
学術情報センター紀要

Research Bulletin of
the National Center for Science Information System

第 2 号

1989年3月

学術情報センター

学術情報センター紀要 第2号

目 次

刊頭のことば	山田 尚勇 (学術情報センター研究開発部長) ……………	i
研究論文		
プログラムデータベースを用いたソフトウェア開発支援システム	山崎高日子 (東京大学)、濱田 喬 (学術情報センター) ……………	1
セルフルーティング形高速パケット交換の一方式	浅野正一郎 (学術情報センター)、酒井 保明、崔 明興 (東京大学) ……………	15
UNIMARC - Authority の課題	内藤 衛亮 (学術情報センター) ……………	25
学術分野における機械可読文書の作成と通信	根岸 正光 (学術情報センター) ……………	43
論文集目次型データベース ISTP&B のための検索システムの設計	牧村 正史 (名古屋大学)、根岸 正光 (学術情報センター) ……………	53
日本語医学用語の構造解析	小山 照夫 (学術情報センター) ……………	85
学術用語データベースの構築	大山 敬三 (学術情報センター) ……………	93
ストリング・インデクシング・システムの現状——PRECIS を例として——	影浦 峡 (学術情報センター) ……………	107
資料		
パトリシア・バットン女史の講演記録	内藤 衛亮 (学術情報センター) ……………	123
学術情報センター日誌	……………	151

**Research Bulletin
of
The National Center for Science Information System**

March 1989 Volume 2

Contents

<i>H. Yamada</i>	i	Preface
 Contributions		
<i>J. Hamada and T. Yamazaki</i>	1	A Support System for Software Development based on Program Database
<i>S. Asano and Y. Sakai M. H. Choi</i>	15	A Study on High Speed Pocket Switching by Self Routing Control
<i>E. Naito</i>	25	Poblems of UNIMARC - authority
<i>M. Negishi</i>	43	Preparation and Telecommunication of Machine - Readable Documents in Japanese Academic Community
<i>M. Makimura and M. Negishi</i>	53	Development of NACSIS - IR Intormation Retrieval System for ISTP&B (Index to Scientific and Technical Proceedings and Books) Database
<i>T. Koyama</i>	85	A Study for Structural Analysis of Japanese Medical terms
<i>K. Oyama</i>	93	Database Construction of the Scientific terms
<i>K. Kageura</i>	107	Recent Problems in String Indexing Systems —An Analysis of PRECIS—
 Material		
<i>E. Naito</i>	123	Report on Lecture by Ms. Patricia M. Battin
Diary	151	

刊 頭 の こ と ば

学術情報センター研究開発部長

山田 尚勇

日本の学術活動、特に理工学系分野における成果が、世界に肩を並べる実力を見せだした 1970 年代の中ごろから、それに伴って、日本を含む世界の学会の学術情報の自由で急速な流通が、どこの国においても、国際的に評価できる研究活動に不可欠なものであることが痛切に認識されだした。

そうした学術情報の国内的、そしてゆくゆくは国際的流通の拠点として、有識者によって構想され、学術審議会や関係省庁の積極的支援のもとに、1983 年に発足したのが、学内共同教育研究施設としての東京大学文献情報センターであった。その後いち早く、翌年には全国共同利用施設に改組された。そして、それらの全 3 年にわたる運用経験と実績とにもとづき、1986 年に文部省直轄の全国大学共同利用機関として設置されたのが学術情報センターである。

学術情報センターの目的は「学術情報の収集、整理及び提供並びに学術情報及び学術情報システムに関する総合的な研究及び開発」と幅広く規定されてある。これらセンターの使命のうち、研究および開発を担当しているのが研究開発部であり、本紀要はセンターを中心としたこの分野での研究活動の成果の主たるものを広く一般に公開してご参考に供し、かつご批判をいただく目的で、1987 年に創刊された。

現在において見通しのきいている情報関連技術の発展動向は、さまざまに新しい、情報流通形態の近未来における実現を示唆している。

近年は学協会誌なども、まず原稿を一度機械可読の形に入力してから編集、割り付けなどを行ない、最後に印刷原版を電算写植によって製版するものが増えてきている。ゆくゆくは大方の出版物がそうした形態をとることになると思われる。現在脚光を浴びている卓上出版などもその簡易版といえるであろう。また著者からして、大方が機械可読の形で入力、推敲したものをそのまま投稿する時代もやがて来るであろう。

それと同時に、パソコンや端末機も急速に普及し、また通信端末としての性能も格段に向上してくる。そうなると学術情報なども、雑誌としていちいち紙に印刷し、配布することは必ずしも必要でなくなり、遠隔地の情報センターにある機械可読の全文章と画像のデータベースとから、通信回線を通じて端末機のディスプレイ上でページ合成すれば、たいいてい目的には充分である。学術情報センターにおいても、今年はず一部化学分野の論文から、通信網を介してこの全文データベースの検索サービスを始める。

こうした通信網を通しての大規模情報サービスが、これから広く展開し、普及していく一方において、最近音楽の LP レコードに取ってかわりつつあるコンパクトディスク (CD) をコンピュータ用と

した、CD-ROM を使った情報サービス技術が急速な発展を見せている。たとえば、アメリカの IEEE は会員 30 万人を数える、電気・情報関連の大学会であるが、今イギリスの同種学会 IEEE およびアメリカのユニバーシティ・マイクロフィルム社と組んで、IEEE が 1988 年度に出版したすべての学会誌、論文集などの 20 万ページを、所載画像まで含めて CD-ROM 化する作業を進めている。今回はそれを CD-ROM 25 枚に収めるが、文字の符号化と画像データの圧縮をした次世代版では、それが 2、3 枚に収まってしまう。しかも全文テキストを、内容によって検索することができる。

そればかりか、昨年から今年にかけて実用化に動きだした DV-I (digital video interactive) では、同じ大きさの円盤のなかに映画約 1 時間分の動画を記録できるばかりか、話の筋に選択肢があり、使用者が筋の流れを対話的に自由に制御することさえできる。

一般に CD や CD-ROM は、一度母型さえ作れば、量産時に 1 枚をプレスする原価は数十円のオーダーにまで下げられるという。その 1 枚が、色彩写真を含む 10 万ページ分の、機械検索可能な情報を運んでくれるのである。パソコンなどに追加するリーダーも、DV-I 用含めて、量産時には原価 1 万円近くにまで下げられる見通しが示されているという。このような製品が家電製品として普及する近い将来には、それが専門家のあいだだけでなく、広く一般大衆のあいだでも多くの情報をどんどん流通させる強力な媒体として使われるようになることは明かである。

たとえば、傾向の似た多くの雑誌の数種から数十種が一緒になって、前年度分の総ページを 1 枚の CD-ROM に収めたものを、おのおの雑誌の年間購読者のみに、付録として配布するというサービスが経済的に可能になるばかりでなく、そうした企画に加わらないものは市場において脱落するかもしれないという、厳しい生存競争状態が生まれることさえ予想される。

ともあれ、近い将来における情報の流通形態としては、比較的需要の大きいものは CD で、残りの特殊な専門的内容のもの、また特に即時性を要求されるものは、学術情報センターのような、データベースセンターからの通信網を通じたサービスで、と二分極化していくことが考えられる。

最近では日本の工業製品の高品質と適正価格とが国際的経済力の急激な増大をまねき、遂には国際摩擦にまで発展しだした。それと同時に、日本は海外から情報を吸収するばかりで、外へは出し惜しみをしているというような国際的意見さえ聞かれるようになってきた。

しかし、客観的にみれば、わが国のように学術用言語である日本語のほかに、外国語でも別途学術雑誌を多量に製作しているばかりか、これほど各種の日刊新聞をわざわざ外国語で出している国は世界でも珍しい。事実、本格的調査の結果によっても、国際的評価の高いデータベース INSPEC、Chemical Abstracts、COMPENDEX、EMBASE の対象としている自然科学の 24 研究諸分野で、日本は国として 2 位か 3 位の論文収録数を誇っている。

こうした事実を国際的に理解してもらうことにより、国際摩擦を解消すると同時に、先端技術を用い、学術情報を世界的に、より一層広範に提供する方向を求め続けるという使命をになう学術情報センターの責任には、まことに重大なものがある。

研 究 論 文

プログラムデータベースを用いたソフトウェア 開発支援システム

A Support System for Software Development based on Program Database

学術情報センター 濱田 喬

東京大学大学院工学系研究科 山崎高日子

要 旨

プログラム開発の効率化をはかるために、既存のプログラムモジュールをデータベース化し、これを能率よくアクセスすることによってソフトウェア作成を支援するシステムを開発した。プログラムモジュールは、意味表現リスト、仕様、ソースプログラムからなり、特定の言語には依存しない仕様による自動合成も行なわれる。

Abstract

A support system for software development using a database of function modules is described. There are three levels of an expression for each function module, that is, a semantic expression to be used for retrieval, a function specification that is independent of specific programming language, and a source program. Program modules can be retrieved to be reused for composing another function, and the general specification is translated into a source program.

1. はじめに

近年、ソフトウェア需要の増大に伴い、プログラム作成の効率化、再利用に関する研究が活発に行なわれている。一方、計算機をより身近で使いやすいものにするため、使用しやすいマン・マシン・インタフェースの必要も認識されている。本報告では、このような目的に沿ったプログラム開発支援システムを紹介する。このシステムにおいては、対話形式で入力を行ない、プログラムを関数単位で部品化して、データベースに蓄積することにより、プログラムの生産性を向上させ、プログラムの有効な再利用を図ることを目的としている。

2. 従来方式の問題点

人手によるプログラム作成は、数行程度の要求仕様から、最終的に何十ステップかのプログラムへとトップダウン的に変換していく作業とみなすことができる。プログラム合成システムを考える場合、どのレベルから先を機械に任せるかということが問題になる。理想をいえば、トップレベルでの要求仕様をシステムの入力にしたところである。しかし、これは要求仕様の理解だけでも膨大な知識ベースが必要なことが予測され、かつ、仮に理解したとしてもこれを個々のステップに分解していくには、推論が不可欠である [1]。

現状では、一番下位のレベル、すなわち、入力に用いられる仕様言語が、生成されるプログラミング言語に極めて近い、仕様言語とプログラミング言語間の言語トランスレータといったタイプのシステムが開発されている程度である [2]、[3]。

これは、上述の理由からも止むを得ないところであろう。

3. システムの特長

より上位のレベルからのアクセスという観点からデータベースを用いて、関数を単位としたプログラムの合成と再利用のシステムを作成することにした。

このような、いわゆる関数の部品化の研究は盛んであるが、本システムでは、一つの関数につき、TOP → DOWN に

- ①関数の機能を簡単に説明した意味表現リスト（検索に利用される）
- ②仕様言語で書いたアルゴリズム（特定言語に依存しない）
- ③具体的なプログラミング言語で実際に書かれたプログラム

という、3つのレベルを導入した（図1）。

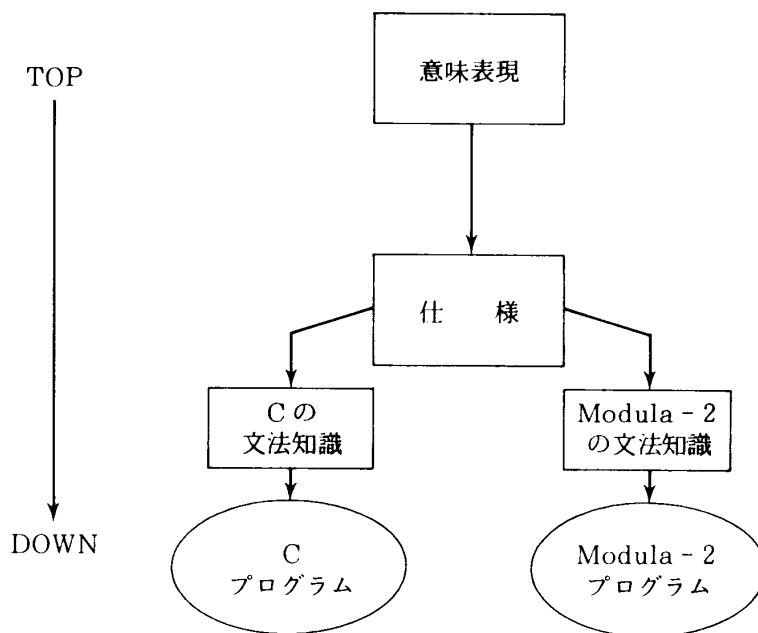


図1 3つのレベル

仕様（レベル②）→プログラム（レベル③）

への変換は、システムがサポートするが、第2章で述べた理由から、

意味表現（レベル①）→仕様（レベル②）

への変換は、このシステムではサポートしない。そのかわり、①と②のデータベースをリンクしておくことにより、①からは上位のレベルに、また、②からは下位のレベルに、と、両面からプログラムの生成を支援しているのが、このシステムの大きな特長である。

4. システムの概要

システムは次のような概要をもっている。

- (1)データベースに標準関数、および今まで作られたユーザ関数を登録しておく。ユーザの要求により、該当する関数を検索する。もしなければユーザが新たに作ることになる。
- (2)ユーザ関数の作成は、一般に、本システム専用の仕様記述言語 SP で記述する。この仕様言語は特定のプログラミング言語に依存していない。
- (3)仕様記述は、システム内のコード生成モジュールによって、目的の言語のプログラムに変換される。
- (4)関数の組み合わせという特徴を持った言語 C および Modula-2 を対象言語としているが、同様な性質を持つ、Ada 等にも応用できる。すなわち、言語に応じて、仕様言語 SP から、その言語へのトランスレータをコード生成モジュールに入れてやればよい。
- (5)ユーザにより作成された関数も新たにデータベースに登録されるので、プログラムの再利用が可能であり、システム自体が拡張していく。
- (6)意味表現を用いて検索を行なう、知的システムである。

5. システムの構成

システムの構成は、図2. のようになる。

- ①ユーザ入力部
- ②関数データベース
- ③コード生成モジュール

について、具体的な仕様を以下に説明する。

5. 1. ユーザ入力部

ユーザの入力モードには、次の3つがある。

- ①検索モード
- ②SP 入力モード
- ③言語入力モード

本システムの目的である、関数単位でのプログラムの合成の支援には、主に①、②を用い、③は、

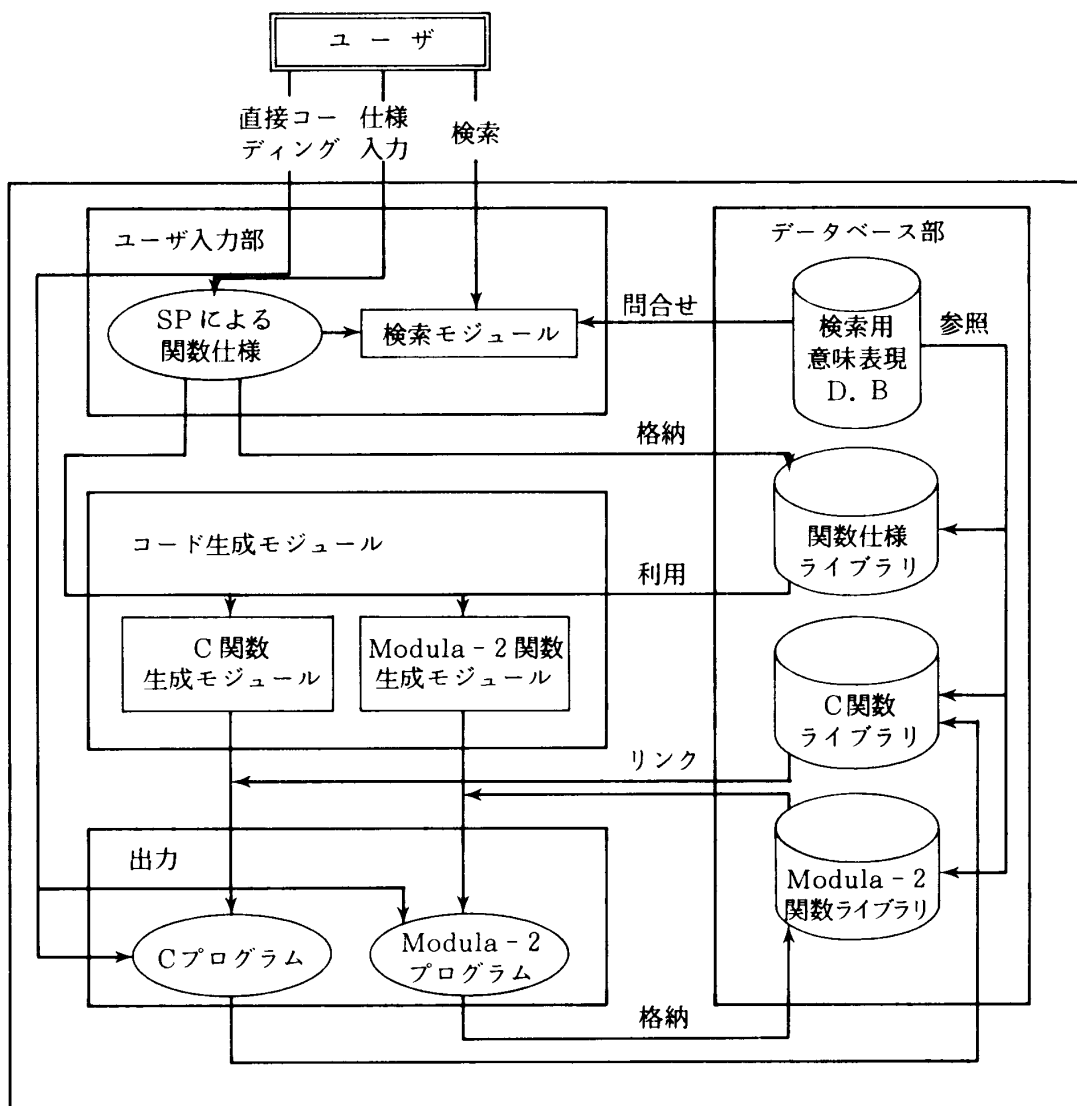


図2 システム構成

main プログラムなど、データや入出力への依存が強く、再利用性の低い部分を記述する時に、通常のプログラミングと同様に使用できるために存在する。

以下に、それぞれの機能を説明する。

①検索モード

検索モードでは、意味表現により、関数を問い合わせることになる。

図3. には、qsort (クイックソート) 関数を検索した例を示す。ユーザが“request” コマンドを送ると、システムは、その関数の意味表現リストを生成するように、対話的に関数の機能を聞いてきて、その意味表現に基づき、意味表現データベースを検索する。一致する意味表現があれば、その関数名を候補として挙げる。ここで言う意味表現のリスト形式については、次節のデータベース部の説明中で述べる。

検索では、図3. に示すように、意味表現中のキーワードを増やしていくことにより、候補を絞り込んでいくことができる。

それでも、候補が複数ある時は、ユーザに選択を任せる。選択のためのガイドとして、標準関数の

```

-> (request)
function_name ? qsort
subject ? qsort
act ? sort
object ? character
else ? y/n y
  attribute ? by
  value ? quick_sort
  attribute ? end
your request function is
(qsort (subject (qsort))
      (act (sort))
      (object (character))
      (by (quick_sort)))
search starts

search by depth-1 (act) ? y/n y
(shell (subject (sort))
      (act (sort))
      (object (integer))
      (by (shell_sort)))

(qsort (subject (qsort))
      (act (sort))
      (object (character))
      (by (quick_sort)))

search by depth-2 (subject) ? y/n y
search starts
(qsort (subject (qsort))
      (act (sort))
      (object (character))
      (by (quick_sort)))
-> (exit)

```

図3. qsort関数の検索例

ときはそのマニュアルを、ユーザ関数の場合は、希望により、SP で記述された仕様、あるいは、ソースコードも出力する。

候補がない場合、あるいは直接ユーザが関数を作りたい場合、次に述べる、SP 入力モード、あるいは、言語入力モードに入り、関数をコーディングする。その途中で新たに関数を作りたい時、検索モードに戻って、先と同様に意味表現データベースにアクセスする。

②SP 入力モード

検索により関数がなかった場合、あるいは関数仕様ライブラリに有ったとしても、SP による仕様に何らかの修正を加える必要がある時（パラメータの型の変更など）に、ユーザは、このモードを利用する。具体的な、仕様言語の記述例、文法規則については、それぞれ後述の、関数データベース、コード生成モジュールの説明において述べる。

③言語入力モード

通常のプログラミングが可能なモードである。使用している計算機システムの一般的なエディタを利用する。

5. 2. 関数データベース

関数データベース部には、前述した、図 1. の 3 つのレベルに対応する 3 種類のデータベースが存在する。上位のレベルから順に

- ①検索用意味表現データベース
- ②関数仕様ライブラリ
- ③関数プログラムライブラリ

である。

ユーザにより新たに作成されたプログラムも、それぞれの属性について、対応する上記 3 つのデータベースに登録され、再利用が可能になる。以下に、3 つのデータベースの機能について説明する。

①検索用意味表現データベース

このデータベースは、意味表現リストの集合である。意味表現リストは次のようなフレーム構造になっている。

意味表現リスト：：（関数名 意味表現）

意味表現：：（subject value）

 （act value）

 （object value）

 ：

 ：

 （その他 optional な属性）

act の属性値は、その関数の機能を示しており、subject の属性値は、更に細かい分類の対象になっている、act の属性値が等しいものは、同じファイル（ファイル名=act の属性値、になっている）に格納するようにして、検索の効率を良くした。

図 3. の、qsort を検索する例では、ユーザの要求中の、act (sort) によって、sort に関するデータが入っているファイル（名前は、sort である）にアクセスして、候補のリストを出している。更に、ユーザが対象を絞りたければ、要求者の、subject (qsort) によって、その中から適合するものを選び出す。

②関数仕様ライブラリ

仕様言語 SP で書かれた関数仕様から、色々なプログラミング言語のプログラムに変換できるというのが、このシステムの大きな特長であるが、そのためには、SP が、特定のプログラミング言語に依存せずに、アルゴリズムを記述できる必要がある。

しかし、全ての言語に変換できるというのは、事実上無理であるし、またシステムが、プログラムを関数単位にモジュール化して扱うという主旨を踏まえると、構造化可能な言語に的を絞っても差し支えない。

そこで、仕様から変換可能な言語は、構造化が可能で、かつ、手続き型の、C、Pascal 系 (Modula-2、Ada など) に限定して、言語 SP を設計した。

SP による、qsort のアルゴリズム記述例を、図 4. に示した。各モジュールは、[] でくくられた単位になっている。

例えば、最初のモジュールは、main 関数である（この部分は、データの個数、入出力に依存するので、プログラミング言語で直接書いた方が一般的であるが）。

モジュール前半部の、

```
PNAME>    value ;
           :
           :
INTFVAR>  value ;
```

までが、モジュールの形式的な型の記述である。

```
FSUB>
```

に対する属性値には、そのモジュール中で呼び出している関数名があてはまる。図 4. の例では、関数名が、\$\$…\$\$ で囲んであるが、これは、その関数名が確定していない仮の名であることを表わす。コード生成部が、これを走査すると、その関数を検索する検索モードに入るようになっている。


```

[
  PNAME> qsort;
  HEADER> stdio.h
  PARAMETERS> ;
  FSUB> $$ qsort () $$ ;;
  INTFTYPE> item::STRUCT (key:INT;);
  INTFVAR> table[100]:item;

  VARDEC number, i:INT;

  BODY
    LOOP infinite (;) {
      WRITE("How many data ? ");
      READ(number);
      EXIT infinite(number < 0);
      $ printf("%ninput data\n"); $;
      LOOP ( i <- 1; ++ ) {
        READ(table[i].key);
        EXIT(i > number);
      }
      $ printf("%nsorted data\n"); $;
      $$ qsort(number) $$;
      LOOP ( i <- 1; ++ ) {
        WRITE(table[i].key);
        EXIT(i > number);
      }
    }
  END
]

[
  FNAME> qsort;
  RETURN> ;
  PARAMETERS> n:INT;
  FSUB> sort();
  INTFTYPE> item::STRUCT (key:INT;);
  INTFVAR> table[100]:item;

  VARDEC ;

  BODY
    sort(1,n);
  END
]

[
  FNAME> sort;
  RETURN> ;
  PARAMETERS> l, h:INT;
  FSUB>;
  INTFTYPE> item::STRUCT (key:INT;);
  INTFVAR> table[100]:item;

  VARDEC w:item;
  l, j, m:INT;

  BODY
    l <- 1;
    j <- h;
    m <- (l + h) / 2;

    LOOP(; ) {
      LOOP(; l < l + 1) { EXIT(table[l].key >= table[m].key)}
      LOOP(; ) { EXIT(table[m].key >= table[j].key;
        j <- j - 1; )
      }
      ?? i =< j --> { w <- table[l].key;
        table[l].key <- table[j].key;
        table[j].key <- w;}
      EXIT(i > j)
      ?? l < j --> sort(l, j);
      ?? l < h --> sort(l, h);}
    }
  END
]

```

図4 qsort関数の仕様記述例

また、

INTFTYPE >

INTFVAR >

には、それぞれ大域的な、データ型、変数が入る。これが、結合されるモジュール同士で整合している必要がある。

このモジュール前半部は、例えば、仕様を

Moldula-2 へ変換する際、definition module を作るのに役立つ。

SP における予約語は、大文字で書く。また、直接プログラミング言語で記述したい部分については、\$...\$ で囲む。

SP の文法規則については、後述のコード生成モジュールの節で示すが、ここでは、個々のプログラミング言語ごとに、いろいろな記述規身が存在する制御構造について、SP での、統一した記述法を説明する。

ループの記述は、基本的に 1 つの形で表現できるので、これを、LOOP-EXIT の形に統一する。

LOOP の構文は次のようになる。[] 内は省略可能を示す。

[ループ名] LOOP

([初期値 <文 1>];

[ループを繰り返す際の実行文 <文 2>])

{

ループ本体

}

例えば、Pascal の、for 文のような単純な繰り返しでは、文 2 は、++ (1 つ増やす)、-- (1 つ減らす) という、省略形を用意する。

また、これに対応する、EXIT の構文は次のようになる。

EXIT [脱出するループ名] (脱出条件) :

これにより、ループに名前がついていれば、ループの入れ子からも抜け出せるようにする。

また、通常 IF 文で代表される条件判断文については、

?? cond1 --> stat1

? cond2 --> stat2

? --> stat3

```

/*****
 * lexical analysis
 *****/
%%
[\\n \\t] ;
[0-9]
PNAME">"
HEADER">"
FNAME">"
RETURN">"
PARAMETERS">"
FSUB">"
INTFTYPE">"
INTFVAR">"
VARDEC
BODY
END
LOOP
EXIT
INT:CHAR:REAL
AND:OR
"="/=":"<">=":"="
"*"/:";MOD
"."
"'"
[A-Za-z] (A-Za-z0-9)*
%%

return( DIGIT );
return( PNAME );
return( HEADER );
return( FNAME );
return( RETURN );
return( PARAMETERS );
return( FSUB );
return( INTFTYPE );
return( INTFVAR );
return( VARDEC );
return( BODY );
return( END );
return( LOOP );
return( EXIT );
return( VTYPE );
return( BOOLOP );
return( EQOP );
return( MULOP );
return( PERIOD );
return( PTR );
return( LETTER );

/*****
 * syntax analysis
 *****/
% (
#include <stdio.h>
#include <ctype.h>

% )
%start spec

%token DIGIT LETTER
%token PNAME HEADER FNAME RETURN
PARAMETERS FSUB INTFTYPE INTFVAR
%token VTYPE
%token VARDEC BODY END
%token LOOP EXIT
%token EQOP BOOLOP PLUSOP MULOP UMINUS

```

図5 SPの文法規則(その1)

plist	:	plm	:	intftys	:	intftys	:
	:	plst plm	:		:	intfty	:
	:		:		:		:
pelm	:	vlist vtypes ':'	:	intfty	:	intfelm	:
	:		:		:	intfty intfelm	:
vlist	:	fvid	:	intfelm	:	LETTER ':' 'struct' '(' plists ')'	:
	:	vlist ':' fvid	:		:		:
	:		:	decbody	:	vardec body	:
fvid	:	vid	:		:		:
	:	dcfids	:	vardec	:	VARDEC plists	:
	:		:		:		:
vid	:	LETTER	:	body	:	BODY `slist` END	:
	:	darray	:		:		:
	:		:	slist	:	stats	:
arsizes	:	arsize	:		:	cntl slist	:
	:	arsizes arsize	:		:	'(' plstats ')'	:
	:		:	plstats	:	stats	:
arsize	:	'(' number ')'	:		:	plstats stats	:
	:	'(' number '->' number ')'	:	stats	:	stat ':'	:
	:		:		:		:
dcfids	:	dcfid	:	stat	:	/* empty */	:
	:	'\$\$' dcfid '\$\$'	:		:	expr '<-' expr	:
	:		:	usfids	:	usfids	:
dcfid	:	LETTER '(' farg ')'	:		:	usfid	:
	:		:		:	'\$\$' usfid '\$\$'	:
farg	:	/* empty */	:	usfid	:	LETTER '(' aarg ')'	:
	:	plst	:		:		:
	:		:	arg	:	/* empty */	:
vtypes	:	' ':' vtype	:		:	explist	:
	:		:		:		:
vtype	:	ptr VTYPE	:		:		:
	:		:		:		:
ptr	:	/* empty */	:		:		:
	:	ptr '^'	:		:		:

図5 SPの文法規則 (その2)

のように、視覚的、直感的にわかりやすい記述法にする。

その他、配列宣言は、Pascal などでは宣言が長くなるが、SP では、

```
table [10] : int ;
```

のように書く。

構造体（レコード）についても、

```
item : : struct {key : int ; name : char ;}
```

のように、明解な記述法にする。

③関数プログラムライブラリ

SP による仕様からコード生成モジュールで、プログラミング言語に変換された関数、あるいは、ユーザが、言語入力モードで直接プログラミングした関数は、ここに格納され、再利用が可能になる。

5.3. コード生成モジュール

SP で記述された仕様を入力して、具体的なプログラミング言語を出力する部分である。

まず、入力された仕様記述が、SP の文法と一致しているかをチェックする。この部分の作成には、UNIX のユーティリティである、プログラムジェネレータ、Lex（語い解析処理用）および Yacc（構文解析処理用）を利用する。

図 5. は SP の語い規則および文法規則を、Lex および Yacc の入力形式にそって記述したものである。これが、Yacc を通ったことにより、SP の文法規則が一意に定まることが示された。トランスレータ部は、C により、現在作成中である。

6. 今後の課題

仕様の記述において、変数の型を定めず、コード生成の時に、ユーザの指定によって型を定める機能をつけること、配列のサイズもコード生成時に指定できるような機能をつけることにより、仕様の汎用性が高まるので、その 2 点を検討中である。

7. おわりに

本システムでは、関数をモジュール化してデータベースに入れることを、基本目的としているが、個々の関数について、検索に用いる意味表現、特定言語に依存しない仕様、実際のプログラムという、3つのレベルを設けていることが、大きな特長となっている。これにより、プログラムの合成と再利用を効率よく行なえることを示すことができた。

〈〈参考文献〉〉

- [1] 伊藤貴康：プログラムの自動作成情報処理、Vol. 19, No. 10, pp. 993-1002 (1978)
- [2] 市川、榎本他：仕様記述言語 TELL/NSL における仕様記述からプロトタイププログラムへの変換、情報研報、86-SW-46 pp. 9-16
- [3] 西塚、松山、伊藤：日本語によるプログラム合成の試み、61年度信学会全国大会論文集 PART8 pp. 223-224

研 究 論 文

セルフルーティング形高速パケット交換の一方式

A study on high speed packet switching by self routing control

学術情報センター 浅野正一郎*

東京大学 酒井 保明

東京大学 崔 明興**

要 旨

本論文では、セルフルーティング制御を採用した高速・大容量パケット交換方式を提案している。本方式により、データ通信のみならず、音声・画像の統合通信網の形成が可能となる。本パケット交換の特性を解析した結果、100 G ビット/秒の交換処理を可能とすることを明らかにしている。

Abstract

This paper proposes a high speed and high capacity packet switching system by self routing control, which accommodates various communication services, such as data, voice and video. This system is composed of Beneš network based on newly proposed routing scheme including transmission strategy. The paper also shows that the proposed system features low packet loss due to collision on the switch, a little delay time as well as adequate amount of hardware. The proposed system will realize a high performance packet switching system which has a throughput of 100 G bits per second.

1. はじめに

現在、世界各国で ISDN の構築が行なわれている。今後、ISDN の進展及び新しいメディアの登場などにより、通信量がますます増大すると思われる。また、画像など大量のデータの伝送等、広帯域交換に対する要望も大きい。これら速度や発生パターンの異なる多様な属性のデータを一元的に扱うものとして、高速大容量パケット交換が注目を集めている^{[1]-[4]}。

これらの実現には、毎秒 10^6 - 10^8 パケットの処理容量を持つパケット交換機を必要とするが、従来の

* Asano, Shoichiro : National Center for Science Information System.

**Sakai, Yasuaki, Choi, Myun Heung : University of Tokyo.

ような複雑なソフトウェア制御による蓄積交換では、プロセッサの処理能力とメモリのアクセス速度の限界により、対応できない。そこで、網全体を高度な処理を行なう端末—網間インターフェイスの部分（エッジ部）と簡素なプロトコルによる網内のパケット転送部分（コア部）に分離し、コア部ではハードウェアスイッチによるセルフルーティング制御を行なうという方式が有望視されてきている¹¹⁾。

コア部における交換機には次のようなことが要求されている。まず、多様な属性の情報（データ・音声・画像・映像）を運ぶため、

- ①様々なトラフィックに対して安定している
- ②遅延及び遅延変動が小さい
- ③パケットの棄却率が小さい

ことが必要である。また、高集積・高速化するため、

- ④ハードウェアスイッチの論理を簡単にし
- ⑤メモリやスイッチ数を極力減らすことによりハードウェア量を抑えなければならない。

本稿では以上のことを考慮して、新たな制御法によるベネス網を用い、交換機の内部で自動再送処理を行なうことを中心とした高速大容量パケット交換方式を提案し、検討結果を示す。

2. 多段リンク接続網における衝突検知再送法

β エlementによる多段リンク接続網でパケットをセルフルーティングさせる場合、パケット同士が衝突が起これることによってパケットが目的の出線に到達できなくなることがある。どのパケットが目的地に到達し、どのパケットが失敗したかを簡単な方法で検知し、再送を行なう方法を示す。

2.1 構成法

この方法は、 β エlementを用いるすべての網に応用できる。

まず、各 β エlementにパケット転送線に加えて逆方向の信号線を設ける。この信号線自体もスイッチングを行ない、その状態はパケット転送線の状態と同じとする。つまりこの new- β エlementは、図1(a),(b)の2つの状態をとる。信号線同士の接続形態も、パケット転送線と同じとする。

システムの全体図を図2に示す。多段リンク接続網と中継回線の間には、入線側に SSRU (Speed-up, Slotting and Retransmission Unit)、出線側に SDU (Speed-down Unit) を設ける。SSRU は、回線速度をあげ、他の回線とのスロット同期をとり、2.2に示すアルゴリズムに従ってパケットの送出及び再送を行なう。SDU はパケットが到着したことを SSRU に知らせ、回線速度を元に戻す。

2.2 動作アルゴリズム

SSRU に到着したパケットは SSRU 内のバッファに蓄えられ、各スロットごとに送出される。送出されたパケットは網内をセルフルーティングされて目的出線に向かう。そして目的出線にパケットの先頭が到着すると、SDU が信号線に信号を送る。この信号はパケットの通ってきた道順を逆にたどり、そのパケットが送出されている SSRU に到着する。SSRU は信号が戻ってくるとそのまま最後ま

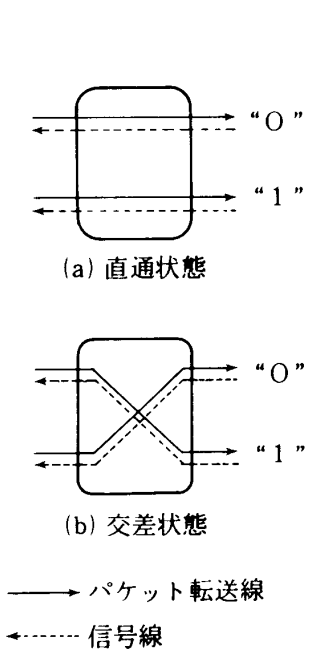


図1 β 要素の状態

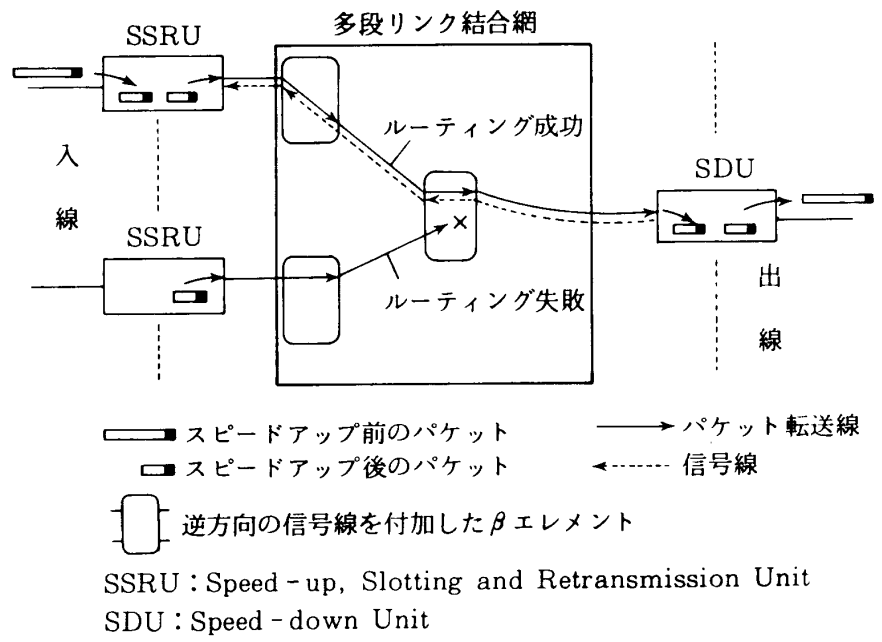


図2 衝突検知再送システム

でパケットを送出し、ある一定の時間内に戻ってこない場合、送出を中止して次のスロットまで待ち、再送を行なう。

(アルゴリズム終)

以上のアルゴリズムに従えば、どのパケットが目的出線への到着に成功し、どのパケットが失敗したかを検知し、ただちに再送を行なうことができる。以後、この方法を CDR (Collision Detection and Retransmission) 法と呼ぶ。

バッファを用いてパケットの棄却率を下げる方法として、各 β 要素にバッファを持たせる方法があるが、ここに提案した方法によりバッファを網の端に集約して設置することにより、バッファ量を大きく減らせると考えられる。また、この方法の有利な点として、回線の高速化とも整合性が高いことがあげられる。すなわち、高速化された後では回線上でパケット同士の間隔が開き、何度かルーティングに失敗しても再送の余裕があり、バッファ溢れによる棄却が小さくなるのである。

3. ベネス網における新たなパケット制御法

パケットをベネス網^[5]中でルーティングさせる場合の新たな制御法を提案する。以下、網の構成とパケットの制御アルゴリズムを示し、他の網制御法との比較を行なう。

3.1 網構成

8×8の場合の全体図を図3に示す。これは次に述べるアルゴリズムの適用上、一般にベネス網と呼ばれているものより、リンク段数が一段多くなっている。すなわち、n×nの場合のリンク段数は、 $2 \log_2 n$ となる。また、網の前半部を分散網、後半部分をルーティング網と呼ぶことが多いが、ここでもこの名前を用いることにする。

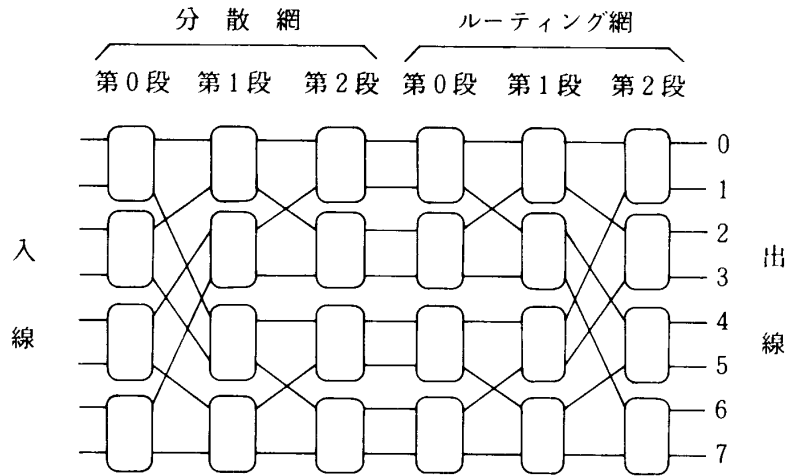


図3 8×8ベネス網

3.2 制御アルゴリズム

$n \times n$ の場合、パケットの目的出線の番号の2進表示は $\{d_{N-1} d_{N-2} \cdots d_1 d_0\}$ と表わされる（ただし $N = \log_2 n$ 、 d_i は0または1）。エッジ制御により、これがルーティングビットとしてパケットヘッダ中に書き込まれているものとする。提案する制御法は以下のとおりである。

[分散網内]

- 分散網内第 i 段 ($i=0, 1, \dots, N-1$) のスイッチは、入力されたパケットの d_i の値に従って出線（図1の"0"または"1"）を選択する。
- 2つのパケットが入力され、どちらも同じ出線を選択した場合は、どちらか1つ（どちらになるかは確率2分の1）を異なる出線に送出する。つまり、1つを出線"0"へ、もう1つを出線"1"へ送出する。従ってここでは棄却はない。

[ルーティング網内]

- ルーティング網内第 i 段 ($i=0, 1, \dots, N-1$) のスイッチは、入力されたパケットの d_{N-i-1} の値に従って出線（図1の"0"または"1"）を選択する。
- 2つのパケットが入力され、どちらも同じ出線を選択した場合は、スイッチの状態を直通状態（図1(a)）とする。この時、選択した出線に行けないパケットは棄却される。

(制御法終)

分散網内で、すべてのスイッチで d_i の出線を選択できたパケットは、分散網とルーティング網の接続部分で $\{d_0 d_1 \cdots d_{N-2} d_{N-1}\}$ 番目の接続線に送出される。このパケットはルーティング網内ではすべて直通状態でルーティングされる。ルーティング網内では直通状態優先であるので、これらのパケットは棄却されることなく目的出線に到着することができる。

分散網内のいくつかのスイッチにおいて d_i の出線を選択できなかったパケットは、ルーティング網内ではそれらに対応するスイッチにおいて交差状態を選ばなければならない。直通状態優先であるのですべて交差状態を選べるわけではないが、回線使用率がそれほど大きくない場合、棄却率は小さくなる。

以上のように分散網において意図した方向以外にルーティングされたパケットも、ルーティング網において回復できる可能性があるので棄却率が小さくなるのである。

3.3 他の網制御法との比較

バンヤン網と異なり、ベネス網は入出線のすべての組合せの経路設定が可能であることが知られている。ベネス網の経路設定のアルゴリズムとしては、Looping アルゴリズム^[6]・Lee のアルゴリズム^[7]・接続網アルゴリズム^[8]等があるが、これらはいずれもすべての入出線端子間の接続要求から網全体の経路を集中して決定するものであり、制御に要する処理量が大きく、高速パケット交換のように即時処理の必要なものには適用が難しい。また、セルフルーティングによる経路設定の方法^[9]も提案されてはいるが、まだ完全なものはない。

