHEADING: CHEMISTRY
 (3415) Each piece of gum contains approximately 0.44 mEq of
 sodium.(2).
 AB (3 OF 35)
 SECTION HEADING: CHEMISTRY AND STABILITY
 SUBHEADING: STABILITY
 (3345) Commercially available nicotine polacrilex chewing gum
 should be protected from light and stored at room temperature less
 than 30DGC.(1).
 :
 :
 AT ADDITIONAL TEXT.
 (1 OF 45)
 SECTION HEADING: CHEMISTRY AND STABILITY
 nicotine.
 AT (2 OF 45) ←段落番号
 (本文) SECTION HEADING: CHEMISTRY AND STABILITY
 SUBHEADING: CHEMISTRY
 (3113) Nicotine polacrilex is a resin complex of nicotine and
 polacrillin.(61) Polacrillin is a unifunctional low-cross-linked
 carboxylic cation-exchange resin prepared from methacrylic acid and
 divinylbenzene.(62) Nicotine is a pyridine alkaloid;(3.12) the
 nicotine present in the cation-exchange resin is a highly purified
 extract obtained from the dried leaves of the tobacco plant
 (Nicotiana tabacum (Fam. Solanaceae/)(2.3) (3133) The drug occurs
 as a basic, colorless to pale yellow, very hygroscopic, oily, volatile
 liquid and has an unpleasant pungent odor and a sharp, burning,
 persistent taste.(3.12) (3143) Nicotine is soluble in water and in
 alcohol.(3.12) (3123) The drug has a pKa of 7.9.(12).
 :
 :

CR REFERENCES.
 (1 OF 72)
 (参考文献) 1. Merrell Dow Pharmaceuticals Inc. Nicorette prescribing
 information. Cincinnati, OH. 1984 Jan.
 :
 :
 CR (71 OF 72)
 71. Lees AJ. Hemidystonia relieved by nicotine. Lancet. 1984; 2:871.
 Letter.
 CR (72 OF 72)
 72. Anon. Dental effects of nicotine gum. Med Lett Drugs Ther.
 1984; 26:98.

出力例 2

AN ACCESSION NUMBER: 582036. 8502.
 OC PARAGRAPH SENTENCE NS-WORD
 AN (1)
 SO SOURCE: HANDBOOK ON INJECTABLE DRUGS.
 TI MONOGRAPH TITLE: Chlordiazepoxide HCl.
 GN GENERIC DRUG NAME: CHLORDIAZEPOXIDE HCL.
 TN TRADE NAME MFR: Librium (Roche).
 RN REGISTRY NUMBER: 438-41-5.
 TC THERAPEUTIC CLASSIFICATION: (AHFS 28-24-08)
 Benzodiazepines.
 DE DESCRIPTORS: (5403) Concentration. (5412) Other ingredient ion
 (Table). (5413) Other ingredient ion. (5453) Reconstitution mixing.
 (5510) Special instruction. (5303) Stability. (5343) Storage.
 (5383) Container compatibility. (5153) pH. (5523) Dosage schedule.
 (5553) Age dosage relation. (5523) Administration route.
 (5332) Vehicle compatibility (Table). (5372) Syringe additive (Table).
 (5353) Equipment compatibility.
 SH SECTION HEADINGS/SUBHEADINGS (PRIMARY TEXT):
 CONCENTRATION.
 STABILITY.
 SORPTION.
 PH.
 DOSAGE.
 COMPATIBILITY INFORMATION.
 SOLUTION COMPATIBILITY.
 DRUGS IN SYRINGE COMPATIBILITY.
 ADDITIONAL COMPATIBILITY INFORMATION.
 Y-SITE INJECTION.
 AB PRIMARY TEXT.
 (1 OF 18)
 SECTION HEADING: CONCENTRATION
 (5403) Chlordiazepoxide HCl (Roche) is available in 5-ml dry-filled
 ampuls containing 100 mg of drug. (5453.5413) Accompanying the
 ampul of drug is a 2-ml ampul of diluent for intramuscular
 administration. (5412) The diluent contains (232p 1601).
 AB (2 OF 18)
 SECTION HEADING: CONCENTRATION
 SEE TABLE FORMAT OF THIS RECORD.
 :
 :
 AB (18 OF 18)
 SECTION HEADING: ADDITIONAL COMPATIBILITY
 INFORMATION. SUBHEADING: Y-SITE INJECTION
 One milliliter of each primary solution was mixed in a test tube
 with 1 ml of chlordiazepoxide HCl (Roche) 10 mg/ml. The mixtures
 were observed for physical compatibility immediately and after four
 hours at room temperature by visual observation and microscopic
 examination. No physical incompatibility was noted (534).
 AT ADDITIONAL TEXT.
 TA TABLE DATA.
 (1 OF 3)
 (5412) The diluent contains (232p 1601):

Benzyl alcohol	1.5%
Polysorbate 80	4%
Propylene glycol	20%
Maleic acid	1.6%
Sodium hydroxide	to adjust pH.
:	:
:	:

 TA (3 OF 3)
 (5372) Drugs in Syringe Compatibility

DRUG	(IN SYRINGE)	MFR	AMT	MFR	AMT	REMARKS	REF	C
Chlordiazepoxide HCl								
Benzquinamide HCL	RR	25 mg/	RC	100	Incompatibility	569	I	
		1 ml		mg/2	observed immedi-			
				ml	ately upon mixing			
Benzquinamide HCL	RR	50 mg/	RC	100	Incompatibility	569	I	
		2 ml		mg/2	observed immedi-			
				ml	ately upon mixing			
Benzquinamide HCL	RR	50 mg/	RC	100	Incompatibility	569	I	
		1 ml		mg/2	observed immedi-			
				ml	ately upon mixing			

図 2 BRS の DIFT ファイル出力例

- AN 84:98 CJACS
 DN OM840541Q
 SO Organometallics, (1984), 3(12), 1916-1917. CODEN: ORGND7. ISSN: 0276-7333
- (タイトル)→ TI Facile C-Mercuration of Bis(diphenylphosphino)methane
 AU (1) Lusser, Maria; (2) Peringer, Paul (*)
 CS (1,2) Institut für Anorganische und Analytische Chemie der Universität Innsbruck, Innrain 52a A-6020 Innsbruck, Austria
- (抄録)→ AB The reaction of $\text{CH}_2(\text{PPh}_2)_2$ with $\text{Hg}(\text{OAc})$ and $[\text{Hg}(\text{Me}_2\text{SO})_6](\text{O}_3\text{SCF}_3)_2$ leads to $\{\text{AcOHg}\}_n\text{CH}_2\text{-n}[\text{PPh}_2(\text{HgOAc})]_2(\text{O}_3\text{SCF}_3)_2$ ($n = 1, 2$), which were characterized with ^{199}Hg and ^{31}P NMR spectroscopy.
- TX TX(1) of 5. Most metal complexes of deprotonated bis(diphenylphosphino)methane (dppm) are coordinated via the phosphorus atoms of the ligand either in a chelating (i) or in a bridging mode (ii). 1 .Footnote. .General Graphic. This is true for transition metals including Au, Ni, Pd, and Pt 2-6 .Footnote. as well as for lithiated dppm that adopts structure i according to ^{31}P and ^6Li NMR measurements. 7 .Footnote. Coordination through the methylene carbon is known only in a few cases. 8-12 .Footnote. The complexes were generated from the deprotonated ligand and appropriate metal compounds or by deprotonation of the coordinated ligand. We report here further examples of C-substitution in various mercury complexes of dppm.
- TX(2) of 5. The reaction of dppm with $\text{Hg}(\text{OAc})_2$ and $[\text{Hg}(\text{Me}_2\text{SO})_6](\text{O}_3\text{SCF}_3)_2$ 13 .Footnote. according to eq 1 .Equation. in methanol at ambient temperature affords the C-metalated product 1 almost quantitatively within a few minutes. 14 .Footnote. The compound was isolated and crystallized as an Me_2SO solvate. 15 .Footnote. The structure is evidenced by ^{199}Hg and ^{31}P NMR spectroscopy. 16 .Footnote. The $^{199}\text{Hg}\{^1\text{H}\}$ NMR spectrum (^{199}Hg , $I = 1/2$, abundance 16.8%) of 1 shows two different mercury sites: one doublet attributable to the mercury bound to phosphorus and a 1:2:1 triplet for the carbon-bound mercury being coupled to two equivalent phosphorus atoms via two bonds. This is confirmed by the ^{31}P NMR spectrum. The NMR data are summarized in Table 1. .Table.
- TX(3) of 5. The reaction of one further equivalent of $\text{Hg}(\text{OAc})_2$ according to eq 2 .Equation. leads under the same conditions almost quantitatively to the dimercurated product 2, which was characterized similarly to 1 (Table 1). 14,17 .Footnote. Particularly conclusive is the nearly doubled intensity of the ^{199}Hg satellite pattern due to 2 (^{199}Hg , ^{31}P) in the ^{31}P NMR spectrum according to a statistical abundance of 19.37% for the isotopomer 2a .General Graphic. compared with that of 1a (11.56%). The mercury- ^{199}Hg resonance of the C-bound mercury is at high frequency relative to that of 1 in keeping with the results for the series $\text{CH}_4\text{-n}(\text{HgX})_n$. 18 .Footnote.
- RN RN(1) of 5. 93254-20-7 CN(1) 1
 RN(2) of 5. 93254-22-9 CN(2) 2
 RN(3) of 5. 2071-20-7 CN(3) $\text{Ph}_2\text{PCH}_2\text{PPh}_2$
 RN(4) of 5. 1600-27-7 CN(4) $\text{Hg}(\text{OAc})_2$
 RN(5) of 5. 76703-09-8 CN(5) $[\text{Hg}(\text{Me}_2\text{SO})_6](\text{O}_3\text{SCF}_3)_2$
- RE RE(1) of 21. 1 Puddephatt, R. J. Chem. Soc. Rev. 1983, 99.
- (参考文献)→ RE(2) of 21. 2 Schmidbaur, H.; Mandl, J. R. Angew. Chem. 1977, 89, 679. Schmidbaur, H.; Mandl, J. R.; Bassett, J. M.; Blaschke, G.; ZimmerGasser, B. Chem. Ber. 1981, 114, 433.
- RE(3) of 21. 3. Bassett, J. M.; Mandl, J. R.; Schmidbaur, H. Chem. Ber. 1980, 113, 1145.