一方、Batcher-Banyan 網^[10]や上松氏等の提案する^[11]一部のスイッチ同士が相互に情報の授受をする網（以後 SR-SW 網と呼ぶ）のように、ある時刻において2つ以上のパケットが同一出線に向かっていることがない場合、すべてのパケットがセルフルーティングによって目的出線に到着できる網がある。これらの網をここでは完全網と呼ぶことにする。

完全網はこのような優れた性質を持っているが、バンヤン網やベネス網に比べてスイッチ数がかなり多くなってしまふ。これは回線数が多くなるほど顕著になる。例えば、回線数 512 本の時 Batcher-Bayan 網のスイッチ数はバンヤン網の 6 倍、SR-SW 網では約 30 倍にもなる。

さらに、パケットの生起が完全にランダムである場合、すなわち 2 つ以上のパケットが同時に同一出線に向かう可能性がある場合、当然目的出線にたどりつかないものが出てくる。平均リンク使用率 ρ 、回線数 n のとき、完全網のパケット棄却率 P_c は

$$P_c = \sum_{i=1}^n {}_n C_i \left(\frac{\rho}{n}\right)^i \left(1 - \frac{\rho}{n}\right)^{n-i} / \rho$$

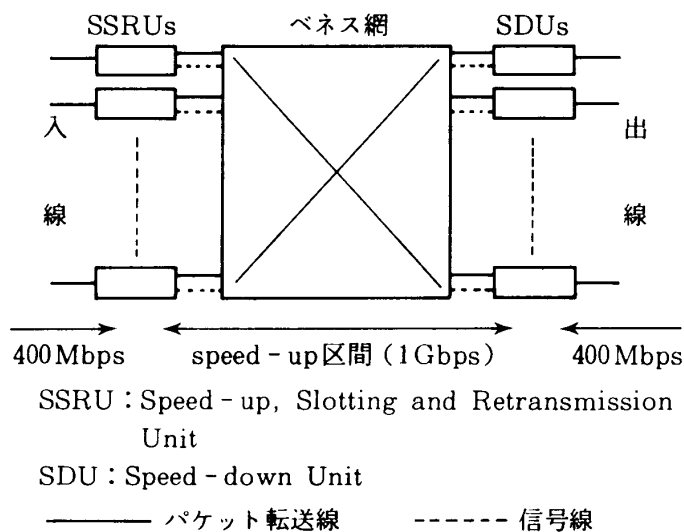
$$= \frac{\rho - 1 + (1 - \rho/n)^n}{\rho}$$

と表わすことができる。従って、平均リンク使用率 20%、回線数 512 本では全パケット中 9.6% も棄却されることとなる。これは、同一条件でのバンヤン網の約 32%、ここで提案する制御法の約 14% に対して極端に少ないというわけではない。

また Batcher-Banyan 網はソーティング網の各スイッチでパケットヘッダ中の全ルーティングビットを比較しなければならず、論理回路が多少複雑になる。

以下、提案した制御法の特徴を述べると次のようになる。

- ①パケットヘッダ内の 1 ビットによるリアルタイム制御であり、 β エレメント自身簡単な論理で構成できる。
- ②ハードウェア量はバンヤン網の約 2 倍で、完全網よりかなり少ない。
- ③パケットは目的出線に到達できないとわかった時点で棄却されるが、棄却率は完全網より多少大きく、バンヤン網よりかなり小さい。



SSRU : Speed - up, Slotting and Retransmission Unit
SDU : Speed - down Unit

——— パケット転送線 - - - - - 信号線

図4 提案するシステム構成

表1 各パラメータの値

処理容量	100 M packet/sec
パケット長	1000 bit 固定
回線数	512本
中継回線使用率	50%
内部高速化比	2.5倍
パケット棄却率	10^{-6}

4. セルフルーティング形高速パケット交換

2章のCDR法と3章のパケット制御法を組み合わせることにより、高速大容量のセルフルーティング形パケット交換が構成できることを示す。

4.1 システム構成

システムの全体構成を図4に示す。各部位の働きは既に述べた通りである。各種パラメータの値は次のようにして決定した。

まず、目標として念頭に置いているのは、広帯域通信網におけるノード交換機、または中小都市規模の通信のトラフィックを収容できる交換機である。従って交換機の処理容量を約100 Gビット/秒とする。また、実際にはパケット長は固定すべきではないかもしれないが、ここでは解析の都合上1000ビット固定とする。よって交換機のパケット処理容量は100 Mパケット/秒とする。

一方、現在実用化され、これから主流となるであろう400 Mbpsの中継回線を直接収容することを考える。中継回線の使用率を平均50%とすると、毎秒100 Gビットの情報量を伝送するには約500本の中継回線が必要になる。ここでは網の形態上、2のべき乗である512本とする。

また、交換機内部での棄却をなるべく少なくするために、内部のリンク速度を高速化することを考える。実用的な論理回路の動作速度の限界値として、内部リンク速度を1 Gビット/秒とする。従って、高速化比は2.5倍である。

さらに、パケットの棄却率の目標を 10^{-6} とする。これは、2~3回のエンドーエンド再送により、現在のパケット交換のパケット紛失率 10^{-12} 程度が達成可能となる値として採用した。

これらの値を表1にまとめる。

4.2 シミュレーションによる特性評価

ここで提案したパケット交換システムとバッファ付きβエレメントを用いたシステムを、棄却率・

バッファ量・遅延の点から特性評価を行なった。評価は主にシミュレーションにより行い、部分的に M/D/1 の待ち行列解析手法を用いた。ただしパケットの生起はポアソン分布に従い、目的出線はランダムに選択するものとする。シミュレータは東京大学大型計算機センタの VAX 8600 上に C 言語で作成した。

提案したシステムと比較するため、バッファ付き β エlement によるシステムは、交換機内部で高速化するものとし、しないものの両方を調べた。以後、次のように呼ぶこととする。

方式 I : 提案するシステム

方式 II : バッファ付き β Element によるシステム (ただし速度交換を行なう)

方式 III : バッファ付き β Element によるシステム (ただし速度交換を行わない)

(1) 棄却率とバッファ量について

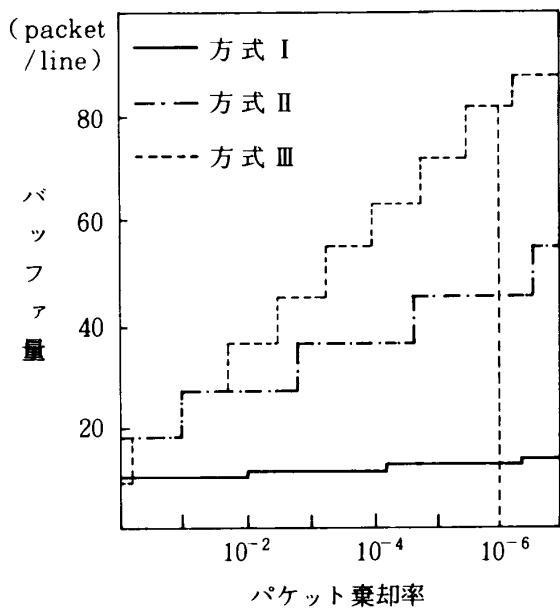
グラフ 1 は、表 1 の設定の下でのバッファ量と棄却率の関係である。バッファの量は、各回線ごとにパケット何個分のバッファが必要かということで表わしている。これには、回線を高速化する部分に必要な 1 個と、SDU において必要な 8 個を含んでいる。

目標値である 10^{-6} の棄却率にするには、各回線ごとに方式 I では 12 個、方式 II で 45 個、方式 III で 81 個のバッファが必要になることがわかる。この差は棄却率が小さくなるにつれて拡大する傾向があり、方式 I ではあと数個の追加で 10^{-12} も達成可能ではないかと思われる。

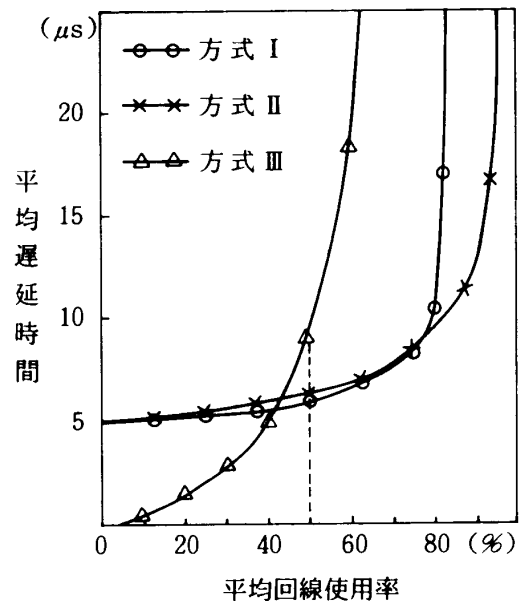
(2) 平均遅延について

グラフ 2 は、表 1 の設定の下での平均回線使用率と平均遅延の関係である。

400 Mbps の回線上を 1000 ビットの情報を伝送するには $2.5 \mu\text{sec}$ 必要であるので、方式 I・II ではスピード変換時に最低 $5 \mu\text{sec}$ 必要となる。方式 III はスピード変換しないので、回線使用率の小さいときには他の方式よりも遅延が小さいが、使用率 45% を越えると大きくなる。方式 I と II はほぼ同じよ



グラフ 1 棄却率とバッファ量



グラフ 2 遅延時間特性

うな遅延だが、方式Ⅱが使用率95%付近まで低遅延であるのに対し、方式Ⅰでは80%を越えたあたりで遅延が急激に大きくなり、それ以上のスループットは実現できないことがわかる。しかし使用率50%で 10^8 パケット/秒のスループットが実現されており、目標は達成している。しかもこのときの平均遅延は約6 μ secであり、非常に小さい。

(3) 総合評価

以上見てきたように、方式Ⅰは他と比べて非常に少ないバッファ量で低棄却率を実現することが可能で、遅延も非常に小さい。従ってここで提案する方法により、 10^8 パケット/秒の処理容量を持つ低遅延・低棄却率のパケット交換機が実現できると思われる。

5. まとめ

本稿で提案したセルフルーティング形高速パケット交換方式の特徴を列挙すると以下のようになる。

- ① β エlementに逆方向の信号線を付加することにより、簡単な制御で交換機内部での自動再送処理を行なうことができる。
- ②ルーティングの制御は各 β Elementでパケットのヘッダ中の1ビットを判定するだけであり、論理回路が簡単である。
- ③ルーティングを行なう部分のハードウェア量はスイッチ数にしてバンヤン網の約2倍であり、そう多くはない。
- ④平均遅延時間は数 μ sec程度で非常に小さい。
- ⑤バッファを網の前に置くことにより、少ないバッファ量で低棄却率が実現できる。

以上、新たな制御法によるベネズ網を用い、交換機の内部で自動再送処理を行なうことによって、大容量・低遅延・低棄却率のパケット交換機が実現できることを示した。特にCDR法はいろいろ応用が可能であり、もしハードウェア量の少ない完全網が作られれば、これと組み合わせることによりさらに優れたものができるであろう。

【参考文献】

- [1] 高見、竹中、野口：「高速パケット交換網アーキテクチャとその性能評価」、信学技報 SE 85—135.
- [2] Thomas A., et al. : "Asynchronous Time Division Techniques : An experimental Packet Network Integrating Videocommunication", ISS'84.
- [3] Paul K., et al. : "Fast packet Switching for integrated Network Evolution", ISS 1987.
- [4] 鈴木、竹内、明石：「高速パケット交換プロトコルの一検討」、信学技報 SE 86—66.
- [5] V. E. Beneš : "Optimal rearrangeable multistage connecting networks", B. S. T. J., vol. 43, No. 4, pp. 1641—1656 (July 1964).

- [6] D. C. Opferman and N. T. Tsao-Wu : "On a class of rearrangeable switching networks, Part 1 : Control Algorithm", B. S. T. J., vol. 50, No. 5, pp. 1579 — 1600 (1971).
- [7] K. Y. Lee : "A new Beneš network control algorithm", Proc. Int. Conf. on Parallel Processing, pp. 51 — 58 (1984).
- [8] 岸本 : 「セルフルーティング制御によるスイッチング回路網構成」、信学技報 SE 86 — 86.
- [9] D. Nassimi, et al. : "A self-routing Beneš network and parallel permutation algorithm", IEEE Trans. Comput., C— 30.5, pp. 332 — 340 (May 1981).
- [10] K. Batcher : "Sorting networks and their applications", Proc. Spring Joint Comput. Conf., 1968.
- [11] 上松、渡辺 : 「相互に情報の授受を行なう空間分割型セルフルーティングスイッチ網の検討」、信学会論文誌 Vol. J 70 —B, No. 6, pp 628 — 637, Jun 1987.

研 究 論 文

UNIMARC-authority の課題

Problems of UNIMARC-authority

学術情報センター 内藤 衛亮*

はじめに

- 1 IFLAの典拠データに対する取り組み
- 2 UNIMARC-authoritiesの要点
- 3 IFLAに対するコメント概要

おわりに：今後の課題

付録

IFLA's UNIMARC-authority is commented by group of Japanese experts from the point of handling Japanese scripts. Multiscript situation should be taken into account in standardization activities in library and information services. East Asian regional standardization, proposed repeatedly in Japan, being also a part of international standardization, is also reviewed.

はじめに

国際図書館連盟（IFLA：International Federation of Library Associations and Institutions）は1988年はじめに、UNIMARC-authorities フォーマットの草案（第4次）を世界各国の関係機関に送付してコメントを要請した。これに関して宮沢彰（学術情報センター）、田村俊作（慶応義塾大学）、田村貴代子（国立国会図書館）そして筆者の4名で調査分析をすすめた。1988年夏にオーストラリアのシドニーにおいてIFLA年次総会が開催され、期間中のセッションの一つとしてUNIMARC Workshopが開催されたので、これに出席しコメントを提示した。本稿はコメントするに至る背景としてIFLAの活動を紹介し、コメントの要点をまず報告し、今後の課題を4人の共同作業者を代表して論ずるものである。参考のためコメント全文を付録として付す。

1 IFLAの典拠データに対する取り組み¹⁾

IFLAは目録・書籍情報の国際流通、国際的共有について長らく活動してきた。世界的書誌調整（UBC：Universal Bibliographic Control）すなわち各国内の書誌調整の組織や制度（NBC：Nation-

* NAITO, Eisuke : National Center for Science Information System

al Bibliographic Control) を確立して、これをもとに世界的な書誌情報の共有を推進し、資料の共有の促進するという IFLA の目的は第二次世界大戦直後からのものである。特に 1960 年代の終わりから世界的な書誌調整をテーマに各国の図書館界にさまざまな標準化を働きかけてきた。その第一は国際標準書誌記述であり、第二は UNIMARC であり、そして第三は典拠に関わる国際的標準化である。

国際標準書誌記述 (ISBDs : International Standard Bibliographic Descriptions) は、目録や書誌のレコードを構成するデータ要素の種類と組み合わせ順序を規定し、それらの要素の区切り記号を規定するものである。ISBD を採用することによって各国間の書誌情報交換のための標準化を図り、読めない言語で書かれた書誌の各要素をも識別可能とし、機械可読形式への変換に役立てることを意図している。1970 年代の後半から単行書、逐次刊行物など文献の種類に対応して次々に公表され、各国の目録規則に影響を与えた。米国、英国、カナダ、オーストラリアなどが共通に使用することを目的とし、わが国でも洋書の目録作業で準拠されている英米目録規則 (AACR) にも採用されている。

UNIMARC : Universal MARC format は 1977 年に IFLA が発表したもので、各国内の全国誌作成機関 (たとえば、わが国では国立国会図書館) が、全国書誌を磁気テープその他の機械可読形式 (たとえば、JAPAN/MARC) で交換する際の標準フォーマットである。この仕様書は内容指示識別子いわゆるタグの規定に特色があり、各国独特のデータ内容を国際交換する際の中間辞書として機能することを目指しており、1980 年に第 2 版が発行されている。UNIMARC はわが国の JAPAN/MARC が準拠しているのをはじめとし、米国、英国などの MARC も UNIMARC 版を開発するなどして次第に普及しつつある。

UNIMARC を採用してデータベースを開発運営するためには、仕様書のほかにも多く情報が必要となる。そのため仕様書を補う資料が次々に発行されてきたが、これらを統合して 1987 年には UNIMARC Manual が刊行された²⁾。仕様書第一版の発行から 10 年を要している。このマニュアルの特色の一つは文字セットの規定 (Appendix J) にあり、ISO-646 を基礎として、ISO DIS 6630 Bibliographic Control Set、ISO 5426 - 1984 Extended Cyrillic Set、ISO 5428 - 1984 Greek Set、そして ISO 6438 - 1983 African Language Set など、1987 年までに ISO が規定した文字セットを豊富に参照している。また、文字コード拡張方式としては ISO の現行の標準方式である ISO 2022 によると規定している。

このように書誌情報処理の必要性に応じて、標準的な文字コード拡張方式のもとに各種の言語と文字に対応しようとする IFLA の方針は歓迎すべきではあるが、一方、いわゆる市販の端末やパソコンにはない文字が、利用者である研究者や学生が図書館に来て利用するのではなく、研究室や自宅から図書館のデータベースにアクセスしようとする時代に対応できるのか、逆に多様な多くの文字をなんらかの方法で普及型の端末に表示するにはどうすべきか、書誌情報の本来的使命を追求する UNIMARC Manual の示唆する課題は多い。

書誌要素の種類と組み合わせを規定し、これを交換する MARC フォーマットを規定すれば、次に問題となるのは典拠情報の交換である。典拠情報の用途は多様であり、たとえば、書誌レコードにア

アクセスするための探索キーや、目録を冊子体その他のハードコピーとして出力印刷するためのソートキーそのものないしはソートキーの材料となるデータを提供して、書誌データベースの整合性を維持管理する役割がある。もちろん、現地規則に基づく現地語による目録書誌情報では、利用者独自の目録規則などの規約があるにしても、利用者は流用できるのであり、第一、他国のデータベースの整合性について事情不明のまま典拠作業をせずにすむ点では大きな効果がある。

典拠コントロールとは何かについて ALA 用語集は、次のような関連用語を収録している³⁾。

典拠コントロール/典拠調整 (Authority Control) : 書誌レコード・ファイル中の標目として使用される名前、件名、統一タイトルの権威ある形を整合性を持って維持する方法。機械可読レコード群 (データベース) に対する権威ある形式と、それらに対する適切な参照を収めた典拠レコードのファイルならびに典拠ファイルその整合性を保つ上ですべてのレコードが自動的に更新されるメカニズムを含む。

典拠ファイル (Authority File) : 書誌レコードの集合において使用する権威ある標目形ならびにこの標目形へと/標目形からの参照を確立する典拠レコードの集合、典拠ファイルのカテゴリとしては、名前典拠ファイル、シリーズ典拠ファイルそして件名典拠ファイルがある。

典拠レコード (Authority Record) : 書誌レコードの集合の中で使用すべく確立された形で標目を示すレコードで、標目確立に際して参照した情報源を引用し標目へと/標目からの参照を示し、情報源で発見された注記情報を標目の形ならびに参照の選択についての正当性を示すものとして収める。

「典拠作業によって典拠データが作られ、それらを利用して典拠コントロールが行われる」⁴⁾ のは、現在ではきわめて当然のことと理解されているが、ここに至るには、米国議会図書館の MARC-Authorities の開発普及や急速に増加するデータベースの量と検索効率、入力費用とこのようなサービスの効果の評価などへの関心の高まりを背景としている。MARC の開発が各国で進行する過程で、その交換流通のために典拠コントロールがますます重要な課題であるとの認識が深まり、ついに典拠関係のプロジェクトが推進されるに至ったのである⁵⁾。

1974 年の Unesco Intergovernmental Conference において、各国の国立図書館・書誌作成機関の責任として “establishing the authoritative form of name for its country’s authors, both personal and corporate, and authoritative lists of its country’s authors, personal and corporate.” のあることが確認されている。

1977 年の IFLA ブラッセル総会では、the Steering and Advisory Committee of the IFLA International Office for UBC は、典拠ファイルの確立・形成のための諸原則と典拠情報の国際交換を実施するための手順に関する新しいプロジェクトの設置に合意した。IFLA UBC Office が予備調査を行うとともに、Standing Committee of the IFLA Section on Cataloguing ならびに当時の Section on Mechanization が合同でこのプロジェクトを担当することになった。⁶⁾

IFLA UBC Office は、典拠ファイルに関して状況 (status)、種類、規模、データ内容、維持管理、

運用、配布状況、目録規則、シソーラス、参考資料 (reference sources) などについて質問調査を実施した。質問状は 85 機関に送付され、40 件の回答があった。これは A survey of authority files and authority control system for catalogue headings : first report (unpublished) としてまとめられた。

1978 年 8 月にチェコスロバキアの Strbske Pleso で開催された IFLA 総会で Section on Cataloguing と Section on Mechanization (当時) は、Joint Working Group on an International Authority System の設置を提起した。IFLA は 1977 年末までに各国の典拠ファイルに関する調査を進めていた。これを受けて設置が提起されたのである。Joint Working Group on an International Authority System が設置された。WG はコペンハーゲン (1979 年 8 月)、ワシントン (1980 年 4 月)、モントリオール (1982 年 8 月) などにおいて会合した。

WG の任務

- 1 図書館の必要条件を充たすための国際的典拠ファイルの仕様の検討と起草。
- 2 典拠データ交換用 UNIMARC フォーマットの開発。
- 3 典拠データの効率的・効果的交換の方式の開発。

WG の課題

- 1 国際的典拠システムが支援すべき機能、書誌的活動、交換媒体などの定義。
- 2 標目中のデータと準拠している目録規則の関係の分析。
- 3 典拠 (authorities) の種類の定義と追加・削除すべき種類の検討。
- 4 UBC Office の調査の見直しとデータに対して新規の分析が必要か否かの検討。
- 5 典拠データ交換のための既存の提案の分析 (INTERMARC, LC MARC, MABI など)
- 6 その国の著者 (national author) と、その国の団体 (national bodies) という用語の定義の検討。
- 7 著者コントロール標目の種類に応じたデータ要素の定義。
- 8 UNIMARC-authorities レコードの定義。
- 9 手作業・機械可読用の国際標準典拠番号 (International standard number for authority) の定義。

1982 年 10 月には草案が完成して 300 以上の機関、国立図書館などに発送された。1983 年 8 月にミュンヘンで草案が合意され、同年 12 月には最終案が確定した。1984 年 9 月には印刷原稿となり、ブタペストの Szechenyi Library が印刷している。

この委員会がまとめたガイドラインは "Guidelines for authority and reference entries." 1984.⁷⁾ で GARE と略称で呼ばれており、UNIMARC-authorities の前提条件とされている。ガイドラインの範囲は次のように規定されている。

- 1 件名標目の典拠は、当面、除外する。これは言語上、概念上の問題が発生するので、後に主

題分析ならびに分類の専門家によって検討されるべきである。

- 2 シリーズ典拠は、IFLA Section on Serial Publications と ISDS の間の協議によるとして、当面、除外する。
- 3 統一標題（書名）典拠は、法令、聖典、楽譜、複数の標題の下に発行された著作などに対する実務上の必要性をさらに検討するまでは、無著者名古典のための標目のみに限定する。

なお、関連する IFLA の資料としては “Name of person : national usages for entry in catalogues.” 1977 / 1980.⁸⁾ および “Form and structure of corporate headings.” 1980.⁹⁾ がある。

1983 年に GARE の完成により WG は解散したが、最終報告においていくつかの勧告をなし、特に UNIMARC 典拠フォーマットの確立に優先度を与え、そのための IFLA Steering Group for an Authorities Format（表1）が発足した。このような発展過程のもとに、UNIMARC-authorities は GARE を前提として、UNIMARC と対応する典拠データの交換に使用するフォーマットとして提案されているのである。

表1 IFLA Steering Group for an Authorities Format

Marie-Louise Bachmann, Kunglia Biblioteket, Stockholm
Christine Bossmeyer, Deutsche Bibliothek Frankfurt (Chairperson)
Diana B. Dack, National Library of Australia, Canberra
Tom Delsey, National Library of Canada, Ottawa
J. M. Feyen, Pica Samenwerkingsverband, Koninklijke Bibliotheek, Den Haag
Francoise Finelli-Lemelle, Bibliotheque nationale, Paris
Gunter Franzmeier, Staatsbibliothek Preussischer Kulturbesitz, Berlin
Paula Goossens, Koninklijke Bibliotheek Albert I., Brussels
Anthony Long, The British Library
Sally McCullum, The Library of Congress, Washington, D. C.

2 UNIMARC-authorities の概要¹⁰⁾

UNIMARC-authority は、情報技術分科会と目録作業分科会合同の作業部会が起草したもので、その主査を西ドイツの Christine Bossmeyer 女史が務めている。起草の主役は、米国議会図書館の Sally McCullum 氏と英国図書館の Anthony Long 氏であることが草案の序文に書かれている。第1草案（1984-26-05）、第2草案（1985-10-21）、第3草案（1987-03-13）と重ねて1988年初めには第4草案となった。

第4草案の章別構成を表2に収めた。当然のことながら、典拠フォーマットも UNIMARC の一つであるため共通するところが多い。注目すべき点は、「フォーマットの使用方法」において規定ないしは解説されている部分にある。すなわち、「典拠記入レコード」、「参照記入レコード」、「一般参照記入レコード」、「composite headings」などの事項であり、また、「レコード内容の概要」、「UNIMARC

表2 UNIMARC-authorities 第4草案の章別構成

まえがき
序説
定義
フォーマット設計上の指針
機能ブロック
使用上の注意
必須フィールド
コントロール機能
フィールドとサブフィールドの繰り返し
サブフィールドの順序
フィルキャラクタ
コードデータの値
句読記号
フォーマットの使用方法
典拠記入レコード
参照記入レコード
一般参照記入レコード
並列データ
並列する文字データ
composite headings
レコード内容の概要
UNIMARC-authorities と UNIMARC の対応
参照記入と典拠記入の表示
説明
フィールド一覧
フォーマット
レコードラベル
ディレクトリ
0—識別ブロック
1—コード情報ブロック
コントロールサブフィールド
2—標目ブロック
3—注記ブロック
4—を見よ参照指示ブロック
5—をも見よ参照指示ブロック
6—標目リングブロック
7—分類番号ブロック
8—出典情報ブロック

-authority と UNIMARC の対応」、「参照記入と典拠記入の表示」などであるが、これらは、典拠データベースに特有の事項に留まらない。

もちろん、フィールドは UNIMARC の考えに従ったグルーピングがなされ、典拠情報を次のように展開している。

- 2—標目ブロック
- 3—注記ブロック
- 4—を見よ参照指示ブロック
- 5—をも見よ参照指示ブロック
- 6—標目リンクブロック
- 7—分類番号ブロック
- 8—出典情報ブロック

これまでに述べたような UNIMARC-authorities のフォーマット上の考え方については、特に大きな問題はないとするのが研究会の理解である。その最大の理由は、わが国の洋書の目録作成が AACR 2 に準じており、LC-MARC を採用しているからである。しかしながら、例示の多くに問題点を見いだすことができた。このことは、ついでながら仕様書の例示の重要性を示している。

UNIMARC Manual は作成者ないしは作成国の拡大と書誌情報に収録される言語ないしは文字の拡大に対応して、多文字・多言語環境を前提とした文字セット、文字コードの規定を盛り込んでおり、UNIMARC-authorities にも踏襲されている。UNIMARC-authorities は「並列データ」、「並列する文字データ」など、多文字・多言語対応のための規定を設け、中国語、韓国語、日本語その他の種別コードやサブフィールドコードを規定し例示に挙げている。問題となるのは、その規定が少なくとも日本語については必ずしもわが国の状況を十分には反映していないおそれのあること、Japan/MARC のデータとは同一とは言い難いことにある。このことがコメントの中心となったのである。

3 IFLA に対するコメントの要点

UNIMARC-authorities に対するコメントとして、宮沢彰、田村貴代子、田村俊作、内藤衛亮が、この問題に関心を持つものという立場からとりまとめたものは、*Parallel Description : one aspect of cataloguing of Japanese materials* と題したもので、A 4 シングルスペース 6 ページである。その論点は 13 項目であり、付録に全文を付したので詳細を検討されたい。日本語版はない。参考のために 13 項目を日本語で表現すれば次のようになる。

- 1 ファイナルキャラクターの直接使用（の提案）
- 2 文字セット種別コード（の問題点）
- 3 (ISO に) 登録された文字セットの完全リストを入手することの困難性
- 4 一つのデータ要素の（いくつかの文字による）並列記述
- 5 記述と読みの関係（と読みの並列記述）

- 6 日本語アクセスポイントの表記に使用する文字
- 7 文字の種別方法における不具合
- 8 複数存在するローマ字化表
- 9 UNIMARC Manual の規定する文字（日本の文字の種別の確認）
- 10 混合文字列で表記する日本のタイトルと著者名
- 11 （欠番）
- 12 外字の取扱い/運営規定の必要性
- 13 草案文章の校正および論理的整合性の不具合の指摘
- 14 例示の提案

約2か月の間に数回の会合によってとりまとめたため、ふりかえって読み返すと文意不明あるいは説明不足の箇所がないわけではなく、コメントとして十分のものがどうか反省せざるをえない。しかしながら、UNIMARC-authorities ないし UNIMARC Manual に当事者の関与ないしはレビューが必要なことは指摘できたと考えている。問題は当事者とは誰であり、関与とは何であり、継続的な貢献がはたして日本の図書館界に可能かという点にある。

おわりに：今後の課題¹¹⁾

IFLA の世界的書誌調整事業とその一環である UNIMARC-authorities には、いくつかの問題がありしかも多岐にわたっている。本稿の主題は1988年夏に提示したコメントであるので、本節では関連する問題については今後の課題として、それらの存在について指摘するに留める。そのような問題として、1) MARC レコードそのものの見直し、2) 文字セット/コードに対する考え方、3) 東アジア書誌調整などである。

1) MARC レコードそのものの見直し

松井幸子は、LC-MARC 典拠ファイルの分析の結果、データ要素の複合的な繰り返しが、ファイルの膨張の原因の一つであるとし、また、一人の著者の典拠レコードが文献の出版に応じて次々に作成されてしまう現在の LC-MARC の作成方針に対して、一著者一レコードとするような方針と方法を提案した¹²⁾。ここで注目すべき点は、松井の分析内容と結果であるよりは、確立されてしまっていると思なしがちな MARC に対して、典拠データベースの作成と維持の効率を求めて、松井が新しい考え方と新しいデータ要素の持ち方を提案している点である。

データベース作成を安定した状態に仕立てあげることの困難さは、わが国でも多くの事例においてよく知られており、したがって、安定しているものをさらに改造することには大きな心理的抵抗があり、何よりもシステム改造とデータベースの書き換えという主に経済的な難題がある。しかも、データベースの仕様が長い期間の要求に耐えるか、あるいは現状の規定が安定した経済的な入力作業を保証しているかという点には、まだ必ずしも客観的な評価があるわけではない。

UNIMARC-authorities は、変換辞書としての役割を果たそうとしているが、他方において UNIM-

ARC そのものが提起したように、新たなシステム開発の際の枠組みを提供しようとするものである。松井提案の意義は、新たな枠組みへの取り組みを示唆している点にある。

2) 文字セット/コードに対する考え方

コメントの大部分が UNIMARC の多文字・多言語規定に関与していることから明らかなように、UNIMARC Manual と UNIMARC-authorities の多文字・多言語への対応は、多とすべきではあるが、なお、それぞれの言語や文字の当事者による評価を必要としている。しかも個々の評価を総合する必要があるかのごとくである。

特に漢字関連では、宮沢彰が論じているように、文字コード論の深化が必要であり¹³⁾、これは北米、東アジア各地域で同時多発的に漢字処理が進行中であるため、先にふれた松井提案と同様に、現実との同期をどのようにとるのかという深刻な側面と、それをしないままにして、円滑な書誌情報の国際交換あるいは東アジア地域での交換が可能となるのかというさらに深刻な問題がある。

3) 東アジア書誌調整

日本の情報に関する国際会議は、最近では英国ウォリック (1987 年)、ロンドン (1988 年)、そしてベルリン (1989 年) と続いている。また、第 4 回日米大学図書館会議 (1988 年ウィスコン州ラシーヌ) では日本語の文字セットに関するセッションが設けられている。

現地語によるアクセスを情報利用の基本とする前提に立てば、海外から日本語あるいは日本文の、あるいは中国、韓国そしていわゆる NIES 各国のデータベースをオンラインで利用できるようにすることは必須の課題である。したがって、漢字を使用した情報ないしはデータベースの生産国である漢字圏諸国とそれ以外の地域にいる利用者の双方に多文字・多言語への対応が迫られている¹⁴⁾。

問題は漢字データベースの国際交換の円滑化である。そのためには国際的には漢字を対象とする各国の標準規格の互換性の検討が必要であり、各国内では標準規格の定めた文字以外の文字すなわち「外字」の扱いが課題となっている。この面での日中韓の国際調整とはどのようなものかについての検討が必要であろう。

JIS はその規定するもの以外には関与せず、ここに標準規格の価値があるかのごとくである。外字についてはガイドラインすらもなく、しかも JIS の利用者である無数のアプリケーションが人名・地名などをデータベースに収める際に JIS がない文字については各自に対応してきた。各国が同様の発展過程をたどるとすれば、漢字版バベルの塔は何年も前から盛大に創られつつあるのである。

漢字ないし東アジア文字によるデータベースが北米で構築・利用されていることは、当事国である東アジア 3 国に、たとえば、データベースの国際交流、文字コード/セットの国際的標準化など多様な側面で影響を及ぼすことになる。ローマ字化の国際規格制定においてわが国と米国とがヘボン式と訓令式をめぐる論争したことと類似の構造がここにある。

漢字データベースは当事国のみならず北米をはじめ世界各国に利用者が存在する。したがって JIS の漢字に限らず、漢字の文字セット/コードも世界各国に利用者が存在するのである。すなわち、文献の第一の利用者はその国の人々であるが、外国にも利用者がいることは、わが国における欧米文献の

状況を挙げるまでもない。日本の情報への海外の関心は文献にとどまらず、情報のあらゆる形態を指すことになるだろう。

いまや各国の言語政策、文化行政が、その国のコンピュータ技術開発と拮抗しているだけでなく、漢字の文字セット/コードについては、1980年代の前半には国際的に調整を必要とする段階に至っていたのである。BLG/LCの「東アジア文字調整委員会」の創設提案、そして1987年に学術情報センターが主催した International Conference Scholarly Information Network¹⁵⁾あるいは国際情報化協力センター(CICC: Center for International Cooperation on Computerization)が開催した「アジア情報技術標準化フォーラム」などがその証拠である。しかしながら、これらの動向はメーカー、ベンダーあるいは行政主体であって、漢字の利用者すなわちデータベースの作成や検索を担当する者の現場経験や現場からの要求を意識しつつも反映するものとはなっていない。このような状況に対して、田辺広は1986年には東アジアの書誌情報に関する常置委員会の設立を提案し、その要件を論じている。たとえば、「英語を共通語にしようかも知れない。しかしわれわれにとって議論を深めるためにはそれでは充分ではない。日本語、中国語、韓国語が、2国語が分かる図書館員の通訳によって公用語となるべきである。」という当然とは言え、わが国の現状ではきわめて困難と思われる課題を組織・制度の問題と予算の問題とあわせて提起している。¹⁶⁾

従来、ISBDやUNIMARCの導入に関して、きわめて多くの人知が投じられてきたことが示すように、IFLAはわが国の図書館界で尊重されてきた。UNIMARC ManualとUNIMARC-authoritiesが多文字・多言語対応を推進している現在、わが国の図書館界の対応ないしは国際的貢献が挑戦されているとするのは過剰反応であろうか。

参考文献および注

- 1 本稿の第1章は昭和63年3月に報告した科学研究費「学術情報システムにおける総会目録の機能と運用に関する研究」による。
- 2 UNIMARC Manual/edited by Brian p. Holt with the assistance of Sally H. McCallum & A. B. Long. London. IFLA UBCIM Programme. 1987. 482 p.
- 3 The ALA Glossary of Library and Information Science. H. Young ed. ALA. Chicago. 1983. 245 p.
- 4 Robert H. Burger 松井幸子・内藤衛亮訳「データベースの典拠作業」丸善 1987 「あとがき」p. 163.
- 5 Authority system. International Cataloguing. Vol. 7, No. 4. Oct/Dec 1978.
- 6 International Cataloguing. Vol. 7, No. 3.
- 7 "Guidelines for authority and reference entries." 1984.
- 8 IFLA International Office for UBC. "Names of person : national usages for entry in catalogues." London. 1977 / 1980.

1961年にパリで開催された目録原則国際会議（ICCP：International Conference on Cataloguing Principles）において名前の扱いが論じられた。1963年に暫定版が発行され、1967年に“a definitive editions”として改訂発行された。IFLA UBC Officeは、1975年中頃に各国の国立図書館、書誌作成機関に名前に関する質問状を送り、1977年2月に締め切り、同年中に第三版を発行した。1980年には第三版の改訂・追加分を補遺を発行した。このマニュアルは名前標目の形に関するもので、選択に関するものではない。

- 9 IFLA Working Group on Corporate Headings. “Form and structure of corporate headings.” London：IFLA International Office for UBC. 1980.
団体の定義、大文字使用法、原則などが盛り込まれている。General recommendation, Territorial authorities, Subordinate bodies of territorial authorities, Conferences, Religious bodiesなどの団体名の規定原則が提案されている。
- 10 Draft UNIMARC/Authorities-Universal MARC Format for Authorities. IFLA Steering Group for an Authorities Format, 1988-01-15. 90 p.
- 11 本稿の「おわりに」の第3節「東アジア書誌調整」は1989年データベース白書「標準化の新しい動き—漢字コードをめぐる問題について」（仮題）による。
- 12 Matsui, Sachiko. Empirical analysis of the Library of Congress name authority file and a proposal for a new name authority file structure. paper presented at the 51st Annual Meeting of the American Society for Information Science, Atlanta, Oct. 23-27, 1988. 5 p.
- 13 宮沢彰「文字コード論の試み」情報処理学会情報学基礎 8-3（1988.2.22）
- 14 橋本萬太郎、鈴木孝夫、山田尚勇「漢字民族の決断」大修館書店 1987 487 p.
- 15 H. Inose ed. Proceedings of International Conference on Scholarly Information Network, Tokyo, 1988. Kinokuniya. 1988. 253 p.
- 16 Tanabe, Hiroshi. Bibliographic Information Interchange in East Asia. paper presented at the IFLA General Conference, Tokyo, 1986. 113-BIBCO-1-E（本論文、田辺広「東アジアにおける書誌情報の交換」の日本語訳は「多言語、多文字資料利用のための図書館自動化システム—問題と解決」雄松堂出版に収録されている。）

付 録

Parallel Description One Aspect of Cataloguing of Japanese Materials —a comment to UNIMARC Authority 4 th Draft— August 1988

prepared by Akira MIYAZAWA (NACSIS); Kiyoko TAMURA (National Diet Library);
Shunsaku TAMURA (Keio University); Eisuke NAITO (NACSIS)

0 Introduction

This document is a comment by a group of Japanese experts to the 4 th Draft of the UNIMARC Authority from the point of Japanese bibliographic data handling. Parallel description of Japanese data elements, adoption of “final character” instead of UNIMARC character set code, and editorial matters of the 4 th Draft are discussed.

1 Straight designation of character set by “final character”

Character sets other than the default set of UNIMARC, ISO 646, can be designated by using the final character assigned to the character sets according ISO 2022 and ISO 2370 and byte length of the character set. Thus the definition of the six character sets may be excluded from the Manual.

2 UNIMARC handling of character set codes

Character sets used are entered in coded form into the character position 26 – 29 and 30 – 33 of subfield \$a in field 100 (p. 60) where character sets applied are defined as follows :

character set code	character set
01	ISO 646 (IRV) , Basic Latin set
02	ISO 5426 – 1980, Extended Latin set
03	ISO Registration # 37, Basic Cyrillic
04	ISO DIS 5427, Extended Cyrillic set
05	ISO 5428 – 1980, Greek set

* Opinions expressed in this document do not necessarily represent those of organizations to which authors belong; revised on October 11, 1988.

06 ISO 6438 – 1983, African coded character set

Positions 26 – 27 designate the G 0 set.

Positions 28 – 29 designate the G 1 set, or contain blanks if G 0 set is not needed.

Additional character set (s) are also designated in character positions of 30 – 33.

Positions 30 – 31 designate the G 2 set, and

Positions 32 – 33 designate the G 3 set, or contain blanks if additional character set is not needed.

This provision of designating character set (s) , though necessary, is not directly related to ISO 2022 that specifies standard technique for handling of multiple character sets. Although character set codes are defined as above, only six character sets assigned the code. There are more than 100 character sets to be handling according to ISO 2022, registered according to ISO 2370. Thus adoption of final character defined by ISO 2022 is recommended as character set code.

3 Difficulties in finding the full of registered character sets

The full list of registered character set from which the final character is obtained is not provided by ISO, nor available as a form of revised comprehensive listing. ECMA is in charge of disseminating information on the registered character set, and there may be some national /local mechanism to announce the latest registration. There may be delays of revising the local list, missing in the local list.

4 Parallel description of data elements

It is a specific feature of Japanese bibliographic data that data description is not a single string but a combination of “hyoki description”and its“reading”in most cases. “Description”is for display of data element, and “yomi reading”is for manipulation of data such as searching, sorting, etc. This combination or parallel data element is inevitable for Japanese bibliographic data to input, store, manipulate, and display since one character or one word may have many readings (pronunciations). Provision should be made in machine-readable record to link the parallel data strings of description and reading, andamong parallel readings as is done in Japan MARC such as :

\$b “description” of data element	{for display}
\$a kana “reading”of data element	{for access}
\$x romaized“reading”of data element	{for access}

5 Relationship between description and readings

Scripts for "description" are mixture of every kind including Latin scripts. Script of reading is usually limited to kana. However, "reading" is described by using Latin scriptas romanized form in the early stage of computer handling depended on machine capability, and still in use for various ends. Then there is a problem of parallel relationship among description "of" "readings". Readings are described either in kana, in Latin alphabet or in romanized form. It is a common practice in Japanese database to accommodate both description of reading, one consisting of kana and the other in romanized form.

6 Script of Japanese access points

Access points in most Japanese database other than coded keys are of three types :

Romanized form

Kana form in Kata-Kana

Mixed form of any script including Kanji, Kana, Latin alphabet, etc.

EXAMPLES :

INOSE HIROSHI	romanized form	"reading"
イノセ ヒロシ	kana form	"reading"
猪瀬 博	mixed form but kanji only	"description"
HURANKI SAKAI	romanized form	"reading"
フランキー サカイ	kana form	"reading"
フランキー 堺	mixed form with kata-kana and kanji	"description"
MARUTA	romanized form	"reading"
マルタ	kana form	"reading"
Malta	mixed form but Latin alphabet only	"description"

Data in romanized form and kana form are recorded either by ISO 646 (Latin script) , JIS X 0201 (compatible to ISO 646) , or EBICDIK (de facto standard character set) , or even by JIS X 0206 which is two-byte Kanji code.

7 Expressing relationships by types of script

These parallel data elements, one between description and readings, the other between

readings, simple for Japanese native speaker but confusing for foreigners, can be categorised by type of script, though not exactly complete. “Description” is a mixture of any scripts, i. e. “unspecified”; “reading in kana” is of kana script, and “reading in romanized form” is of Latin script. It is not important for Japanese database to designate / distinguish whether the data string is “Kanji / Chinese script only” or not.

8 More than one romanization / transliteration schemes

It is not important in Japan to specify types of romanization in case of Japanese bibliographic data where either Hepburn or Kunrei methods are used. Software assimilate differences and produce the same effect. However, it may be required in system to indicate the type / scheme of romanization employed to transcribe data since it is well known fact that various romanization schemes exist in East Asian countries.

9 Scripts defined in UNIMARC Manual

Scripts are also a subject in the 4th Draft (p. 33; p. 44) as specified in UNIMARC Manual (§ 6 and § 7 p. 11 - 12) for interfield linking data as well as designation of script code; for script of title (Manual p. 60; p. 61), etc, and script code is defined (Manual p. 61; 4th Draft p. 44) as follows :

- ba = Latin
- ca = Cyrillic
- da = Japanese — script unspecified
- db = Japanese — kanji
- dc = Japanese — kana
- ea = Chinese
- fa = Arabic
- ga = Greek
- ha = Hebrew
- ia = Thai
- ja = Devanagari
- ja = Korean [should be ka? in Manual]
- la = Tamil
- zz = Other

As shown above, three types, “da”, “db” and “dc” are defined for Japanese scripts.

10 Japanese titles and names in a mixture of scripts

Japanese titles of literature are a mixture of scripts such as :

PL / 1 入門	L = Latin script
LLLNCC	N = numeric
C 言語	C = Kanji or Chinese script
LCC	k = kata-kana
	h = hira-kana

学術情報システムにおける総合目録の機能と運用に関する研究

CCCC k k k k h h h h CCCC h C C h C C h C h h C C

Three examples include “kanji=db”, “kana=dc”, and “Latin alphabet=ba”, thus these titles are categorized as “Japanese script unspecified”, and this situation is quite common in Japanese literature.

Japanese names, mostly consist of “kanji=db”, are also mixture of scripts such as :

アグネス・チャン	dc	Kata-kana
いしい・ひさいち	dc	Hira-kana
フランキー堺	da	unspecified but kata-kana and kanji
浅野ゆう子	da	unspecified but kanji, hira-kana and kanji
司馬遼太郎	db	kanji
Malta	da	unspecified Japanese but Latin script

The third and fourth cases are mixture of scripts. The last case is a name of a musician, spelled in Latin script but he is a Japanese by origin and his literature is written in Japanese language.

Thus, description of Japanese titles and names on the literature are not limited to Japanese scripts, but language is cataloguing is Japanese even it partly uses Latin script as appeared in the item.

UNIMARC categorization of Japanese scripts may be somewhat misleading. It may be stemmed from the presumption on the difference of character sets relative to scripts. Kana (kata-kana) is recorded by JIS X 0201, but also recorded by JIS X 0208 of two-byte kanji code. Kanji character is only recorded by JIS X 0208 that includes kana in kanji-code, but data strings may and can include not only kanji but any scripts as appeared in the literature.

12 Treatment of extention

New characters to the established standard character set may be added in multi-scripts ituation. These new characters are called as "extension" or "external character". Provision may be needed 1) to indicate the existence of extention(s) in a record; 2) to provide a standard method of recording the extention(s) on the record; 3) to inform the recipient on the extention by means of printed form(?); 4) to transfer, if possible and required, fonts/graphics of extention to the recipient.

13 Editorial matters

p. 47 (subfield definition) : \$x; \$y; \$z

definition not clear; explanation may be needed

p. 49 (subfield) \$y Geographical subdivision

definition or explanation may be needed

p. 53 3rd para. : ... embedded data.

definition or explanation may be needed

p. 62 EXAMPLE 1 ; EXAMPLE 2 ;

p. 63 EXAMPLE

layout should be identical to other example

p. 66 last exampie

usage of \$x may be questionable

研 究 論 文

学術分野における機械可読文書の作成と通信

Preparation and Telecommunication of Machine-Readable Documents in Japanese Academic Community

学術情報センター 根岸 正光*

要 旨

わが国学界における文書・文献の機械可読化および電気通信によるその授受の現況を、関連「白書」類により確認される一般的・社会的状況と対照しつつ、総括して報告する。まず、研究活動における学術情報の生成から受容にいたる再生産過程の図式を提示した上で、「学術情報システム」構想に基づく、大学環境での最近のコンピュータ・ネットワークの展開を通観する。つぎに、こうした通信路で伝送されるべき、文書・文献の機械可読化の状況、さらにデータベース化の動向をまとめる。すなわち、日本語ワープロ、FAX、デスクトップ印刷、全文データベース等の実用状況であり、終わりに、最近における文書ファイルの論理構造、割付構造の規格化の動きを、学術情報センターでの全文データベース作成計画とあわせて検討する。

本稿は、ベルリン日独センター（Japanisch-Deutsches Zentrum Berlin）により、1988年11月30日～12月2日に西ベルリンで開催されたセミナー「経済および学問におけるテレコミュニケーション」（Telekommunikation in Wirtschaft und Wissenschaft）のための日本側報告として書かれたものである。

ABSTRACT

The present state regarding preparation of machine readable documents and transmission of those by telecommunication at Japanese academic community is summarized in contrast with the general situation formally depicted in the governmental white papers. A reproduction model for scientific information is proposed to lead the discussion which follows. As "Science Information System" is the basic concept governmentally approved to develop an efficient information distribution system for the university people, computer networks have now been constructed within and among universities. Preparation of scientific documents in machine

* Negishi, Masamitsu : National Center for Science Information System

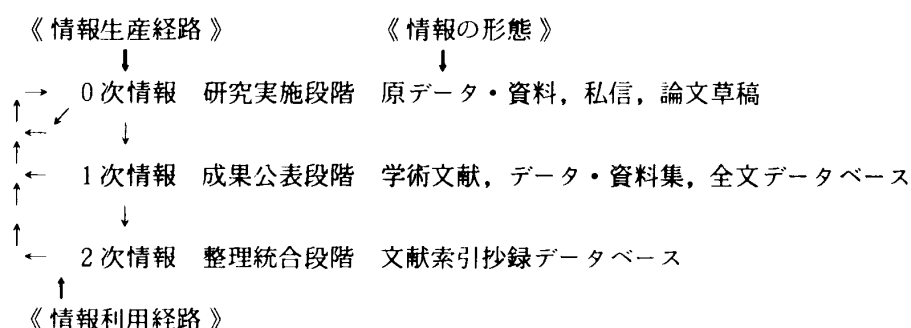
readable forms and creation of databases are the next topics to be discussed, where the current practice in utilizing Japanese word processors, facsimile, desktop publishing, full text databases, etc. is touched upon. Standardization aspect for document files is discussed along with the full text database project at the National Center for Science Information System (NACSIS).

The paper was prepared for the seminar, "Telecommunication in business and science", Nov. 30-Dec. 2, 1988, at Japanese-German Center Berlin (Japanisch-Deutsches Zentrum Berlin), West Berlin, and was presented as a report from Japanese delegation.

1. 学術情報の発生・流通

学術分野における情報伝達の最も基本的なものは、論文・文献の流通であろう。学術文献の主な媒体は学術図書・雑誌であり、学界内でのその流通によって、学術は進歩してきたと考えられる。ところで、電子計算機の出現以降、こうした学術文献の作成・流通にも電算技術を応用し、研究をより効率化しようとする試みが続けられている。本報告では、まず学術文献およびその原材料たる各種研究用データ・資料・文書など、研究活動に関わる情報のあり方をとりあえず簡単に整理してから、それらが、電子計算機と電気通信技術の進歩にともなって、どのような形態で作成され、保持され、また伝送されるかを、わが国学界での現状を踏まえつつ、検討してゆきたい。

学術研究に関わる情報を段階的にモデル化すると、およそ次のようになろう。



まず、研究の実施段階では、研究の進行の中で生み出され、使用される情報、例えば各種の測定値などがあり、またそこで作られる様々の文書がある。つぎに研究の成果をとりまとめて公表する段階で、学術文献あるいはデータ集のような公刊物形態の情報が作られる。こうした文献について、研究者の参照に便利なよう、抄録誌・索引誌が作成されるという、情報の整理統合段階が次にくる。ここにのべた研究実施→公表→整理統合という順序は、学術情報が生成され、広められてゆく過程である。一方、整理→実施、公表→実施という逆の流れは、まさに情報が研究者間に受け入れられて、利用される過程となる。

本報告では、上のようなモデルを念頭におきながら、わが国学界において、文献・文書の作成と、流通すなわち上図の矢印の過程で、電子計算機と電気通信がどのように応用され、また、今後そこにはどのような発展が望まれるかをまとめるものとする。

2. コンピュータ・ネットワークと「学術情報システム」

(1) 大学間コンピュータ・ネットワーク

研究における電算機の利用は、数値解析型の科学技術計算から始まったことはいうまでもない。わが国ではこうした計算需要を賄うための共同利用施設として、東京大学に大型計算機センターが1965年に初めて設置された（以後、いわゆる旧七帝大すべてに設置された）。当初はカード・ベースのバッチ処理であったが、1973年の機種更新にともなって本格的なTSSが導入され、1975年には国産電算機初の公衆電話網接続が可能になった。処理内容は相変わらず数値計算主体であったが、1970年頃から米国Chemical Abstracts ServiceのCA Condensatesデータベースのオンライン検索システムについて研究が進められ、1974年にはこのシステムが全国の研究者に公開された。

各大学の計算機センターを連結する高速のコンピュータ・ネットワークの必要性は、データベースの相互利用の面からも認識され、1974年から東京大学、京都大学、NTT（電電公社、現日本電信電話会社）の共同研究として大学間コンピュータ・ネットワーク“N-1”の開発が開始された。N-1は、各大学の計算機センターに導入されている異なる機種の電算機を相互に接続するネットワークであって、1980年に開設されたNTTのパケット交換網（DDX）を利用して、1981年に7大型計算機センター間での公式サービスが開始された。その後N-1には多くの大学が加入し、研究者はいろいろな大学の計算機を使うことができるようになった。

ネットワーク利用の目的には、大型機での高速科学技術計算もあるが、各センター固有のデータベースの検索やソフトウェアの利用なども重要な動機になる。すなわち1975年以降、多くの大学の計算機センターで、情報検索システムが開発され、これに主として外国製の文献抄録データベースを投入してセンター利用者に公開したり、また研究者のグループが独自のデータベースを作成して公開するという、データベース活動が展開されたからである。

(2) 学術情報システムと学術情報ネットワーク

このような状況において、電算機とネットワークを用いた学術情報の生成・保持・流通が、研究の促進に極めて有効であるとの認識が深まり、1980年には学術審議会から文部大臣に対して、「今後における学術情報システムのあり方について」と題した答申が提出された。この答申を実現するため、1983年に東京大学に文献情報センターが設置され、ここでは全国大学図書館の蔵書目録を構築するための、オンライン共同分担目録システムが開発され、運用されている。これにともなって、N-1は漢字データをも伝送できるよう拡張され、大学間での日本語データの通信が可能になった。

文献情報センターは、答申にいう「学術情報システム」の中核センターとして本格的に活動すべく、1986年に国立大学共同利用機関・学術情報センターに改組され、各種学術情報データベースの編集にも着手した。すなわち文部省科学研究費補助金を受けた研究の成果概要報告や、各種学会での研究発表要旨のデータベース化であり、1987年からこれらをオンライン検索システムにより公開した。また1988年には電子メール・システムの運用を開始したが、これはOSIのMHS (Message Handling System) 規格に準拠するもので、OSI普及への先導的役割をも果たすことが期待される。

このような学術情報システムにおける基幹的な通信路として、「学術情報ネットワーク」の構築が1986年から進められている。これは、大学の研究者や図書館など間での、全国的な情報流通を促進することを目的とする自営パケット交換網であって、既述のN-1などをすべて収容する学術情報システムの基盤施設である。NTTの高速デジタル回線(1.5 Mbps)を借り上げて、全国各地域の大学にノードを開設しており、ノード数は1989年1月には18大学になる。各大学は近隣のノードに回線を接続することで全国網に参加できる。学術情報センターには、処理能力500パケット/秒のパケット交換機と回線多重化装置を置き、これを介して大学間のデータ転送を行う。この網は利用者に無料で開放されており、通信費用に由来する地域間情報格差の解消に資するところは大きい。

(3) 学内 LAN

上のような全国ネットワークに呼応して、各大学内でもLAN(Local Area Network)の構築が進んでいる。1987年末の調査によれば、国立大学108校中93大学が何らかのLANを持っており、全学的なLANも47大学に達する。LANへの接続機器数はシステムあたり平均90台ほどである。大規模なLANの稼働は比較的最近のことで、1988年3月東京大学工学部、4月早稲田大学理工学部、5月東北大学全学、9月筑波大学全学、1989年3月京都大学全学といった状況である。ところで、これらLANの利用目的であるが、やはり手元のワークステーション、パソコンからメインフレームやスーパーコンピュータを高速アクセスして、科学技術計算、具体的には画像処理などを能率よく行なうことに重点があるようである。電子メールによる研究者間の連絡事務も目的に掲げられるが、文献・文書の作成、共有、配信への指向はまだ薄いと見ざるを得まい。

このように、学術情報の伝送路としてのコンピュータ・ネットワークは着々と整備されつつあるが、これを用いた文書・文献の流通という点での進展には未だしの感がある。これには何らかの隘路が存在するとみられ、この点、わが国での文書処理やデータベース関連の技術の現況に即して、次項で検討してみる。

3. 機械可読文書への動向と動機

文書・文献を電気通信技術により伝送するには、それらを機械可読形態(machine readable)にしなければならない。そこで、まず日本語ワープロ(word processor)などの実用状況を確認した上で、学術分野における文書機械可読化の推進/阻害要因を検討する。

(1) 文書・文献の機械可読化

1) 日本語ワープロ以前——先に回顧したとおり、計算機センターを利用する研究者らの文章型データへの接触は、文献データベース検索システムに始まるといえる。これに触発されて、研究者グループでのデータベース作成活動が推進されるようになり、文部省では1982年から科学研究費補助金でこれら活動を助成している。学術情報センターによる調査では、大学等で作成されているデータベースは現在約750件である。しかし研究者によるデータベース作成は、数値データを主体にするものであったといえる。これは、研究に直結する数値データの整理作業を、ある程度組織化することで

進められ、また文献データベースのため文字データにくらべて入力容易であるという点で、研究者にとって取り組み易いからであろう。文字データについては、大型機の TSS のエディターと ROFF などのフォーマッターを用いて、国際会議用のカメラ・レディーの英文論文を印刷するということだが、相当以前から行なわれた。ただし、ここには進んで機械可読文書を作ろうという動機はみられず、これはワープロやパソコンの普及まで待たなければならない。

2) 日本語ワープロ——電算機における日本語(漢字)処理は、1975年頃から大型機で可能になったが、それは、企業での名簿管理など定型大量事務処理への漢字データの導入という形態にとどまった(漢字の JIS コードは 1978 年制定)。電算機を用いた文書作成は、1978年にカナ漢字変換方式による日本語ワープロ(専用機)が開発され、その後低価格の製品が市場に現われるようになって、はじめて普及をみた。昨今は 10 万円前後の実用機の販売が好調のようである。16 ビット・パソコンで稼働する日本語ワープロ・ソフトウェアも 1983 年頃から機能、速度面でワープロ専用機に匹敵するようになった。大学の研究者にとっては、パソコンで稼働するワープロ・ソフトの方が総体経費的に有利であり、ここに、ワープロによる文書作成が研究者間で本格的に普及することになった。

1986 年以降、有力なワープロ・ソフトのカナ漢字変換モジュールは、これを OS に組み込んで、日本語入力フロントエンド・プロセッサとして常時動かすことができるようになった。これにより、表計算ソフト(spread sheet)、通信ソフトなど各種適用プログラムでの漢字データ入力を、効率的に行なうことができる。ここに至って、機械可読文書の作成がようやく一般利用者のものになったといえるであろう。

3) ワープロ文書ファイルの交換・通信——日本語ワープロ専用機は各社独自のファイル形式を持っており、同一系列機種間でフロッピー・ディスクの互換が保たれる程度である。通信機能を持つワープロでもやはり同一系列機種間でファイル転送ができるだけである。パソコン用ワープロ・ソフトでも、ファイルの互換は当初考慮されなかったが、1985 年頃から、平文(テキスト)を収容する MS-DOS の ASCII ファイルと書式データを収めたファイルを別建てにする、あるいは ASCII ファイルを出力する機能を備える、という方法が採られるようになり、ともあれテキストだけは、MS-DOS 対応の各機種、各種プログラムの間で確実に相互利用できるようになった。

4) ファクス通信アダプター——パソコンの付加装置として、ワープロ文書ファイルからページ・イメージをドットに展開し、これを G3 ファクシミリ方式で送信するものが 1986 年から市場に登場した。ファクシミリは、手書き文書を送信できることから、わが国ではとくに普及し、1987 年には 10 万円以下の G3 機など、個人用機種も出回って、設置台数は 220 万台(1988 年 3 月、北米 110 万台、欧州 60 万台)に達し、事務所用電話の 4 台に 1 台の割りでファクスが置かれるという状況にある。つまり現状では文書通信の主力は、ファクシミリであり、従ってワープロ文書のファクス送信は大いに有効といえよう。

5) パソコン通信——パソコンの通信ソフトウェアでは、ファイルのアップロード、ダウンロードが可能であり、これにより文書ファイルを伝送できる。パソコンの端末としての活用は、大学の計算

機センター利用者の間では1980年頃から行なわれていたが、その本格的普及はやはり1985年の通信回線の規制緩和、NTTの民営化以後のことである。民間に多数のVANが出現し、またNTTのDDX-TP(電話網—パケット網接続)サービスも開始された。こうした環境に即して、日本語機能を含む使いやすい通信ソフトが出回るようになったのである。

6) 高精細度印刷、デスクトップ・パブリッシング(DTP)——パソコン用印刷機は現在、漢字1文字24×24ドット、180ドット/インチのものが普通であり(ワープロ専用機では48×48ドットの印刷機搭載のものが事務用高級機として出回っている)、さらに50万円程度のレーザー・ビーム・プリンター(240ドット/インチ)が普及しつつある。これにより一応の文字品質は確保されるが、さらに品質の高い印刷文書を得るために、DTPのシステムがにわかに注目されている。

DTPソフトにはWYSIWYG(What You See Is What You Get)の系統と、Tex、ROFFなどのフォーマッターの系統がある。一般には前者の系統が歓迎されるが、これは、ワープロ文書ファイルを取り込んで、画面上でページ割付の編集操作を加えることができるからであろう。フォーマッターの系統は、カメラ・レディーの英文論文作成用として、研究者には以前からなじみのあるものであるため、和文原稿にもそれらの日本語化版を使おうとする傾向が研究者の一部には根強い。WYSIWYG系統はスタンド・アロン型で互換性への配慮はないが、フォーマッター系統は書式情報も通常の文字で記述するから、ファイルの交換・転送は可能である。ただし、その書式記述はかなり面倒であって、やはり一部の理工系研究者向きと考えざるを得まい。

7) 図表・数式の入出力——学術文献では図表や数式が重要な意味をもつことが多い。日本語ワープロでは、表と線図の作成機能をもつのが普通で、操作はやや面倒だが一応のことはできる。またイメージ・スキャナーで図を読み込んで、これを文書内に配置できる。ただしこれら図表の内部表現形式に互換性はなく、ファイルの交換・転送はできない。複雑な数式は一般人には縁がないため、ワープロ側の対応は悪い。この面では、元来研究者向きに設計されたTexやROFFは有効で、これらが研究者に支持される要因の一つになっている。

(2) 機械可読化の動機・目的

機械可読文書に関する上のような現状を、研究者が、文書・文献を機械可読の形態で持ちたいとする動機の面から整理してみると、つぎのようになろう。

1) 文書作成の能率化——日本語ワープロ利用の動機としてまず考えられるのは、ワープロで原稿を「書く」方が、手書きより速いのではないかという点である。しかし実際には、タイピングとワープロ操作のよほどの熟達者でなければ、草稿走り書きの速度を凌駕するのは難しいようである。またカナ漢字変換のキー操作による思考の中断を嫌う向きも多い。そこで走り書き草稿を推敲しながらワープロ入力するなど、いわば清書の道具としてワープロを使う例が多いとみられる(もっとも昨今、研究者の中にはワープロの前に座らないとそもそも考えがまとまらないという、ワープロ愛好者も出現している模様である)。ワープロによる文書作成の大きなメリットは、保存された文書ファイルの一部書き換え、編集を加えて、再利用できるところにある。これは初期入力時の非能率を補ってあまり

があり、これがワープロ普及の主たる要因であろう。

2) 文書の印刷——手書き文書主体の時代には、一部の公式・重要文書だけが専門和文タイピストによって印刷文書化された。こうした印刷文書への価値意識から、作成時の非能率を顧みず、何でもワープロ文書にする傾向が生じ、現在に至っている。ところで研究者の活動は、論文、資料などを、ごく小規模かつ簡便に印刷文書にして配布したい場合が多く、この点でワープロは有用である。今や小規模の研究会誌や学会の年次大会予稿集は、カメラ・レディーのワープロ文書で投稿をもとめるのが普通になり、従って印刷の高品質化をもとめて、DTP への関心が高い。しかし、こうした著者による機械可読原稿の作成という方式を本格的な学会誌等にも適用して、印刷費用の圧縮と発行期間の短縮をめざそうとすると、文書ファイルと印刷会社の電算写植システムとの間での互換性が必要になり、問題はにわかに面倒になる。なお、英文の学会誌については、論文の投稿を TeX 形式の文書ファイルに限定し、TeX を受け付ける外国の印刷会社に印刷を依頼する学会も現われている。

3) 文書の伝送——文書の印刷イメージでの配布方法は、最も原始的にはコピーの郵送あるいはファクス送信であるが、ファクス通信アダプターを用いれば「紙」を介さずに送信することができる。文書ファイルのフロッピー・ディスクを郵送したり、あるいはパソコン通信で転送して、受け手側で印刷イメージを再生するには、同種のワープロが受け手側にも必要である。文書を単に配布するだけなら、郵送やファクスで十分であり、ファイル交換がことさらに必要とされるのは、受け手側でその文書に何らかの処理を加えたい場合である。たとえば研究者グループで報告書を分担執筆する際、あらかじめ有力ワープロを特定し、分担者から原稿の文書ファイルを集めて、全体的編集を施すなどは、よく行なわれている。平文ファイルは書式情報を欠くので、編集側での手間がかかるが、ワープロの種別に依存せずに文書の交換が可能である。

上述のとおり、文書の機械可読化が研究者個人の範囲内でもたらすメリットや、研究者のグループ活動において発揮される利点に対しては認識は浸透しつつある。しかし、これは冒頭の図にいう研究実施段階までのことである。これを、文書・文献の流通という一般的なレベルに進めようとする、次項にみるような標準化の問題に逢着する。

4. 文書ファイルの流通——印刷物と全文データベース

文書ファイルの交換を広範囲に行なおうとすれば、その形式の標準化が必要である。ところで、流通する文書ファイルでは、受取り側でこれを様々に編集することが前提になるが、その場合、文書の内容に即した処理を加えようとする、ワープロ文書に挿入されているような書式制御情報だけでは、それがたとえ標準化されていたとしても役には立たない。標題、著者名、章別など、文書の論理的な組み立てを明示する情報が別に必要である。つまりここで、文書の機械可読化は、印刷物という考え方だけでなく、いわゆるデータベースという観点を含んで進めなければならないことがわかる。

(1) 全文データベース作成の事例

学術情報センターでは、わが国の各学会との協力のもとに、学会誌の図表をも含む全文データベー

ス (full text database) 作成を進めている。ここではこれを事例として、学術文献における印刷物形態と全文データベースとの関連をみでみる。現状では多くの学会誌が電算写植によって印刷されており、その過程で機械可読な論文ファイルが作成される。しかし、それは特定の写植機用のデータ・ファイル (CTS ファイル) であって、標準形式といったものとは無縁である。従って、これらファイルを全文データベースに変換するために次のような手順をとっている。

1) 各種写植機用ファイルから書式制御コードの類を削除して、平文の文書ファイルに変換する (写植機種別に個別対応のプログラムを作成する)。なお、最終校正では、必ずしも写植ファイルの更新によらず、版下に校正部分を直接貼込むという便法が行われることが多い。こうした校正部分は校正原稿をみながら、次の段階であらためて入力せざるを得ない。

2) 平文ファイルをエディター上に呼出して、論理項目識別子 (tag) を挿入する。この作業には、著者名とその所属機関を正しく対応させるとか、本文と脚注との対応を設定するなど、編集者の能力が要求される。また、文字データとして処理可能な簡単な表や数式と、画像データとして蓄積すべき図表や数式を仕訳して、後者にはデータベース上の図表番号を割付け、同時に本文中の該当個所にこの図表番号を挿入するなどの作業を伴うので、相当に手間がかかる。

3) 図表や数式を、図番をキーにして光ディスクに焼込む。現状ではこれらを G3 ファクスで出力するので、小さい図は拡大コピーした上で、光ディスク入力装置にかけている。

4) 2) で作られた本文 (文字) データベースを、オンライン情報検索システムにロードして、利用者に公開する。

5) 利用者は情報検索用コマンドを用いて、本文データベースを検索する。

6) 図表の説明文は本文データベースにも入っているので、見たい図表がある時は、その図番を指定して出力コマンドを出せば、その図表がファクスで送られてくる。つまりこのシステムでは、利用者の手元には通常の端末の他に、ファクスとそのための回線が必要である。

ここにのべたシステムは、1989 年からの公開をめざして、実用レベルで当面利用可能な技術に依拠して構成されたものである。従って、必ずしも満足できるものではなく、今後の研究開発に期待するところが大きい。

(2) 文書ファイルの国際標準と実質 (de facto) 標準

文書ファイルの受け手の側で、これに処理を加えられるようにするには、印刷物としての書式、すなわち「割付構造」 (layout structure) を表す情報だけでなく、文書内容の「論理構造」 (logical structure) を記述する情報が付加されていなければならない。そして、文書ファイルの交換・流通を活発にするには、こうした構造情報の記述方式が標準化され、これに対応する製品が市場に出廻る必要がある。

文書構造の国際規格化は、ISO では ODA (Office Document Architecture) と SGML (Standard Generalized Markup Language) の 2 系列で進められている。ODA は企業などにおける事務用の文書ファイルの相互交換を目指すものである。規格化にあたっては、論理構造、割付構造等一連の機能

要素を規定するが、それらの記述形式は規格化せず、製品開発者に任せている。ただし、交換の際の形式である ODIF (Office Document Interchange Format) を規定して互換性を確保する。一方 SGML は、印刷物の電算ファイル形態での交換、再編集を目的とする規格とみられ、ODA よりも精密な項目設定が可能になっている。もっとも、SGML 自体はシンタックス記述のための meta 言語であるから、実質的な規格はこれを用いてさらに規定する必要がある、米国では AAP (American Association of Publishers) の国内規格などが成立している。なお、SGML は文書の論理構造のみ扱い、これにつづけて割付構造記述のための DSSSL (Document Style Semantics & Specification Language) と、ページ記述の SPDL (Standard Page Description Language) を規格化して、一連の規格が整うことになるが、後二者についてはまだ成案にいたっていない。

上記両系列の規格に対するわが国の対応は、まず ODA についてみると、民間企業におけるオフィス・オートメーションの進展に伴う文書通信への需要を踏まえて、各電算機メーカーは、ODA の製品化に積極的に取り組むであろう。各メーカーとも、既に文書管理のための独自のシステムを開発・販売しているから、当面はこれらの ODA 規格への変換ソフトウェアから手を付け、徐々に ODA 準拠の一連の製品を開発してゆくものと思われる。

一方、SGML の系列については、これを具体的印刷物に結びつける DSSSL、SPDL の規格化が進んでいないこともあってか、この規格の当面の推進者と目される印刷会社を含めて、わが国での対応は鈍いように見える。印刷会社には以前から電算写植が導入され、辞典類など、更新・再版を前提にする出版物については、データベース的なファイルの設計、すなわち論理構造の導入が個別に行われてきた。この方向は CD-ROM での出版の進展につれて、さらに明確に意識されるようになってきている。しかし、現状では文書ファイルの社内的な転用・再利用可能性を考えるにとどまっており、AAP のような標準規格設定への動きに至っていない。むしろ、学界方面で一定の地位を獲得している TeX への対応の方に関心が向けられている。いずれにしても、この種国際規格は、わが国での実用化に際して、これに準拠した日本語向きの規格をさらに決める必要がある。

こうした状況から、有力メーカーでは独自の文書ファイル交換形式を提案する例がみられる。ソニーは、国産ワークステーションのベストセラーである "NEWS" での DTP の普及を目指して、関連各社と協議体を作り、CDFD (Common Document File Format) を設定した。これには、有力ワープロである "一太郎" や "OASYS" との相互変換や TIFF (Tag Image File Format)、PostScript など現時点での実質的な標準を取り込んで、早期の普及をもくろんでいる。また、特許庁では特許出願の効率化を目指し、ファクス形式の図面ファイルと文字ファイルを組合わせた、電子出願のシステムを開発しつつあり、1989 年から稼働させる予定である。これは基本的に OSI を採用したシステムで、図表の入った文書ファイル送受信の大規模システムの先行事例として、わが国の関連業界に一定の影響を与えるであろう。

(3) 論理構造の記述

上述のように、文書・文献を、単に機械可読にするだけでなく、「機械可処理」のものにするには、

文書の作成にあたって、その論理構造が明示的に記述されていなければならない。学術情報センターでは、先に述べた CTS ファイルを利用する、いわば後処理型の全文データベース事業に加えて、電子化された学会誌作成のシステムを計画している。この場合、著者には論文の論理構造をも記述してもらうことになる。しかし、著者の当面の関心は文書の仕上がり形式、つまり割付構造の方にあるので、論理構造がごく自然に記述できるような仕組みがワープロやエディターに用意されない限り、論理構造付きの文書・文献の作成が一般化するとは考えられない。アウトライン・エディター、アイデア・プロセッサあるいは HyperCard など、文書の論理構造を組み入れたエディターが発売されるという昨今の状況は、こうした方向への進展に期待を持たせる。とくに、ODA あるいは SGML のような標準規格に則ったエディターの開発が、文書ファイルの流通性確保のいみから待たれることになる。

5. 結 言

NTT は、1988 年 4 月から ISDN (Integrated Services Digital Network) のサービス「INS ネット 64」(2B+D) を、月額 5400 円 (現行の電話は 2350 円)、通信料は電話と同一という低料金で開始した。1989 年には「INS ネット 1500」(23B+D) も開始される予定で、通信の高速化、多様化は一層進展する。一方、パソコンはまさしく文房具の一種として、文科系の研究者間にも浸透してゆき、機械可読の文書・文献は、ごく普通のことになるであろう。こうして、冒頭に掲げた 0、1、2 次情報の処理のあらゆる面で電子計算機の能力が活用され、また、それら学術情報の流通に、テレコミュニケーションが支配的役割を果たすという時代は、そう遠くないといつてよいと思われる。

《基礎的参考資料》

通信白書 1988 郵政省編

科学技術白書 1987 科学技術庁編

情報化白書 1988 (財)日本情報処理開発協会編

ニューメディア白書 1988 郵政省監修・(社)日本情報通信振興協会編

データベース白書 1988 通産省機械情報産業局監修・(財)データベース振興センター編

情報サービス産業白書 1988 通産省機械情報産業局監修・(社)情報サービス産業協会編

OA 白書 1987 (社)日本経営協会・(社)日本事務機械工業会

金融情報システム白書 1988 (財)金融情報システムセンター編

図説・日本のネットワーク 1988 郵政省ネットワーク化推進会議編

コンピュータ・ネットワーク (先端科学技術の現状と展望第 1 巻) 科学技術庁編 1987

大学における情報ネットワークの現状と動向 (社)文教施設協会 1988

電子メールシステム調査報告書 学術情報センター 1988

全文データベースのシステムに関する現状調査報告書 学術情報センター 1987

学術情報データベース基本調査報告書 学術情報センター 1987

データベース利用動向調査報告書 学術情報センター 1987

研究論文

論文集目次型データベース ISTP&B のための検索システムの設計

Development of NACSIS-IR Information Retrieval System for
ISTP&B (Index to Scientific and Technical Proceedings and Books) Database

名古屋大学附属図書館 牧村 正史*

学術情報センター 根岸 正光**

要 旨

学術情報センターでは、昭和 62 年度から研究者・図書館員を対象として、オンライン情報検索システム NACSIS-IR (National Center for Science Information System-Information Retrieval System) のサービスを開始した。そこで運用されるデータベースの一つである ISTP&B は、科学技術関係の会議録と論文集の内容を目次的に収録したものであり、これは、巻冊レベルと個別論文レベルの 2 階層の書誌レベルに関わるデータをもつ。本稿では、このデータベースのための情報検索システムの設計について、書誌階層構造の扱い方を中心に報告し、進んで構造性を内包する文献データベースに対する検索システムの構成方法を検討する。

ABSTRACT

NACSIS : National Center for Science Information System has begun an information retrieval service "NACSIS-IR" for university researchers and librarians in 1987. The paper describes a method devised in incorporation of ISTP&B database into NACSIS-IR, which contains information of monograph level and individual paper level.

Conventional information retrieval systems do not provide functions to search and display records with hierarchical structure as seen in ISTP&B. A design of the online database developed here, providing users with the facility to easily locate and output data on monographs and papers, will be applicable to conventional information retrieval software packages. The paper also touches upon the problems in introducing bibliographic databases with hiera-

* Masashi MAKIMURA (Nagoya University Library)

**Masamitsu NEGISHI (National Center for Science Information System)

archical record organization into DBMS software and information retrieval packages.

《目 次》

要旨・Abstract	
1. はじめに	55
2. ISTP&B のオンライン化設計	57
2. 1 ISTP&B の概要	57
2. 2 オンライン化設計上の問題点と解決の方針	59
3. データ設計とレコード構造設計	60
3. 1 データベースの内容とデータ設計	60
3. 2 レコード構造の設計と出力方式	63
4. オンライン索引の設計	66
4. 1 目次型データベースにおける索引の構成	66
4. 2 親子2重化索引	68
5. 検索コマンドの設計	69
5. 1 検索コマンドの全体	69
5. 2 多種索引向き検索—SEARCH コマンドの機能拡張	69
5. 3 多種索引向き AND 検索—PHRASE コマンドの機能拡張	70
5. 4 正規化著者名向き検索—AUTHOR コマンドの内部処理	70
5. 5 親子関連表示—DISPLAY コマンドの開発	72
5. 6 親子構造検索—EXPAND コマンドの応用	73
5. 7 レコードの親子構造検索例	73
6. 情報検索システムにおける階層構造データベース	73
付表1 ISTP&B データベースの内容	80
付表2 言語コード表	81
付表3 論文種別コード表	81
付表4 ISTP&B 検索コマンド一覧	82

1. はじめに

学術情報センターでは、学術情報システムの一環として、従来から行ってきた目録所在情報サービスに加え、情報検索サービス（NACSIS-IR）を昭和62年4月に開始した。前者が、全国大学図書館ネットワークを目指した、図書館における目録業務を主対象にするサービスであるのに対し、後者は、大学等の研究者による直接利用や図書館での代行検索利用を目的にするものである。

本稿では、米国ISI社発行の科学技術関係の会議録・図書索引データベースISTP&B（Index to Scientific & Technical Proceedings & Books）について、NACSIS-IRでのオンライン化設計に関わる諸問題を、具体例を踏まえて報告する。これまでの文献情報検索のためのシステムは、文献抄録集データベースを前提にして、検索・表示機能の設計がなされていた。これに対してISTP&Bは、会議録、論文集などの内容目次に相当するデータを収録した、いわば「目次型」データベースであり、このための検索システムには、従来のものとは異なる新たな設計が要求される。その要点は、巻冊レベルと個別論文レベルという2階層構成の処理方式にあるといえるが、その場合においても、利用者から見たときには、NACSIS-IRの他の文献抄録型データベースと同様の考え方、方法で検索が行われるような設計が強く望まれる。

〈図1〉は、本稿で報告する検索システムによる、ISTP&Bデータベースの検索例である。これから分かるとおり、ISTP&Bも、見かけ上、他の文献抄録型データベースと同様の手法で検索できる。そのコマンド体系は東京大学大型計算機センターにおいて運用されてきたTOOL-IRに準拠したもので、NACSIS-IR全般についてこの方針を維持している〈注1～4〉。つまりここでは、データベース個々のデータ構成上の特徴を活かしつつ、なお統一的な利用者インターフェースを提供するという観点でのシステム設計が必要になるわけである。この点、汎用情報検索ソフトウェア製品は、単に規格化された検索・表示機能を提供するのみであるから、データベース個別の特性を敢えて阻却してオンライン化するか、あるいはデータ構成上の差異をそのまま露出してしまうことしかできない。前者ではデータベースの特性に応じた検索が不可能であるし、後者の場合には、データの構成を理解できかつそれに即して汎用コマンドを「応用」できる、特殊な一部利用者のためのシステムになってしまうであろう。いずれも、ここにいう「統一的」インターフェースとは異なるものである。

上のような役割を担うシステムは、各々のデータベースに対応して個別に設計されなければならず、また現状では相当の開発量を伴う。本稿ではISTP&Bについて、まず原データの内容を確認した上で、そのオンライン化における問題点を指摘し、NACSIS-IRにおける解決の方針を提示する。その後、NACSIS-IRでの具体的なデータの設計とレコード構造の設計、オンライン索引の生成方式、検索・表示のためのコマンド設計という順で報告する。おわりに、この種「目次型」データベースをはじめとする、書誌階層構造を内包するデータベースの「情報検索システム」での扱い方を、いわゆる「データベース管理システム」との対比を踏まえつつ検討し、今後における両者のあり方への示唆を得るものとした。

>>ISTP ← I S T P & B データベースの検索開始
 Welcome to NACSIS-IR ISTP database. (Rel. 880527) ←更新日付
 Copyright Institute for Scientific Information.
 This database contains 1021960 records. ←レコード数
 For further information, enter ?INFO subcommand.

TYPE IN COMMAND

1/ S_ALGEBRA ←「代数学」の検索

* 1543 1/ K.ALGEBRA ←該当1543件(文献集合1)

TYPE IN COMMAND

2/ S_POISSON ←「ポアソン」の検索

* 232 2/ K.POISSON ←該当232件(文献集合2)

TYPE IN COMMAND

3/ AND 1,2 ←「代数学」と「ポアソン」の組合せ

* 25 3/ 1 AND 2 ←該当25件(文献集合3)

TYPE IN COMMAND

4/ D M.B ←該当文献の表示(Bモードによる)

 (1)
 ACCN:840117397 RTYP:2 ←該当論文の標題等(子レコード)
 AUTH:GUEDIRA F
 AFFL:FAC SCI RABAT,DEPT PHYS/RABAT//MOROCCO
 TITL:STAR-PRODUCTS AND FORMAL LIE-ALGEBRAS ASSOCIATED WITH REGULAR AND
 TANGENTIALLY EXACT POISSON MANIFOLDS LANG:FR
 PAGE:0025-0028

ACCN:840117390 ←上出論文収載の会議録書誌事項(親レコード)
 CONF1:SESSIONS OF THE FRENCH ACADEMY OF SCIENCES ON MATHEMATICS, PAPERS
 PRESENTED
 TITL:COMPTES RENDUS DE L ACADEMIE DES SCIENCES SERIE I-MATHEMATIQUE
 LANG:EN
 YEAR:1984
 VISS:0299 N1 NDOC:10

.....

TYPE IN COMMAND

4/ END ←検索終了

End of NACSIS-IR ISTP database.
 Copyright Institute for Scientific Information.