図4 STNのCJACS ファイル出力例

DIALOG呼び出し番号

AZ = 323270000 Summary RT =
Chapter CH=32327 CH =
Type TY=323270 TY =
Unit UN=323270000 UN =

/TI Chapter Title: Vaccine Technology

AU = Author: Jegede, V. A., Kowal, K. J., Lin, W., Ritchey, M. B.
CS = Institution: Lederle Laboratories, American Cyanamid Co.
SO = Source: Kirk-Othmer. 3rd ed. Vol 23. 628-643
Number of Sections = 5 Tables = 5 Descriptors = 131 References = 132

Abstract: 抄録

/AB Vaccines are preparations used to prevent disease by inducing immunity to the disease in the vaccine recipient. Currently licensed vaccines provide protection against poliomyelitis, diphtheria, tetanus, pertussis, measles, mumps, rubella, influenza, pneumococcal pneumonia, and hepatitis, among others. Indications for use, composition and methods of manufacture, and standardization and testing are described for each vaccine. Techniques used to manufacture these vaccines include propagation in tissue culture for live viral vaccines and biochemical fractionation, purification, and inactivation for killed preparations.

Section Headings:

/SH "Untitled" <SELECT SN=32327S1>
Licensed Vaccines <SELECT SN=32327S2>
Vaccines Being Developed <SELECT SN=32327S3>
Future Technology <SELECT SN=32327S4>
Economic Aspects <SELECT SN=32327S5>

SN =

Tables: 表

/TT Table 1. Schedule for Immunization of Normal Infants and Children**a
<SELECT TN=32327T1>
Table 2. Selectively Used Vaccines <SELECT TN=32327T2>
Table 3. Vaccine Doses Distributed in 1978**a <SELECT TN=32327T3>
Table 4. Vaccination Costs**a <SELECT TN=32327T4>
Table 5. Ranges of Total Costs for Some Vaccine-Associated Injuries for Persons Less Than One Year of Age (Assuming a Discount Rate of 2.5%)**a,b <SELECT TN=32327T5>

TN =

表の標題検索

Descriptors: キーワード(索引語)

/DE ACIP. See Immunization practices advisory committee., #23:629
Adenovirus vaccine, #23:632
Aluminum hydroxide [21645-51-2], vaccine adjuvant, #23:633
Ammonium sulfate [7783-20-2], fractionation in vaccine, #23:633
Ammonium sulfate [7783-20-2], pill purification, #23:634
Antibodies, monoclonal, #23:637
Australia antigen, #23:635
Bacillus subtilis, genetic engineering, #23:637
Bacteriophages, in vaccine mfg, #23:637
Beef-heart infusion, in vaccines, #23:630
BETA -Propiolactone [57-57-8], inactivation of viruses, #23:632

辞典の巻数・ページ付けを表示

CAS Registry Nos:

RN = 50-00-0; 50-70-4; 50-99-7; 54-64-8; 57-09-0; 57-13-6; 57-50-1;
57-57-8; 69-53-4; 108-95-2; 123-86-4; 126-73-8; 137-16-6; 141-78-6
143-74-8; 144-55-8; 151-21-3; 868-18-8; 1306-06-5; 7647-14-5;
7647-15-6; 7647-17-8; 7783-20-2; 9000-71-8; 9001-67-6; 9001-75-6;
9002-61-3; 9005-65-6; 9016-12-0; 21645-51-2

図5 DIALOGのKIRK-OTHMER ONLINE ファイル例

3. 1. 5 国内のフルテキストデータベースに関する動向

(1) 日本国内で作成・提供されているフルテキストデータベースは、日本経済新聞社の日経テレコンや平和情報センターのHINETなど新聞記事情報を対象としたものが中心となっている。

日経テレコンでは、1982年以降の日経4紙(日経本紙、産業、流通、金融)約110万件の全文化を図っている。キーワードは自動切出しを行ったものと統制語を併用している。検索はメニュー方式。図、表、写真等(オンラインでの検索不可)については画像イメージでファックスを用いて伝送している。この他、朝日新聞、毎日新聞、読売新聞、日刊工業新聞等のサービスも行っている。

平和情報センターでは、日経4紙、朝日新聞、読売新聞、毎日新聞、日刊工業新聞等のサービスを行っている。HINETは富士通のFAIRSをベースにした検索システムである。キーワードは自動切出しを行ったものを使う。

新聞記事の場合には、科学技術系雑誌ほどには必要性がないためか、両者の検索システムでは、レコード単位で検索が行われており、BRS、STN、DIALOGなどでフルテキスト検索用に開発された記事中の文(センテンス)やパラグラフ(段落)単位での検索はできない。

(2) 日本特許情報機構(JAPIO)では、現在、PATOLISによる特許情報検索で得られた公開特許公報番号をホストよりデータ転送し、CD-ROM端末から公開特許公報(原文)を出力表示する、いわゆる原文提供機能に主体を置いたサービスを行っている。また、光ディスクによる原文(公開・公告特許、同実用新案)提供も行っている。