図1 ISTP&Bデータベースの検索例(下線は利用者入力)

2. ISTP&B のオンライン化設計

2.1 ISTP&Bの概要

《ISTP&B データベース》 ISTP&B データベースは、次にのべる索引誌、ISTP と ISBC の電算化編集工程の中で得られる磁気テープ版の出版物であり、1982 年以降のものが電算ファイル化されている。NACSIS-IR では、1982 年以降の磁気テープを導入して、オンライン・データベース化している（レコード件数：約 96 万件。ただし、親〈巻冊レベル〉と子〈論文レベル〉レコードの合計値）。

版元の ISI (Institute for Scientific Information) 社は、引用索引 (Citation Index) の考案者として有名な Eugene Garfield 博士の創業にかかる会社で、その主要製品たる引用索引誌 SCI (Science Citation Index)、SSCI (Social Sciences Citation Index)、A&HCI (Arts & Humanities Citation Index) (上記 3 点とも、昭和 63 年度から NACSIS-IR でオンライン・サービス開始) や、雑誌の目次情報を収める Current Contents 誌は、文献検索の重要なツールとしてよく用いられている。

ISTP&B は、以前 ISI 社の直営システムでオンライン化されていたが、採算上の問題から、このシステム自体が数年前に廃止された〈注 6、10〉。ISTP&B は、オンライン化の難しさもあってか、現在、NACSIS-IR のほかには、西ドイツの DIMDI (Deutsches Institut für Medizinische Dokumentation und Information) でオンライン提供されるのみとなっている〈注 11〉。

《ISTP 誌、ISBC 誌》 ISTP&B は、Index to Scientific & Technical Proceedings (ISTP) と Index to Scientific Book Contents (ISBC) 両誌の合綴磁気テープ版である。1978 年に創刊された ISTP 誌は、世界各国で発行される図書および雑誌形態の会議録に対する索引誌（月刊および年刊累積版）で、農学、生物学、工学、物理学、化学、生命科学、臨床医学、環境科学、エネルギー等、科学技術分野全般を対象とし、1986 年時点では、年間約 4,000 件の会議についての論文約 127,000 件を収録している。これは、全世界の重要な会議論文の約 75～90% に相当するとされ、類似の会議論文索引誌 Conference Paper Index (Cambridge Scientific Abstracts 社)、Ei Engineering Conference Index (Engineering Information 社) 等に比べて、収録率はきわめて高い。また個別論文に関するデータまでは収録しない会議録出版情報誌 InterDock: Directory of Published Proceedings (InterDock 社) や Proceedings in Print (Proceedings in Print 社) と比較しても、その収録会議数において殆ど差がないとされている〈注 5～9〉。

1985 年創刊の ISBC 誌は、論文集等、複数著者による科学技術関係図書を対象とし、その図書中の論文（章）レベルの情報を収録する索引誌（年 3 回刊および年刊累積版）であり、1986 年現在では年間約 2,200 点の図書から約 31,000 件の論文（章）を収録している。

《ISTP 誌、ISBC 誌の内容構成》 ISTP&B データベースのオンライン化設計にあたっては、まず現行の冊子体 ISTP 誌および ISBC 誌が、どのような構成で編集・出版されているを確認しておく必要がある。両誌とも編集方法は基本的に同一で、(1)会議録または図書の標題等、全体的な書誌的事項と、各論文（章）ごとの著者・標題等とを目次的に示す、会議録の内容 (Contents of Proceedings) また

は図書の内容 (Contents of Books) (以下、「記入の本体」と呼ぶ) と称するセクションと、(2)それら記入の本体を指示する各種の索引群のセクションの2部構成になっている。〈図2〉および〈図3〉に、ISTP 誌における雑誌形態の会議録についての記入の本体と、ISBC 誌における共著図書の記入の本体の記載例を示す。

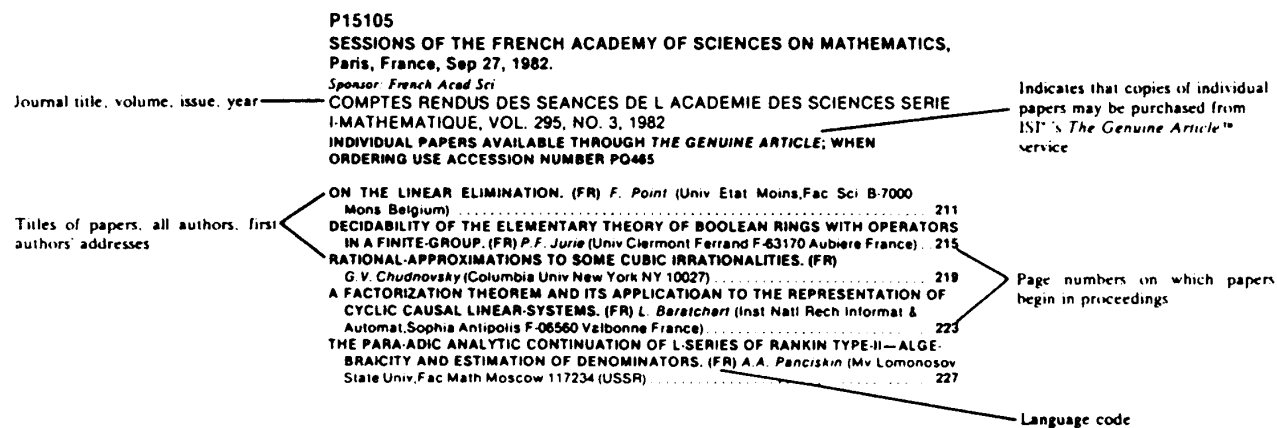


図2 会議録の「記入の本体」(ISTP 誌)

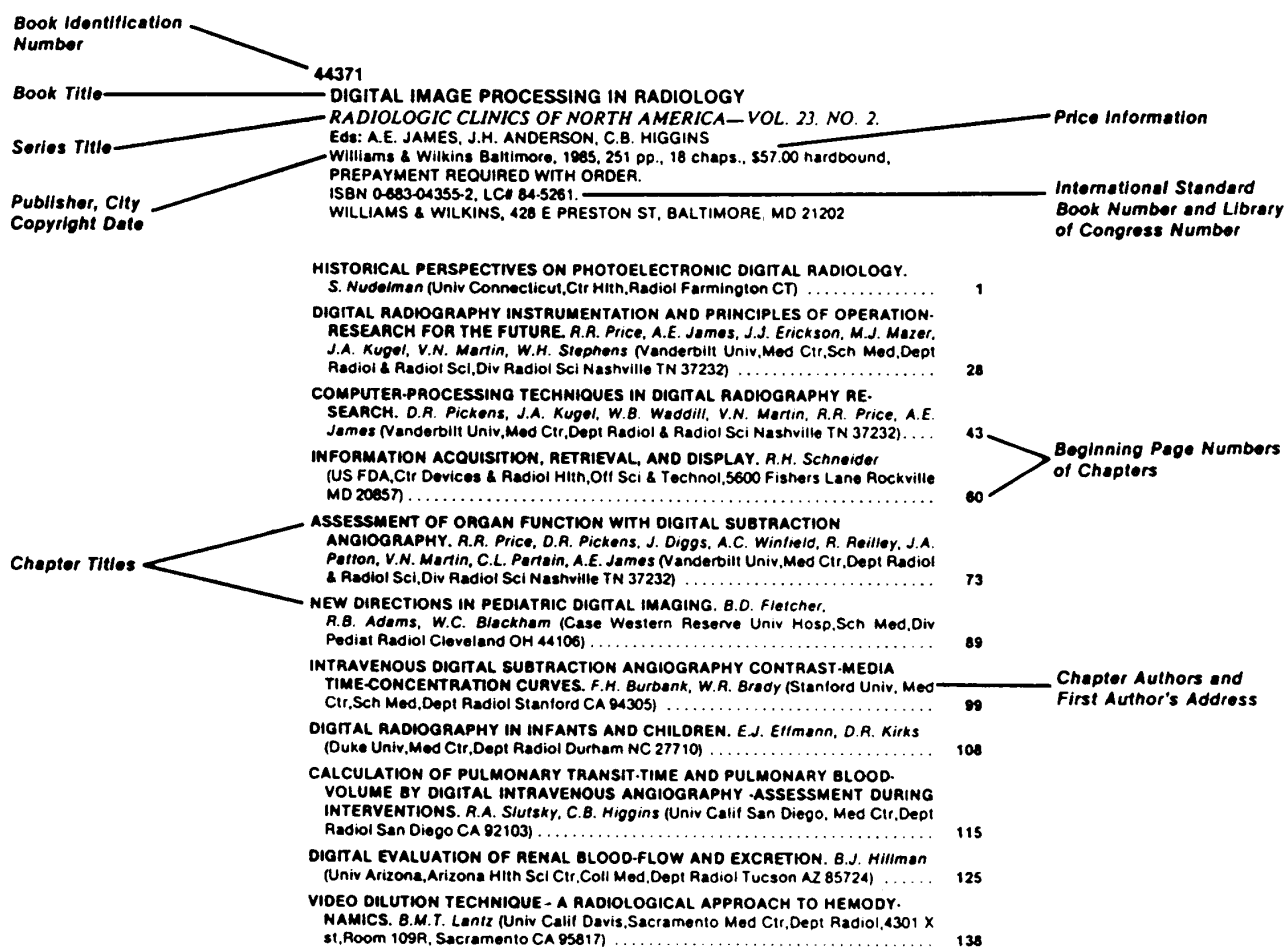


図3 共著図書の「記入の本体」(ISBC 誌)

索引セクションには、ISBC 誌の場合、著者・編者索引、主題語組合せ索引、団体索引の 3 種があり、ISTP 誌については、これらに主催団体索引、開催地索引、主題索引を加えて、多様なアクセスに備えるという構成になっている。各索引の概要は以下の通りである。

- (1) 著者・編者索引 (Author/Editor Index) —— 各論文の著者、および会議録・図書の編者の索引。
- (2) 主題語組合せ索引 (Permuterm Subject Index : PSI) —— 会議名や書名、あるいは個々の論文標題から抽出したキーワード (重要語) を 2 語ずつ組み合わせて構成した索引。自然語による検索が可能。
- (3) 団体索引 (Corporate Index) —— 論文の第 1 著者の所属機関の索引で、所属機関別の機関セクションと機関の所在地別の地名セクションに分かれている。
- (4) 主催団体索引 (Sponsor Index) —— 会議の主催団体名の索引で、一部団体著者を含む。
- (5) 開催地索引 (Meeting Location Index) —— 会議の開催地別の索引。
- (6) 主題索引 (Category Index) —— 会議の主題を約 200 のカテゴリーに区分した索引。

これら索引にみるとおり、ISTP、ISBC は、冊子体においてすでに相当のアクセス経路を具備したものになっている。こうした冊子体の既利用者が、NACSIS-IR の ISTP&B を利用するであろうことを考慮すれば、オンライン化設計にあたっては、ともあれこれら索引に匹敵する検索機能を包含することが、一つの条件になるであろう。

2.2 オンライン化設計上の問題点と解決の方針

データベースをオンライン検索システムに乗せるにあたっては、(1)原データベースの変換、(2)オンライン索引の生成、(3)検索コマンドの機能という 3 面からの検討が必要である。これらは相互に関連した事項であり、利用者からみて同一の機能を実現するにあたって、互いに代替可能な場合も多い。簡単な例を示せば、年号表記が原データでは 19 を省略して「88」などになっているが、検索コマンドでは「1988」と指定して検索できるようにしたい場合に、次の 3 つの方法が考えられる。すなわち、(1)原データベースをオンライン・データベースに変換する際に、88 → 1988 の変換を施す、(2)年号データからオンライン索引を生成する際に、「1988」として索引を作る、(3)検索コマンドで 1988 が指定された時、これを「88」に変換して検索を実行する、という 3 法である。もちろんこれらの手法の間では、プログラム開発の工数、計算機処理時間などについて長短があるし、副次的効果あるいは副作用にも差異が生じるから、その選択あるいは組合せはオンライン化設計上の重要な検討点になる。

文献抄録型データベースと比べたときの ISTP&B の特徴は、次の 2 点に要約できよう。

- (1) 論文要旨がないから、要旨中の単語を手掛かりとする「主題検索」はできず、書誌的データによる「書誌検索」にだけ頼らざるを得ない。このことは、オンライン索引を充実させて、各種のデータ項目についてきめ細かな検索ができるような設計を示唆する。すでに見た ISTP 誌、ISBC 誌における豊富な索引構成が、この際参考になろう。
- (2) 会議録・図書全体に関わるデータと個々の論文を記述するデータという 2 階層のデータを、効果的に検索しかつ表示しうる機能が必要である。文献抄録型データベースの場合でも、これらデータが収

録されてはいるが、検索システム設計上は、各文献レコードを相互に独立のもののみなして不都合を生じない。しかし、ISTP&B は元来が「目次」のデータベースであり、従って利用者にとって、会議録・図書を探してその目次を通覧する機能と、個別の論文を探してその関連データだけを表示する機能の双方ともに必要になる。

他の NACSIS-IR データベースと同様に、ISTP&B の場合も、日立製汎用情報検索ソフトウェア製品 ORION を基底システムとして用い、これに TOOL-IR 型の仲介プログラムを介在させて、検索システムを実現することを設計の前提にした〈注 12〉。これにより、他のデータベースのために開発した既存ルーチンの多くがそのまま、あるいは軽微な改変で、ISTP&B 向きに適用できるからである。この前提のもとに検討を加えた結果、(1)については、データ項目毎の多種のオンライン索引の生成と多種索引組合せ検索向きのコマンドの設計、(2)に関しては、親子 2 階層のレコード構成、親子 2 重化索引の生成、親子関連表示コマンドの設計という方法により解決を図った。以下、これらの点を、データ設計・レコード構造設計、オンライン索引設計、検索コマンド設計の 3 項に分けて述べることにしたい。

3. データ設計とレコード構造設計

3.1 データベースの内容とデータ設計

ISI から供給される ISTP&B データベース（原データベース）は、上述冊子体における記入の本体に相当するデータをそのまま収めたものである。これに対して、レコードの再構成、索引の切り出し等、オンライン検索向きに加工を加えて、NACSIS-IR・ISTP&B オンライン・データベースとするのである。〈付表 1〉に加工後の ISTP&B オンライン・データベースのデータ項目一覧を示す。この表からわかるように、NACSIS-IR における ISTP&B データベースでは、巻冊レベルのレコードと論文レベルのレコードに分けて、レコードを格納するようにしている（以下、これを各々「親レコード」、「子レコード」と表現する。詳細は後述）。〈図 4〉は NACSIS-IR・ISTP&B の出力例である。

付表の見出し欄は、レコードを表示したときの各データ項目の見出しを示し、項目内容欄は収録されているデータの内容を、索引欄は各項目から作成されるオンライン索引を示している。また、出力モード欄は、DISPLAY コマンドによる出力項目の詳細度レベルを示し、後述のように A モードから D モードへと次第に詳細な表示が行なわれる。各データ項目の内容は次のとおりである。

(1)アクセシオン番号 (ACCN) 《親》・《子》——アクセシオン番号は、文献レコードをオンライン・データベースに收容する際に割り振った NACSIS-IR での固有レコード番号であり、9 桁の年度ごとの一連番号である。

各会議録・図書について、親レコードに続けて子レコードを配し、これらに一連の番号を付与している。そこで後出の掲載論文数 (NDOC) と組み合わせれば、子レコードのアクセシオン番号は、親レコード自身のそれから計算できることになる。

(2)標題 (TITL) 《親》・《子》——親レコードの場合は会議録・図書の標題であり、子レコードの場

(1)

ACCN:820014996 RTYP:1 FTYP:ISTB 《図書レベル書誌事項、親レコード》
 TITL:TRACING NEURAL CONNECTIONS WITH HORSERADISH PEROXIDASE LANG:EN
 SERS:IBRO HANDBOOK SERIES : METHODS IN THE NEUROSCIENCES
 EDTR:MESULAM MM
 PBSR1:JOHN WILEY & SONS LTD
 PPBL1:CHICHESTER SUSSEX
 ADDR1:JOHN WILEY & SONS LTD,BAFFINS LANE,CHICHESTER, SUSSEX PO19 7UD UK
 YEAR:1982
 PAGE:0001-0251 NDOC:3 PRCE:52.00;26.00
 LCCN:81-14692 ↑
 ISBN:0-471-10028-5 以下の論文3件を収載

ACCN:820014997 《収載論文レベル事項、子レコード1》
 DTYP:1(Articles) RTYP:2 PACN:820014996
 AUTH:MESULAM MM
 AFFL:BETH ISRAEL HOSP,CHARLES A DANA RES INST,DEPT NEUROL,NEUROL UNIT/
 BOSTON//MA/02215
 TITL:PRINCIPLES OF HORSERADISH-PEROXIDASE NEUROHISTOCHEMISTRY AND THEIR
 APPLICATIONS FOR TRACING NEURAL PATHWAYS - AXONAL-TRANSPORT, ENZYME-
 HISTOCHEMISTRY AND LIGHT MICROSCOPIC ANALYSIS LANG:EN
 PAGE:0001-0151

ACCN:820014998 《子レコード2》
 DTYP:1(Articles) RTYP:2 PACN:820014996
 AUTH:CARSON KA, MESULAM MM
 AFFL:OLD DOMINION UNIV,DEPT BIOL SCI,ELECTRON MICROSCOPY LAB/NORFOLK//VA/
 23508
 TITL:ELECTRON-MICROSCOPIC TRACING OF NEURAL CONNECTIONS WITH HORSERADISH-
 PEROXIDASE LANG:EN
 PAGE:0153-0184

ACCN:820014999 《子レコード3》
 DTYP:1(Articles) RTYP:2 PACN:820014996
 AUTH:BISHOP GA, KING JS
 AFFL:OHIO STATE UNIV,DEPT ANAT/COLUMBUS//OH/43210
 TITL:INTRACELLULAR HORSERADISH-PEROXIDASE INJECTIONS FOR TRACING NEURAL
 CONNECTIONS LANG:EN
 PAGE:0185-0247

図4 ISTP & B の出力例

合は各論文の標題である。

- (3)言語 (LANG) 《親》・《子》——標題の言語を2桁のコードで示す〈付表2参照〉。
- (4)著者 (AUTH) 《親》・《子》——当該図書または論文の著者名。共著の場合は、コンマで区切って並記する。名前は姓、名の順で、名はイニシャルのみ。冊子体では、イニシャルの後ろにピリオドが付加されているが、データベースではこれら頭字そのままとし(中間名のある時はその頭字を接続)、「姓_名頭字」という「正規化」された形式にしている。著者名のオンライン索引も同じ形式で作成される。
- (5)ページ (PAGE) 《親》・《子》——親レコードの場合は、図書・会議録全体の開始—終了ページ。子レコードの場合は、当該論文の開始—終了ページ。
- (6)レコード区分 (RTYP) 《親》・《子》——親と子のレコードを識別するため、1桁の数字コードを割り当てた。
- 親レコード——”1” (21,526件、図書・会議録件数、1988年2月現在値)
- 子レコード——”2” (615,804件、上記に含まれる論文件数)
- (7)会議名等 《親》——会議録の場合の、会議名 (CONF 1, CONF 2)、開催地 (LOCN 1, LOCN 2)、会議開催日付 (DATE 1, DATE 2)、開催者 (CORG 1, CORG 2)、後援者 (SPSR 1, SPSR 2)。2つの会議に対応する会議録の場合には、それぞれ CONF 1, CONF 2 等に収容される。
- (8)翻訳標題 (TTIL) 《親》——英語以外のタイトルの英訳標題。
- (9)シリーズ名 (SERS) 《親》——図書の場合のシリーズ名。
- (10)編者 (EDTR) 《親》——図書、会議録、シリーズの編者名。前出「著者」欄と同様、正規化人名の形式。
- (11)団体著者 (CORP) 《親》——団体著者名。
- (12)出版者等 《親》——出版者 (PBSR 1, PBSR 2)、出版地 (PPBL 1, PPBL 2)、住所 (ADDR 1, ADDR 2)。
- (13)出版年 (YEAR) 《親》——出版年。
- (14)巻号等 (VISS) 《親》——雑誌等の巻号。巻は4桁の数字で、号はNにつづけて表示している。「VISS:0299 N 1」は、VOL. 299 NO. 1の意味。
- (15)掲載論文数 (NDOC) 《親》——その会議録・図書中に含まれる論文の数。これは子レコード検索のための重要項目であるから、オンライン・データベース作成時に、実際に存在する子レコード数を勘定して、挿入している(原データには誤りも見られる)。
- (16)価格 (PRCE) 《親》——図書、会議録の価格。
- (17)米国議会図書館カード番号 (LCCN)、国際標準図書番号 (ISBN) 《親》
- (18)主題区分 (CLAS) 《親》——対象とする主題分野(カテゴリー)を英字2文字の分類コード(約200分類)とその説明語句で表わしたもの。利用者に分かりやすいよう、説明語句をオンライン・データベース作成時に、コード表から挿入した。

(19)会議録・図書区分 (FTYP) 《親》

会議録—”ISTP”、図書—”ISTB”

(20)著者所属、住所 (AFFL) 《子》——第1著者の所属機関名称、住所。

(21)論文種別 (DTYP) 《子》——論文の種別を示す1文字の英字または数字〈付表3参照〉。説明をコード表から挿入。

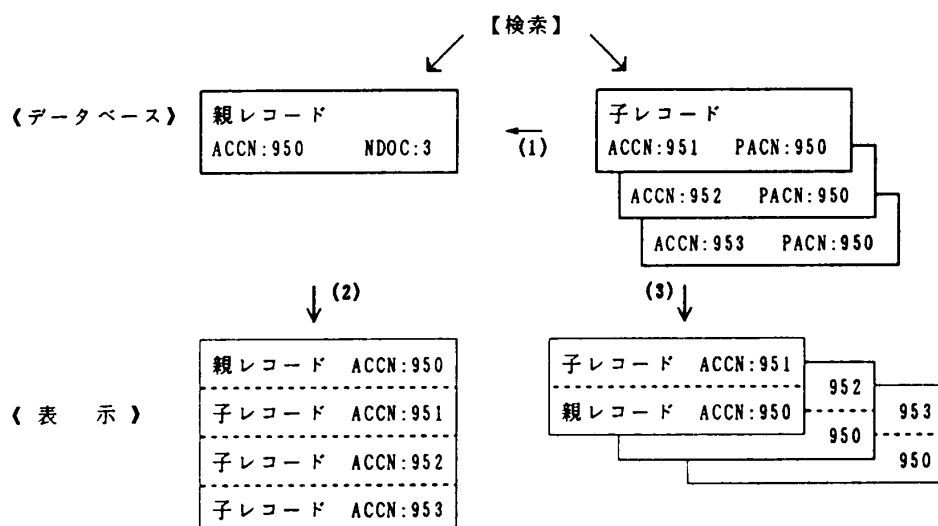
(22)親レコード・アクション番号 (PACN) 《子》——子レコードから親レコードを参照するため、親レコードのアクション番号をここに挿入した。

上に見るとおり、1～6が親・子レコードともに存在する項目、7～19は親独自の項目、20～22は子レコードのみの項目である。これらのうち、オンライン・データベース作成時に当方で新たに生成したり、また原データに相当の加工を施している項目は、1、6、15、18、21、22である。

3.2 レコード構造の設計と出力方式

〈図2〉、〈図3〉で示したように冊子体のISTP誌、ISBC誌では、記入の本体にその会議録・図書中に収録される全ての論文の情報が目次的に列挙される。この論文の数は大きな会議になると100件を超え、1会議（1記入）で数ページにもわたることがある。冊子体における一覧性はこのような場合とくに有利である。

オンライン検索の場合は、端末画面の大きさが制約されており、目次型の表示ばかりでなく、図書・会議録レベルのデータあるいは特定論文のデータだけが表示できるような機能も当然必要になる。こうした利用者側での自由度の確保と、データベース容量やオンライン・データベース更新作業の効率性を考慮して、NACSIS-IRでは、会議録・図書のレベルのデータと各論文レベルのデータを、「親レコード」と「子レコード」という、データベース上では別々のレコードとして格納することとした。このとき、親子のレコード間をアクション番号によって相互に結合しておき、DISPLAYコマ



- (1) 子(論文)レコードの著者・標題の親(巻冊)・子2重索引化(図8参照)
- (2) 親レコードヒット時の子レコード連続表示(図書目次型)
- (3) 子レコードヒット時の親レコード付帯表示(論文+掲載図書)

図5 ISTP & B データベースの構造と表示形式

(1) Aモード表示 (子レコードのみ簡略表示)

TYPE IN COMMAND
1/ D I.5, M. A

(5)
ACCN:880109076 RTYP:2 ←子レコード簡略表示
AUTH:NEGISHI M, NAITO E
TITL:INTER-UNIVERSITY INFORMATION NETWORK IN JAPAN - A DEDICATED VAN FOR
SCHOLARLY INFORMATION-FLOW (SIG/III, INFORMATION INSTITUTIONS AND THE
TRANSFORMATION OF SOCIETY)

(2) Bモード表示 (子レコード+親レコードの主要項目表示)

TYPE IN COMMAND
1/ D M.B, I. 5

(5)
ACCN:880109076 RTYP:2 ←子レコード
AUTH:NEGISHI M, NAITO E
AFFL:NATL CTR SCI INFORMAT SYST/TOKYO//JAPAN
TITL:INTER-UNIVERSITY INFORMATION NETWORK IN JAPAN - A DEDICATED VAN FOR
SCHOLARLY INFORMATION-FLOW (SIG/III, INFORMATION INSTITUTIONS AND THE
TRANSFORMATION OF SOCIETY) LANG:EN
PAGE:0181-0185

ACCN:880109044 ←親レコード
CONF1:50TH ANNUAL MEETING OF THE AMERICAN SOC FOR INFORMATION SCIENCE :
INFORMATION - THE TRANSFORMATION OF SOC
TITL:ASIS 87: INFORMATION : THE TRANSFORMATION OF SOCIETY LANG:EN
SERS:PROCEEDINGS OF THE ASIS ANNUAL MEETINGS
EDTR:CHEN CC
PBSR1:LEARNED INFORMATION INC
YEAR:1987 VISS:0024 NDOC:46
LCCN:64-8303 ISBN:0-938734-19-9

(3) Dモード表示 (子レコード+親レコード詳細表示)

TYPE IN COMMAND
1/ D M.D, I. 5

(5)
ACCN:880109076 ←子レコード
DTYP:1(Articles) RTYP:2 PACN:880109044
AUTH:NEGISHI M, NAITO E
AFFL:NATL CTR SCI INFORMAT SYST/TOKYO//JAPAN
TITL:INTER-UNIVERSITY INFORMATION NETWORK IN JAPAN - A DEDICATED VAN FOR
SCHOLARLY INFORMATION-FLOW (SIG/III, INFORMATION INSTITUTIONS AND THE
TRANSFORMATION OF SOCIETY) LANG:EN
PAGE:0181-0185

ACCN:880109044 RTYP:1 ←親レコード
FTYP:ISTP
CONF1:50TH ANNUAL MEETING OF THE AMERICAN SOC FOR INFORMATION SCIENCE :
INFORMATION - THE TRANSFORMATION OF SOC
LOCN1:BOSTON, MA DATE1:87/10/4-8
SPSR1:AMER SOC INFORMAT SCI
TITL:ASIS 87: INFORMATION : THE TRANSFORMATION OF SOCIETY LANG:EN
SERS:PROCEEDINGS OF THE ASIS ANNUAL MEETINGS
EDTR:CHEN CC
PBSR1:LEARNED INFORMATION INC PPBL1:MEDFORD
ADDR1:LEARNED INFORMATION INC, 143 OLD MARLTON PIKE, MEDFORD NJ 08055
YEAR:1987 VISS:0024 PAGE:0001-0287 NDOC:46 PRCE:40.00
LCCN:64-8303 ISBN:0-938734-19-9
CLAS:EY(COMPUTER APPLICATIONS & CYBERNETICS);PE(OPERATIONS RESEARCH &
MANAGEMENT SCIENCE)

図6 子レコードがヒットした場合の表示例

(1) Bモード表示 (親レコードの主要項目表示)

TYPE IN COMMAND
1/ D I.9, M. B

(9)
ACCN:840126746 RTYP:1
CONF1:2ND SYMP ON MARKER PROTEINS IN INFLAMMATION
LOCN1:LYON.FRANCE DATE1:83/6/27-30
SPSR1:GRP ETUDE RECH MARQUERS INFLAMMAT/ INST PASTEUR LYON & SUD EST/ FDN
MERIEUX/ SOC FRANCAISE IMMUNOL/ ASSOC ANCIENS INTERNES & INTERNES
PHARMACIE LYON/ COMMISS MEDICALE CONSULTATIVE HOSPICES CIVILS LYON/
BIOMERIEUX/ TRAVENOL FRANCE/ DU PONT CO/ HOECHST BEHRING
TITL:MARKER PROTEINS IN INFLAMMATION, VOL 2 LANG:EN
EDTR:ARNAUD P, BIENVENU J, LAURENT P
PBSR1:WALTER DE GRUYTER
PPBL1:D-1000 BERLIN 30
YEAR:1984
PAGE:0001-0687 NDOC:100 PRCE:102.50
LCCN:84-9462
ISBN:3-11-009872-5

(2) Dモード表示 (親レコード+子レコードの詳細表示、目次同等)

TYPE IN COMMAND
1/ D I.9, M. D

(9)
ACCN:840126746 RTYP:1 FTYP:ISTP ←親レコード
CONF1:2ND SYMP ON MARKER PROTEINS IN INFLAMMATION
LOCN1:LYON.FRANCE DATE1:83/6/27-30
SPSR1:GRP ETUDE RECH MARQUERS INFLAMMAT/ INST PASTEUR LYON & SUD EST/ FDN
MERIEUX/ SOC FRANCAISE IMMUNOL/ ASSOC ANCIENS INTERNES & INTERNES
PHARMACIE LYON/ COMMISS MEDICALE CONSULTATIVE HOSPICES CIVILS LYON/
BIOMERIEUX/ TRAVENOL FRANCE/ DU PONT CO/ HOECHST BEHRING
TITL:MARKER PROTEINS IN INFLAMMATION, VOL 2 LANG:EN
EDTR:ARNAUD P, BIENVENU J, LAURENT P
PBSR1:WALTER DE GRUYTER
PPBL1:D-1000 BERLIN 30
ADDR1:WALTER DE GRUYTER, 200 SAW MILL RIVER ROAD, HAWTHORNE NY 10532
ADDR2:WALTER DE GRUYTER, GENTHINER STRASSE 13, D-1000 BERLIN 30, W GERMANY
YEAR:1984
PAGE:0001-0687 NDOC:100 PRCE:102.50
LCCN:84-9462 ↑論文100件を収録
ISBN:3-11-009872-5
CLAS:M ; Q ; IO; HE(EDUCATION, SPECIAL); IS; RY(NUCLEAR SCIENCE & TECHNOLOGY); & ;
OL; CU(BIOLOGY)

ACCN:840126747 ←子レコード1
DTYP:1(Articles) RTYP:2 PACN:840126746
AUTH:KUSHNER I
AFFL:CASE WESTERN RESERVE UNIV, CLEVELAND METROPOLITAN GEN HOSP, DEPT MED/
CLEVELAND//OH/44109
TITL:INDUCTION AND CONTROL OF ACUTE PHASE REACTANT SYNTHESIS LANG:EN
PAGE:0003-0014

ACCN:840126748 ←子レコード2
DTYP:1(Articles) RTYP:2 PACN:840126746
AUTH:FELDMANN G
AFFL:FAC MED XAVIER BICHAT, BIOL CELLULAIRE LAB, 16 RUE HENRI HUCHARD/F-75018
PARIS//FRANCE
TITL:USE OF IMMUNOMORPHOLOGICAL METHODS AND HEMOLYTIC PLAQUE TESTS TO
INVESTIGATE THE SYNTHESIS AND SECRETION OF PLASMA-PROTEINS BY THE
LIVER-CELLS DURING THE ACUTE INFLAMMATORY REACTION LANG:EN
PAGE:0015-0024

.....

図7 親レコードがヒットした場合の表示例

ンドの表示モード指定により、必要に応じて関連する親ないし子のレコードが同時に表示されるようにした。また、子レコードの索引を親レコードの索引にも二重にポスティングすることにより、親子の組合せ検索も可能にしている（後述）。このような ISTEP&B データベース上のレコード、および検索・表示の関係を概念的に表わせば〈図5〉のようである。

〈図6〉は子レコードがヒットした場合の表示例である。(1)は DISPLAY の A モードの出力で、ここでは該当する論文(子)レコードのみが表示される。(2)は B モードの出力であり、論文(子)レコードに続けて、その論文が含まれる会議録・図書(親)レコードが表示され、通常の文献データベース的出力が得られる。この例の場合、親レコードの NDOC は「46」となっているので、この会議録には他に 45 論文含まれていることがわかる。それらの論文も参照したい場合には、親レコードの ACCN を使って、

DISPLAY A. 880109044.M. C

とすれば、目次風の表示が得られる。

〈図7〉は親レコードがヒットした場合の表示例である。(1)は B モードの出力で該当の会議録・図書のレコードのみを表示している。この会議録・図書にどのような論文が含まれているか参照したい場合には、(2)のように D モードまたは C モードの出力を行うと、親レコードに続けてそれに含まれる論文のレコードが列挙され、目次型の出力が得られる仕組みにした。このように、DISPLAY コマンドに、親→子または子→親という関連出力の機能を内蔵させ、利用者が「構造」を特段意識しなくても、必要な情報が得られるよう配慮した。

4. オンライン索引の設計

4.1 目次型データベースにおける索引の構成

NACISIS-IR における ISTEP&B の検索は、他のデータベースと同様一次検索と二次検索の二様によって行われる。一次検索とは図書の索引を引くのと同じように、索引ファイル上に予め抽出されている索引語(オンライン索引)を指定して、その語を含む会議録・図書あるいは各論文の集合(文献集合)を一時的に作ることである。二次検索には、一次検索によって作られた複数の文献集合について、それらの和集合や積集合を作ることと、1つの文献集合内の各文献について、特定の項目の文字列を走査し、適当な文献だけを選択して新たな文献集合を作ることの2つがある。

オンライン索引は、2.1に示した冊子体の索引セクションに相当するものであり、その設計によって、基本的検索機能である一次検索の機能範囲が左右される。ISTEP&B は、文献抄録データベースにくらべて、1レコード当りのデータ量が少ない。論文要旨やキーワード等々、主題からのアプローチのためのデータがないからである。そして、このことは各種の書誌的データ項目を意識的に使い分けて検索を行う必要を意味し、これは索引の編成上では、データの種別を明示するような索引編成、つまり細分化されたオンライン索引を意味することになる。そこで、既述の冊子体の索引構成にも留意して、次の15種類のオンライン索引を用意することにした。各索引は、その種別を表わすプレフィク

スを索引語の前に置いて、「プレフィクス、索引語」の形式で編成されている。〈附表1〉のオンライン索引欄には、各データ項目から作成される索引のプレフィクスを示した。

(1)一般語索引 (K.) —— 会議名 (CONF 1, CONF 2)、会議開催地 (LOCN 1, LOCN 2)、標題、論文標題 (TITL)、翻訳標題 (TTIL)、シリーズ名 (SERS)、団体著者 (CORP) 中の各「単語」を切り出し、索引語として収録している。

単語の区切りは空白やコンマ等の記号類であるが、- (マイナス、ハイフン) については、ハイフネーションによる複合語 (ISI では、特定の対象を表わす句は単語間にハイフンを補うという編集方針であり、したがってハイフネイテッド・ワードがデータベース中に頻出する) を有効に検索できるよう、次のような特別の処置を行なっている。すなわち、BLOOD-FLOW からは、K. BLOOD、K. FLOW、K. BLOODFLOW という3語の索引を生成するというように、各構成単語と全体の連結形の双方を索引化する。また、会議開催地については、都市名のみ表示形のまま索引化している (例: K. NEW YORK)。空白区切りによる単語化を行なわない。

各索引語は、語尾変化による検索漏れを防止するため、4文字以上の語の語尾の「S」は複数形の S とみなし切落している (SYSTEMS→SYSTEM)。また、語尾による検索が有用な場合もあるので、会議名、標題、論文標題、翻訳標題、団体著者からの索引語については、語尾からの逆引索引を作っている。後述の SEARCH コマンドや PHRASE コマンドで、プレフィクスを省略した場合は、この一般語索引が検索される。なお、子レコードの標題 (論文標題) については、親・子レコードの2重化索引を生成している (詳細は次項参照)。

(2)著者索引 (A.) —— 著者 (AUTH)、編者 (EDTR) を収録する索引。前述のように著者、編者の名前は「姓 名頭字」の記法で正規化されており、著者索引も同じ形式で作成してある (例: A. RICHTER SL)。子レコードの著者については、親・子レコードの2重化索引を生成 (詳細は次項)。

(3)後援者索引 (SP.) —— 会議の後援者 (SPSR 1, SPSR 2) である団体名を収録する索引。その名前がデータベース中で統制されていないので、単語別に索引化した (例: SP. IEEE)。

(4)著者所属機関索引 (AF.) —— 論文の第1著者の所属機関 (AFFL) を収録する索引で、単語別に索引化 (AF. INSTITUTE)。なお、この項目には全般的に省略形が用いられているので、検索上注意を要する。

(5)出版者索引 (PB.) —— 会議録・図書の出版者 (PBSR 1, PBSR 2) を収録する索引。統制されていないため単語別に索引化 (PB. JOHN、PB. WILEY)。

(6)会議開催年・出版年索引 (Y.) —— 会議の開催日付 (DATE 1, DATE 2) 中の開催年に「19」を補って索引化 (Y.1985)。また出版年 (YEAR) を索引化。

(7)米国議会図書館カード番号索引 (LC.) —— 米国議会図書館カード番号 (LCCN) からハイフンを除去してを索引化 (LC.90456114)。

(8)国際標準図書番号索引 (B.) —— 国際標準図書番号 (ISBN) からハイフンを除去して索引化 (B.

0571117597)。

(9)掲載論文数索引 (NDOC.) ——会議録・図書中に掲載された論文数 (NDOC) の索引 (NDOC.100)。

(10)主題区分索引 (C.) ——主題区分 (CLAS) の2桁の英字コード (例:C.AA)。

(11)レコード区分索引 (R.) ——レコード区分 (RTYP) から作成した親と子のレコード区分索引。親レコード—”R.1”、子レコード”R.2”。

(12)会議録・図書区分索引 (F.) ——会議録・図書区分 (FTYP) についての索引。会議録—”F.ISTP”、図書—”F.ISTB”。

(13)論文種別索引 (D.) ——論文種別 (DTYP) のコード索引 (例:D.1)。〈付表3参照〉

(14)言語索引 (L.) ——言語欄 (LANG) の言語コード索引 (例:L.EN)。〈付表2参照〉

(15)親レコード・アクセス番号索引 (PACN.) ——子レコード中の親レコード・アクセス番号 (PACN) の索引 (PACN.870000133)。

4.2 親子2重化索引

索引ファイルは通常、索引語とそれを含むレコードのアクセス番号の対から構成される。これによって、ある索引語で検索すれば、それを含むレコード群 (文献集合。内部的にはレコードのアクセス番号の群) が即座に得られることになる。ところで ISTP&B の場合、会議録・図書レベルのデータからなる親レコードと個々の論文に対応する子レコードという、いわば階層的なレコード構成を行なった。ここでは、親レコード、子レコードとも各々独自のアクセス番号を持ち、検索上は相互に無関係のものとして扱われることになる。

しかし実際には論文レベルのキーワードと図書レベルのキーワードを組合せて、検索したい場合も生じる。すなわち、会議論文について特定の著者名と開催地名を手掛かりにして検索を試みるような事例である。このとき上述の親子レコード分離のままでは、論文著者名 (親レコード項目) と会議開催地 (子レコード項目) を別個に検索することはできるが、その両者に適合するレコードは得られない。

そこでこのような場合にも、ともあれ図書・会議録 (親) レコードがヒットするよう、

(1)子レコード・論文標題からの単語索引 と

(2)子レコード・論文著者名索引

については、当該子レコードのアクセス番号に加えて、その親レコードのアクセス番号を持つ索引をも生成することにした (図8)。つまり索引編成上は、会議録・図書レベル・レコードにも、論文標題やその著者名が、あたかも収容されているというような扱い方である。これにより、親データと子データを組み合わせた自由度の高い検索が可能になる。

なお、これにともない、例えば論文著者名だけで検索すれば、常に該当の子と親の2レコードがヒットすることになり、これをそのまま表示すれば、ほぼ同一の内容データが重複して出力される。これを避けたい場合には、親子レコード区分 (RTYP) と組合せて、子レコードのみを対象に検索すればよい。

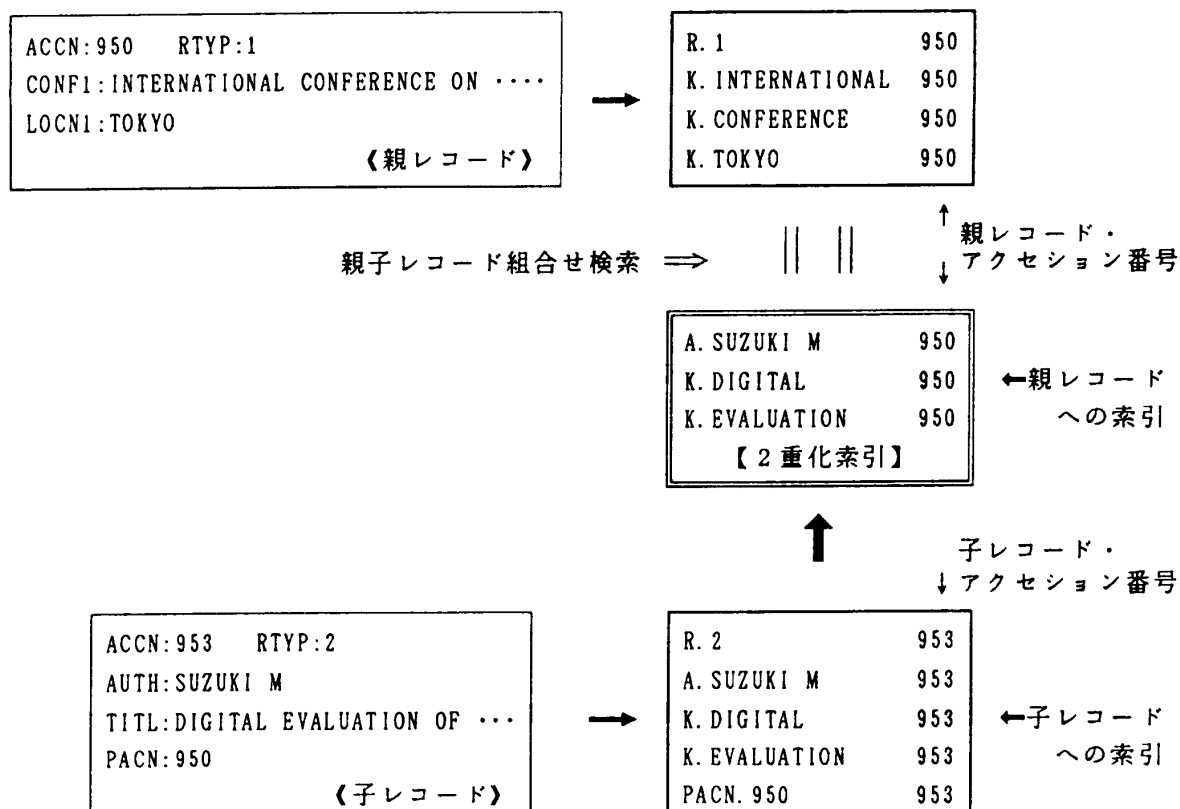


図8 親・子レコード2重化索引の方法

5. 検索コマンドの設計

5.1 検索コマンドの全体

検索コマンドには、機能的に、一次検索コマンド、二次検索コマンド、表示コマンド、補助的コマンド、終了コマンド等がある。〈附表4〉にISTP&B データベースで利用できる検索コマンドの一覧を示す。なお、コマンドの書式記述中、角括弧 [] は、その中に記された項目が省略可能であることを示す。中括弧 { } 中には択一すべき相互背反項目を列挙する。「……」は繰り返し指定が可能であることを示す。これらコマンドの多くは、通常の文献抄録型データベースである COMPENDEX 等に対して開発されたものが、そのまま適用できたものである〈注1〉。以下には、目次型データベース ISTP&B のオンライン化に際して、新たに設計・開発されたコマンド機能とその関連事項を摘記する。

5.2 多種索引向き検索——SEARCHコマンドの機能拡張

SEARCH term [, term,……] [{ AND, NOT } term [, term,……]]

各 term の形式は [prx.] [*] kw [*] [- - [prx.] [*] kw [*]]

ここで kw は検索キーワード、prx は索引プレフィクス (標準 K.、既出指定の省略可)

SEARCH コマンドは、基本的に一般語索引を検索するコマンドである。しかし、ISTP&B では既述のとおり、細分化されたオンライン索引が必要になるので、SEARCH コマンドではこれら索引の種類、具体的にはプレフィクスの指定が、簡便に行えるよう配慮した。

まず、プレフィクスの標準値を一般語索引 (K.) として、この指定の省略を可とした。そこで例えば、

SEARCH AUTOMO*, CAR. AND. ENGINE, POWER

などは、一般語索引を対象とした検索になる。次にプレフィクスを指定して、

S ECOLOG*, C. GU. AND. A. SUZUKI S

とすれば、一般語索引、主題区分索引、著者名索引を組み合わせた検索が行われる。このようなとき、同一のプレフィクスを繰り返して指定するのはわずらわしいので、左に指定したプレフィクスと同じものの指定は省略できるようにして、

S C. MM, MQ, MR. AND. Y. 1987,1988

は「S C. MM, C. MQ, C. MR. AND. Y. 1987, Y. 1988」と解釈される。これにより、多種の索引を使い分けた検索指定を敢えて1コマンドで実行したい場合にも、これを相当程度簡便に行い得るようになったと思われる。

5.3 多種索引向きAND検索——PHRASEコマンドの機能拡張

PHRASE [prx.][*]kw[*] [prx.][*]kw[*]……

プレフィクス(prx)は標準K.で、既出prx指定の省略可。ストップ・ワード除去処理内蔵。

PHRASE コマンドも、SEARCH コマンド同様、一般語索引を対象に、その多重 AND 検索を主眼にするものであった。しかし、ISTP&B におけるオンライン索引の細分化に応じて、上記 SEARCH と同様のプレフィクス指定とその省略法を導入した。すなわち、

PHRASE ADVANCES IN CHEMICAL PHYSICS

のようにすれば、標準プレフィクス K. が仮定されて、各語について一般語索引が検索された後、その間の AND がとられる。つぎに、上例に出版者と刊年を検索条件として付加して、

PHRASE ADVANCES IN CHEMICAL PHYSICS PB. JOHN WILEY Y.1988

と指定すれば、これは、

PHRASE K.ADVANCES K.CHEMICAL K.PHYSICS PB.JOHN PB.WILEY Y.1988

と解釈され、JOHN WILEY 発行の ADVANCES IN CHEMICAL PHYSICS の 1988 年刊行のものが検索されることになる(図9)。これにより、各種索引の組合せ検索が能率よく実行できるようになったといえよう。

5.4 正規化著者名向き検索——AUTHORコマンドの内部処理

AUTHOR 正規化著者名[*][, 正規化著者名[*]………]

AUTHOR コマンドは著者索引を検索するコマンドである。著者名については、前述のように「姓名の頭字連結」という形式(正規化著者名)で索引化されている。この点を考慮し、AUTHOR コマンドでは適当な内部処理を加えて、次のような検索ができるようになっている。これは既に COMPENDEX において導入したもので、下例のようである。

AUTHOR SMITH——姓のみ指定 (SMITH A, SMITH AGR, SMITH XXX 等が該当)

```

TYPE IN COMMAND
  1/ PHRASE ADVANCES IN CHEMICAL PHYSICS PB. JOHN WILEY Y.1988

*   5282      K. ADVANCE
*   14818     K. CHEMICAL
*   4640      K. PHYSIC
*   587       PB. JOHN
*   484       PB. WILEY
*   3043      Y. 1988
*   2         2/ "K. ADVANCE" AND "K. CHEMICAL" AND "K. PHYSIC" AND "PB. JO

TYPE IN COMMAND
  2/ D M. B
    
```

```

-----
(      1)
ACCN:880022554   RTYP:1
TITL:ADVANCES IN CHEMICAL PHYSICS, VOL 72   LANG:EN
SERS:ADVANCES IN CHEMICAL PHYSICS
EDTR:PRIGOGINE I, RICE SA
PBSR1:JOHN WILEY & SONS INC
PPBL1:NEW YORK
YEAR:1988
VISS:0072
PAGE:0001-0345   NDOC:3   PRCE:75.00
LCCN:58-9935
ISBN:0-471-63626-6
    
```

```

-----
(      2)
ACCN:880023117   RTYP:1
TITL:ADVANCES IN CHEMICAL PHYSICS, VOL 70, PT 1: EVOLUTION OF SIZE EFFECTS
      IN CHEMICAL DYNAMICS   LANG:EN
SERS:ADVANCES IN CHEMICAL PHYSICS
EDTR:PRIGOGINE I, RICE SA
PBSR1:JOHN WILEY & SONS INC
PPBL1:NEW YORK
YEAR:1988
VISS:0070
PAGE:0001-0556   NDOC:10   PRCE:100.00
LCCN:58-9935
ISBN:0-471-62784-4
    
```

図9 PHRASE コマンドの使用例

AUTHOR SMITH A——名の頭字も指定 (SMITH A のみ該当。SMITH ASR 等は該当せず)

AUTHOR SMITH ASR——同上 (SMITH ASR のみ該当。SMITH A は該当せず)

AUTHOR SMITH A *——名の頭字前方一致 (SMITH A, SMITH ASR, SMITH AX 等が該当)

AUTHOR SMITH *——姓の前方一致 (SMITH XX, SMITHERS XX, SMITHSON XX 等が該当)

つまり、コマンド中に前方一致の指定 (*) がなければ、内部処理で空白と * を補って検索するので ("SMITH" → "SMITH *", "SMITH A" → "SMITH A *"), 利用者はいえて "SMITH *" のような指定をする必要はない。

5.5 親子関連表示——DISPLAYコマンドの開発

```

                { A.アクセション番号 } {A}
                {                       } {B}
DISPLAY {      {N.n      } } [M. {C}]
                {[S.文献集合番号][, {I1[,I2・・}]]} {D}
                {I1-I2      } {X}
    
```

N.n は集合の最初の n 件を表示、I₁,I₂・・・は集合中の i₁,i₂・・・番目を表示、I₁-I₂ は集合中の i₁ 番目から i₂ 番目までを表示、M.A 等は出力モードで表示データ項目の詳しさを指示

DISPLAY コマンドは作成された文献集合内の各文献の内容を表示させるコマンドである。また、アクセション番号を直接指示して、そのレコードの内容を表示させることもできる。上記のコマンド書式をみる限りは、徒来のもと同であるが、既に示したように、ISTP&B における親子レコード構造を活かした表示形式を得るよう、〈表1〉のコマンド内部処理を行っている。すなわち表示対象とされるレコードの親・子区分と出力モード指定の組合せに応じて、適切なデータが表示されるよう設計した（細目は〈附表1〉参照）。

DISPLAY コマンドについては、データベースごとに、その内容に即してこの種の特殊処理が必要になる場合が多いものである。例えば学術雑誌総合目録データベースのための DISPLAY には、所蔵データを走査しながら、該当のものだけを表示する機能を内蔵させた〈注4〉。ただし、その場合は、コマンド書式上に、表示すべきデータを指定するパラメーター（オペランド）が追加されており、この機能の存在が外見的にも明らかになっている。ISTP&B では、こうした外見的な相違は現われない。こちらの方が利用者にとって好ましいことはいうまでもなく、設計側としては出来る限りこの「見えない努力」方式を追求するべきであろう。

表1 DISPLAY コマンドでの親子レコード組合せ表示

出力 モード	表示対象レコードの種別	
	親レコード〈図7〉	子レコード〈図6〉
M. X	親レコード(自身)のデータ表示	子レコード(自身)のデータ表示
M. A		
M. B		
M. C	親レコード(自身)のデータ表示 +	子レコードのPACN項目から 親レコードのアクセション番号 を得、これを検索して、 書誌型表示
M. D	NDOC項目により子レコード(後 続レコード)のアクセション番 号を計算し、これを順次検索し て、目次型表示	

5.6 親子構造検索——EXPANDコマンドの応用

EXPAND ACCN, LINE=文献集合番号, OR, PREFIX="PACN."