JAPIOでのフルテキストデータベースの構築計画は、特許庁のペーパーレス計画と連動して進められている。特許情報は権利情報と技術情報の2面を持っており、権利情報としての側面では請求範囲(クレーム)、実施例あるいは図面等が重要になってきている。

JAPIOでは、これらを二次情報に過不足無く収録することは難しく最終的には一次情報である公

報全文(フルテキスト)による確認が必要となる、との認識の下にフルテキストデータベースの開発に着手している。

現在、抄録文はコードデータ(約350字/件)で、抄録図面はイメージデータ(50mm×70mm 平均数キロバイト/件)で作成し、蓄積している。

本年開始の電子出願により公報全文のコードデータ(電算写植データ)を用いてのフルテキストデータベース化がどの程度進むか注目される。

(3) 学術情報センターでは、化学関係雑誌(高分子学会の和文誌「高分子論文集」および英文誌「Polymer Journal」、日本農芸化学会の英文誌「Agricultural and Biological Chemistry」の3誌)を対象に、1989年から提供サービスを開始している。今年は、電子情報通信学会の英文誌が一つ加わる。

本文は、CTS変換により、また図表については、版下の光ディスクへの焼き込みとファックスでの送信により(ELと同様の方式で)提供している。図を出力するコマンドがあるが、オンライン上での実現は、現在の技術では無理なため、この方式を採用している。同センターによると、本文は、CTSにうまく反映されないことが多く、データベース化の際、後処理にも手間がかかっているようである⁶⁾。

(4) エレクトロニック・ライブラリー(EL)は、1988年1月よりELNETにより、新聞(33紙)、雑誌(約230誌)の記事原文を光ディスクにイメージ情報として収録し、原文提供サービスを行っている。キーワードによる検索はキーワード集の用語を用いている。

(5) JICSTでは、情報科学技術委員会を設置し当面、医学・薬学分野を対象に事業化を図ることを目指して、昭和63年度より、フルテキストデータベースの検討を行っている。同委員会により、この2年間にフルテキストデータベースに関する調査報告書及び検討報告書(部内資料)⁷⁾⁸⁾が出されている。この5月には、日立のフルテキストサーチマシン(TSM-1)を用いての実験的デモが行われており、本年より、開発に着手する方向で進

められている。

(6) その他、財団法人規格協会でも規格情報のフルテキストデータベース化の検討が行われている。

3.2 CD-ROM によるフルテキストデータベース

CD-ROM は、辞書、辞典、便覧、年鑑、電話帳、地図、技術文献、特許等あらゆる分野の光記憶媒体として利用されている。しかし、フルテキストデータベース（広い意味での）として活用されているものは、原田も指摘しているように限られてくる³⁾。

オンラインデータベースとの関係でみると、現在、図・表、写真等のグラフィック情報はオンラインデータベースのテキストから省かれている。このため CD-ROM や光ディスク等のオフライン系の大容量記憶媒体を利用した出力やファックスを用いた伝送などの方法が試みられている。

前述した JAPIO の CD-ROM による公開特許公報の原文提供がその代表例である。一次雑誌(200 誌以上)の原文を CD-ROM にすべて収録して原文サービスを行っている ADONIS(大手出版社 10 社による)のサービスに代表されるように、現在の CD-ROM でのフルテキストデータベースサービスとされているものは、どちらかというと膨大なデータ量からなる原文の複写代替機能的な役割を果たしていると言えよう。

したがって、第 2 章で述べたような本来のフルテキストデータベースに期待されている機能・内容の中でも、「多様かつ詳細な検索が可能」という最大の利点については、CD-ROM と言えども与えてくれない、というのが現状であろう。

大量情報を瞬時に多数の者が同時に検索できデータの更新も早いというオンラインデータベースの長所と、時間および通信コストを意識しないで個人で気軽に検索でき、図、表を含めた詳細な大量情報の提供(出力・表示)が可能な CD-ROM の長所のどちらが好まれ、市場を獲得していくかは注目される場所である。

3.3 プライベートデータベースでのフルテキスト化——文書検索

最近の情報産業における新しいシステムや市場は、ほぼ 2～3 年のタイムラグをもって米国から日本に渡来するというパターンが定着していると言われている。1985 年の米国におけるデスクトップ・パブリッシング(DTP)の登場による電子出版(EP)ブームも例外ではなかった。

現在、日本での文書作成のワープロ化は個人レベルにまで普及しているが、オフィスでの OA 化による文書作成や文書管理(文書検索等)の電子化は、1987 年以降の日本での DTP および電子出版ブームの流れの中で加速した。

ここでは、どちらかというオフィスレベル向けに製品化もしくは試作・研究されている文書検索システムの現状について述べる。

文書検索システムといっても、従来型の書誌事項、キーワード、抄録等を対象にするものから、原文そのものを検索するもの、あるいは、キーワードによる質問(検索)から文章そのものを入力しての検索など様々なレベルがある⁹⁾。

① DIALBASE

パソコン用リレーショナル DBMS(RBase)と DIALOG コマンドを使用。キーワード自動切出し、インバーテッドファイルによるパソコンでのフルテキスト検索。

紀伊国屋/ビーコンシステム。

② METAMORPH (英語対象)¹⁰⁾

キーワードによるインデックス無しの自然語検索。英語の語尾変化(～ing, ～ic 等)に配慮したパソコンでのフルテキスト検索。

1986 年サンダートン/EPI 社開発。

医学情報電子出版システム(Oncomed-MM)に応用。

③ Free Base

質問文での検索が可能なパソコンでのフルテキスト検索。エム・シー・ワードセンター発売。

④ GESCAN

キーワード等によるインデックス不要のシーケンシャルな自然語によるフルテキスト検索⁴⁾。

日本語（英語も可）を対象に、大量データを高速検索する日本での最初のミニコン等での本格的製品として注目されている。

GE 社開発のテキスト・アレイ・プロセッサを VAX/VAX Micro 上で稼働。日商エレクトロニクス㈱販売。

⑤ IRIS

質問文で入力し、意味マッチングを行い、推論により文書照合を行うフルテキスト検索。

富士通が開発。

⑥最近発表された高速のフルテキスト・サーチ・マシン（試作機）¹¹⁾¹²⁾

最近、文字列検索関係のハードウェアの試作の発表が相次いでいる。大量の文字情報（文書等）のなかから任意の文字列を含む文書を検索するフルテキスト・サーチ・マシン（TSM-1）を日立製作所中央研究所が試作した（1989年）。処理速度は、1,000 万文字以上/秒を可能にしている。また、同様の応用をねらって、日本電気は文字列検索 LSI（ISSP：Intelligent String Search Processor）（1987年）を、松下電器産業は専用プロセッサ SDP（Stream Data Processor）（1988年）を開発している。

これらは、いずれもハードウェアは完成しているが研究開発の段階で、製品化については各社とも模索中である。今後、大規模システムを対象とするのか小規模システムを対象とするのか、また、検索精度（検索速度はかなり達成されているが）の向上を如何に図っていくかなど課題はあるものの、今後の動向が注目される。

⑦その他、研究・開発中のシステム

文書を直接入力し、その入力文書と最も類似した文書をキーワードや係り受け関係を利用して検索する文書入力型のシステム（NTT ヒューマンインタフェース研究所 本誌 5 月号で紹介）や藤原、福永らのものがある⁹⁾。

4. フルテキストデータベースの利用および波及効果

第 2 章でフルテキストデータベースへの期待やニーズが高まっている理由、いわば、その利点、長所について既に述べた。ここでは、冊子体や電子出版など出版形態に関わる問題、端末操作や画面——ユーザインタフェース、二次情報との関係、原報の出力と複写の問題等について述べる。

4.1 冊子体と電子化時代の出版形態

1985 年頃、デスクトップ・パブリッシング（DTP）の登場で米国に電子出版（EP）のブームが起こり、以後、日本においても EP と DTP のコンセプトおよび市場が急速に広がった。原稿や文書が電子的に作成されることにより、編集、印刷、出版の電子化が進むこととなった。電子出版と CTS（Computer Type-setting System）の結合の一例を図 6 に示す。

フルテキストデータベースにおいては、冊子体印刷物と電子媒体の今後の動向が、データベースの作成およびその利用に大きな影響を及ぼす。

現在でも、パソコンで小説を読む電子小説などが出てきており、論文をハイパーテキストの形にしているフロッピー販売もある。電子手帳も売れているようである。文庫本など文庫手帳、ローンカードなどを使って読むなどの環境になるものと思われる。

冊子体印刷物の果たしてきた役割をフルテキストデータベースや電子出版物が代替していく時代になると、端末画面、端末操作等のユーザインタフェースに関わる問題が重要になってくる。例えば、画面に同じ情報が一度に出ること（印刷体では一度に全体を把握できる）、画面の大きさ、画面の拡大縮小、文字のポイント、レイアウト、字体など視認性の問題等が挙げられよう。

最近、マウスやキーボードを使わないで、画面を指で指して操作したり、利用者側の目の動きをキャッチして端末操作する技術などが出てきている（ゼロックス、ATR 等）。

図 6 TOPPAN CTS 文字情報処理システム

図、表は論文に入れることが大変であるが、コンピュータで段々できつつあり、そのような発行ができるようになってくると、科学技術の論文中に図や表がこれまでよりも多く掲載されるようになるという波及効果も出てこよう。図や表を見れば、論文の内容は大体わかるということを考えると、フルテキストで、図や表を扱うことの意義は大きいと言える。

4.2 二次情報とフルテキストデータベース

データベースの作成の面から考えると、本来、二次情報データベースは、フルテキストデータベースのバイプロダクトと言える。しかし、現状では、二次情報データベースが先行して提供されているので、次のような効果の利用も考えられる。

データ量、検索速度、コスト等に関わる問題点を解決する一手段として、二次情報データベース（または二次情報データ）と連動して検索することの効果は大きいと言える。二次情報の活用により大量情報を絞り込み、次にフルテキストの詳細な検索を行うことが効果的であろう。

現在の二次情報データベースでの索引は、量・質、程度、速度等の不確定要素に対しては、索引者の恣意が入るために、扱いきれないで無視されているのが現状である。このことが二次情報のみでは、詳細な必要情報がすべて得られない理由の一つとなっている。また、逆に、原報の出力表示（複写代替）のみでなく、原データのダイレクトな検索が必要とされる所以でもある。