EXPANDは、本システムの基底に用いている汎用情報検索ソフトウェアORIONの標準コマンドの一種で、あるデータ項目に着目して、その項目の内容を検索語として、検索を行わせるコマンドである〈注12〉。つまり検索語を直接指定するのではなく、検索語を含むデータ項目を指定することになる。ISTP&Bの場合、これは親レコードからその子レコードを検索するのに有効で、上記のように指定すればよい。これは、指定された文献集合(LINE=)中の各レコードのアクセション番号(ACCN)にプレフィクス"PACN."を付加したもの、すなわち"PACN.accn"を検索キーとして検索を行い、結果を順次に論理和(OR)せよという指令であり、従って、親レコードを元にして、その子レコードのすべてが検索されることになる。オンライン索引編成の関係で、検索結果として親レコードのみが得られる場合も多いが、所望の論文はその子レコードの中の一つであるというようなとき、このEXPANDコマンドにより子レコード群を検索し、これにLIMITを施して、所期の論文のみを抽出することなどができる。EXPANDを用いた親子構造関係の検索例を次項に掲げておく。

5.7 レコードの親子構造検索例

先にも記したとおり、NACISIS-IR・ISTP&Bデータベースでは、「目次型」データベースの特性に鑑み、図書・会議録の巻冊レベルの書誌的データとそれらに含まれる論文個別の書誌的データを別個のデータベース・レコードとするという方法を開発した(親レコード、子レコード)。そして、一部データは親子二重に索引し、またDISPLAYの際には親子を連結して表示するなど、利用者のごく自然にアクセスできるような方策を講じた。一方、こうしたレコード間の親子関係を積極的に用いた検索も可能であり、この点を次の例題を使って示す。

《例題》 1) 会議録標題、会議名に「超伝導」とある会議録の全所収論文(論文標題には「超伝導」がなくてもよい)と、2) 論文標題中に「超伝導」を含み、会議録標題、会議名には「超伝導」を含まない会議発表論文を検索して表示する。

この例題は、1) 親レベルで検索して、その子レコードを全件表示する、2) 子レベルで検索した上で、そのうちの1)に該当しないもののみ選択して表示する、という2つの手順で検索する。1)はDISPLAYの親→子連結表示機能でそのまま可能である。そして、2)の手順において、1)に該当するものを除去する際に、親→子関係の構造に沿った検索が必要になり、これにEXPANDコマンドを用いる〈図10〉。

6. 情報検索システムにおける階層構造データベース

いわゆる「データベース管理システム」(DBMS)の世界では、本稿にみたようなレコード間の構造を操作することに主眼が置かれており、ISTP&Bについても、この種ソフトウェアを適用しさえすれば、大きな問題は生じないかに見える。ところが、既存DBMSの類では、全単語索引とその組合せ検索という、文献データベース向きの肝心の機能、すなわち「情報検索システム」の世界における眼目

- 1) 会議録標題、会議名に「超伝導」とある会議録の全所収論文(論文標題に「超伝導」がなくても可)
- 2) 論文標題中に「超伝導」を含むが、会議録標題、会議名には「超伝導」を含まない論文

TYPE IN COMMAND

```

1/ LOOK SUPERCONDUCT* ←「超伝導」索引語の点検

A 1215 K. SUPERCONDUCTING
B 2 K. SUPERCONDUCTINGFERROMAGNETIC ←ハイフン連結語
C 2 K. SUPERCONDUCTINGINSULATING
D 2 K. SUPERCONDUCTINGMAGNETOMETER
E 2 K. SUPERCONDUCTINGNORMALSUPERCONDUCTING
F 2 K. SUPERCONDUCTINGSUPERHEATED
G 2 K. SUPERCONDUCTINGTC
H 2 K. SUPERCONDUCTINGTONORMAL
I 65 K. SUPERCONDUCTIVE
J 475 K. SUPERCONDUCTIVITY
K 2 K. SUPERCONDUCTIVITYINDUCED
L 643 K. SUPERCONDUCTOR
M 2 K. SUPERCONDUCTORMAGNETIC
N 2 K. SUPERCONDUCTORMETAL
O 4 K. SUPERCONDUCTORSEMICONDUCTOR
P 2 K. SUPERCONDUCTORSEMIMETAL
Q 2 K. SUPERCONDUCTORSTHETA

1/ S SUPERCONDUCT* ←超伝導で検索
* 2236 1/ K. SUPERCONDUCT ( 17 TERMS COMBINED)

2/ S FTYP.1STP ←会議録
* 832554 2/ F.1STP

3/ S RTYP.1 ←親レコード
* 23821 3/ R.1

4/ S Y.1987,1988 ←1987、1988年
* 5832 Y.1987
* 725 Y.1988
* 6335 4/ "Y.1987" OR"Y.1988"
    会議名、会議録標題、所収論文標題のいずれかに「超伝導」
5/ AND 1,2,3,4 ←を含む1987、1988年の会議録(親レコード)
* 88 5/ 1 AND 2 AND 3 AND 4 ←88巻あり

6/ LIMIT 5,TITL:SUPERCONDUCT ←会議録標題に「超伝導」とあるもの
* 4 6/ SCAN 5 TITL *"SUPERCONDUCT"*

7/ LIMIT 5,CONF1:SUPERCONDUCT ←会議名に「超伝導」とあるもの
* 6 7/ SCAN 5 CONF1 *"SUPERCONDUCT"*

8/ OR 6,7 ←会議録標題か会議名に「超伝導」とあるもの
* 6 8/ 6 OR 7 6巻あり
    
```

図10 レコードの親子構造検索例 (EXPAND コマンド)

9/ D M C ←「超伝導」会議録の目次型出力(子レコード全件表示)

 (1)
 ACCN:870055290 RTYP:1 (親レコード)
 CONF1:1986 CONF ON APPLIED SUPERCONDUCTIVITY
 LOCN1:BALTIMORE.MD DATE1:86/9/28*10/3
 SPSR1:APPL SUPERCONDUCT CONF INC/ IEEE/ USN, OFF NAVAL RES/ USN, NAVAL RES
 LAB/ NBS/ NAVAL SURFACE WEAP CTR/ DAVID TAYLOR NAVAL SHIP RES & DEV
 CTR/ NATL SECUR AGCY/ CATHOLIC UNIV/ UNIV MARYLAND
 TITL:IEEE TRANSACTIONS ON MAGNETICS LANG:EN
 YEAR:1987 VISS:0023 N2
 PAGE:0329-1850 NDOC:326 ← 3 2 6 論文収載
 CLAS:M;K;HY(EMBRYOLOGY);IC:ON;EN;ED;MA(HEMATOLOGY);M;IQ(ENGINEERING,
 ELECTRICAL & ELECTRONIC);EN;IN;ER;NG

ACCN:870055291 ←親レコードに連続する番号 (子レコード1)
 AUTH:PIPPARD AB
 AFFL:UNIV CAMBRIDGE, CAVENDISH LAB/CAMBRIDGE CB3 OHE//ENGLAND
 TITL:EARLY SUPERCONDUCTIVITY RESEARCH (EXCEPT LEIDEN) LANG:EN
 PAGE:0371-0375

ACCN:870055292 (子レコード2)
 AUTH:COOPER LN
 AFFL:BROWN UNIV, DEPT PHYS/PROVIDENCE//RI/02912
 TITL:ORIGINS OF THE THEORY OF SUPERCONDUCTIVITY LANG:EN
 PAGE:0376-0379

ACCN:870055293 (子レコード3)
 AUTH:ROWELL JM
 AFFL:BELL COMMUN RES, 331 NEWMAN SPRINGS RD/RED BANK//NJ/07701
 TITL:SUPERCONDUCTING TUNNELING SPECTROSCOPY AND THE OBSERVATION OF THE
 JOSEPHSON EFFECT LANG:EN
 PAGE:0380-0389

.....

↓論文標題に「超伝導」とある論文を収録しているが、
 9/ DIF 5-8 会議録標題等には「超伝導」のない会議録 8 2 巻あり
 * 82 9/ 5 AND NOT 8

↓上記会議録に含まれる全論文を検索(親→子。EXPAND)
 10/ EXPAND ACCN,LINE=9,OR,PREFIX="PACN."
 * 7582 10/ EXPANSION OF FIELD 1,LINE 9. ←7 5 8 2 論文
 11/ AND 10,1 ←そのうちの「超伝導」の論文 4 7 3 件
 * 473 11/ 10 AND 1

図10 (つづき)

12/ D M B ←会議名等に「超伝導」がうたわれていない会議録に収められた「超伝導」の論文

(1)
 ACCN:850356942 RTYP:2 (子レコード)
 AUTH:NAKAYAMA W, OGATA H
 AFFL:HITACHI LTD, MECH ENGN RES LAB, 502 KANDATSU/HITACHI/IBARAKI 316/JAPAN
 TITL:COOLING OF SUPERCONDUCTING ELECTRIC GENERATORS BY LIQUID-HELIUM
 LANG:EN ↑論文標題
 PAGE:0357-0374

ACCN:850356913 ←
 CONF1:1ST INTERNATIONAL SYMP ON TRANSPORT PHENOMENA
 TITL:HEAT TRANSFER AND FLUID FLOW IN ROTATING MACHINERY LANG:EN
 EDTR:YANG WJ
 PBSR1:HEMISPHERE PUBL CORP
 PBSR2:SPRINGER-VERLAG (親レコード)
 YEAR:1987 NDOC:43
 LCCN:86-9775
 ISBN:0-89116-572-X

(2)
 ACCN:850364023 RTYP:2
 AUTH:AONO S, NISHIKAWA K, KIMURA M, KAWABE H
 AFFL:KANAZAWA UNIV, FAC SCI/KANAZAWA 921//JAPAN
 TITL:SUPERCONDUCTING AND OTHER PHASES IN POLYACENIC SKELETONS LANG:EN
 PAGE:0167-0172

ACCN:850363996
 CONF1:1986 INTERNATIONAL CONF ON SCIENCE AND TECHNOLOGY OF SYNTHETIC METALS
 (ICSM 86)
 TITL:SYNTHETIC METALS LANG:EN
 YEAR:1987 VISS:0017 N1-3 NDOC:115

(3)
 ACCN:850366691 RTYP:2
 AUTH:KUBO Y, YOSHIZAKI K, HASHIMOTO Y
 AFFL:MITSUBISHI ELECT CORP, MAT & ELECTR DEVICES LAB/SAGAMIHARA/KANAGAWA 229/
 JAPAN
 TITL:SUPERCONDUCTING PROPERTIES OF CHEVREL-PHASE PBMO6S8 WIRES
 LANG:EN
 PAGE:0851-0856

ACCN:850366544
 CONF1:1986 INTERNATIONAL CONF ON SCIENCE AND TECHNOLOGY OF SYNTHETIC METALS
 (ICSM 86)
 TITL:SYNTHETIC METALS LANG:EN
 YEAR:1987 VISS:0018 N1-3 NDOC:149

(4)
 ACCN:850366705 RTYP:2
 AUTH:ALYAN MAAS, RAHIM YH
 AFFL:KING SAUD UNIV/RIYADH//SAUDI ARABIA
 TITL:THE ROLE OF GOVERNOR CONTROL IN TRANSIENT STABILITY OF SUPER-
CONDUCTING TURBOGENERATORS LANG:EN
 PAGE:0038-0046

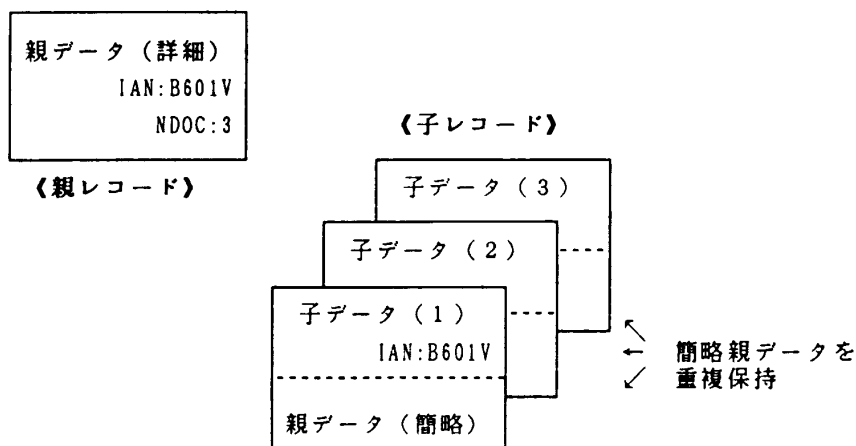
ACCN:850366699
 CONF1:1986 WINTER MEETING OF THE IEEE POWER ENGINEERING SOC
 TITL:IEEE TRANSACTIONS ON ENERGY CONVERSION LANG:EN
 YEAR:1987 VISS:0002 N1 NDOC:20

の方が、充足されないのが実態であり、こうした状況の概略は別途論じておいた〈注4〉。

そこで、ISTP&Bのようなデータベースについても、レコードのオンライン化の際に原レコードを加工して、構造を取り払った上で「情報検索システム」の上に導入するという行き方が採られることになる。実際、DIMDIのシステムでは、〈図11〉のようなレコード構成、すなわち親データの膨大な重複保持によって、各子レコードを「完結」させることにより、ISTP&Bを通常の情報検索システムの上に展開しているようである〈注11〉。

次に考えられるのは、全文データベースのための検索システムを用いることである。これは、雑誌の各論文について、本文全体を収容してオンライン検索に供するシステムであり、米国BRS Information Technologies社のものや米国化学会Chemical Abstracts ServiceのCJACSなどがある〈注13、14〉。この種システムでは、本文における「本文全体>段落>文」という階層構造を扱うよう設計されているから、この点を応用して、段落をISTP&Bにおける論文とみなし、各段落の第1文を論文著者、第2文を論文標題、等々に充当する。そして、いわゆる近接演算により著者、標題等、論文レベルの検索を試みることになろう。もっとも、利用者にとって自然で扱い易いシステムとするには、こうした読み替えをカバーするために、相当の工夫と工数を要することは想像に難くない。これには、ブール演算と近接演算との関係という、一般利用者には理解しにくい全文データベース検索システム特有の問題も含まれる。また、全文データベース用のシステムでは、その性格からして、収容可能な文献数(全文論文数)の「実用的」上限値が比較的小さいとみられるが、ISTP&Bについては、これを会議録・図書件数と読み替えるので、その膨大な件数を収容しうるか否か、疑問である。

ISTP&Bデータベースのオンライン化設計の際に、また本稿のとりまとめにあたって、ISTP&Bのような階層型文献データベースの情報検索システムでの扱い方について、LISA(Library and Information Science Abstracts)データベース等を用いて文献探索を試みた。結果的には、クラスター分析の手法により、文献を予め階層的クラスターにまとめて検索効率を向上させる方法や、階層型主題



- 1) 簡略な親データを各子レコード内に複製して、重複的に保持し、子レコードを完結
- 2) 会議・図書識別番号 (IAN: ISI Accession Number) による親子一括検索可

図11 レコード非構造化 (完結化) の方法 (DIMDIでのISTP & Bの例)

分類やソーラスを利用した検索システムの構成法を論じたものから、ゲーデルの不完全性定理などを援用しつつ情報処理システム一般の階層構成について哲学的考察を巡らすものまで、各種の論文が発見されたのであるが、書誌的階層構造を取り込んだ情報検索システムについて論じたものはみられなかった。現在、国際規格化が進められつつある情報検索システム用「共通検索言語」(Common Command Language)でも、構造検索的側面はとり上げられていない<注15>。どうやら、情報検索システム側からの構造的データベースへの接近は、あまり活発ではないようである。

NACSIS-IRのISTP&Bは、本稿にのべたような方法で、「情報検索システム」であるところのORIONの上に乗せ、一応の「構造検索」機能を実現してみたものである<注12>。しかし昨今における全文データベースの発展などをみても、文献型でありかつ構造性を内包する各種のデータベースは、今後一層増加するとみられる。こうした状況からして、従来のような「データベース管理システム」と「情報検索システム」の分立状況は甚だ不都合であり、両者を包括した機能を持つソフトウェアの開発が早急に望まれるところであろう。

<注>

- 1) 根岸正光、小澤 宏「TOOL-IR 工学文献データベース (COMPENDEX, Ei ENGINEERING MEETINGS) 検索システムの開発」学術情報センター紀要、第1号、p.25-59 (1987)。
- 2) 「情報検索サービス利用の手引—利用申請から検索まで」学術情報センター、1987. 70 p.
- 3) 「情報検索サービス利用の手引 (補遺版)」学術情報センター、1988. 48 p.
- 4) 根岸正光「情報検索システムの構築における利用者の特性とデータベースの特性——NACSIS-IRの構成を通して」大学図書館研究、第32号、p.1-11 (1988)。
- 5) Garfield, E. "ISI's new Index to Scientific & Technical Proceedings lets you know what went on at a conference even if you stayed at home." Current Contents. No.40, p.5-10 (1977)。
- 6) Garfield, E. "Introducing ISI/ISTP&B (Index to Scientific & Technical Proceedings & Books) —online access to the conference literature and multi-authored books." Current Contents. No.34, p.5-9 (1981)。
- 7) 仁木恵美子ほか。「理工学分野における会議資料の2次資料への収録率」三田図書館・情報学会1985年度研究大会、p.44-48 (1985)。
- 8) 青木美恵ほか。「会議文献——効果的利用のために (2)会議文献検索のための二次資料」科学技術文献サービス、No.78, p.20-26 (1986)。
- 9) 上田修一ほか。「理工学文献の特色と利用法」勁草書房、1987. p.67-79 (図書館・情報学シリーズ 8)。
- 10) "ISI Online Data Bases." Institute for Scientific Information, 1982.
- 11) "ISI databases in DIMDI—A brief guide." Institute for Scientific Information (ISI) /Deutsches Institut für Medizinische Dokumentation und Information (DIMDI), 1984. 38 p.

- 12) 「プログラムプロダクト VOS 3 情報検索システム ORION 利用の手引」日立製作所.
- 13) "BSR/Search Service User's Manual." BRS Information Technologies, 1986. 154 p.
- 14) "The CJACS file database description." American Chemical Society, 1987. 66 p.
- 15) "Documentation—Commands for interactive text searching. Draft International Standard."
International Organization for Standardization (ISO) , ISO/DIS 8777,1988. 34 p.

付表1 ISTP & B データベースの内容

見出し	項目内容	オンライン索引	出力モード (DISPLAY)	
			親レコード ヒット時	子レコード ヒット時
			X A B C D	-- B C D
ACCN	アクセシオン番号		X A B C D	-- B C D
CONF1	会議名(1)	K. 単語 (逆引可)	- A B C D	-- B C D
LOCN1	開催地(1)	K. 都市名	-- B C D	--- C D
DATE1	開催日付(1)	Y. 19xx	-- B C D	--- C D
CORG1	開催者(1)		-- B C D	--- C D
SPSR1	後援者(1)	SP. 単語	-- B C D	--- C D
CONF2	会議名(2)	K. 単語 (逆引可)	- A B C D	-- B C D
LOCN2	開催地(2)	K. 都市名	-- B C D	--- C D
DATE2	開催日付(2)	Y. 19xx	-- B C D	--- C D
CORG2	開催者(2)		-- B C D	--- C D
SPSR2	後援者(2)	SP. 単語	-- B C D	--- C D
TITL	図書・会議録標題	K. 単語 (逆引可)	X A B C D	-- B C D
TTIL	翻訳標題	K. 単語 (逆引可)	-- B C D	-- B C D
LANG	言語種別 (付表2)	L. 言語コード	-- B C D	-- B C D
AUTH	著者名	A. 姓名頭字	- A B C D	-- B C D
SERS	シリーズ名	K. 単語	-- B C D	-- B C D
EDTR	編者名	A. 姓名頭字	- A B C D	-- B C D
CORP	団体著者	K. 単語 (逆引可)	- A B C D	-- B C D
PBSR1	出版者(1)	PB. 単語	-- B C D	-- B C D
PPBL1	出版地(1)		-- B C D	--- C D
ADDR1	住所(1)		--- C D	--- D
PBSR2	出版者(2)	PB. 単語	-- B C D	-- B C D
PPBL2	出版地(2)		-- B C D	--- C D
ADDR2	住所(2)		--- C D	--- D
YEAR	出版年	Y. 19xx	-- B C D	-- B C D
VISS	巻号等		-- B C D	-- B C D
PAGE	頁		-- B C D	--- C D
NDOC	掲載論文数	NDOC. xxxx	-- B C D	-- B C D
PRCE	価格		-- B C D	--- C D
LCCN	LCカード番号	LC. カード番号	-- B C D	-- B C D
ISBN	国際標準図書番号	B. 図書番号	-- B C D	-- B C D
CLAS	主題区分	C. 分類コード	--- C D	--- D
RTYP	親子コード区分(1)	R. 1	X A B C D	--- D
FTYP	図書・会議録区分 (ISTB, ISTP)	F. ISTB F. ISTP ★	--- D	--- D



【論文(子)レコード】				
ACCN	アクセシオン番号		--- C D	X A B C D
AUTH	著者名	A. 姓名頭字 ★	--- C D	- A B C D
AFFL	著者所属、住所	AF. 単語	--- C D	-- B C D
TITL	論文標題	K. 単語 (逆引可) ★	--- C D	X A B C D
LANG	言語種別 (付表2)	L. 言語コード	--- C D	-- B C D
PAGE	頁		--- C D	-- B C D
DTYP	論文種別 (付表3)	D. 種別コード	--- D	--- C D
RTYP	親子コード区分(2)	R. 2	--- D	X A B C D
PACN	親アクセシオン番号	PACN. 親番号	--- D	--- D

★：オンライン索引抽出の際、対応する親と子両方のアクセシオン番号のポスティングを生成する<図8参照>。

付表2 言語コード表 (件数は1988年2月現在)

コード	言語	件数	コード	言語	件数
AF	アフリカース語	44	IT	イタリア語	959
AR	アラビア語	-	JA	日本語	8
BE	ベンガル語	-	LA	ラテン語	-
BU	ブルガリア語	-	MA	マライ語	-
CA	カタロニア語	-	NO	ノルウェー語	16
CH	中国語	14	PE	ペルシャ語	-
CZ	チェック語	33	PO	ポーランド語	82
DA	デンマーク語	18	PT	ポルトガル語	84
DU	オランダ語	24	RM	ルーマニア語	-
EN	英語	611,587	RS	ロシア語	1,732
FI	フィンランド語	6	SC	セルボクロアチア語	60
FL	フラマン語	2	SK	スロバキア語	17
FR	仏語	12,187	SL	スロベニア語	-
GA	ゲーリック語	-	SP	スペイン語	263
GE	独語	9,983	SW	スウェーデン語	105
GR	ギリシャ語	-	UK	ウクライナ語	-
HE	ヘブライ語	2	WE	ウェールズ語	-
HU	ハンガリー語	76	XX	多言語	28
IC	アイスランド語	-			

付表3 論文種別コード表 (件数は1988年2月現在)

コード	種別	件数
1	Articles	706,965
C	Correction, Addition	62
D	Discussion	535
E	Editorial	982
I	Item about an Individual	68
K	Chronology	22
L	Letter	7
M	Meeting Abstract	2,382
N	Note	143
P	Scientific and Technological Proceedings Paper	192,091
R	Review, Bibliography	21,666
U	Proceedings Paper	14
≠	Combined Scientific and Technological and Social Science Proceedings Paper	151

付表4 ISTP & B 検索コマンド一覧

	コマンド名	機能	指定形式
一次検索	SEARCH (Sも可)	キーワードによる検索	SEARCH term [, term, …] [(AND) term [, term, …] .NOT. “. ”はOR結合 各termの形式は [prx.][*]kw[*] [--[prx.][*]kw[*] kw:検索キーワード kw*:語頭一致 *kw:語尾一致 kw--kw:範囲指定 prx:索引プレフィクス(標準 K. 既出指定は省略可)
	PHRASE	キーワードの論理積(AND)による検索	PHRASE [prx.][*]kw[*] [prx.][*]kw[*] …… stopword除去処理内蔵 kw, prxの指定はSEARCHに同じ
	AUTHOR	著者名による検索	AUTHOR 正規化著者名[*] [, 正規化著者名[*] ……] 正規化著者名とは「姓 名の頭字」
	ISBN	国際標準図書番号による検索	ISBN isbn[*] [--isbn[*]] [, isbn[*] [--isbn[*]], ……]
二次検索	AND	文献集合間の積集合を作成	AND 文献集合番号, 文献集合番号[, 文献集合番号…]
	OR	文献集合間の和集合を作成	OR 文献集合番号, 文献集合番号[, 文献集合番号…]
	DIF	文献集合間の差集合を作成	DIF 文献集合番号-文献集合番号 “-”は“, ”でもよい
	LIMIT	文献集合中の指定項目に特定文字列を含む文献のみ抽出	LIMIT [文献集合番号,] 項目見出し : { [*]文字列[*] { [*]文字列1*文字列2[*] ALL { 空白 * } 集合番号省略時は最近の集合を仮定。文字列1、2はこれらが指定項目にこの順序で含まれることを指示。文字列前後の「[*]」と「[*]」は、空白の存在を明示して、単語区切りを指示。文字列を指定しない(空白)と当該項目を持たない文献、*のみで項目を持つ文献を指示。
表示	DISPLAY (Dも可)	文献内容の表示	{ A. アクセション番号 } (A) DISPLAY { (N. n) } [, M. (B)] { (S. 文献集合番号) [, (I. i1 [, i2…])] } (C) { (I. i1-i2) } (D) (X) 集合番号省略時は最近の集合を仮定。N. は集合内の最初のn文献、I. はi番目の文献を表示。N. , I. とも省略時は全件表示。M. は出力モード指定。Aから順次詳しい項目が表示され、Dは全項目表示。省略時はX、検索結果点検用。
	LOOK	オンライン索引の通覧および検索	LOOK { [prx.]kw[*][*] [prx.]*kw kw末尾の*は語頭一致、**は後続索引表示。 *kwは語尾一致検索用の逆索引表示。

付表4 ISTEP & B 検索コマンド一覧 (つづき)

	コマンド名	機能	指定形式
検索手順・文献集合の保存呼出	QSAVE	文献集合番号1から2に対応する一連の検索コマンドにプロファイル名を付けて保存	QSAVE プロファイル名 [, 文献集合番号1[, 文献集合番号2]] 文献集合番号省略時は、それまでの全検索コマンドを保存。
	QUSE	QSAVEで保存されたコマンド群を取出して、検索を実行	QUSE プロファイル名
	KEEP	文献集合にプロファイル名を付けて保存	KEEP プロファイル名[, 文献集合番号] 集合番号省略時は最近の集合を仮定。
	ENTER	KEEPで保存された文献集合を取り出す	ENTER プロファイル名
補	REMIND	検索経過の表示	(文献集合番号, 文献集合番号, ...) REMIND {文献集合番号-文献集合番号 } {[ALL] } 各文献集合を生成した検索コマンドと文献数を表示。
	SET	LOOKコマンドでの表示索引語数の変更	SET { ADJ } n ADJは前後隣接語、TERMSは語頭一致・後続語表示数をn(4~26)に変更。
		検索対象範囲の限定	{=文献集合番号} 以後の検索をこの集合中の文献に限定。ONは直前の集合。OFFは限定解除。 例:SEARCH R.1 / SET UNIV ON とすれば親ノードに限定
	?INFO	項目一覧表の表示	?INFO
助	SORTO	集合中のレコードの排列	SORTO [LINE=文献集合番号] 項目見出し(n:m) { /A } /D } ... 排列キーとする項目の見出しを指定。n, mにキーの左端・右端文字位置を指定可(省略時は全桁がキー)。 /Aは昇順(A→Z, 標準)、/Dは降順。
	EXPAND	親ノードの集合を元に子ノード群を検索	EXPAND ACCN, LINE=文献集合番号, OR, PREFIX="PACN."
終了	END	検索セッション終了	END

研 究 論 文

日本語医学用語の構造解析

A study for structural analysis of Japanese medical terms.

学術情報センター 小山 照夫*

要 旨

多様な研究分野における各種の専門用語は、情報検索において重要な役割を果たしているが、これらを用いて正しい検索を行なうためには用語の間のさまざまな概念関係を活用する必要がある。専門用語の多くが複合概念であることを考えるならば、用語を一旦基本要素に分解し、その構造を明らかにすることにより、専門用語の間を合理的に解析することができるという予想が立てられる。日本語はこのような解析に本来適している言語といえるであろう。本論文は医学の分野においてこのような予想の正当性を確認するための予備的な解析として、基本概念要素の設定、計算機上の辞書の実現、基本要素から複合概念を生成するいくつかの規則に関する考察などを紹介するものである。

Abstract

In variety of research fields, scientific terms are important for information retrieval systems. To attain proper retrieval, the relationship among terms should be administrated systematically. Because most of scientific terms represent composit concepts, it is prospective that the administration can be made efficient, if terms can be decomposed to the elementary concepts. Japanese terms can be regarded relatively easy for these kind of decomposition. In this paper, the auther introduces a preliminary study about structural analysis of Japanese medical terms. A morphological analysis method, the internal structure of a computerised dictionary for quick retrieval of elementary terms, consideration about some rules constructing composit terms from elementary ones are discussed.

1. はじめに

各分野の情報化の進展にともない、自由テキスト形式で記述された情報が、計算機によって蓄積・管理・検索されるようになってきている。計算機の数と容量が増すにつれて、その応用範囲も拡大

* Koyama, Teruo : National Center for Science Information System

されてきているが、高度の判断などを支援することを目的とするシステムでは、一定の形式に規定できない情報の管理が必要となる。このような情報を記述する一つの枠組みが自由テキストであると考えられることができるであろう。学問研究はまさにこのような情報を必要とする分野であり、研究の実施に際して、自由テキスト形式で記述された情報を計算機内に蓄積し、必要に応じて検索を行うことがますます重要なものとなりつつあるということが出来るであろう。とりわけ医学・医療の分野は、完全には整理しきれない情報を大量に扱う必要のある分野であり、知識の活用という観点から自由形式で書かれたテキストをデータベース化する試みが行なわれたり、あるいは自由形式で書かれた病歴をそのままデータベース化する試みが行われたりしている点、この種の情報の管理に対するニーズは大きいと考えられる。

自由テキストデータの活用にあたって問題となるのは、適切な情報に効率よくアクセスする情報検索の問題である。自由テキストデータの検索手段として従来から利用されてきているものにキーワード検索があるが、ここでは検索の対象となるデータに対する索引として、検索の鍵となり得るキーワードをなんらかの形（多くは人手による）で付与し、これらのキーワードの組合せによって適切な情報を選択することが行われる。文献検索システムなどはその代表的なものといえるであろう。今後とも自由テキスト情報を検索する一つの有力な手段としてキーワードが利用されていくことが予想される。

情報検索一般に要求される機能として、関連する情報を確実に検索すると同時に、関連の低い情報は可能な限り切り捨てる必要があるとされる（図1）。このようなことが可能であるためには、索引付けを行なうにあたっての対象となる情報の内容を的確に示すキーワードの選択と、検索にあたっての適切なキーワードの選択が重要な課題となる。理想的にはさらにキーワードの間の関係ないしは対象データにおけるキーワードの役割等の情報までを索引として付与し、検索に利用することが望まれるが、これはもう少し将来の課題であろう。

キーワード検索を利用するにあたっては、原情報に対する索引付けをどのようにするか、及びキーワードをどのように利用するかが大きな課題となるが、以下では検索にあたってのキーワードの選択を中心を絞り、この問題に対してキーワードの持つ概念構造がどのように利用される可能性があるか

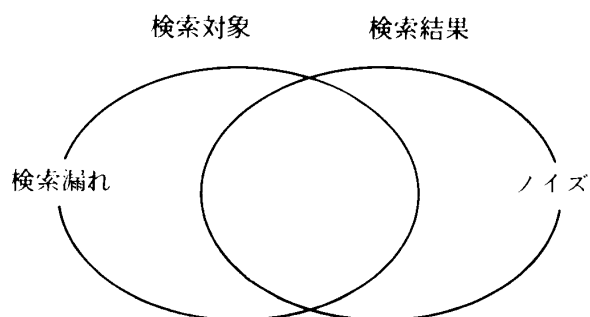


図1 情報検索の要件
検索漏れとノイズの最小化が要求されている

を考察し、つづいて医学分野のキーワードに範囲を絞ってキーワードの概念構造を解析する予備的な試みについて述べることにする。

2. 検索キーワード選択と概念構造

キーワード検索において、必要な情報に適切にアクセスするという情報検索の要件を満足するためには、検索に利用されるキーワード自身がある程度まとまった概念を表わすものである必要がある。このようなキーワードで問題となるのは、表現される概念が複雑になるほど、キーワードに多くの同義語や関連語（上位語、下位語など）が存在するようになり、検索に際して指定されたキーワードがそのままの形で現われているものだけを検索すれば良いとはいえなくなることである。特に日本語ではいくつかの概念を表わす用語を連結して新しい用語を作り上げることが頻繁に行われる結果、同義語、関連語が生じる可能性が高いといわざるを得ない。このような問題に対処するために一般にはキーワードとして認める語彙を制限するとともに、キーワードの間の関連を整理した関連語辞書（シソーラス）を用意することにより、検索の精度を向上させることが行なわれる。このようなシソーラスの編集は、一般には大きな労力を必要とする作業となる傾向にあり、自由テキスト情報の役割が大きくなるに連れて、この作業を効率化する手段が望まれるようになるであろう。特に専門用語辞書の編集を、少なくとも部分的には機械化し、編集作業の効率化を促進する方法が必要となる。またこれまでシソーラスのもう一つの問題として、辞書に登録されていない用語をキーワードとして原データの索引としたり検索キーワードとして指定することが無効となるという問題もある。これは特に新しい概念が多く出現する可能性のある分野において大きな問題となる可能性がある。

医学の分野において複雑な概念をみると、その多くはより基本的な概念を組み合わせて作られた複合概念であることがわかる。先に述べたように、日本語では基本概念を表わす用語を連結して複合概念を形成することが多いため、専門用語の多くではその表現形態の中に、元の基本概念を表わす要素の形態が保存されている。これは複合概念を表す文字列を、概念を構成する基本概念を表す文字列に形態的に分解可能な場合が多いことを意場している。例えば「先天性代謝異常症」は「先天/性/代謝/異常/症」のように分解することが可能である。このような分解を見て、例えば「先天」は原因を表わすものであり、「代謝」は全身的機能を表わすし、「異常」は機能の異常を示すものであるから、全体の表わす概念はこれらの基本概念の複合したものであるという解釈が可能となる。一般に診断を表わす用語は、SNOMEDなどで提唱されている4軸ないしは5軸の記述からの解釈が可能なものが多く、図2に示すように部位、形態、機能、原因、形容（時間、量、性状）、物質、状態などの記述に分解することが可能である。物質そのものや組織/細胞レベルの記述は更に細かな形容を伴うものもあり、例えば「陰イオン交換樹脂」は「陰/イオン/交換/樹脂」と分解され、属性は「性状形容/物質/作用/物質」という構造にを持つと解釈することが可能である。また、形態を表わすものの中には例えば「胃十二指腸吻合」→「胃/十二指腸/吻合」＝「部位/部位/相互形態」のように、二つ以上の部位を同時に参照するものも存在する。その他検査に関する用語では検査機器の参照や検査技法の参照が多

- 部位 (PO)–器官 (OG)
 - 部分構造 (PT)
 - 組織 (TS)–細胞 (CL)–細胞器官 (MO)
- 機能 (FN)–機能異常 (FTAB)
 - 作用 (EF)
- 形態 (MP)–形態異常 (MPAB)
 - 相互形態 (MPNT)
- 原因 (ET)
- 状態 (ST)–疾患 (DG)
- 物質 (MT)–化学物質 (MTCM)
 - 生物物質 (MTBO)
- 形容 (AJ)–時間形容 (AJTM)
 - 量的形容 (AJQT)
 - 性状 (AJQL)
- 変換 (CV)–形容変換 (CVAJ)
 - 概念変換 (CVCT)

図2 基本概念要素分類属性の例

くみられるし、治療に関する用語では薬剤の参照や、外科的手技の参照も見られる。このような用語の構造を明確に意識しておくことにより、用語間の関連を整理する上で有用な情報を明らかにすることができると期待できる。場合によっては、基本要素の間の関連だけで用語間の関連が十分に記述できる場合も考えられる。また、全く新規に現われた用語についても、その構造が解析できるならばおおよそのところは他の用語との関連を明らかにすることができると期待される。このような形でキーワードの整理が行えるならば、一つにはシソーラスの編集にあたって大いに参考となることが期待されるし、また一つには、必ずしも辞書に登録されていない用語であっても、同義語や関連語となる可能性のある用語を推定することができる可能性もある。以下では医学分野の専門用語に関して、専門用語としての複合用語の構造解析の試みを示す。

3. 医学専門用語の構造解析

専門用語、特に複合概念を表す専門用語の構造解析にあたって問題となるのは、第一に基本概念要素としてどのようなものを用意すればよいかであり、第二にそれらの概念要素がどのように分類でき、概念要素から複合概念を形成していく生成規則にどのようなものがあるかを明らかにすることである。

第一の問題については、一つの考え方としてはそもそも漢字自体が表意文字であり、もともとはなんらかの基本的概念に対応するものであることを考えるならば、全ての用語を個々の構成漢字にまで分解することが考えられる。しかしながら例えば「細胞」を「細/胞」と分解するとすれば、医学のような特定の応用分野を仮定する限りでは、多くの場合個々の基本概念要素に分野特有の意味付けを与