4.3 原報の出力と複写

二次情報データベースの場合には、検索結果中の標題、抄録、キーワード等を見て、原文献に記載してあるであろうという仮定（推測）のもとに、原報を複写している。フルテキストであれば、記載箇所、内容等の確定ができるために、出力件数をある程度絞ってから複写を行うことが可能となるという大きな利点がある。

現在サービスされているフルテキストデータベース（あるいはフルテキスト的な原文サービス）

では、原データの量が多い場合には光ディスク、CD-ROM、ファックス伝送の方式を採用しているところが多い（JAPIOの光ディスク・CD-ROMでの公報提供、日経テレコン等）。

いずれにしても、原報をその場ですべて端末出力するのか、あるいは、ファックス等で送付するかは、対象データの種類（文章、数式、図、表、写真等）、データ量、時間、経費等に関連している。

4.4 その他（著者、査読者への影響等）

学協会での今後の考え方や体制等とも関係するが、著者がフルテキスト向きのための抄訳を作成したり、多段刊行（5.1項(1)参照）のためのフラグ化を行ったり、あるいは査読者によるチェック作業（特定分野での予め用意したチェック項目によるチェック等）をフルテキストへ活用していくことなども、関係機関に共通する今後の課題として挙げられる。

5. フルテキストデータベースをめぐる技術的課題

フルテキストデータベースは、未解決の多くの技術的課題を抱えている。その主な課題、問題点についてデータ入力関係とデータ管理関係（検索、データ圧縮等）の二つに分けて述べる。

5.1 データ入力関係の課題

5.1.1 フォーマット

各雑誌などに掲載されている事項は、一般には、表紙、目次、本文、図、表、著者によるキーワード、奥付、広告等から構成されている（図7参照）。学術雑誌、単行本、辞典等で構成要素は異なる。

フルテキストデータベースや全文電子出版では、特に冊子体の書誌事項としての「頁」は、意味が異なってくる。現在のオンライン用のフルテキストでは、図1～図4で例示したように段落（パラグラフ）単位や文章単位にフラグ（番号等）を付けたりして利用に供している。

表 紙

雑誌名：
欧文雑誌名：
巻：
号：
通号：
頁一頁：——→段落番号
発行年：
発行月：
発行日：
出版社：
出版社住所：
出版社Tel：
印刷所：
第3種郵便物認可年月日：
ISSN 番号：
編集委員会名：
編集委員名：

目 次

雑誌名：
欧文雑誌名：
巻：
号：
通号：
頁一頁：——→段落番号
発行年：
発行月：
発行日：
広告索引
広告

欄 名 1)：
著者名 A：
論題：
頁一頁：——→段落番号

著者名 B：
論題：
頁一頁：——→段落番号
.....
欄 名 2)：
.....
欄 名 N)：

本 文

論文 1

論題：
英文論題：
記事の種類
著者名：
著者名 (ローマ字)：
著者の所属・住所：
原稿受理日：
キーワード：
抄録：
英文抄録：
序文：
材料と方法：
結果：
考察：
参考文献：
1)：著者名：論題：雑誌名：巻：号：頁一頁：年
.....

図・写真：1)

論文 2

.....
和文索引：キーワード，頁——→段落番号
欧文索引：キーワード，頁——→段落番号

図 7 学術雑誌の構成要素の概略

フルテキストデータベースの高度利用・多目的利用のために、さらにテキスト内に、例えば重要度を示すようなフラグを付けて、ニーズに応じた幾通りもの刊行を半自動的に行う「多段刊行」のように、内容の識別用付加情報なども重要となろう。中村が指摘しているように著者をも含めた今後の課題であろう¹³⁾。

フルテキストデータベースの作成・提供者にとっては、利用形態、出版形態等を念頭に置いてのフォーマットの設定が課題となる。

5. 1. 2 文献構造

学術雑誌の文献などのように、その使い方・利用の仕方がある程度限られているようなテキストの場合には、文献の構造を考慮したデータ入力が行われていることが望ましく、使いやすくなる。元のデータ自身を尊重するとともに、検索の手法を工夫することが必要となる。

これにより、例えば、文献中の「材料と方法、結果、考察、参考文献、図、表」等の部分を特定した検索および表示やテキスト中での文章と、「引用文献」との組み合わせ表示・出力やブラウジング機能の充実等が可能となる。

したがって、データ入力に際しては、検索の要求仕様に基づく実現機能を念頭に置いて、文献構造の扱いを検討することが大切となる。

実際の科学技術系の文献では、文献構造は雑誌ごとにまちまちであるのが通常であるが、特許などのように、文書構成や記載項目等が予め定められ標準化されているものもある。

フルテキストデータベースでは、このように文献構造や記載内容・表現の違いがデータ入力時の技術的課題や処理量／処理能力等に大きく影響する。

5. 1. 3 インデックス(用語抽出、自動索引等)

フルテキストデータベースは、従来の二次情報データベースでは限界のある原著の生データへの直接アクセスを可能にするところに意義がある。したがって、インデクシング、用語抽出等に関わるデータへのアクセス手法は、全体の利用方式とも深く関連する。最近は、3. 3 項でも紹介したような様々な手法が研究されつつあるが、日本語を

対象とする場合には、英語の場合と異なり、辞書が重要な役割を果たす。

膨大なデータを対象とする場合は、オンライン時のパーシャルマッチや全体をナメル方式ではハード面での限界もあり、用語抽出等の何らかの前処理が必要となると言われている。その点、試作機として昨年発表された日立製作所のフルテキスト・サーチ・マシン (TSM-1) のように、全体をナメル方式であっても文字をビット情報で表したテーブル(文字成分表)や凝縮本文を利用した段階を経た検索により高速化を図っている手法は、一考に値する。

いずれにしても、インデクシングの仕方や体制、データ構造にも大きく関係する問題、課題が含まれている。シソーラスをはじめとする各種用語辞書等を用いてのキーワードなどの自動付加(自動索引)も今後の課題となっている。また、フルテキストデータベースを利用したシソーラスの作成やメンテナンスなどもフルテキストの副次的効果である。将来的課題の一つとして、自動抄録も挙げられる。

5. 1. 4 マルチメディア(文章、数式、図、表、写真、カラー等)

フルテキストデータベースでは図、表を含めての実現が望ましいが、内外のいずれのデータベース関係機関も実現するには至っていない。入力の問題のみならず、検索、ハード、伝送、標準化等あらゆる技術的問題や流通体制上の問題が山積みしている。今後の技術開発課題も、この図、表等の処理関係に多い。

図形データはヒトの概念やイメージの認識結果のデータであり、意味情報が盛り込まれている。共通にその意味要素の分解可能なものはベクターデータに変換し、データベースとして、フルテキスト処理可能となるが、要素分解できないものあるいはしないほうがベターな場合、イメージ画像として扱うしかないのが現状である。ただし、一般イメージ画像(写真等)からの内容抽出は、現在不可能で、予め意味の情報を付加して置くしかない。また、カラーか白黒かでも課題が異なる。

5. 1. 5 キャラクターセット、フォント、レイアウト、その他

キャラクターセットは、出版物、電子媒体のいずれにおいてもすべてが標準化されているわけではない。データ入力時の大きな問題となっている。原文に現れる文字の種類と JIS 文字種（あるいは外字文字種）の関係、表記できない字種や記号等の翻字規則などでの処理方法、使用するハードウェアのサポート範囲、情報流通上での標準化の問題（出版、CTS 化データ、使用文字種・コード、データの交換・インタフェース等）など、データベースの作成と利用の双方にとって解決しなければならない課題がある。

JIS 外字が多いと言われている医学分野でみると、例えば、医学大辞典（4 万語：医歯薬出版（株））での外字登録は約 360 キャラクターとなっている。

文字デザイン、書体（明朝、ゴシック、イタリアック、縮小文字等）、上つき、下つき、数式等のサポートなど、フォントに関わる課題がある。

レイアウトも印刷体では重要な要素となってきたものであり、フルテキストデータベースにおいても、出力時の表示方法などユーザインタフェースの点からも無視できない。データベース作成に際しては、原データの意味上の区分、論理構造を勘案し、入出力系の標準形式 (SIST 等) を定める必要がある。

これらの標準化に関連するものとして、以下の 3 系統がある。

- ① TeX（印刷物作成等）
- ② SGML（文書構造、論理構造等）
- ③ ODA（文書通信等）

国際標準化機構 (ISO : International Organization of Standardization) のなかに、専門委員会である TC (Technical Committee) があり、TC 46 で「情報及びドキュメンテーション」を扱っている。日本国内の受け皿は工業技術院（標準部情報規格課）となっている。

5. 2 データ管理関係

ここでは、実現されるべき技術的課題、検索システム、データ圧縮について述べる。

(1) 実現されるべき主な技術的課題

- ① 従来システム（二次情報）との併用
- ② コンピュータ技術（端末機、大容量記憶、ディスプレイ技術）
- ③ 端末独立性の達成（外字、図、表が同じように出せるか。色、解像度、音声データ）
- ④ 利用者インタフェース（使いやすさ、マルチウィンドウ機能、ハイパーテキスト——付加情報／多段加工／ダウンロード）
- ⑤ データ圧縮（大量データの処理、圧縮と操作との関連）
- ⑥ 論文の標準化（分野別の形式、コンピュータ可読の論文形式、抄録（多段要約、自動抄録））
- ⑦ データ比較、評価機能

(2) データモデル

フルテキストデータベースでのデータモデルの適性については、以下のように言われている。

- ① 使い途の自由度を重視した場合は、不定長文字列をデータとして扱え、文字列処理機能を付加した関係モデル。
- ② 文章の構造がある程度決まっており、使い方も関連する部分をたどるだけである場合には、網モデルに近いハイパーカードのようなもの。
- ③ マルチメディアを扱うとき、ある程度はハイパーカードで扱えるが、図表、写真、グラフ等の意味に立ち入った検索を必要とするときには、オブジェクト指向モデル。

(3) 検索システム

フルテキストデータベースでは、検索機能として、従来の二次情報データベースで実現されている機能に、新たな機能を追加する必要がある。いずれのシステムにおいても共通する主な機能として、次のようなものが挙げられよう。研究や技術開発を伴うものが多々ある。