えることが困難となるであろう。個々の漢字から熟語が形成される場合の生成規則や結果に対する意味付けなどは、より一般的な観点からは確かに興味深い問題を提起しているが、特定の応用分野の問題を想定する場合、とりあえずはもう少しマクロな概念要素を設定する方が分野固有の概念形成モデルを明らかにできるとも考えられる。ただしこのような観点に立った場合には、基本概念と考えられるレベルをどこに設定するかの問題を生じるのであり、例えば「前頭葉」のような概念については、これを基本概念と考えるか「前/頭/葉」と分解できると考えるかが問題となるであろう。前者では概念そのものが「臓器」または「部位」を表すと考えるのに対して、後者ではこの概念を「(制限範囲内)相対位置/臓器/部位構造」と分解した上で、これらが構造的に結合することによって全体としての概念を構成していると考えることとなる。この例では分解を行なうことによってより詳細なかつ意味付けが可能な概念構造を記述できるが、一方では構造解析を機械化する場合には解析の曖昧度を極端に増加させる可能性もある。同様のことが「血圧」/「血/圧」、「症候群」/「症候/群」などについても検討される必要がある。ただし以下では最初の試みとして、将来のいずれかの時点で基本概念設定レベルの見直しをすることを前提とし、とりあえず著者の主観的な判断による基本概念の設定に基づいて検討を進めることとする。

第二の問題については、先に述べたように SNOMED などに見られる部位、形状、機能、原因などの系列に属するものや時間、量的、質的形容を表わすもの、物質の系列に属するものに加えて、疾患概念や状態などの属性を仮定することができる。

また、複合概念構成のため一般法則の内のいくつかのものは、少数の例を見ることによりただちに理解することができる。例えば「臓器(部位)/組織」という構造によって構成される「組織」(PO, TS→TS、または OG, TS→)として「副腎/皮質」や「心/内膜」などを挙げるができる。また、「吻合」やこれに類似の「癒着」などは、「[胃/十二指腸]/吻合」に見られるように、並列する複数の臓器(部位)と対応づけられることによって「形態異常」を生成する場合がほとんどである(PO, PO, 吻合→MPAB)。「臓器/機能異常」ないしは「臓器/機能/機能異常」という構造によって「機能異常」を構成する例(OG, FN→FNABまたは OG, FN, FNAB→FNAB)としては「腎/不全」や「下肢/運動/不能」などがあり、また、「臓器(組織)/形態異常」という構造によって「形態異常」を構成する例(OG, MPAB→MPABまたは TS, MPAB→MPAB)としては「眼球/突出」などがある。「機能異常」、「形態異常」、「疾患概念」は「性」に代表される変換要素によってしばしば「原因」または「性状形容」を構成する(FNAB, 性→ETまたは FNAB, 性→AJQLなど)が、構成された結果がどちらであるかの判定は必ずしも容易ではない。

より複雑な構造を持つ用語、例えば「移植/片/対/宿主/反応」などでは、「対」という要素が二つの組織(「移植/片」と「宿主」)を対比させているという解釈が必要である。この種の「対」は、しばしば省略されることがあり、上記の用語を「移植/片/宿主/反応」で置き換えても不自由ではないという問題もある。また、「性」を用いて名詞を形容詞化するものの中には解釈の複雑なものが含まれており、例えば「遺伝/性/非/球状/赤血球/性/溶血/性/貧血」では、「遺伝」、及び「溶血」は直接「貧血」

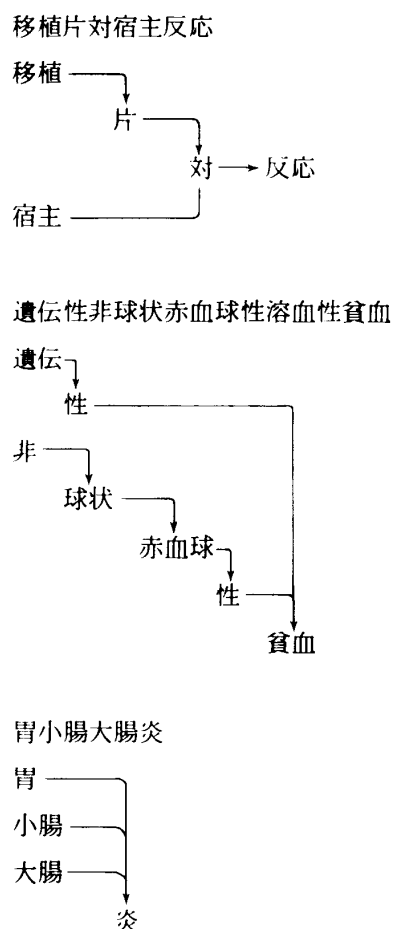


図3 複雑なかかり受け構造の例

にかかると、「非/球状/赤血球」は「溶血」にかかると解釈するのが自然であろう。さらに「胃/小腸/大腸/炎」などに見られる類似用語の単純羅列をどのように解釈するかの問題もある。これらのかかり受けの構造を図3に示す。

4. 構造解析の実例

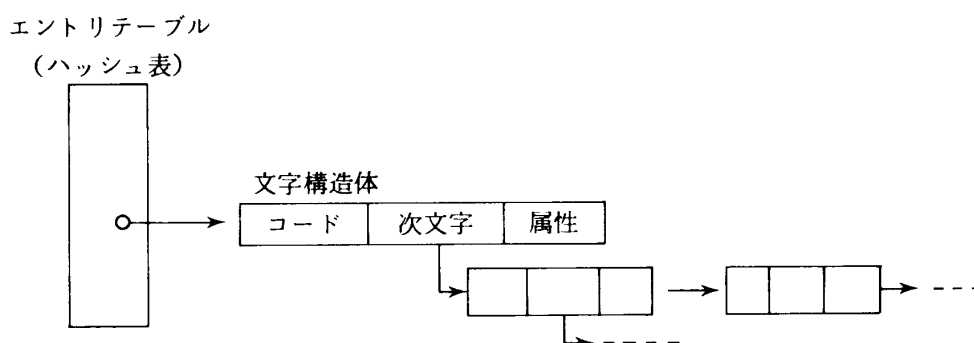
以上のような問題を実際に検討するために、現在、ある程度まとまった量の医学用語を実際に計算機に入力し、その構造解析を試みている。内科関連の用語集から収録した約 25,000 の見出し語から、手作業で約 7,000 の基本要素を抽出し、辞書のプロトタイプを作成した。ただし現段階では基本要素の分類属性は一部しか入力していない。この基本要素辞書を用いてもととの用語を基本要素へ分解するプログラムを作成した。基本要素への分解にあたっては、対象語の文字列を先頭から基本要素辞書文字列と比較していき、対象文字列が辞書文字列の並びに分解されれば辞書とのマッチングが取れたと判断している。このような方法では、対象文字列の分解は一意には決定できない場合があるが、現段階ではマッチングの取れるものすべてを生成し、その中から最も少数の基本要素数からなるものを正解と考えている。この基本要素分解プログラムは HITAC—M 680 上の Lisp を用いて作成した。図

遺伝性発作性脱力症
 (“遺伝” “性” “発作” “性” “脱力” “症”)
 (“遺伝” “性” “発作” “性” “脱” “力” “症”)
 —上側を正解とする—

一過性脳虚血発作
 (“一過” “性” “脳” “虚血” “発作”)

急性細菌性心内膜炎
 (“急性” “細菌” “性” “心” “内膜” “炎”)

図4 基本要素分解の例



実例

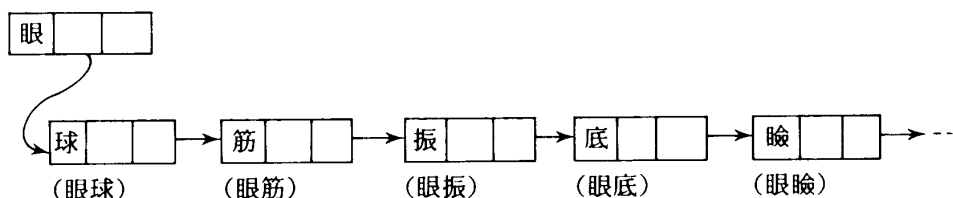


図5 辞書の構造

4に基本要素分解の例を示す。このような基本要素分解を行なう場合、実際にはかなりの解釈上の曖昧さを示すものもあるが、現在までのところ構成要素数最小という条件で不自然な構造が生成される場合はごくわずかである。

複合語を基本語に分解するにあたっては、複合語の部分文字列を基本語辞書と比較する処理が必要となる。この処理にあたって辞書の検索を効率化するために、図5に示すような辞書の構成を採用している。すなわち基本要素はその最初の文字によってハッシングを行い、エントリーテーブルに登録される。ここで各文字は、文字コード、次文字リスト、分類属性からなる構造体として定義されている。次文字リストは文字構造体のリストとなっており、先頭文字から次文字リストの要素をたどることによって、基本要素の文字列を求めることができる。文字構造体の分類属性がNILでない場合に

は、先頭文字から次文字リストをたどることによって得られる文字列が基本要素文字列に属するものであり、分類属性がNILであれば、先頭文字からこの文字までの文字列は基本要素文字列の部分文字列ではあるが、辞書に登録されたものではないことがわかる。

実際にはこのようにして分解された文字列について、その分類属性を手がかりにして概念構造を解明する必要がある。この問題は基本的には自然言語処理にみられる文法パーザと同様な方法で可能であると考えられるが、解析を始めるにあたってまず第一に文法規則を整備する必要がある。現在、基本要素に対する分類属性の付加と文法規則の整備を行なっている段階であり、まとまった結果は得られていない。部分的には先にも紹介したいくつかの生成規則が明らかになっているが、今後とも辞書の分類属性の整備と、生成規則の追加を行なった上で概念構造の解析を行なっていく予定である。

5. 今後の課題

今回の発表はキーワードを基本要素に分解する初歩的な解析に留まっている。材料としては相当量の語彙が入力済みであるから、これを材料として、今後各基本要素への分類属性の付加と、用語生成規則を明らかにしていく予定である。また、シソーラス作成支援や検索システムインターフェースへの応用の有効性について検討を進めていきたい。

先にも述べたように、現在の基本要素の設定単位は暫定的なものである。この問題についても実際に解析システムの運用を通して、どの程度のレベルが基本概念要素として最適なのかを検証していく必要がある。

今回は主として検索キーワードという観点から考察を進めてきたが、キーワードの取り扱いに関しては、そもそも原資料へのインデックス付けも大きな問題他なる。現在キーワード付加の自動化の試みとして、KWICが試みられているが、この問題に関しても、キーワード切り出し範囲の適切な設定などの観点から、キーワード構造解析の応用の可能性について検討を進めたいと考えている。

参考文献

1. 小山照夫、日本語医学文献情報検索キーワードに関する一考察、第27回日本ME学会大会論文集、p. 478、1988。

研 究 論 文

学術用語データベースの構築

Database construction of the scientific terms

学術情報センター 大山 敬三*

要 旨

学術用語データベースの構築過程として、筆者は多くの分野の学術用語集冊子体を詳細に調査し、データベースに格納するためのデータ形式と既存資料をデータベース入力形式に変換する手順の設計を行ない、実際にデータベースの構築を行なった。本論文では、この結果として明らかとなった用語に付随する情報の種々の特徴と、これらを表現する方法を示している。また、UNIX ベースのエンジニアリングワークステーションを用いた校正方式とその効果についても述べている。

Abstract

As processes of database construction of the scientific terms, the author closely examined almost all the existing books titled "Japanese Scientific Terms" in various fields, designed the data formats to store informations in the database and the procedures to convert both printed and CTS-format data sources into the database input format, and actually carried out the procedures. As the result, various features of the informations accompanied with each term and the method to represent them are made clear. The author also describes the proof-reading method using a UNIX-based engineering workstation and its effects.

0. はじめに

学術用語集の制定の効率化と高度な利用のためには、学術用語データベースの構築は欠かせない。その初期段階では、冊子体の形態をとっている既存の用語集を機械可読形式に変換し、一元的に計算機処理できるようにすることが必要である。将来の学術用語集の制定作業の基礎となすためにはデータの精度を十分に高くするとともに、情報のもつ属性を洗い出して定式化しておく必要がある。

本論文では、実際に学術用語データベースを構築した経験を基礎に、すでに制定されている学術用語集を調査して明らかとなった学術用語のもつ情報の属性について最初に述べ、次に学術用語集の

* Oyama, Keizo : National Center for Science Information System.

初期入力・校正作業の手法とその効果について述べることにする。

1. 学術用語データベースのデータの単位

学術用語集では用語が表す概念を直接定義することはせず、用語として採択された日本語（以下では日本語用語という）とそれに対応する外国語（主に英語、以下では外国語用語という）を示すことにより記述している。冊子体では利用者が検索するために各学術用語を和英の部および英和の部の両方について辞書順に配列してある。本来ならば双方に記された内容は、同義語と異義語の区別に関する情報以外は、基本的には同一の情報を持つと考えられる（ただし、一部の分野ではある種の付加的情報が一方のみに記されている場合もある）。実際、既存の学術用語集の多くは、英和または和英のどちらか一方で用語の選定作業を行い、最後に配列を整えるという方法で編集されている。冊子体で用語集を利用する場合は英和と和英が必要であるが、編集作業を行う段階、およびデータベースを利用する段階においては論理的には一つの単位の情報が重複して記憶されていることは望ましくない。

また、冊子体では同義語がある場合は、同一の概念を表すにもかかわらず、それぞれを見出し語として複数個所に配列しており、これも同様に一つにまとめて管理する必要がある。逆に日本語あるいは外国語の用語の表記が同じでも、対応する用語が異なる意味を持つ場合は異義語として区別しており、データベース上でも異なるものであることを明確にしておく必要がある。

さらに、辞書順の配列では、見出し語に対応する語に同義語や異義語があることはわかるが、見出し語自体の同義語や異義語については特別な表記をしていない限りその存在を知ることはできない。つまり、対応関係が完結しない形式となっている。この形式をそのままデータベースに持ち込むと、情報としては日本語から外国語、あるいは外国語から日本語の一方向の対応関係を個別に示すものの集合となってしまう、同一の実体であるはずの用語が対応する側と対応される側の双方に出現することになり、情報の一貫性の維持のためには望ましくない。

これらの条件を満足させるための一つの方式としては、日本語用語と外国語用語の同義語をすべて集めたものをデータベース上の論理的な情報の単位とする、すなわち1つの概念を表す用語の集合をひとまとめにして取り扱うということが可能である。

同義語と異義語のすべてを集めたものを情報の単位とすると異義語の対応関係が際限なく広がる可能性があり、データベース上のデータの単位としては不適格である。逆に、日本語用語と外国語用語の1対1の組合せを情報の単位とするとすべての同義語と異義語の組合せが必要になり、さらに、同義語と異義語の区別を表記するために別の仕組みが必要になって、データの単位としての単純さが失われ取り扱いが複雑になる。

以上のような理由から、今回のデータベース構築にあたっては日本語用語と外国語用語の同義語の集合をデータの単位とし、これをデータベースのレコードとして実現することとした。

なお、同義語には分野をまたがったものもあり、全分野の学術用語集を全体としてとらえた場合はこの扱いについても配慮する必要があるが、既存の学術用語集では異なる分野の用語間に関する関係について

ては情報がないので当面は異義語とみなして取り扱うのが適当である。

上述のように冊子体とデータベースとの間には情報の単位の取り扱いに関する相違があるが、個別の学術用語集の情報の属性を詳細に定義することにより、冊子体に必要な情報をすべてデータベース上に保持することが可能であり、機械的な処理によってデータベースから冊子体の印刷に必要なすべての情報を作成することができる。

2. 学術用語に付属する情報の性質

学術用語集では基本的には日本語用語と外国語用語の対応を示すだけであるが、これだけでは不十分なために学術用語に付属する種々の付加的情報が付記されている場合も多い。この付加的情報にはさまざまな属性があるため、データベース化する場合に概念上の整理が重要となる。

既制定の学術用語集を広範かつ詳細に調査、検討した結果として、学術用語に付属する付加的情報には各種のものがあ、これらは種類だけでなく階層についても整理する必要があることが明らかとなった。このためデータベースにおける情報の階層性を表現するためにレコード、フィールド、サブフィールドという構成をとることとした。本章ではこの付加的情報のデータベース上での取り扱いの考え方と各種の付加的情報の種類について述べ、次の章でこれを収容するための入力データ形式の定義について述べることにする。

2.1. 付加的情報の取り扱い

冊子体では、印刷された表記のみによって用語の検索と情報の取得の両方の要求を満たさなければならないために、用語の付加的情報については簡略な記法を取らざるを得ず、その種類を厳密には表現できない。このために、利用者の常識的な解釈を頼りにして、情報の内容を十分に定義しないままに記述している場合もある。しかし、データベースとして管理、利用する場合には情報をさまざまに加工して、それぞれの目的に合った形で提供できることが要求され、このため、可能な限り情報の構造と種類を明確にしておくことが必要である。今回のデータベース構築に当たっては、データベースのレコードを構成する項目であるフィールドとしてこれらを定義することにした。フィールドの種類は冊子体の表記よりも詳細に定義しておく必要がある。

2.2. 付加的情報の種類

既制定の学術用語集の調査から、学術用語に付属する付加的情報には各種のものがあ、これらは種類だけでなく、階層についても整理する必要があることが明らかとなった。ここでは抽出した情報の種類について述べることにする。

日本語用語 日本語用語自体の表記。

ローマ字読み 日本語用語の訓令式によるローマ字読み。分かち書きされている。

外国語用語 外国語用語自体の表記。

同義語関係 日本語間、または外国語間の同義語に関する情報。参照すべき語として扱われている場合もある。

異義調関係 日本語間、または外国語間の異義語に関する情報。

説明・注記 説明ないし注記。用語の概念（同義語の集合全体）に関係するもの、個別の用語（同義語の集合のうちの特定の用語）に関係するもの、用語中の一部の文字、単語、記号などに関するもの、などがある。

複数形（外国語用語のみ） 外国語用語の複数形。既存の学術用語集では複数形を記述する場合、1）特に何も断わらない、2）注記に複数形である旨を記す、および3）単数形と複数形の両方を用語として記す、といった記法がある。

品詞 用語の品詞。特に意識する必要のある場合に記されている。日本語の形容詞、英語の名詞複数形（複数形のみを記し注記を付ける場合と単数形に付記する場合がある）などがある。接頭辞、接尾辞、動詞などは明示されていないが、用語の表記を見るとわかる場合がある。

原語（外国語用語のみ） 外国語の原語。外国語は通常は英語を用いるため、特に断わらない場合は英語である。ただし、英国語と米用語を区別する必要のある場合は英語の場合でも記述されている。既制定の学術用語集では英語を含めて全体で18種類の原語が出現している。用語中の一部の単語の原語が英語以外である場合はイタリック体が用いられているが、原語の種類は明記されていない。

使用上の制限 用語の使用上の制限。使ってもよいもの、当分の間使ってもよいもの、日本語用語に対応する外国語が暫定訳であるもの、などがある。

属性 用語の属性。常用漢字表外の漢字表記、その他分野ごとの特殊な属性（例：化学一字訳通則の例外、動物学一絶滅した類）などがある。

分類 用語の分野内での分類。すべての用語の分類を表記しているのは機械工学だけであるが、他の分野でも、説明ないし注記として扱っているもので分類として形式的に整理できるものがある。

ハイフネーション（外国語用語のみ） 長い外国語用語を、印刷の都合上、単語の途中で折り返すために挿入されているハイフネーション。外国語用語中の本来のハイフンとは異なる扱いを必要とする。規則と辞書だけで機械的に完全なハイフネーションを行うことは不可能であるので、情報としては保存しておく必要がある。

文字そのものとして読む文字（ローマ字読みのみ） 日本語用語のローマ字読みの一部を文字そのものとして読むことを表現するためにイタリック体が用いられている（例：X線/X-sen）。イタリック体が用いられていなくとも文字そのものとして読むこともある（例： α 線/ α -sen；レ形/レ-gata）。

省略してよいもの 用語の一部を省略してよいもの。

適宜置き換えて使うもの 用語の一部を適宜置き換えて使うもの。冊子体では置き換える部分のみを示しており、置き換えられる部号が明示されていない。データベースではこれを明示する必要がある。

常用漢字表外の漢字表記を付記するもの（日本語用語のみ） 用語の常用漢字表外の漢字表記を付記

するもの。用語の一部分に対する表記のみを記して対応する部分が明示されていないものがある。データベースではこれを明示する必要がある。

常用漢字表外の漢字に印を付けるもの（日本語用語のみ） 用語中の常用漢字表外の漢字に文字ごとに印を付けるもの。

3. データベースの論理構造

前節で述べたような各種の付加的情報をデータベースに格納するための論理構造として、レコード、フィールド、サブフィールド、および文字表記の四階層を定義し、それぞれの階層について記述の内容を示す。

3.1. レコード

原則として一つの日本語用語とそれに対応する外国語の対に関連する情報を一つのレコードとする。ただし、日本語、外国語にかかわらず同義語のある場合はこれを合わせて一つのレコードとする。異義語のある場合はそれぞれについて日本語用語とそれに対応する外国語の対を別個のレコードとする。同義語と異義語の判別は冊子体における表記に準ずる。

3.2. フィールド

レコード中の情報の論理的な単位をフィールドとして形式化する。情報の詳細な種類ごとにフィールドの種類を分けることもできるが、論理的な階層性を正しく表現し柔軟性を確保する目的から、フィールドの種類は必要最小限にとどめ詳細はサブフィールドにより表現することとする。レコードに直接付属する情報をフィールドとして定義し、各用語に付属する情報はサブフィールドとして定義する。フィールドの種類を以下に示す。

ID フィールド 全分野を通じて各レコードが一意に定まるような識別子を記述する。学術用語集の改訂の場合、新規のレコードには新しい識別子を付与する。レコードの修正の場合は改訂前の識別子のままとする。レコードの廃止の場合はレコード自体を削除することはせず、制御フィールドに廃止を指示する。

制御フィールド レコードの情報の制御状態を記述する。状態には、通常、廃止、未定の3つがある。

出典フィールド レコードの情報の出典となった学術用語集の分野、制定年などを記述する。必要に応じて複数回繰り返すことができる。

コメントフィールド レコード全体に付属する情報を記述する。冊子体上には印刷されない。

日本語フィールド 日本語用語とその関連情報を記述する。同義語がある場合は繰り返す。一つの日本語用語の後に、それに付属する情報があればそれらをすべてサブフィールドとして区別して記述する。

外国語フィールド 外国語用語とその関連情報を記述する。同義語がある場合は繰り返す。一つの外国語用語の後に、それに付属する情報があればそれらをすべてサブフィールドとして区別して記述する。

説明・注記フィールド レコードに付属する説明ないし注記を記述する。冊子体に印刷される形式で記述する。文字の記述規則にしたがう。

3.3 サブフィールド

日本語および外国語フィールドに記述された用語に付属する情報を同一のフィールド中に記述するため、論理的な単位ごとにサブフィールドとして形式化する。サブフィールドとして記述された情報を冊子体で表記する場合は一定の規則にしたがって変換することが必要であり、この規則は分野ごとに異なる場合もある。

レコードに直接付属するかあるいは各フィールドに付属するかを定義しにくい情報（例えば分類、典拠、履歴など）は混乱を避けるため、フィールドとしては定義せず、サブフィールドとして定義する。したがって、場合によっては複数のフィールドに重複して同一内容のこれらのサブフィールドを記述する必要がある。サブフィールドの種類を以下に示す。

用語サブフィールド 日本語用語自体、あるいは外国語用語自体の表記を記述する。文字の記述規則にしたがう。

カナ読みサブフィールド（日本語フィールドのみ） 日本語のカナ読みを記述する。文字の記述規則にしたがう。冊子体にはカナ読みが振られていないが利用上の便宜、編集の精度などを考慮してカナ読みを付することが望まれる。ローマ字読みとの比較により読みに関する大部分の誤りが除去できる。

ローマ字読みサブフィールド（日本語フィールドのみ） 日本語のローマ字読みを記述する。文字の記述規則にしたがう。

複数形サブフィールド（外国語フィールドのみ） 外国語の複数形を付記する場合に記述する。文字の記述規則にしたがう。冊子体では複数形を記すために単数形と複数形の両方を記す記法をとっている場合がある。この場合、単数形と複数形を区別するため、複数形は必ずこのサブフィールドを用いる。

品詞サブフィールド 語の品詞を記述する。

原語サブフィールド 外国語の原語を記述する。外国語は通常は英語を用いるため、このサブフィールドが省略された場合は英語であることを意味する。

説明・注記サブフィールド 他のサブフィールドで表記できない説明ないし注記の情報を記述する。文字の記述規則に準ずる。

状態サブフィールド 語の状態あるいは種類を記述する。状態には、使ってもよい、当分の間使ってもよい、暫定、参照、廃止、未定、参考の7つがある。参照は同義語ではないが特に参照すべき用語である場合にこれを指示する。廃止、未定は編集のための履歴としてフィールドを保存しておきたい場合に用いる。参考は他の情報源からの語を参考として記述する場合にこれを指示する。なお、参照は一応定義しておいたが、これまでに制定された分野の用語集を調べた範囲では参照すべき語が見出し語の同義語でない例は見つかっておらず、用語から同義語の用語へか、あ

るいは使ってもよい語または当分の間使ってもよい語から同義語の用語への参照のいずれかになっている。参照の指示は、例えば反対語やより広い概念を表す語などを参照する場合に行なうべきであり、このような参照を記述する必要がある分野においてのみ使用すべきものである。

属性サブフィールド 用語の属性を記述する。属性の種類は常用漢字表外表記のほかは分野ごとに定義して用いる。冊子体の上での表記に合わせるのが望ましい。

分類サブフィールド 用語の分野と分野内での分類などを記述する。

履歴サブフィールド 用語の典拠、状態変化の履歴などを記述する。

3.4 文字表記

用語に付属する情報についてはいくつかのサブフィールドにより形式的な記述ができるように定義した。しかし、サブフィールドの中において任意の形式が許されているもの、特に日本語用語および外国語用語の部分についてはこれまでの定義では不十分である。

冊子体においては様々な書体、字種を用いて語を表記しているが、用語集を計算機処理し広く高度な利用を図るためには、一般に流通可能かつ相互に変換可能な形式を定めて、これに基づく記述をすることが必要である。学術用語集の性格上、一般の学術出版物ほどは変化に富んだ書式がないので、必要最小限の表記ができる程度の簡易な形式を定義することをここでの目的とする。

原則として、語に付属する情報はすべてサブフィールドの形式によって表現し、語自体の表記の中には付加的情報は記述しないこととする。例えば、英語以外の外国語を記述するときには冊子体ではイタリック体を用いる分野が多いが、注記として原語を表記する場合にはこれは言語サブフィールドを用いることにより表現できるので、書体の情報を語の表記の中に記述する必要はない。

ただし、個々の文字、あるいは個々の単語に付属する情報はサブフィールドの形式をとることができないので、語の文字列中にこれらの情報を埋め込むことができるような表記も定めることにする。この表記の使用はあくまでもサブフィールドを用いて表現できない情報に限るものとする。例としては常用漢字表にない文字に付する印、あるいは歯学分野におけるドイツ語の単語に付する品詞の記号がある。以下に文字の表記に関する考え方を述べる。

使用文字種 使用する文字はすべて JIS 2 バイトコード (JIS X 0208-1983) に含まれる文字とし、文字集合の制御のためのエスケープシーケンス等は使用しないものとする。

一般的な記号の表記 用語の表記の際に紛らわしい一般的な記号として、ハイフン、スリードット、ダッシュなどがある。これらの表記の揺れがないように使用法を定めておく。

特殊な文字・記号の表記 JIS 2 バイトコードに定義されていない特殊な文字・記号として、表記を定めておく必要のあるものには以下のようなものがある。

- ・ J I S 外字 : 特に外字として扱う必要のあるもの
- ・ 特殊アルファベット : 音標符号付き文字およびその他の特殊アルファベット
- ・ 重ね合わせ : 文字を重ね合わせるにより表現できる記号類
- ・ 上・下付き文字 : 通常文字の横上および横下に付く文字

イタリック体 イタリック体の文字の扱いには分野ごとに若干の相違がみられるが、統合した場合に混乱をきたす可能性はないと思われるので、冊子体に印刷される通りの記述をするものとし、イタリック体の部分を特別の記号で囲んで表記する。

文字そのものとして読む文字 日本語フィールドに記述されたカナ読みおよびローマ字読み、サブフィールドに記述される読みの中で、文字そのものとして読む文字列を特別の記号で囲んで表記する。

用語中に埋め込まれる文字列 以上に掲げられていない記号類で、印刷あるいはその他の用途の目的で、用語中の各文字あるいは各単語などに付加するために用語中に埋め込む必要がある文字列は、その部分を特別の記号で囲んで表記する。この文字列はデータベースの管理のための操作においては無視される。

4. データ校正方式

4.1. データ校正の特性

データベースが構築されてこれをもとに用語集の編集作業が行われるようになった時点では、データベースへの入力形式によりデータを受け入れ、機械的な処理を経てデータベースにロードすることになる。しかし、データベースの初期構築の過程では入力データの品質が十分高いとは言い難い。この主な理由としては、データ源がデータベース化を前提としていない、既存のものによらざるをえないという点が挙げられる。

実際のデータ源としては冊子体の学術用語集が基本であり、分野によっては、印刷の過程で副産物としてできた電算写植機 (computerized typesetting system : CTS) の入力データが利用できる場合もある。

冊子体の場合は、外部の業者によりパンチする過程を経て入力データが作成されることになるが、この過程でパンチミスなどの誤りが混入することは避けられず、さらに冊子体の印刷の擦れや汚れによる誤りの可能性もある。

CTS の入力データを用いる場合は、写植機の制御用のコードを削除したり、通常の文字コードで表現できないものを表現できる形に変換したりする処理を経て入力データが作成される。しかし実際には、CTS の入力データには冊子体の最終的な内容と一致しない部分が含まれる。これは印刷の最終工程において誤りなどが発見された時に、全体の版を組み直すことが時間的、コスト的にできないため、誤りの部分だけを組み直して切り貼りするという処置が取られるために起こる。また、印刷したときに見栄えが良くなるように適当に空白を入れたり、ハイフネーションしたりしている場合があり、これらを完全に自動的に削除することが難しい。

いずれにしてもデータベースにロードする前に入力データを校正する必要がある。校正の内容はデータ源の性質と分野ごとの記法により違いがあるが、一般にパンチデータにはあまり偏った特性は見られないのに対し、CTS 入力データには同じような修正を各部に施す必要がある。切り貼りにより

修正された部分に関する情報は残っていないことが普通であり、この修正にはパンチデータの誤りの校正と似た性質の作業を行う必要がある。また、冊子体にも明らかな誤りが見られ、校正の過程でこのようなものが見つかる場合もある。

さらに、いずれのデータ源にも読みはローマ字でしか振られておらず、今後のデータベースの維持と利用を考えた場合、カナ読みが欠かせないため、これを付加するという作業が必須である。

以上のような内容の校正作業を行うためには、操作性のよい計算機を用いた機械的処理（対話形式とバッチ形式を含む）と人手による逐次的な確認作業の両方が欠かせない。そこで今回の校正過程においては、第1段階として計算機を直接利用することにより定形的な校正処理を行い、第2段階としてこの確認と非定形的な校正をリスト上で人手により行って、パーソナルコンピュータのワードプロセッサを用いてデータに反映するという手法を用いた。

4.2. 計算機処理による校正

この処理では計算機の使い勝手が作業の効率に非常に大きな影響を与える。そこで今回は日本語を含んだ文字あるいは文字列処理用ユーティリティが非常に豊富で、かつマルチウィンドウによる高い操作性が備わった UNIX をベースとしたエンジニアリングワークステーションを用いて作業を行った。

ワークステーションで行ったバッチ形態の処理の主なものを以下に示す。

- (1) 入力データを扱いやすい形式に変換する。
- (2) ID を付与する (CTS 入力データのみ)。
- (3) 出典を付与する (CTS 入力データのみ)。
- (4) 分類を付与する (予めパンチされたもの以外)。
- (5) JIS 2 バイトコードの文字 (全角文字) 中で JIS 1 バイトコードの文字 (半角文字) で代用できるものについて置き換える。
- (6) 括弧類と、空白、ハイフンなどとの対応関係をチェックして修正する。
- (7) 訓令式ローマ字読みから機械的にカナ読みを生成する。

(5)は計算機の画面上でできるだけ多くの情報を見るために必要な処理である。(7)は単純に行ったのでは外来語の長音などに対して適切な変換を行うことができないので、日本語フィールドの内容と対応を取りながら行った。ただし、「おおきい」、「ひづけ」や助詞の「を」、「は」などには対応していない。これらは頻度があまり高くないので、後で対話処理により修正を行った。

学術用語集の入力データ単独では、このような形式的な処理を行う程度が限界であるが、用語自体の誤りについても可能性のあるものを機械的に調べることができれば、人手による校正の効率化を図ることができる。そのためには、何らかの参照用のデータとの比較を行う必要がある。文部省学術用語集のいくつかの分野においては、カナ読みを付加したものを機械可読形式として流通しているものが入手可能であったので、これと入力データとの突き合わせを行うことにより上記の調査を行った。突き合わせには以下のような方式を採った。

- (1) 外国語用語については、入力データの外国語用語のみを取り出し、各種の括弧類を展開して全ての可能な組合せを生成し、正規化したのち、参照データの外国語部分と比較する。同一のものが参照データ中になればチェックリストに出力する。
- (2) 日本語用語については、入力データの外国語用語と日本語用語を取り出し、外国語用語は組合せのうちの最も短いものを生成し、日本語用語は括弧類の展開や正規化を行って、これらを組合せて参照データと比較する。同一の組合せがデータ中になればチェックリストに出力する。
- (3) 読みについては、入力データの日本語用語とローマ字読みを取り出し、日本語用語は組合せのうちの省略のない形のもの生成し、ローマ字読みはカナ読みに変換して、これらを組合せて参照データと比較する。同一の組合せが参照データ中になればチェックリストに出力する。

しかし、参照データは括弧類の扱いが元の学術用語集ほど厳密でなく、また、読みのふり方に関する方針が異なる（例えば ATR などの外国語の省略語を「エーティーアール」と読むなど）ため、単純に比較すると全く問題のないものまでチェックリストに出力してしまうため、比較の前処理としてさまざまな細かい操作を行っている。

参照データとしてはその他に UNIX 自体が持っている英語のスペルチェック用の辞書がある。対象となるのは外国語用語部分だけであるが、辞書自体の精度はかなり高いので、用語集全体の誤りを発見するのにかなり有効に利用できる。入力データの外国語用語を取り出し、これに対して UNIX のスペルチェックユーティリティを適用して、誤りの可能性のある単語のリストを出力させた。ただし、化学の物質名のように一般的でない単語が多い分野ではチェックリストが大量になりすぎて実際的でない。

これらの調査の結果得られた誤りのある可能性を含んだ用語のリストを元に、ワークステーション上のテキストエディタを用いて、対話処理により入力データと冊子体の比較を行い、誤りがある場合は原データに修正を加えた。参照データにも誤りがある可能性があるので自動的に置き換えるようなことはできない。したがって、この部分は目視による確認の過程を経ざるを得ない。

チェックリストには処理の都合上、必ずしもレコードを特定するための ID がついてはおらず、配列順も処理の途中で入れ替わっているため、チェックリストと入力データの照合は紙の上で行うことは不可能であり、エディタの検索機能を使う必要がある。さらに、作業用にできるだけコンパクトな形式に変換しているとはいえ、用語数の多い分野（化学、機械工学）では入力データが 12,000 レコード、1 M バイト以上にもおよぶ、このように大きなテキストはパーソナルコンピュータなどではとても扱える大きさでなく、大型汎用計算機でも難しい。その他にも、チェックリストの項目を探す場合に、いちいち用語の文字列をキーボードから入力するのもかなりの労力である。ワークステーションを用いることによりスクリーン上に複数の作業用ウィンドウを開いておいて、一方でチェックリストを表示させ、また一方で元データをエディットし、ウィンドウ間のコピー機能を使って検索文字列を与えるといったような機能を用いて作業の効率を数倍に高めることができた。作業中のワークステーションの画面例を図 1 に示す。

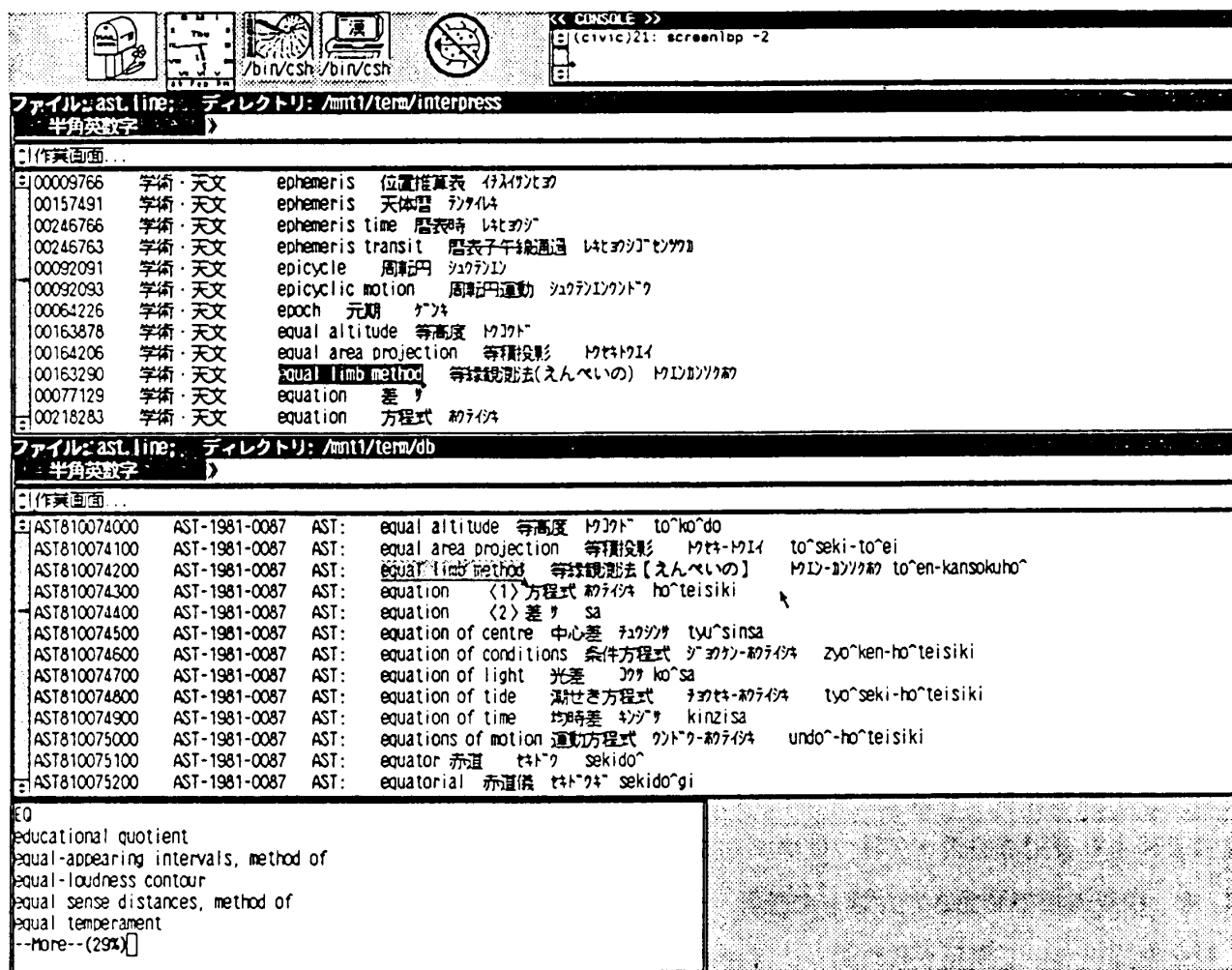


図1 校正作業中のワークステーションの画面

これらの処理の結果、さまざまなパンチミスや元の学術用語集の誤りが発見された。これらに関する統計については4.4節で述べる。

4.3. 人手による校正

上述の対話処理の間にも、チェックリストに含まれていない誤りを偶然発見することはあったが、基本的にはチェックリストにある用語の確認を行うものであり、校正の精度を高めるためにはやはり全用語に関する目視による確認作業が必要である。この場合、作業者がその用語集の分野の専門にある程度通じたものであれば、作業の成果は大きいものとなるが、そうでない場合は元の冊子体と入力データとの一致の確認程度しか期待できない。

作業の方式としては、入力データに4.2節で述べた処理を施したのを見やすいように形式を整えて出力し、これを冊子体と見比べ、訂正をリストに赤で書き込むという方法を取った。リストの配列は基本的には冊子体と同一である(入力データの配列を崩さないようにした)。時間的、あるいは人手の関係で全ての分野について目視による確認を行うことはできなかったため、参照用の用語集のデータが入手できなかった分野(キリスト教学、地理学、海洋学)と冊子体の形式が特殊な分野(図書館学)、それに参照データの版が古い分野(化学、機械工学)についてだけ作業を行った。

リストに加えられた訂正の入力データへのフィードバックは、パーソナルコンピュータを用いて以下のような手順で行った。

- (1)ワークステーション上でパーソナルコンピュータの文字コードに変換する。
- (2)LAN を用いてファイル転送で入力データをパーソナルコンピュータに取り組む。
- (3)パーソナルコンピュータ上で処理可能な大きさのファイルに分割する。
- (4)パーソナルコンピュータの日本語ワードプロセッサで処理可能な形式に変換する。
- (5)リストの訂正をワードプロセッサを用いて修正する。
- (6)元のファイル形式に再度変換する。
- (7)分割したファイルをまとめる。
- (8)再びワークステーションにファイル転送して元の文字コードに戻す。

この作業においては、リストは入力データの順になっており、ファイルを分割しても全く問題ない。また、実際の修正作業には、専門的知識を持たない者でも扱えるようにワードプロセッサを用いたが、より適当なものがあればエディタでも何でも構わない。修正が大量になるようであれば、外注により修正データだけパンチさせて元のデータに組み込むことも可能である。

ローマ字読みに関する部分については、そのままでは非常に校正作業が難しく、カナ読みに変換することにより表記法の誤りが見つけやすくなっている。

4.4 校正過程で発見された誤りの統計

校正過程において発見された誤りにはさまざまなものがある。種類としては、用語自体に関するものと括弧や記号類に関するもの、パンチないしCTS データで混入したものと冊子体自体に含まれるものがそれぞれある。誤りを発見した工程の特徴的なものを以下に示す。

外国語用語：参照データとの比較/辞書を用いたスペルチェック

日本語用語：参照データとの比較/人手による校正

ローマ字読み：カナ読みの生成（変換不能な文字列の調査）

カナ読み：人手による校正

各工程別の統計はとっていないので比率は不明であるが、全体としてかなりの部分が機械的処理の過程で見つけられている。

表1にパンチデータ及びCTSデータのそれぞれに関する各フィールドごとの誤りの頻度を示す。ここで注意が必要なのは、パンチによる入力データの誤りには各種の記号類の使い方に関するものがかなり多く含まれているのに対し、CTSデータでは元々このような記号類はなく、校正過程で付加しているために誤りとなりえない点である。また、パンチデータはパンチ後確認作業を経たものであり、CTSデータも粗い校正を経たものである。なお、全分野における各フィールドの平均文字数は、外国語用語：15.3、日本語用語：5.6、ローマ字読み：12.3である。

日本語の入力データの誤りの頻度に関しては、パンチデータの方が高いのが目につくが、漢字見誤り、打ち誤りが主な原因と考えられる。これに対してCTSデータでは日本語の誤りが少ない。おそら

表1 入力データの誤りの頻度

(a) パンチデータ		(単位：レコード数)	
	外国語	日本語	ローマ字読み
全レコード数	46722	46722	46722
入力データの誤り	188(0.36%)	300(0.64%)	68(0.15%)
原データの誤り	24(0.05%)	25(0.05%)	121(0.26%)

(b) C T S データ		(単位：レコード数)	
	外国語	日本語	ローマ字読み
全レコード数	33776	33776	33776
入力データの誤り	113(0.33%)	52(0.15%)	47(0.14%)
原データの誤り	104(0.31%)	4(0.01%)	86(0.25%)

く日本語は専門家が粗く校正しただけでかなりの精度が得られることが理由であろう。

一方、外国語とローマ字読みに関してはパンチデータとCTSデータとの間に有意な差がない。アルファベットで表現されたものに関しては専門家と一般的な作業者との間にさほどの差がないことを示していると考えられる。同じアルファベットであるのに、外国語の誤りの頻度がローマ字読みに比べて2倍以上高いのは、発音に直したときにローマ字読みの方が確認しやすく、外国語は単語自体の知識がないと確認しにくいことに起因していると考えられる。

原データの誤りとしては、ローマ字読みはパンチデータ0.26%、CTSデータ0.25%とほぼ等しい値となっており、かなり大きい数値と言わざるをえない。一方、外国語はパンチで0.05%、CTSで0.31%と明らかな相違が見られる。CTSによっている化学、電気工学、機械工学の3分野の中にも特異な性質を示すものはなく、自然科学系という特異性によるものなのか、CTSという組版方式によるものなのか、この原因は不明である。逆に、日本語の誤りはパンチで0.05%、CTSで0.01%と頻度が逆転している。自然科学系という特異性により誤りが少ないとも考えられるが、ここではむしろパンチデータ中の書式制御のための記号類の誤りが主要なものであると考えられる。なお、この校正過程で発見された原データの誤りは、専門以外の者が一般的知識と形式をたよりに作業を行った結果であり、特に外国語と日本語の内容についての誤りは発見されずに残っているものがあると考えられる。

これらの誤りの統計を見る限り、通常は入力データの誤りの方が頻度が高いのに対し、ローマ字読みについては原データの誤りの方が頻度が高いのが問題となる。最終的な用語集の冊子体の体裁はともかくも、少なくとも編集過程においてはカナ読みを併用することが望ましいであろう。また、外国語に関してはほとんどが英語であることから、特に専門的な単語は別にして、一般的な英語の綴り誤りなどは単純なスペルチェックユーティリティを用いれば取り除くことができるであろう。

5. おわりに

本研究においては、学術用語データベースの構築の初期段階として行った情報の分析、データベースの構造の設計、および校正の作業環境に関する検討を行った。既制定の学術用語集を対象とした情報の分析により、学術用語のもつ情報の属性やレベルには種々のものがあり、構造としても用語間の相互関係や付加的情報の階層性などがあり、さらに用語の表記の任意性や読みなどに関連する特異性があることも明きらかとなった。この結果を受け、学術用語に付属する情報を、データベースとして、もれなくかつ整然と格納するためのデータの構造や項目の設計を行い、その概要について述べた。これらの検討は、単に学術用語データベースを構築するための準備というだけでなく、今後の学術用語集の選定、編集作業における情報の整理の基本的考え方にも資するものである。また、既存の学術用語集に基づいた初期データベースの構築作業の経験から、計算機、特に UNIX 系のエンジニアリングワークステーションを用いることにより、データの校正作業が質的にも効率的にも大幅に改善されることが明らかになった。この結果は今後の学術用語集の制定における用語の選定、編集過程における機械化の有用性を示唆している。将来的にはこれらの検討をさらに推し進め、学術用語集の高度利用を想定したデータ項目の整備や改訂作業の方式などの検討を行っていく必要があるだろう。

研 究 論 文

ストリング・インデクシング・システムの現状
—PRECIS を例として—

Recent Problems in String Indexing Systems
—An Analysis of PRECIS—

学術情報センター 影浦 峽*

目 次

1. はじめに
2. ストリング・インデクシング・システムの定義と意義、現状
 2. 1 定義
 2. 2 意義
 2. 3 現状
 2. 4 オンライン検索に対する意義と現状
3. 問題の所在と要因
 3. 1 視点
 3. 2 問題の所在—PRECIS を例として—
 3. 3 PRECIS における実例
 3. 4 まとめ
4. 今後の方向

要 旨

ストリング・インデクシング・システムの現状と問題とをまとめる。さらに、PRECIS の索引データをもとに現在ストリング・インデクシング・システムが直面している問題の原因について分析する。最後にストリング・インデクシング・システムの主題アクセスの問題に対する可能な貢献について考察する。

* Kageura, Kyo : National Center for Science Information System

Abstract

This paper summarizes the state of the art and some recent problems in string indexing systems and makes clear the cause of problems by analysing PRECIS index data. The possible contribution of string indexing systems to the problem of subject access is then considered.