- ① 文献構造への自由なアクセスを可能とする機能
論文の章だて、章のつながり等がわかること。

自分の欲するテーマ、内容に関する文章や段落等が取り出せること。参考文献、図、表等へのアクセスができること。図、表、写真等を含んだマルチメディア情報源に対する機能は、研究や技術開発の余地が大きい。

②文字列検索機能

文章中のある文字列を精度よく検索できること。日本語の場合は、英語と異なり、日本語特有の技術的課題が伴う。

③検索結果の出力表示

特に、図、表、写真等の出力表示の方法。

④検索の高速化

大量のデータを対象とする場合は、単純な検索であっても高速処理のための技術開発が必要となる。

⑤検索結果の解析、比較、編集処理機能

フルテキストデータベースの高度利用と関連。新たな利用の可能性へつながる。

現在、パソコン用のフルテキストを入れるデータベースが出ているが、図形は大抵の場合扱えない。

(4)データ圧縮（大量データの処理）

フルテキストデータベースでは、大量のデータの処理・蓄積が必要となるために、データ圧縮技術が最も重要な技術的課題の一つとなっている。圧縮の方法が検索時の使い勝手や操作性に関係する。検索対象ファイルの作り方、検索時の速度に大きく関連する。情報蓄積媒体の性能／コスト、関連技術の進展度合いに応じた手法が採用されているのが現状である。あらゆる技術を総動員しての開発となっている。

ちなみに、現在 JICST で収集している雑誌等の総頁数は、年間で約 1,200 万頁にも及ぶ。これだけの量を一挙にフルテキスト化し、自由自在なオンライン検索を可能にすることは、ハードウエ

ア的にもコスト的にも、また検索速度や検索精度等の面からも不可能である。したがって、特許 (MEAD, JAPIO) の例にみられるように、大量データの場合には光ディスクや CD-ROM にデータを蓄積しておき、オンラインデータベースで検索した結果を特許番号等を介して CD-ROM 等から詳細な全文データを引き出し、郵送またはファックスで伝送する手法が採用されている。

6. お わ り に

「これからの情報担当者」というテーマのなかで、一次情報としてのフルテキストデータベースが、その利用が現在必ずしも多くない割りには期待され、また、何故にデータベース提供機関をはじめとし少なからぬ情報処理関係の研究者が注目しているかを念頭に置いて述べた。

このため、現状のサービスの紹介よりも、フルテキストデータベースに立ちはだかっている技術的課題やその秘めている可能性、しいて言えば、「フルテキストデータベース」そのものの理解に重点を置いて述べてみた。

現場の実務での利用に重点を置いたデータベースの紹介は、別の機会に譲りたい。

いずれにしても、3.3 項で述べたオフィスレベルや個人レベルでの文書検索にみられるように、当面この世界でのフルテキスト化が先行していくのか、あるいは大規模な商用オンラインデータベースが遅れることなく期待に答えていくのか、90 年代には結論が出ることは間違いないと思われる。

最後に、JICST の情報科学技術委員会（委員長 中村幸雄氏、分科会主査 藤原譲教授）の各委員をはじめ諸氏に多くの貴重な知見を頂いたことを付記しておく。

参 照 文 献

- 1) Herty J. A., Rohrbaugh V. K., Current state of full text primary information online with recom-

- mendations for the future. Online Rev. 13(2) 135-140(1989)
- 2) Herty J. A., Smith C. G., Marketing strategies for promoting online full-text primary information. Proc Natl Online Meet. Vol. 1988 91-97(1988)
- 3) 原田 勝. 全文データベースドキュメント・デリバリーの観点から. 情報の科学と技術, 39(7)247-256(1989)
- 4) 科学技術庁. 総合レビュー 先端科学技術の現状と展望 第3巻 データベースの先端技術 142-167 1989
- 5) 学術情報センター. 全文データベースのシステムに関する現状調査報告書(部内資料). 1987
- 6) 根岸正光. CTS とデータベース. JAGAT テックセミナー「印刷メディアからマルチメディアへ」講演. 1989
- 7) 日本科学技術情報センター. 情報科学技術委員会報告書「フルテキストデータベースに関する調査・検討報告書」(部内資料). 1989
- 8) 日本科学技術情報センター. 情報科学技術委員会報告書「フルテキストデータベースに関する検討報告書」(部内資料). 1990
- 9) 稲垣博人. 文書内容検索システムにおける現状と課題. 情報管理, 33(2)123-135 1990
- 10) 米田健二. メタモルフ (METAMORPH) ーテキストを理解する人工知能ソフトウェア. オンライン検索, 10(3)117-126. 1989
- 11) 星 暁雄. 文字列検索ハードウェアの試作相次ぐ, 処理速度は 1,000 万文字/秒以上. NIKKEI ELECTRONICS no. 480. 109-111. 1989.8.21
- 12) 情報処理学会. 情報処理学会研究報告 89-FI-14 特集「フルテキスト・データベースのご利益はどこにあるか?」. 1989.7.27
- 13) 中村幸雄. 一次資料の復権 一つのトレンド分析. 情報の科学と技術, 39(1)19-22(1989)