1. はじめに

近年、コンピュータの導入によって、情報の蓄積や処理、検索の過程の機械化が促進されてきた。このような機械化の促進は、情報伝達過程の技術面側面にのみでなく、理論的側面にもさまざまな影響を与えている。そして、理論的な側面では、機械化された情報検索システムの枠内に留まらず、冊子体の索引のような伝統的な情報伝達の手段にもコンピュータは影響を与えてきた。

このような状況は主題索引法に関してもあてはまる。主題からの情報へのアクセスに関しては、機械による検索だけではなく、基本的に冊子体を想定した索引システムの理論的発達もコンピュータの登場によって促されてきた。いわゆるストリング・インデクシング・システムはその1つで、コンピュータの導入をきっかけとして生まれた比較的新しい主題索引法である。

ストリング・インデクシング・システムは、主題索引法の理論に対して重要な貢献をしたといわれるが、基本的に冊子体の索引として登場したものであり、その点で様々な実際上の制約を受けている。したがって、オンライン情報検索システムにおける主題アクセスという今日状況の中でストリング・インデクシング・システムの貢献を考えるならば、そこには再考しなくてはならない問題が含まれている。

本稿では、ストリング・インデクシング・システムについてこれまで言われてきた意義を整理し、次に、オンライン検索システムとの関わりで現状を概観する。そして、ストリング・インデクシング・システムの中で最も広く用いられている PRECIS (PREserved Context Index System) を対象に問題点を整理し、今後の主題アクセスの問題に関するストリング・インデクシング・システムの可能な貢献について方向付けを試みる。

2. ストリング・インデクシング・システムの定義と意義、現状

2.1 定義

ストリング・インデクシング・システムという用語は比較的新しいものであるため、図書館情報学分野の基本的な用語集には収録されていない。早い時期にその定義を明確にしたのは Svenonius と Schmierer である。彼女らは、ストリング・インデクシング・システムを、主題の構成する各概念の関係を保ち、その関係に基づいて、1つの主題がそれぞれの概念を見出し語とする複数の記入のもとで過不足なく表示されるシステムと定義している¹⁾。その後、より整理された定義が Craven により与えられている。それによると、ストリング・インデクシング・システムとは、次の2点を特徴として備えている主題索引システムの総称である。

1. 索引の対象となる文献その他の1つ1つについて、通常、索引の記入が複数作られ、それぞれの記入には共通したタームが用いられ、同一の主題を表現する。

2. 明示的に示された系統だった統合規則にしたがって、コンピュータで記入を生成する²⁾。

主題から情報にアクセスするために用いられる索引の主なものとして、ストリング・インデクシング・システムのほかに、件名標目表やチェイン・インデクシングがある。以下では、簡単にこれらとストリング・インデクシング・システムとの相違点をまとめ、ストリング・インデクシング・システムの特徴を明確にしよう。

件名標目表は、主題を表すタームのリストであり、各タームは、たとえそれが複合的なものであっても、それぞれあらかじめ決められている。したがって、件名標目表は、ストリング・インデクシング・システムのように、系統的な統合規則により索引記入を生成するという手続き中心の体系ではない。一般に、ストリング・インデクシング・システムは、何らかの方法で抽出されたタームを調整し、規則にしたがって統合するという索引語抽出型の体系であるのに対して、件名標目表は、文献の主題に一致するタームを表から選択するという索引語付与型の体系である。さらに、件名標目表により1つの文献に複数の記入が作られることもあるが、それぞれの記入は通常、同一の主題を表現するものではない³⁾。

チェイン・インデクシングは、言葉で表現された主題と特定の分類表とを対応付けるもので、複合的なタームで表現される主題を特定性の高いタームから一般的なものへと配列し、それによって分類表中では分散してしまう概念を同一箇所にとめる。このとき、タームをはじめから順に削除しながら主題中の各タームを見出しとする記入を作成する。これは、各記入がそれぞれ同一の主題を表わすというストリング・インデクシングの特徴を備えていない。また、タームの関連付けの規則は分類表に依存しており、明示的な統合規則を持っているわけではない⁴⁾。

ストリング・インデクシング・システムの定義を満たすシステムの主なものとして、Luhn が考案した KWIC (Key Word In Context)、Armitage と Lynch がまとめた ASI (Articulated Subject Index)、Bhattacharyya らによって開発された POPSI (POstulate-based Permuted Subject Index)、Craven によって開発された NEPHIS (NEsted PHrase Indexing System)、Farradane が考案した Relational Indexing、Austin らによって開発された PRECIS (PREserved Context Index System) などがある。これらのうちのいくつかは、すでに日本でも紹介されている⁵⁾。

2.2 意義

主題アクセスの分野におけるストリング・インデクシング・システムの意義としてこれまでに言われてきたことは、実際の側面に関するものと理論的側面に関するものとに分けることができる。

実際の側面での意義は、冊子体をはじめとする一覧形式の索引において、作成過程で人間が行う部分とコンピュータが行う部分とを区別し、コンピュータを有効に利用する方法を示したことにあるといわれる⁶⁾。

一方、理論的側面での意義は、主としてストリング・インデクシング・システムの持つ統合規則と

表示にある。一般に、ストリング・インデクシング・システムの表示は、

1. 適切な文献をどの位置で見つけ出すことができるかという予測を可能にする
2. 類似した主題の文献を近くに配列する
3. 主題を曖昧さなく表現する
4. 適切な主題特定性を保つ
5. 適合しない文献の判断を容易にする⁷⁾

といったことを目的に構成される。そして、このような表示を生成するための統合規則は、従来の件名標目やチェイン・インデクシングの持つ限界を越えて、主題を表現するにあたってより柔軟でありながら索引として求められる一貫性を維持するためのものである。

上述の条件を満たす表示形式が研究の対象となり、そして、そういった表示を生成するための統合規則を定式化するために主題を表わす概念間の関係を整理しようという試みが様々なレベルで行われ、それらは異なるストリング・インデクシング・システムとして具体化されている。

2.3 現状

ストリング・インデクシング・システムについては、様々な研究が行われている。異なる表示形式を持つ多様なストリング・インデクシング・システムの誕生は、索引記入の表示形式に関する関心を高め、その結果、索引記入の有効性に関する比較評価がいくつか行われた⁸⁾。また、主として理論的側面から統合規則の基本となる概念の構造化についての比較研究が行われたし、特に PRECIS を対象として、LCSH との構造や機能の比較も行われている⁹⁾。

ストリング・インデクシング・システムは主に英語を対象として生まれたものであるが、PRECIS では比較的早くから英語以外の言語への適用可能性が研究されてきた。フランス語の PRECIS が現在 *Film Canadiana* で利用されているほか、ドイツ語や中国語などへの適用も考えられている¹⁰⁾。

現在までに作られたストリング・インデクシング・システムのうちいくつかは実用化され、冊子体の索引で利用されている。PRECIS は 1974 年以来 *British National Bibliography* の主題索引として利用されているほか、*British Education Index*、*British Catalogue of Music*、*Film Canadiana* などでも利用されているし、ASI は、*Chemical Abstracts*、*Analytical Abstracts*、*World Textile Abstracts* などでも用いられている。そのほかに、*MLA International Bibliography* で用いられている CIFT (Contextual Indexing and Faceted Taxonomic Access System) のように、特定分野で用いられているものもある。

一方、実際の利用の中から、索引作成に要する時間や費用が大きすぎるのではないかという問題も起きている。この点は、特に PRECIS のような複雑なシステムで顕著である。PRECIS は、実用化以来、複雑すぎるといった批判と、他の索引と比べて特に必要な時間や労力が多いわけではないという意見とが対立してきた¹¹⁾。実際に *Australian National Bibliography* では、1985 年に PRECIS の利用を取りやめている。これには、オーストラリアの図書館の多くが LCSH を持つ LC MARC を利用しているという理由もあったにせよ、PRECIS が索引作成者の負担を増大させたこと、索引作成に熟練し

た技術を要したためスタッフの固定化を促したことが大きな理由となっている。

時間や費用の増加に対して得られる利益の大きさという点は現実的な問題であるが、特に複雑なストリング・インデクシング・システムに対して、こうした疑問が投げかけられている。手続き的な面でのストリング・インデクシング・システムの利点と考えられていた人間と機械の分業については再考を要する段階にあるようである。

2.4 オンライン検索に対する意義と現状

オンライン検索システムにおいて、主題アクセス機能の高度化として求められている主なものとして、例えば次のような機能がある。

1. 重み付け検索
2. ブール演算よりも高度なターム間の関係付け
3. 索引に用いられているタームやタイトル、件名標目などのブラウジング機能
4. 利用者の用いたタームとシソーラス中のタームとのマッピングなどによる検索式の自動的な拡張¹²⁾

このうち、2と3に関わる点で、ストリング・インデクシング・システムはオンラインによる情報検索に対しても有意義なものであると言われている。すなわち、

1. ストリング・インデクシング・システムの表示機能の研究が、オンライン情報検索システムにおける索引語の表示とブラウザビリティに関する研究に貢献する。
2. ストリング・インデクシング・システムが規則化している概念間の関係が、オンライン情報検索システムの索引語について求められているブール演算より複雑な概念間の関係について有益な知見を与える¹³⁾。

オンライン検索におけるブラウザビリティは最近特に重要性が認められてきている問題であるし、ブール演算よりも複雑な概念間の関係付けはこれまで継続的に関心が持たれてきた問題である。

では、オンライン検索に対する実際の適用はどのような状況にあるのであろうか。ストリング・インデクシング・システムは索引記入を作成するためにコンピュータを利用するが、オンライン検索システムで実用化されているものは今のところほとんどない。

最も著名なのは、CASearch で用いられている Articulated Subject Index である。しかしながら、この CASearch で用いられている検索手法は、ブール演算であり、ストリング・インデクシング・システムとしての ASI の機能をオンライン検索で発揮しているとはいえない¹⁴⁾。

オンラインへの適用が研究されているもう1つのシステムは PRECIS である。PRECIS は BNB の機械化にともなって実用化された索引システムで、当初からオンライン検索での利用が示唆されてきた。

カナダの National Film Board の FORMAT システムでは、PRECIS の索引フィールドをオンラインで検索する機能を利用者に提供しており、また、ブラウジングのために PRECIS の索引記入を表示する機能も備えている。しかしながら、検索において概念を組み合わせる方法としてはブール演算

しか許されておらず、また、記入の表示においてタームは入力ストリングの順に表示されるだけであり、PRECIS の特徴である 2 行にわたる表示が検索語に応じて行われるわけではない¹⁵⁾。

結局、これまでにストリング・インデクシング・システムのオンライン検索に対する貢献についていくつかの一般的な主張がなされてはきたが、実際には、ストリング・インデクシング・システムが“どのようにオンライン環境での検索に貢献しうるかについては明確な理解がない”¹⁶⁾というのが現状である。

オンライン化の中での実際の役割が明確にならないならば、ストリング・インデクシング・システムの今後の存在意義は薄れていくであろうし、索引作成に必要な作業の量を考えても正当化されることはないであろう。

3. 問題の所在と要因

3.1 視点

ストリング・インデクシング・システムは、現在オンライン検索システムに実際に貢献するかどうか不明確でないという問題に直面している。また、ストリング・インデクシング・システムのうち複雑なものについては索引作成者の作業負担が大きいという問題を持っている。こうした状況については、例えば以下のような解釈が可能であろう。

1. これらの問題は純粋に手続き上の問題を原因としており、したがってその点が解決されれば問題はすべて解決され、オンライン検索システムへも適用されて、その改善をうながす。
2. 索引作成者の負担が大きいという問題は手続き上の問題であり、それを乗り越えさえすれば冊子体の索引システムとしてはよい。しかしながら、冊子体の索引とオンライン検索システムにおける索引とは基本的に異なるため、オンライン検索システムに対する貢献はない。
3. 現在みられる問題は、ともに本質的なものであり、オンライン検索システムへの適用どころか冊子体の索引システムとしてもストリング・インデクシング・システムは基本的に誤った方向に進んできたものであることを示している。したがって、もうストリング・インデクシング・システムは捨てて、別の方向で索引システムを考えていったほうがよい。

どのような立場を取るにせよ、まず手続き上の問題から考えていったほうがよいように思われる。というのは、現に問題として現れているということは、それが本質的な問題に関わっているかどうかによらず、実際に手続き上の障害として存在しているということを示しているからである。また、手続き上の問題は背後にある本質的な問題を見失わせることがあるし、その一方で手続き上の問題が背景にある理論の欠点を表わしていることが多いからである。

3.2 問題の所在—PRECISを例として—

以下で具体的な検討を行うことにするが、それにあたっては、ストリング・インデクシング・システムの 1 つである PRECIS を検討の対象とする。これは PRECIS がストリング・インデクシング・システムの現在の問題を典型的に含むものである一方、最もよく研究され、また比較的広く実際にも

利用されているシステムだからである。

PRECIS による索引作成は、まず資料の主題をタイトルのような句で表わすことから始まる。索引作成者は次に、その句の中のそれぞれの概念を表わすタームにその性質や役割に応じてオペレータを付与し、入力ストリングを作成する。ここまでの索引作成者の作業である。ここで、オペレータとは主題を表わす句の中の各タームを構造化するものである。表1にその種類と基本的な役割を示す¹⁷⁾。この入力ストリングをもとに、表示を生成するアルゴリズムによって索引記入が作られる。

PRECIS をオンライン検索に用いる場合には、オペレータ付与の対象となるタームのレベル、すなわち入力ストリングで表わされるレベルが検索の対象となる。ところが、オペレータがタームの役割を示すものとしても用いられず、また、該当する索引記入の表示も冊子体のそのようなかたちでは行われていない。これには、検索システムの利用者にどのようにしてタームの役割を指定させるか、冊子体のようなアクセス・ポイントの問題がないのに冊子体の表示を行うことは必要かという検索シ

表1 PRECIS の基本的なオペレータ

中心となるオペレータ		
基本分類	オペレータ	役割
システムの環境	0	場所
基本システム	1	キー・システム 他動詞的動作の対象, 自動詞的動作の動作主
	2	動作, 効果
	3	他動詞的行為の遂行者, 側面, 要因
観察者に関するデータ	4	著者の視点や主題の捉え方
	5	標本集団や研究地域
	6	対象となる利用者, 文献の形態

補助的なオペレータ		
基本分類	オペレータ	役割
依存要素	p	部分や属性
	q	準類集団のメンバー
	r	集合
概念を連結する要素	s	役割決定子
	t	著者により与えられる関係
並列概念	g	並列概念

システムに関わる問題もある。

しかしながら、これらと同時に、PRECIS 自体が持っている問題が、ストリング・インデクシング・システムの貢献としてあげられた点をオンライン検索システムに適用する障害となっている。入力ストリングは主題を表わす概念の構造をオペレータにより表示するレベルであるが、実際に入力ストリングをもとに表示を生成するためには、入力ストリングに概念の構造化とは無関係の余分な要素を付け加える必要がある。そして、それによってオペレータの基本的な役割が変更されてしまうことがある。すなわち、概念の構造化とそれをもとに表示を生成するアルゴリズムとが体系的に整理されていない。

現行の PRECIS システムの特徴をオンライン検索システムに強引に適用することは不可能ではないが、そのためには、検索システムのアクセス・ポイントとしては役に立たず、しかも実際に用いられるかどうか明らかでない表示生成用の句をかなり大量に記憶させる必要がある。ただ現在システムがあるから利用するというのではなく、貢献があり得るかもしれないと考えるならば、逆に、PRECIS が持っているこうした問題が PRECIS の基本的な欠陥を反映したものでないかどうか検討する必要がある。

入力ストリングにおいて表示の整形のために導入されるものは次のように整理できる。

- a) 上方代替ストリング これは、索引記入の曖昧さや冗長性をなくすために導入される。実際にどういう場合に上方代替ストリングが導入されるかについての明確な規定はないが、例えば主題に動作概念が2つ含まれ、第二の動作が第一の動作を含む一群のタームと関連し、かつ第二の動作が導入語となる場合や、オペレータ q によって集合を表わすタームが導入され、その集合の動作が導入語となる場合に導入される。
- b) 空のストリング 代替ストリングの1つとしての空ストリングの他に、表示における述語変形を避けるために挿入される空ストリングがある。これは、オペレータ 3 で示された概念が、入力ストリング中で前にくるオペレータ 2 で示された動作の直接の動作主ではないときに必要になる。この空ストリングには必ずオペレータ 2 が付与される。
- c) 複数の入力ストリングの作成 これは、主として、主題にももの移動が含まれ、異なる視点から見ると異なる表現が必要になる場合におこる。
- d) 索引記入の特別な位置にのみ表示されるタームの挿入 これは、例えば、3つ以上の概念が等位接続関係にあるとき、それぞれを導入語とするために挿入される¹⁸⁾。

以上4つの状況が、PRECIS において構造化された概念を表示するレベル、すなわち入力ストリングにおける不純要素と考えられる。このように表示の整形のために概念の構造化のレベルで余分な要素をアド・ホックな規則により付け加えることが、オンライン検索に対する適用の問題の1つとなっているばかりでなく、索引作成者の作業負担を大きくしている。

3.3 PRECISにおける実例

ここでは、前節で整理した問題について実際にデータを分析し、それらの問題が何を原因としてい

るか明らかにすることを試みる。表示の問題は言語の理解の問題と密接に関わっており、そこから生じる問題は、言語表現における基本的な構造を捉えていないことに起因する場合が多い。そこで、検討にあたっては、対応する言語表現を参考にする。

分析に用いたデータは、*British National Bibliography* の 1980 年から 1986 年末までの主題典拠レコード累積版から抽出した、デューイ十進分類標数 153「知能及び知能と意識のメンタル・プロセス」に関するデータである。分析対象となった主題の数は 298、オペレータの延べ数 1018 である。用いられているオペレータとその頻度を表 2 に示す。

前節で述べた、入力ストリング中で表示の整形を目的としてなされたオペレータの導入や反復に関する基本的な数値を表 3 に示す。表からわかるように、前節の a と b に該当する 경우가非常に多い。

まず、a の状況について考える。分析の対象となった 37 の入力ストリングに見られる 40 の上方代替ストリングのうち、37 個は非常に規則的なパターンに従っている。典型的な例を以下に示す¹⁹⁾。

主題：Mathematical models of role of mental images in human memory : Applications of computer systems

入力ストリング：

- (1)man
- (P)memory
- (su2) (1)human memory
- (s)role \$v of \$w in
- (3)mental images
- (su5) (2)role of mental images in human memory
- (2)
- (3)mathematical models
- (su8) (3)mathematical models of role of mental images in human memory
- (s)applications \$v of \$w in
- (3)computer systems

これらの場合については、上方代替ストリングが導入される理由は非常に簡単である。すなわち、適切な表示を維持するためには埋め込まれた構成単位をひとまとまりとして扱う必要があるにもかかわらず、そもそも規則の基礎となる概念の構造化の段階で、自然言語では節や句により表されるような、埋め込まれた単位について全く考慮されていないことである。しばしば PRECIS は埋め込みを表わすシステムであるといわれるが²⁰⁾、むしろ全く逆で、索引記入の理解のために埋め込みを適切に表示することが必要なのは自然なことであり、そのために必要な規則を体系化していないために例外的なものとして処理される。

残りの 3 例は、固有名詞が表わす概念とその説明からなる主題表現の表示に関わっている。例えば、次のような場合である。

表2 オペレータの使用頻度

オペレータ	頻度	相対頻度 (%)
0	5	0.5
1	238	23.4
2	283	27.8
3	177	17.4
4	23	2.2
5	2	0.2
6	54	5.3
g	5	0.5
p	134	13.2
q	9	0.9
r	2	0.2
s	65	6.4
t	21	2.1
総 合	1018	100

表3 余分なオペレータの挿入と複数の入力ストリングの作成

a) 代替ストリング

		頻度	相対頻度
用いている主題		37	12.4
挿入されたオペレータの種類と数	1	14	5.9
	2	24	8.5
	3	2	1.1
挿入されたオペレータ総数		40	3.9

b) 空ストリング

		頻度	相対頻度
用いている主題		39	13.1
挿入されたオペレータ総数		39	3.8

c) 複数の入力ストリング

		頻度	相対頻度
用いている主題		4	1.3
作られる余分な入力ストリングの数	1	3	/
	2	1	
余分な入力ストリングの総数		5	

d) 特別なタームの挿入

		頻度	相対頻度
用いている主題		6	2.0
挿入されたオペレータの種類と数	1	1	0.4
	2	3	1.1
	g	2	40.0
	p	4	3.0
挿入されたオペレータ総数		10	1.0

主題：Assessment of De Bono (CoRT) Thinking Program : Teaching aids for development of school students' thought process

入力ストリング：

- (1)schools
- (p)students
- (p)thought processes
- (2)development
- (2)
- (3)teaching aids

(su6) (1)

- (q)De Bono (CoRT) Thinking Program
- (2)assessment

これらに対する代替ストリングの挿入は、主題における構造的、意味的な視点を反映しているのではないように思われる。むしろ、固有名詞が導入語となったときにその機能や構造の説明を記入から削除しなくてはならないというのは、完全に索引の運用面に関わる問題であり、概念の構造化の不備が直接現れたという性質のものではない。

次に、b の状況を考える。これらの例におけるオペレータ 2 を付された空のストリングは、すべて、述語変形を避けるためにある。述語変形とは、オペレータ 3 を付与されたタームが導入部に来て、かつ入力ストリング中でその前に来るタームに動作を示すオペレータ 2 が付与されているときに用いられる特別な記入生成アルゴリズムである。これに対して、オペレータ 3 を付与されたタームが前に来る動作の直接の動作主ではない場合、すなわち側面や要因を表わす場合、述語変形を行うと意味のない記入が生成されてしまうので、それを避けるために、オペレータ 2 を付与された空のストリングが入力ストリング中のオペレータ 3 を付与されたタームの前に挿入される。以下に例を示す。

主題：Physiological aspects of human mental process

入力ストリング：

- (1)man
- (2)mental processes
- (2)
- (3)physiological aspects

問題はこのような全ての例について、側面や要因を表わすタームが、表面には現れない動作について、意味的には明らかに対象に対する動作主となっていると主張されていることである。すなわち、要因や側面を動作主と区別しないのは、'Physiological aspects of human mental process' においてオペレータ 3 を付与される 'physiological aspects' は、'human mental process' を「コントロールする」という動作の動作主であるからとされる。

この主張の欠点は明らかである。対応する言語表現の形式の分析を無視して、抽象的な意味分析を強引に行ったことである。生成意味論の破綻を指摘するまでもなく、このような理由付けは、どこまで許容されるかの明確な基準を失い理論化を妨げる。少なくとも、このように頻出する側面や要因を示す概念は、オペレータ3とは区別されるべきであろう。

cの問題、すなわち、同一の主題に対して複数の入力ストリングが作成される場合についての実例は、前節で示したマニュアル中の説明とは大きく異なる。そのうちの2つは、次の例のように、埋め込みの概念の欠如に関わっているもので、比較的簡単に改善可能である。

主題：Effects on literature and language of human perception of reality

入力ストリング1：

- (1)reality
- (s)perception \$v by \$w of
- (3)man
- (2)effects of literature & language

入力ストリング2：

- (1)literature
- (p)language
- (2)effects on man's perception of reality

残りの2つは、次のように、並列概念と導入語の問題に関わっている。

主題：Human learning and memory

入力ストリング1：

- (1)man
- (lo) (p)memoly
- (p)learning & memory

入力ストリング2：

- (2)learning \$v by
- (3)man
- (p)learning & memory

これらについて複数の入力ストリングを作らなくてはならない理由は、概念の構造化の欠如を反映した表示の欠点を避けるためのものではない。それ以前に、このような主題について、どういった表示が求められているかについて基本的な規則がないのである。したがって、今のところ概念の構造化と関連付けることのできない索引の運用面に関わっているといえることができる。

dの問題を持つ6例のうち3つの入力ストリングは、次の例のような相互に論理的に関連する並列概念を扱うものである。

主題：Development of cognition, communication and perception

入力ストリング：

- (1)man
- (lo) (p)cognition
- (lo) (p)communication
- (lo) (p)perception
- (p)cognition, communication & perception
- (2)development

これらについては、PRECIS マニュアルの第2版で導入されたオペレータより解決が示されている。すなわち、並列概念が相互に論理的に関連しており、両者の関連が重要である場合には、独立した概念の並列とは異なるオペレータを用いるのである。

残りの3つは、索引記入の運用上の操作に関わっている。例を次に示す。

主題：Relation of age to human learning

入力ストリング：

- (2)learning \$v by
- (3)man
- (su2) (1)man
- (nd) (2)learning
- (t)related to
- (p)age

基本的にこれらの例では、索引記入の表示の一貫した形式を保つための基礎となるオペレータが、その役割とは無関係に無秩序に利用されている。すなわち、表示の整形の一貫性を保つ規則が表示のために変更されている。これらについては、表示の問題から再考を要する。

3.4 まとめ

オンライン検索システムにうまく適用できない原因の1つである入力ストリングの混乱のうち、主なものは理論的整理の不足によるもので、オペレータに変更を加えることで避けることができる。より厳密な体系化は必要であるが、オンラインシステムへの適用における問題の1つは解決できる。

また、このような体系化により、表示の整形に必要な余分なものを入力ストリングに付加するという人間の作業を規則的なものに置き換えることが可能となろう。したがって、PRECIS が直面している「索引作成者の負担が多すぎる」という問題もある程度改善することができる。

結局のところ、直面している問題の原因の1つは、必要な体系化の欠如から起こる問題を例外的なものとして処理し、作業の大部分を索引作成者の知的作業に負担させていたという、PRECIS 自体が持っていた欠点であったということが出来る。これは、多かれ少なかれ他のストリング・インデクシング・システムにも共通する問題であり、したがって、それぞれのシステムについて再検討が必要となろう。

4. 今後の方向

ストリング・インデクシング・システムが直面している問題の原因の1つを、PRECISの検討を通じて明らかにした。しかしながら、結局システムの体系化が不十分であることを指摘しただけで、オンライン検索システムを含む今後の主題アクセスの問題に理論的に貢献し得るかどうかについてはまだ何も答えていない。最後に、これらの問題について簡単に考える。

まず、それぞれのシステムを直接オンライン検索システムに適用することには余り意味がないであろう。そもそもストリング・インデクシング・システムはそれ自体分解可能なものであり、いくつかの可変の要素からなっているし、しかもそれら相互の整理に問題があるからである。したがって、システムの構成要素を整理した上で、それぞれの体系について検討する必要がある。

最初に、ストリング・インデクシング・システムの表示が、オンライン検索システムの表示、ブラウザビリティの問題にどれだけ貢献しうるかという点を考える。この点については、冊子体とオンライン検索では大きなギャップがある。冊子体の索引システムでは、必要なアクセス・ポイントを保ちながら類似の主題を一箇所に集め、しかも理解が容易な記入を作るにはどうしたらよいかというのがCutter以来の基本的な問題であった。ところが、オンライン検索システムでは、アクセス・ポイントの問題と表示の問題は原理的に区別することができる。このような基本的な相違にも関わらず、冊子体のために考えられた線形2次元配列がオンラインでのディスプレイにも有効であるかどうかは疑問の残るところである。

一方、概念の構造化の問題についてはどうであろうか。基本的に冊子体を目的とした索引システムでは、適切な表示を生成する規則を定式化するために概念の構造化が行われた。表示がオンライン検索システムには適切でないとしても、そのために行われた概念の構造化は考慮する価値があるかどうかは明確ではない。しかしながら、これらは「主題」をとらえるにあたっての1つの視点であり、しかも実際に大量のデータを実際に処理してきたものである。したがって、ストリング・インデクシング・システムに見られる概念の構造化の視点が索引独自のオントロジーを示しているかもしれない。

例えば、言語表現と対応付けることのできる主題記述の範囲において索引システムにみられる体系が、実際の索引作成の場で一貫して適用されており、しかもそれが言語学的一般化や言語学的記述のレベルとは異なるものであるならば、それらは真剣に検討する価値があるであろう。

現在、PRECISにみられる概念の構造化とその分析レベルとについてデータを分析中である。しかしながら、PRECISの中心となるオペレータは、言語学でいう主題役割の不完全なものでしかないようである。ストリング・インデクシング・システムが今後の主題アクセスの問題に提供するものはそれほど多くはないかも知れない。

注・参考文献

- 1) Svenonius, E. and Schmierer, H. "Current Issues in the Subject Control of Information," *Library Quarterly*. vol. 47, no. 3, 1977. p. 326 - 346.

なお、string という用語はこれ以前に、次の用語集に現われている。

Buchanan, B. *A Glossary of Indexing Terms*. London, Clive Bingley, 1976.

- 2) Craven, T. C. *String Indexing*. Orlando, Academic Press, 1986. p. 3 - 4.
- 3) Chan, L. M. 目録と分類. 上田修一他訳. 勁草書房, 1987. [*Cataloging and Classification : An Introduction*. New York, McGraw-Hill, 1981.] p. 145 - 177.
Craven, *op. cit.*, p. 50.
- 4) Burkett, S. G. "Strings or Chains?" *Australian Academic and Research Libraries*. vol. 5, no. 4, 1974. p. 195 - 200.
Craven, *op. cit.*, p. 52 - 53.
Mineur, B. W. "Relations in Chains," *Journal of Librarianship*. vol. 5, no. 3, 1973. p. 175 - 202.
- 5) たとえば、以下のようなものがある。
川村敬一. "KWIC 索引の問題点と改良の試みに関する展望," *情報管理*. vol. 23, no. 7, 1980. p. 571 - 589.
川村敬一. "Articulated Subject Index の構造的特性と機械化プログラム," *情報管理*. vol. 24, no. 5, 1981. p. 447 - 456.
川村敬一. "NEPHIS—「入れ子式」ストリング索引法の原理と応用の紹介," *情報管理*. vol. 25, no. 4, 1982. p. 283 - 292.
三浦逸雄. "PRECIS : Preserved Context Index System," *社会教育学・図書館学研究*. no. 3, 1979. p. 87 - 94.
郡司良夫. "PRECIS について—紹介の試み—," *大学図書館研究*. no. 17, 1980. p. 1 - 10.
- 6) Craven, *op. cit.*, p. 6.
Richmond, P. A. *Introduction to PRECIS for North American Usage*. Littleton, Libraries Unlimited, 1981. p. 16.
- 7) Craven, *op. cit.*, p. 7.
- 8) Keen, E. M. "On the Processing of Printed Subject Index Entries during Searching," *Journal of Documentation*. vol. 33, no. 4, 1977. p. 266 - 276.
Keen, E. M. *On the Performance of Nine Printed Subject Index Entry Types : A Selective Report of EPSILON*. Aberystwyth, College of Librarianship Wales, 1978. 77 p.
Hunt, R. "The Subject Catalogue in Australian Academic Libraries : PRECIS, LCSH and KWOC and the Findings of the Wollongong University Subject Catalogue Study," *Australian Academic and Research Libraries*. vol. 9, no. 2, 1978. p. 61 - 70.
Jamieson, A. J. *Comprehension of Entries in Printed Indexes*. Ph. D. Thesis, University of Maryland, 1981. 249 p.
- 9) Gabbard, P. B. "LCSH and PRECIS in Music : A Comparison," *Library Quarterly*, vol. 55, no.

- 2, 1985. p. 192 - 206.
- Micco, H. M. "A Comparison of Subject Access Systems in Medicine : LCSH, MeSH, PRECIS," *Proceedings of the 48 th ASIS Annual Meeting*. vol. 22, 1985. p. 41 - 53.
- 10) Austin, D. "PRECIS." [Kent, A. ed. *Encyclopedia of Library and Information Science*. vol. 42 (supplement 7). New York, Marcell Dekker, 1987.] p. 419 - 421.
- 11) 例えば、以下の文献を参照。
- Bakewell, K. G. B. *A Study of Indexer's Reactions to the PRECIS Indexing System*. Liverpool, Liverpool Polytechnic Department of Library and Information Studies, 1978. 63 p.
- 12) Salton, G. and McGill, M. J. *Introduction to Modern Information Retrieval*. Auckland, McGraw-Hill, 1983. 448 p.
- Markey, K. *Subject Searching in Library Catalogs*. Dublin, OCLC, 1984. 176 p.
- 13) Craven, *op. cit.*, p. 5 - 6.
- Hunter, E. J. "The United Kingdom Contribution to Subject Cataloguing and Classification since 1945," *International Cataloguing*. vol. 16, no. 3, 1987. p. 34.
- 14) ただし、ASI の前置詞を手がかりに格関係を復元し、検索に用いる実験が行われている。
- 石塚英弘他. "科学文献情報データベースの知識ベース・システム化：設計と試作," "1989 情報学シンポジウム講演論文集. p. 53 - 61.
- 15) Cain, J. "PRECIS at UTLAS." [Dodson, S. and Mentis, G. L. eds. *Academic Libraries : Myths and Realities : Proceedings of the Third National Conference of the Association of College and Research Libraries*. Chicago, American Library Association, 1984.] p. 165 - 176.
- 16) Travis, I. L. and Fidel, R. "Subject Analysis," *Annual Review of Information Science and Technology*. vol. 17, 1982. p. 133.
- 17) Austin, D. *PRECIS : A Manual of Concept Analysis and Subject Indexing*. London, Council of British National Bibliography, 1974. p. 423.
- なお、1984 年には第 2 版が出版されているが、今回分析の対象としたデータは初版に従っているので、初版に基づいた。
- 18) *Ibid.*
- 19) 入力ストリングの例で、全てのストリングに付されている (X) は、オペレータを示す。また、(suX) は、それよりも下に来るタームが導入語となるときに、そのストリングが上位 X 個のストリングと置き代わることを示す。(lo) と (nd) は、そのストリングがそれぞれ導入部、限定部にしか現われないことを示す。なお、導入語を示すコードはここでは無関係なので省略した。
- 20) Hutchins, W. J. *Languages of Indexing and Classification*. Stevenage, Peter Peregrinus, 1975. p. 94 - 98.

(資 料)

LIBRARIES, TECHNOLOGY AND SCHOLARSHIP

The Library as an Information Center

May 1987

Patricia Battin*

Vice President for Information Services and
University Librarian, Columbia University

Thank you for inviting me to come to Japan to share with you my thoughts about the changing role of libraries and librarians in the electronic society. I am particularly honored to have this opportunity to discuss these issues with you, because I think the future quality of research and instruction in our respective countries will be largely dependent upon our universities' collective capacity both to preserve our traditional knowledge centers and to develop the framework for the intellectually responsible utilization and management of the new information technologies.

In the United States there is currently enormous interest throughout the country centering on the merging of libraries and computer centers' and the emergence of a phenomenon called the Computer Czar. Unfortunately, I think the focus has been subverted to a turf battle rather than a thoughtful consideration of our objectives in such an organizational transformation. I would like to give you my thoughts and perceptions about the need for a new information infrastructure to support research and education.

I speak to you from my perspective as an American university research librarian. I know that the many differences between our cultures are reflected in the traditional philosophy, organization and financing of our libraries. The use of the new technologies by our scholars is challenging our cultural and intellectual assumptions as that use shapes the development of a new information in both our societies.

I very much appreciate the opportunity to talk to you about the changing role of libraries and technology in the scholarly process and to dispel, I hope, the much too prevalent notion that libraries will be somehow irrelevant in the brave new world of the wired university. Information technology has blurred, I think, our perception of the distinction between knowledge and

* President, The Commission on Preservation and Access

format, between substance and tool. We should remember that libraries are the original academic information systems, and in our zeal to adopt new information technologies, we seem to be in danger of concentrating on the system and overlooking the information and the needs of those who use it.

Today in one of those curious paradoxes of history, the library –long the image of tradition and order–now reflects the revolutionary changes in our society, as we have moved from an agricultural society through an industrial society to what is now called the Information Age. The impact of computer and communications technology has presented the world of higher education with possibly the most critical challenge it has ever known.

And because our universities through our libraries have long supported and sustained for our societies at large convenient access to our intellectual heritage, the manner in which we meet this challenge is crucial for our economic, political and cultural survival.

The rapidity and speed of the electronic revolution and the enormous enthusiasm for the extraordinary capabilities of the new information technologies have tended to obscure the true function of both the libraries and information technology and have contributed to the indiscriminate equation of information and data with knowledge and wisdom, to the detriment of the educational process.

In our enthusiasm for the “wired campus,” we have tended to overlook the importance of what will be transmitted on all those wires, and in our passionate if somewhat non-analytical devotion to the book, we have neglected the needs of the central figure, for whom we exist – the knowledge seeker.

In the United States, there is an interesting new phenomenon – the creation of the position of Computer Czar–Vice–Provost for Computing and Information Services – on most university campuses. These positions are being created in full recognition of the fact that the overwhelming proportion of the incumbent’s responsibility will be the management of “academic information services,” yet few if any of these positions include responsibility for libraries. Most of the current Czars come from technical, hardware-oriented backgrounds and have had little involvement with research disciplines outside the sciences, the process of scholarly communication, and systems of bibliographic control and information retrieval.

Historically, the University has defined the responsibility for providing sources of information to faculty and students as a central function, mainly in the form of libraries. Once resources have been acquired by the university, they have been available as a free good for its members, by virtue of their appointment or their admission as students, and in the case of public institutions, society at large, including the corporate sector. Costs have been budgeted

centrally and users have been charged little or, most commonly, nothing. Subsidized browsing—the capacity for broad access to information sources at no charge to the user—has been the primary characteristic of our educational and research enterprise.

Funding for computing, because of its magnitude and rapid growth, has been generally addressed in an ad hoc manner, and because initial uses of computing were not for information processing, cost recovery from the individual became the basic funding strategy. The rapid growth of computer and communications technology for information exchange among the community of scholars and in society at large has created a chaotic environment and rendered the traditional formulas obsolete. The most troubling trend has been the tendency to focus on the format, rather than the function, of the information process. Librarians and scholars have tended to think in terms of systems for printed materials, while computer specialists have concentrated on the capacities of the technology. In most instances, the needs and the information habits of the user have been ignored.

Our new challenge to conceptualize, develop, and finance the new infrastructure necessary for the full and productive use of the new technologies, which require compatibilities of hardware, software, and network gateways on an international scale to meet the needs of the research community. In addition to these organizational and policy barriers, the characteristics of information technology, which on the one hand permit an extraordinary expansion of the ability to create, store and disseminate information, also encourage the chaotic proliferation of information resources and permit unprecedented control of access to that information. The ability to control access to information by charging a fee per use has serious implications for the individual researcher as well as the university's capacity to provide the necessary information services.

If we are to link successfully the promise of the new technologies with the intellectual heritage of the past to provide continuing unfettered access to knowledge, we must look at information as a function, not a format, and the organization of information services from the perspective of the wired scholar, the information seeker at the individual workstation. It is essential to distinguish between the perspective of the workstation, i. e., the technology, and the perspective of the wired scholar, i. e., the process of scholarship. To me the essential focus is the quest for knowledge by the human mind—not the capacity of the technology. And since the process is an interactive one, that quest will change over time as technological capacities expand the borders of intellectual inquiry. Such a perspective now requires a new conceptual approach to the provision of broad access to information both within our institutions and throughout the research community. That challenge, to my mind, is far more

difficult and far more important to our continued economic, political and cultural survival than filling our institutions with cables, wires, and other peripheral technologies.

The needs of scholars have always transcended local barriers. The American national library is a decentralized system composed of the Library of Congress, the National Library of Medicine, the National Library of Agriculture, and approximately one hundred private and public research libraries across the nation.

The Library of Congress provides via its MARC tape service the records of its cataloging in nationally and internationally standardized machine readable format. The LC Name Authority File is also available on line as are the bibliographic records from the Government Printing Office and the National Library of Medicine. There are three bibliographic utilities which distribute these machine-readable records to libraries across the country. At present, the transfer to the utilities from LC is by means of magnetic tapes. In turn, the participating libraries, connected to the utilities' computers, contribute bibliographic records with location symbols prepared for materials not yet recorded in the data base. The phenomenal growth of these data bases has resulted in a vastly increased capacity to share cataloging responsibilities and thus reduce local institutional expenses. The existence of these large data resources has revolutionized interlibrary loan capacities and made possible the potential for developing a coordinated national collection through new means of access to decentralized collections. The two major bibliographic networks—OCLC and RLIN—provide information currently on a combined total of approximately 30 million unique records for books, maps, manuscripts, periodicals, audiovisual materials, sound recordings, music scores, and machine-readable data files with numbers increasing daily at a rapid pace.

The Council on Library Resources, a privately funded foundation, launched some years ago a Bibliographic Services Development Program to help bring into existence a comprehensive logically consistent, non-redundant data base of bibliographic records. To insure comprehensiveness, the data base must be built by a set of cooperating, contributing institutions adhering to a common set of standards. The element of non-redundancy requires the use of an authority file to record the entities that have been created according to the set of accepted rules. The objective of this program is to create a widely available, cost-effective bibliographic record service which will incorporate the resources of the major shared cataloging services and provide access to a variety of bibliographic data bases in a manner transparent to the end user.

For the past several years, the Library of Congress, the Research Libraries Group, the Western Library Network and OCLC have worked on a cooperative project funded by the

Council on Library Resources to develop a standard network interconnection based on ISO (International Standards Organization) standards. The first phase of the project—distribution of LC records via the link rather than tape—is now in operation between LC and RLG.

The Research Libraries Group represents a focussed effort by thirty-six research universities and their scholars to reshape information services for scholars. In contrast to OCLC, which is a mass-market driven enterprise, RLG derives its direction from the program needs of its owner-member research institutions.

In the library profession today in the United States, we speak of three generations of computing – not to be confused with the five generations of computer design. The first generation—which has essentially occupied us for the past fifteen years—has been the application of computer and communications technologies to library processing activities. We used the computer to print catalog cards and the telecommunications capacities to share the by-product of machine readable information. We essentially revolutionized the manner in which we run our internal operations. The resulting information is not generally available to the public, except through mediated terminals at reference desks. The technical environment is homogeneous—a large central computer linked by long communications lines to terminals across the country. That era is generally referred to as “library automation” and it is basically over. During this period much of the work which will make the workstation useful to scholars was accomplished.

A number of commercial vendors have emerged from the first generation with software and hardware products to support the maintenance of integrated local systems, comprising the cataloging, acquisitions, and circulation functions. Records from these operations can then be made available to local researchers at workstations on local area networks. However, the central data resource—the on-line national union catalog—is essential for scholarly purposes, so the major challenge for the second generation—where we are today—is to plan an orderly path for the distribution of appropriate software applications to the local scene, to develop a new set of technical and staff capabilities for the central node, and to establish effective financial strategies for sharing the costs of the new services.

We have begun the planning and development efforts for the Third Generation, which will move a set of software applications to the scholar's workstation, enabling the selective downloading and interactive manipulation of this wealth of information by individual users. We assume that all institutions will, by the early nineties, have local area networks to accommodate the workstations and that these local networks will be linked via gateway software. The local on-line catalog will evolve into a university information system with

gateways to a broad range of knowledge resources.

At the same time we are building this national communications capacity, we are also engaged in transforming our large research libraries in order to provide convenient access at scholars' workstations to these important research resources. These large libraries can be compared to other infrastructures developed in the industrial age to provide essential services: highways, subway systems, public utilities, to name a few. Large libraries, like many transportation systems, suffer from the effects of massive deferred maintenance and revolutionary changes in their users' environment and life style, while at the same time critical to their users' ability to conduct their essential business.

In many ways, the research library represents a nineteenth century conception of an information system, heavily dependent upon labor-intensive internal operations and demanding labor-intensive efforts from its users, upon which approximately 90% of the information needs of the academic and research programs depend. It coexists with an emerging twenty-first century information system which serves approximately 10% of those needs. The fact of the coexistence contributes to a high degree of frenetic schizophrenia among our scholars, who expect the efficiency and convenience of electronic facilities from traditional library services and the comprehensive literature coverage of traditional library services from electronic system. The fact that the technical potential exists to provide such services serves to create a very high level of unrealistic expectations, further compounded by the fact that the bulk of the scholarly literature still resides only in the traditional print formats.

At the present time, we are coping with essentially three parallel systems, i. e. traditional manual activities, electronic services, and the management of the project-oriented applications of technology. The need for talented staff who combine both generalist and technical knowledge and skills grows ever more acute. The decade ahead of us will require individuals who are capable of managing the traditional labor intensive services, who also have the imagination and creativity to design new systems and who have the talent and skills to put them into place within a very complex environment.

Within that general background, what are some of the specific changes in scholarly information habits which are creating the need for a new infrastructure in our universities and what are the implications for national and international system of access to information?

- 1) The creation and use of machine-readable data files continue to increase substantially, expanding scholarly output as such files are copied and analyzed for purposes secondary to those for which they were originally created; as they allow the gathering, manipulation, and analysis of data in ways not possible before; and as they allow the verification

of original research results by re-analysis of primary material such as field notes, interviews, etc. The implication of this forecast is, of course, the essential need to organize, store, and make universally accessible vast new amounts of machine-readable data. It makes no sense to create new systems of organization and standards of control of machine-readable information when the skills, standardized formats, and on-line catalog capacities are already available in our libraries.

- 2) The text editing capabilities will simplify the production of books and articles. The ease of producing scholarly reports may increase their production and radically change our current publishing structures.
- 3) The development of software may be considered part of the scholar's contribution to his or her field of research and may be disseminated as are other research results.
- 4) For the foreseeable future, the final results of scholarly research will continue to be published in monographs and traditional refereed journals. There is no indication at present that scholars are prepared to accept electronic formats as a substitute for printed materials. This means universities, through their libraries, will need to assume incremental, not substitutional costs, as they make materials available simultaneously in more than one format.
- 5) It is likely that well within five years, networks of scholars working within the same narrow research areas will establish and maintain files of full text articles, preprints, information networks for communication of research results or ideas, and bibliographic citations and abstracts relevant to their subject interests. Storage capacities, indexing / formatting softwares, and standardized compatible network links will be required to support such communication. The need is recognized, but the conceptual work remains to be done.

A major concern among scholars in all disciplines is the lack of good software. To our researchers today, the capacity to customize access is of paramount importance, and they stress that the available technology should be as adaptable and flexible as possible because use differs by discipline, by individual, and by the nature of the inquiry. Finally, our scholars emphasize the need to integrate all kinds of information on the electronic workstation: internal and external data bases, collegial internal and external communication networks, software, user assistance, consultation services, and the ability to share information easily. In sum, scholars and students want a variety of information and communication resources through one source and they want to be able to use it as easily as possible.

What are the financial and service implications of these new information habits for the

information providers? What kind of new infrastructure do we need? Some of the significant characteristics of the new environment are the following:

- 1) The book is no longer the sole means of communication from one mind to another. Nor will the computer replace the book. The diversity of scholarship—both among disciplines and within them—will require a mixture of formats and hardware for a very long time. The challenge for the next decade—and perhaps longer—is to provide the necessary linkages between formats. The range of expertise and hardware / software required to use machine readable information efficiently and effectively is now far greater than the simple ability to read.
- 2) The technological capacities, formerly controlled by the library and computer center have now moved into the hands of the user. Everyone now becomes a publisher, with extraordinary implications for our traditional control of incremental costs to the institution and our systems for access, retrieval, dissemination, and storage. The use of technology by faculty and students means that librarians and computer specialists will become involved in the research process at a much earlier stage as we help our colleagues define their intellectual inquiries within the new knowledge context of multiple information resources and technical capacities.
- 3) Our bibliographic and computer networks were designed to be mediated through centrally managed service organizations—libraries and computer centers—and their protocols reflect that design. Despite the sophisticated power of computers and telecommunication networks, the frustrations of incompatible hardware, software, and networks plus the lack of standardized protocols for organizing and retrieving information in machine-readable data bases lead many scholars to pick up the phone and call a colleague in their search for information. All too often, as one frustrated scientist reported, “Research today is 90% search and 10% research.”
- 4) There is more technology than we can afford. Our choices must be based on clear understandings of the institutional mission, informed recognition of the alternatives, and a thorough knowledge of the consequences of those choices to the quality of instruction and research. It is critical that we find a way to use productively the combined talents of scholars, librarians, technical specialists and institutional officers in planning for the future.
- 5) The tenure process in American universities appears likely to continue to drive scholarly output into greater fragmentation and over-publication which becomes economically impossible to support. The breakdown of the refereeing process has created insupport-

able costs for libraries. New text editing capabilities will simplify the production of books and articles and most certainly add to the publication pressure.

- 6) The increasing privatization of information resources and the direct marketing of knowledge bases to scholars and students threaten our traditional system of providing institutionally subsidized information through libraries. We must develop new strategies to insure equity of services among all disciplines and individual members of our constituencies.

The needs of the wired scholar-coordinated and subsidized gateway access to the universe of knowledge from an individual workstation-will require, in addition to the hardware, a dramatic reformation of our traditional information services.

Every institution has to design its own organizational structure to meet the needs of the new environment. At Columbia, our initial change has been to position ourselves for the future by removing the organizational barriers and obstacles to change by combining the Libraries and academic computing into one organization - the Scholarly Information Center - so that we can shape our response to the wired scholar's demand for information regardless of format.

The Scholarly Information Center is a logical, conceptual framework which seeks to coordinate and integrate the wide array of university talents-technical, disciplinary, and managerial-into a one-stop shopping service for the user. We have found that libraries-in contrast to computer centers-are extraordinarily user friendly to the discipline oriented researcher. At Columbia, we have an infrastructure of twenty-six subject-oriented libraries deployed throughout the university, in contrast to one centralized computer center, whose inhabitants have been traditionally isolated from close involvement in the academic enterprise. Seeking a user services consultant who has the relevant disciplinary background is a frustrating process, while it is a simple matter to walk into a departmental library and find the reference desk.

The first step we took was to combine the computer center User Services Group and the Libraries Services Group into an Academic Information Services Group under the direction of the former Director of Library Services. That move has been an extraordinary success, driven by the logic of the user's need for both disciplinary and technical assistance. Librarians are continuing to develop substantial technical expertise and the technical consultants are beginning to specialize in discipline-oriented technical requirements.

Some examples of our cooperative efforts to transform the library into an information center are the following:

- 1) The software problems of a Hebrew scholar who wanted to do linguistic mapping on a personal computer were solved by the Geology librarian and a technical consultant. The

Geology librarian's experience in dealing with the digital mapping problems of her geologists was the key to solving the software problems.

- 2) The Social Sciences librarian and the chief technical consultant are working with a political scientist to create an on-line data base of required readings for his graduate seminar. He teaches via electronic mail and wants the students to have access to their readings from their personal computers.
- 3) We are experimenting with a wide range of uses for CD-ROM technology, including both the purchase of commercial products and the development of a WORM (Write Once Read Many) data base of geological data by a joint team of geologists, librarians and technical specialists. This project is designed to evaluate whether this technology can alleviate data storage problems faced by researchers in many disciplines.
- 4) A cooperative planning effort of librarians and computer center specialists is exploring the development of CLIO, the on-line catalog, into a university information system with gateways to both local and external data bases and knowledge resources. The success of such a system will have considerable implications for central CPU capacity, operating systems, storage capacities, relational data base availability and expertise and networking requirements. The needs of the wired scholar require this kind of basic cooperation.

Our goal is to evolve toward a functional organization which would possibly be composed of the following divisions:

- Services, comprising all the public interfaces
- Systems of access, including retrieval systems such as cataloging, indexing and abstracting, software development, networks, collection development in all formats;
- Storage management including print, disks, tapes, etc;
- Support operations such as the machine room and the business functions of acquisitions;
- Applications, which we now call library systems and administrative data processing.

We hope this new infrastructure positions us to respond to the following issues, which I think form the Information Agenda for higher education for the next decade:

- 1) The archival obligations for scholarly information, regardless of format.
- 2) The introduction of high technology with its corollary built-in obsolescence.
- 3) Strategies for managing the magnitude of the capital costs required and control of expenditures, at the same time facilitating the kind of open, unobstructed scholarly inquiry necessary to the distinctive quality of the academic enterprise.
- 4) Equity of services in an environment of increasing privatization of information.

- 5) The integration of services offered through book and journal collections, mainframes, microcomputers, and internal and external networks.
- 6) The transition from collection-based services to access-based services.
- 7) The enormous cost and labor-intensive effort to convert the bibliographic records for our large research collections into machine readable format.
- 8) A national plan for preservation of deteriorating materials from the perspective of limited resources, both dollars and time and the provision of greatly enhanced access through technology.
- 9) Achievement of an effective electronic publishing system. There will be an enormous capital investment required to provide broad electronic access and until everyone has it, we will have to provide printed formats as well.
- 10) The capacity of higher education to develop information technology for scholarly applications. Universities are faced with unprecedented requirement of basic research in the very substance of their existence-information technology. In the past, the existence of absence of the capacity for cutting-edge research in a particular field threatened only the survival of a particular department or discipline rather than the institution itself. Now that information is valued as a commodity in the society, competition with the private sector is unquestionably beyond the financial capacity of individual institutions to support.

According to a traditional American, and possibly Japanese, cliché, the library is the heart of the university. I think it is time for a new metaphor – and that metaphor is more appropriately the DNA (deoxyribonucleic acid). The new process will be a helix – we provide a basic set of services and technical capacities, users interact and experiment with the new technical dimensions and develop new information requirements which then influence the evolution of a new shape for the infrastructure. As the genetic code of our societies, the character and quality of our information systems will determine the character and quality of our societies. And that is why it is so important that we re-invent the library in the electronic age from the imperatives of research and scholarship rather than impose upon our creative energies the artificial capacities of the technology.

付

- 1 翻訳「図書館、技術そして学術研究—情報センターとしての図書館」外山良子訳
- 2 日本学術振興会あて報告書（1987年7月2日付け）
- 3 センターニュース報告

付>1 第3～4回学術情報センターシンポジウム講演記録

図書館、技術そして学術研究 —情報センターとしての図書館—

Patricia M. Battin

Vice President for Information Services and University Librarian
Columbia University

訳 外山 良子*

はじめに

私を日本にお招き頂き、皆様とお話できる機会を与えて下さったことを、光栄に思い、日本学術振興会ならびに学術情報センターに感謝申し上げます。

今日のような電子化社会にあって、図書館及び職員役割をどのように変貌、発展させるかというテーマは、学術研究の将来、特にその質を考慮した場合、日米両国にとって共通の、しかも重要な課題であります。

いくつかの前提

学術研究の質は、次の二つの事項に関する大学側の総合能力如何にかかわっている。一つは伝統的な知識センターの保全能力、もう一つは情報技術を、知的に責任のある方法で利用し、管理するために必要なインフラストラクチャ（infrastructure=支持基盤）を形成する能力である。

今、アメリカでは図書館と計算機センターの統合化という現象が、大きな注目を浴びている。「計算機帝王」(computer czar) といった捉え方もあるほどである。これは、あまりに表面的な捉え方で、統合化の本質には触れていない。組織上の地盤闘争に焦点が向けられていて、統合化、それに付随する体質変貌の意図するところ、その目的などについての考慮に欠けている。この組織変形(transformation) が真に目標としている点、すなわち今後の研究及び教育を支えていくための、新しいインフラストラクチャが必要である。

アメリカの大学図書館の一員として見た場合、両国の文化の違いは、とりも直さず図書館における伝統的な物の見方、考え方、組織、経営面の違いに表われている。しかし、新しい技術を学術研究者が取り入れていく過程で、文化的、知的な学問という作業の基盤に、両国の文化の違いを越えた共通のチャレンジが訪れようとしている。すなわち新しいインフラストラクチャ発展の必要性である。

* Ryoko TOYAMA : Head, Access Services, Humanities and History, Columbia University

大学における一機能と形態そして運営

学術研究過程における図書館と技術の役割の変貌において、図書館の影が薄くなって来ているという見方もあるが、私はそれに反対である。ワイヤー（回線）で連結された大学機構の中で図書館の果たす役割は、逆に大きくなってきている。情報技術の発展は、知識とフォーマット（記録様式）の違い、本質である内容と手段の違いに関する、従来の知的区分をあいまいにしまった。このような現状において、図書館こそオリジナルな学術情報システムであるという事実を再認識しなければならない。新しい情報技術を取り入れるに当たって、システムそのものに集中するあまり、情報とそれを応用する人達の真の必要性を見落としてしまわないように心がけなければならない。

ある意味では、今日の図書館は歴史のパラドックスの興味ある一コマを演じているのではなかろうか？ 図書館は伝統と秩序というイメージを長い間持ち続けてきた。今、図書館は新しいチャレンジを受けて、農耕社会から工業社会を経て、一気に情報時代に入りつつあるのである。計算機と通信技術の影響は実に大きく、学術研究および大学教育界にこれまでになかった改革を引き起こすに至っている。

大学は、従来、図書館を通じて社会全体に貢献してきた。すなわち、知的遺産を手ごろな方法で流通し、政治、経済、文化に長く影響を及ぼしてきた。

新しい情報技術は、その電子改革の速さと、従来の感覚からすれば異常とも思える可能性の拡張、それに寄せられる一般大衆の熱意で知られている。しかし情報処理に関係する者として、もっとも大切な面である、図書館と情報技術との関係、役割、機能などを見落としてはならない。単なる情報と、知識と知恵に裏付けられた情報とは、教育という一つの過程にあって、はっきり区別されなければならないのである。

回線で連結された大学キャンパス（wired campus）では、回線を通じて送られてくる情報の量と処理に注意が向けられ、その内容と質は見落とされがちである。それと同じことは、図書館の日常業務にも見受けられる。例えば書物に対する盲目的な愛着のあまり、それを利用する人達への便宜、あるいは図書館業務のもっとも大切な対象すなわち知識の探求者を見落してしまうということはないだろうか？

アメリカでは、興味ある新しい現象が広がりつつある。大学におけるいわゆる「計算機帝王」すなわち学術情報サービス副学長という地位がそれである。新しい役職設置の前提として、大学では、学術情報サービスの運営管理に対する重要性が全面的に認識されるようになった事を指摘できる。しかしながら、残念なことに、図書館の責任、役割については言及されていない。計算機帝王達のほとんどが、ハードウェア出身の技術専門家で、科学以外の分野をあまり知らないというのが現状である。また、彼らのほとんどが書誌アクセス、情報探索に経験を持っていない。

歴史的に、大学は、教授、学生に対する学術情報資料の提供を、図書館の責任とみなしてきた。一旦、資料が大学予算で購入されると、それは大学に所属する全教授学生の共有資源、一種の財産になる。州立、公立大学の場合には、資源の共有範囲はさらに広がり、企業体も含めた地域社会をも包括

する。一切の経費は、大学予算で賄われ、利用者が資料利用の都度、支払うことはほとんどない。このように、教育機関が代表となって収集した資料を、利用者が無料で利用するというのが、研究教育機関一般のこれまでの傾向であった。

情報処理のための計算機関係の経費をどう賄うかに関しては、その対象範囲の広大さと進展のあまりの迅速さのために、今までのところ、暫定的な方法しかでき上がっていない。計算機利用の当初の目的が数理処受にあり、情報処理ではなかったせいもあるだろう。経費回収を個人に依存する基本戦略によってきた。情報交換のための計算機および通信技術の急速の普及は、学術研究者の間および社会一般に、従来の経費回収法を無意味なものとし、一種のカオスを生み出しつつある。もっとも気がかりな点は、情報処理のフォーマットにばかり注目が集まり、その役割が無視されそうな傾向である。図書館員と学術研究者は、情報過程を、印刷物を仲介として把えがちであるが、一方、計算機の専門家は、技術の能力そのものに注目しがちである。その間であって、利用者が本当に必要とするところ、また利用者の情報収集の習性などは無視されがちである。

デジタル化した情報は新しい生活習慣になりつつある。これは、いわゆるチャレンジとなっており、あまり速く進み過ぎて、そのインフラストラクチャとなる機構の形成が、未だになされていない。例えば新しい技術を取り入れた生活習慣を広い意味でどう定義づけるのか、どのように発展させるのか、運営に必要な経費をどう賄うのか。これらの問題が解決されない限りは、新しい技術を有効に応用することは無理だろう。ハードウェア、ソフトウェアの互換性、国際的レベルで広く学界で共有できるネットワーク化への出発点の設定、これらはすべて新しいインフラストラクチャに組み入れられるべき事項であり、組織上、政治上の壁である。情報技術固有の性格についても言及しなければならない。すなわち一方では、データの収集、保全、流通を大々的なスケールで可能にし、反面、情報資源の野放しの増大、先例のない情報アクセス手段の増加を甘受しなければならない事実である。料金制に依る情報入手が、新しい技術のもとに可能になったが、個々の学術研究者に与える影響や、大学が料金を立て替えるとすれば、どの程度まで可能か、真剣に考えてみる必要がある。

新しい技術を知的遺産と連結し、継承し、知的遺産への手ごろなアクセスを成功させるためには、情報をフォーマットではなく機能として把えるべきである。また、情報サービスやシステムが回線で連結されてはいても、情報利用者の個々のワークステーションから眺めた上で把えなければならない。回線連結プロセスと、いわゆる学術研究プロセスとは区別されなければならない。もっとも大切なことは、知識への探求心であって、技術の可能性ではないのである。知的プロセスが、外部への反応作用の中で育っていく以上、探求の対象も、技術の可能性の拡張に従って変化することになる。この展望の実現のためには、学界において、情報収集に関する概念上の規約を必要としている。どのような規約、規定と取り組まなければならないのかは、難しく、しかも重要なチャレンジである。政治、経済、文化のよりよい継承のために、研究機関は、ケーブルやワイヤーなどの付属物で、周囲をいっぱいにしてしまう以前に、この点と取り組まなければならない。

米国における研究情報ネットワーク

学術研究者の要求は、手近な障害物を常に乗り越えてきた。アメリカの研究図書館は分散形式をとっており、国立では議会図書館 (LC: Library of Congress)、医学図書館 (NLM: National Library of Medicine)、農学図書館 (NAL: National Agricultural Library) があり、その他には約 100 館にわたる公私立の研究図書館がある。

議会図書館は標準化された目録 (MARC) テープを国の内外に頒布しており、名前典拠 (LC Name Authority File) も、他の政府刊行物書誌、国立医学図書館書誌と同様に、磁気テープで頒布されている。目録作成管理機関として RLG (Research Libraries Group)、OCLC (Online Computer Library Center) そして WLN (Western Library Network) の三つのネットワークが活動しており、全米の研究図書館のほとんどが、いずれかに加入している。加入館は、ネットワークのホスト計算機に新しい目録レコードを入れ、他の加入館がすでに入力し作成したレコードを自館のシステムに導入したり、自館の所蔵情報をレコードに付与している。各々のネットワークのデータベースは、短期間に驚くばかりに膨張し、加入館は、データベースの拡大と保全の責任を全うすることで、目録作成費の節約に成功している。膨大なデータベースのおかげで、加入図書館間の相互貸し出し業務は、革命的に飛躍し、しかも分散している資料を総合的に眺めることで、ネットワーク加入館の間で相互に有益な総合資料収集方針を打ち出すことも可能になった。目録管理ネットワークのうち大きなもの、すなわち OCLC と RLIN を合わせて、現在、約三千万点の固有目録レコードがあり、その範囲も書籍の他に、地図、原稿、定期刊行物、メディア資料、楽譜、ソフトウェアに及んでいる。

民間財団である図書館振興財団 (CLR: Council on Library Resources) では、数年来、三大ネットワークの目録データベースを包括した上で、重複を避けた、固有データ群を抽出するという意欲的な計画に取り組んできた。重複レコードを避けるためには、典拠コントロールが必要になると同時に、全ネットワーク加入館の間に、一定の目録の様式、規定が必要となる。CLR の計画の最終目標は、広範囲に応用できるコストの安い共有データベースの作成である。全加入館が規定に沿って入力することにより、最終利用者は、個々のネットワークの特殊性に影響されずに、自由にデータベースを参照できる。

過去数年、LC、RLG、WLN、OCLC は、CLR の助成金をもとに、ISO (International Standards Organization=国際標準化機構) の規定に基づいた、目録ネットワークの連結作業を試みている。第一段階として、LC と RLG との間に、磁気テープの交換に依らない、データベース間の直接連結が始まっている。

RLG (RLIN システムの親機関) は、36 大学図書館の共同体で、共通問題の解決を、その目標として 1978 年に発足、現在に至っている。活動の重点として、加入館全体の学術研究者への情報サービス増進を常に計っている。RLG と OCLC の主な違いは、OCLC がネットワークのより広いマーケット化を目指しているのに対して、RLG は加入館がオーナーシップを共有し、共同運営に当り、共同事業の推進を目標としている点にある。

研究者の生活様式の変化と図書館の変化

現代アメリカの図書館界では、計算機世代を三つに区分している。過去 15 年間に第一世代と呼ぶ。第一世代では、計算機ならびに通信技術の応用は、目録の作成と探索作業に限られてきた。すなわち計算機を使って目録を印刷し、磁気テープから読み取ったデータを通信技術によって専用回線を通して加入館に配布したり、新しい目録を入力したりした。その結果、館内活動は、根本的に改革された。しかし、目録データそのものは、直接、利用者の手にはわたらず、参考業務（レファレンス）を通して探索する方法であった。技術環境は、全米を通じて画一的であり、専用回線を介して多数の端末とホスト計算機がつながるといった状況である。この世代は別名「図書館自動化」の時代とも呼ばれ、そして今、図書館自動化の時代は終わろうとしている。この世代に、学術研究者に有益なワークステーションを開発するための基礎的な準備は完了した。

第一世代からは、目録作成、資料収集、貸出業務などの機能をローカル・レベルで一括提供するためのソフトウェア、ハードウェアが商業的に開発された。続いて、LAN（Local Area Network）を通して、研究者の個人ワークステーションに目録情報が直接届くシステムが開発された。これが第二世代である。この世代では、大きいネットワークを経由した総合目録の自由探索、ローカル・データベースの確立が可能となったが、同時に、これらを円滑に連結する基礎となるべきソフトウェアの企画、開発、さらにシステムをワークステーションすなわち個々の利用者に近づけるための図書館職員的能力開発、新しいサービスに必要なコスト分担など、具体的な運営方針が必要になっている。

第三世代への準備が、現在、開始されている。すなわち応用ソフトウェアを研究者のワークステーションに移し、データを利用者各自の必要に応じてダウンロードし、再編成させる能力の実現である。1990 年代の始めまでには、各研究機関とも目録データを個人のワークステーションで再編利用するために、LAN の開発と LAN を交換台とする他のソフトウェアとの連結などの機能を完成させているだろう。結果として、利用者は、個々のワークステーションを出発点として、例えば、個々のオンライン目録が、大学全体の情報システムへと進化し、いっそう広い知識と資源につながることになる。

このようにローカル・レベルから外へ向かって広がっていくシステム開発と並行して、我々は、国内レベルの通信機能の拡張向上を目指している。すなわち個々のワークステーションで仕事をしているシステム利用者が、他の研究図書館の資料を自由に参照し、自分のデータベースに取り入れられるようなシステムの開発である。そのためには、何よりもインフラストラクチャの改善が必要となる。大きな研究図書館は、かつて産業勃興期に不可欠な機構として発達したハイウェイ、地下鉄、その他の公共機関と同様に、情報流通という役割を果している。交通機関は、利用者の生態変化、環境の全般的变化に対応して、相応の改善を重ねていかなければならない。

多くの側面において、現在の研究図書館は、利用者に負担の大きい、人力に頼り切った 19 世紀の情報システムを保持している。需要の 90% までは、19 世紀的な方法で満たされ、残りの 10% が、現に浸透しつつある 21 世紀の技術を応用した方法で満たされている。19 世紀と 21 世紀のサービス方法の共存が、多くの学術研究者を分裂症に近い状態に引き入れていることは事実である。彼らの多くは、

高度な電子化技術に基づく総合書誌探索を、伝統的な図書館サービスの延長上に求め、しかも最大の便宜性を要求している。技術的な可能性があるというだけで、利用者は、ともすると未だに学術文献の大半が伝統的な印刷物フォームである現状を見落としがちである。

我々は三つの基本的な方法で図書館業務と取り組んでいる。伝統的な人力に頼った活動、電子化技術を取り入れた画一的な活動、そしてプロジェクト方式による技術の応用である。このような状況にあって、普遍的なサービスができ、しかも技術上の知識と応用技術を持った人材の育成が、是非必要になってきた。十年後の図書館界では、伝統的な労力に基づくサービスを管理できる人材、新しいシステムをデザインするイマジネーションと創造力の豊かな人材、それらの人材を非常に複雑な機構の中に組み入れ、生かす才能を兼ね備えた人材が必要となるだろう。

新情報インフラストラクチャの構成要因

このような背景の下に、大学における新しい状況と研究情報の収集処理の習性の変化は、具体的などのようなインフラストラクチャを必要としているのか、それが国内、国外を通じて情報収集にどのような結果をもたらすのか、次のような予測が可能である。

- 1) 機械可読形式のデータ・ファイルの作成は確実に増大し、それを写し取り再編成し、分析する頻度も増える一方だろう。これは、従来にはない強力な方法であり、オリジナル・データに使われた一次資料の再検討をも可能にする。新しいアプローチが示唆するものは、膨大な機械可読形式のデータを収集整理し、保存し、広範囲にわたるアクセスを可能とする必要性である。すでに、オンライン目録も始動し、人材の訓練もしてある図書館では、改めて新しい組織や規定を作る必要はない。
- 2) 原稿編集能力は、本や論文の刊行を容易にし、その結果として学術論文は急激に増え、現在の出版流通機構そのものを変えていくだろう。
- 3) ソフトウェアの開発は、学術研究の貢献と見なされ、研究結果の一部として流通されるようになるかもしれない。
- 4) しかし当分の間は、研究発表の最終様式は、印刷物に限られ、電子化された内容がそのまま受け入れられることはないだろう。それはとりも直さず、大学が図書館に注入する経費の倍増化を意味している。研究成果の発表が、二重性を帯びる場合が予想されるからである。
- 5) 次の5年間に、同じ専門分野の研究者が、小論文の全文（フルテキスト）、概略、書誌、研究成果の要旨などを、一つのデータベースに入れて共有することに大きな可能性がある。しかしながら、今のところ、上記の事項を実現するための基盤はでき上がっていない。

学術分野を問わず、研究者の関心事は、良いソフトウェアの欠如にある。研究者にとって、固有の要求に合わせた情報入手方法、個人、専門分野の研究法の独自性に適応し順応できるシステムは欠かせないものである。また、研究者は、最終的には、様々な性格を持ったデータベース・ファイルを、個々の電子化されたワークステーション上でデジタル結合できる能力を要求している。要するに、

利用者側からは、データベース通信、標準フォーマットなどの、仕様の違いを無くし、大学の内部、外部、公的、私的を問わず、レファレンス・サービス機能も含めて、一括したシステムが望まれている。電子化されたワークステーションを利用している研究者や学生にとって、多様な情報を一つの窓口を通して収集、保存、整理、再編するシステムは、本当に必要なものである。

学術研究過程の理解

では、新しい情報利用習慣の、運営面、サービス面に与える影響はどのようなものだろうか？ 顕著なものを以下に挙げる。

- 1) 書籍は、もはや、広く思考を伝える唯一の手段ではなくなった。しかし、計算機が全面的に代行することも、近い将来にはないだろう。学問分野の多様性は存在し続け、情報管理方法にも、ハードウェア改善も含めた様々な要求を出し続けるだろう。次の十年間のチャレンジは、多様な情報のフォーマット間に必要な連結を提供することである。ハードウェア、ソフトウェアは益々複雑化し、それを利用して情報をより有効に収録するためには、単なる読み書き以上の能力が必要となるだろう。
- 2) かつて図書館や計算機センターの職員によって管理されていた技術は、利用者の手に移るだろう。誰でもが、容易に著者になることができ、簡単に発表し、それを流通する能力を得る。その結果、研究機関にかかる経費面で負担は増大するばかりだろう。教授、学生が技術を使い始めたことは、とりも直さず、図書館や計算機センターの職員が利用者の研究過程に早くから参加しなければならないことを意味している。
- 3) 従来、情報、計算機、ネットワークはサービス機関が仲介の労をとるという前提の下に育ってきた。プロトコルもこの点を反映している。計算機通信の高度化にもかかわらず、ハードウェア、ソフトウェアの非互換性は、学術研究者を混乱状況に陥れている。現状は、ある利用者がいみじくも言ったように「90%が情報収集作業、10%が研究作業そのもの」である。
- 4) 使いこなせない程、技術が氾濫している。その中から我々は、学術研究者と一緒に将来のための選択をしなければならない。選択の基礎として、研究諸機関の目的、技術を広い視野で比べること、研究、教育両面に与える影響の把握、その他の明瞭な認識が必要である。
- 5) 大学における永久雇用制度は、今後も続いて行くことだろう。結果として、判定過程に使われる専門化された論文出版は続き、新しい技術の導入で増大するかもしれない。
- 6) 図書館を通じて政府刊行物を広く利用者に流通するという伝統的な方法が変化しつつあり、政府刊行物配布過程の企業化が始まっている。図書館利用者のために専門分野を問わず、公平でしかも簡単な情報の入手方法、対策を立てなければならない。

回線で連結された学術研究者群に、個々のワークステーションから広い知識の海へ船出できるような機能的な港を開発して提供するためには、ハードウェアだけではなく、伝統的な組織の一大改革が必要である。

SIC—コロンビア大学における改革

研究教育機関は各自、目的と環境に合わせた組織を構成しなければならない。コロンビア大学では、将来のための必要性に応じて、図書館と計算機センターの境界線を取り除き、一括した学術情報センター（SIC: Scholarly Information Center）を創設した。

SIC は、論理的かつ概念的枠組であり、技術員、学術文献専門員を結合し、センターの利用者には、One Stop Shopping つまり一ヶ所で買物が間に合うようなサービス提供を目的としている。コロンビア大学図書館は、26 の専門分館から成り立っている。この分散形式のおかげで、構内細部にわたり、いろいろな面で緊密な接触を維持してきた。それに比べて計算機センターは一ヶ所にあり、大学の学部活動とはあまり縁が無く、従来は孤立していた。二つの性格の違う組織を統合することで、利用者は、最寄りのレファレンス・デスクに行くだけで、計算機関係の技術面での助言を受けることができるという仕組みである。

統合は、各組織の部分合併から始まった。第一歩として、計算機センターの利用者サービス・グループと図書館のそれが合併し、学術情報サービス・グループ（Academic Information Services Group = AISG）が形成され、図書館側から部長が任命された。これは、専門情報に関する助言と計算機利用者での技術的な助言を一ヶ所で得たいという利用者の要求に呼応した合併として、大きな成功をおさめた。新しい変化の下で、図書館の職員は益々、実質的な技術の習得に専念し、一方、技術畑の出身者は、専門情報処理に重点をおいた勉強にいそしんでいる。

以下、図書館が情報センターに変貌していく過程で、かつて二つであった部門間でどのような協力が行われているのか、具体例を挙げる。

- 1) ヘブライ語研究者がパソコンを使って言語分布図を作ろうとした。地理専門の図書館員と技術相談員がチームとなり、相談にのった。地理学に広く使われている計算機化された分布図形式についての図書館員の経験が鍵になり、チームは無事に、研究者に満足のいく問題の解決方法を提供できた。
- 2) 社会科学担当の図書館員と主任技術相談員が共同で、政治学教授の大学院生セミナーのための、必読小論文（指定図書）をまとめたデータベースを開発中である。この教授は、データベースを電子メールを介して、学生達に各自のマイコンで読み取らせたいのである。
- 3) CD-ROM の広い活用に努力している最中である。既存の商品（例えば専門辞書、百科辞典）の購入だけではなく、WORM（Write Once Read Many）を使って、地理学者と図書館員が共有データベースを開発中である。このプロジェクトの成果を見て、CD-ROM の新しい用途—資料収録方法—をさらに検討することができる。資料収納スペースの限界は、専門分野に関わりなく共通の問題であるから。
- 4) 図書館と計算機センターの職員は、組織統合が始まる以前から、CLIO（Columbia Libraries Information Online）プロジェクトを共同開発してきた。単なるオンライン目録としてではなく、CLIO を出発点として、他の学術データベースをアクセスできるように設計している。データ

ベースは商業的に既存のもの、構内用に開発されたものなど、種々である。このプロジェクトの成功は、二つの組織の有効な協力体制にかかっている。CPU、オペレーティング・システム、情報収集能力と限界、ふさわしいデータベースの有無、ひいては学内外とのネットワーキングなど、問題は多くしかも多様であるから。

さし当たって、組織上、到達しなければならない目標は、機能中心の組織再編成で、下記のような構想を掲げている。

- サービス部門： SIC 利用者との直接交渉のすべて。
- アクセス部門： 目録、索引作成、ソフトウェア開発、ネットワーク、資料収集など。
- 資料保全部門： フォーマットを問わず、収集した全資料の保全部門など。
- 業務管理部門： 機械室から経理、経営、一般事務。
- 応用部門： 図書館の計算機システム管理から給与計算、業務統計も含めて一般管理に関するデータ処理。

高等教育における情報課題

以上の新機構が、次の世代の教育過程が必要とする課題に取り組むために、十分、適応していることが望まれている。以下、課題のいくつかにふれる。

- 1) フォーマットに関わらない学術情報収集保存の義務。
- 2) 未解決問題も含めて、ハイ・テクの導入。
- 3) 一時投資および維持のための経費の確保対策。同時に学術研究の一般傾向を反映した細部規制のないオープンな研究助成金の確保。
- 4) 情報私有化の傾向にあって、サービスの公平化への努力（支払い能力に依らない情報アクセス方法）。
- 5) 単行本、雑誌、大型計算機、パソコン、構内外のネットワーク、それらを統合した上での情報サービスの強化。
- 6) 資料中心のサービスからアクセス中心のサービスへの移行。
- 7) 膨大な従来の目録を機械可読形式に換えるための費用と人材資源の確保。
- 8) 希少また貴重な学術資料保存とハイ・テクを応用したアクセス方法の開発。これは国家レベルで遂行されなければならないプロジェクトである。
- 9) 電子出版システムの開発。学術書、論文の全読者がワークステーションを持つに至るまで、このプロセス変化は、現状では二重負担であり、大変な投資を要する。
- 10) 大学教育機関が学術的応用を目的とした情報処理技術を開発する能力。

大学はかつて無かったほど、情報技術の基礎的開発という課題と取り組むことを余儀なくされている。かつて、ある専門分野の研究助成を多少削減しても、大学全体への影響は見られなかった。現在、情報は、社会的な立場からは、商品として見なされている。情報確保上の無能力は、直ちに、大学教

育の質につながり、ひいては大学の存在そのものをも脅かすことになる。情報収集への支援、助成を怠ることは、資金の裏付けのある企業に対して、大学の立場を弱めることになる。

「図書館は、大学の心臓である」というアメリカの古くからの慣用句がある。日本でも同様に言い伝えられているのかもしれないが。私はここに、新しい慣用句を紹介したい、「図書館は大学の DNA である」と。すなわち新しい情報処理過程は、複数機能の連鎖反応である。すなわち我々は、基本サービスと技術援助を提供し、利用者はそれらの応用範囲を広げ、さらに新しい情報と処理能力を要求し、その結果、インフラストラクチャそのものが変化していくという一連の過程が考えられる。社会の自然原則として、我々が接する情報の性格と質は、とりも直さず、社会そのものの性格と質を決定する。この原則を省みる時、電子社会にあって、何故、我々が図書館を再生させなければならないのか、お分かりいただけると思うのである。人工的な技術に使われるのではなく、新しい技術の利用に対する主導権を握って、生来の創造力を学術研究の進展に向けなければならない。

付記： 本稿は、Battin 女史の訪問調査計画の一環として、東京では日本学術会議（62年5月20日）ならびに京都では京都大学薬学部講堂（62年5月26日）にて開催した学術情報センター・シンポジウムにおける講演のうち、日本学術会議で通訳した外山良子氏（Head, Access Services, Humanities and History, Columbia University）の翻訳原稿をもとにしたものである。

付) 2

Patricia M. Battin コロンビア大学副学長・図書館長の外国人 研究員招へいに対する日本学術振興会宛報告書 1987年7月2日

アメリカの図書館人という観点からすれば、日本における図書館界はリーダーシップの深刻な危機に面していると言える。大学の伝統的な組織構造では、大学図書館長には教官が順送りに就任しているが、このやり方は才能と創造力のある人材をこの分野にリクルートする際の強力な抑圧要因となっている。加えて、学界 (academic circle) におけるこの職業の相対的に低い地位は、情報技術に支援された急速な変化が進行中の環境下で、強力なリーダーシップを発揮するための能力を弱体化するものとなっている。年功制の強さもこの職業における保守主義と図書館サービスの拡張を構想することへの非積極性 (reluctance) を助長している。

その結果、リーダーシップは技術企業の協力を得て工学系と計算機科学系の人々によって担われている。図書館のアプリケーションが、全国的な目録作成システムの開発を指向した米国とは対照的に、日本における開発は、組織内に限定された縦形であった。これら個々独立の、一般に非互換性のシステム群を水平的に連結するという NACSIS の目標は、研究者の情報要求にとっては望ましく、また先見の明のある理想的 (visionary) な対応策である。

大学図書館では、関心の多くが研究者、学生による技術の利用ならびにそれらの変貌が情報に対する要求条件に及ぼす影響などよりは、図書館業務の自動化に置かれている模様である。印刷資料の迅速な整理業務と貸出業務のために効果的、効率的な自動化システムが開発されてきた。次の段階の重点目標として学術研究への新技術の有効な応用にいっそうの関心を払うべきである。互換性のある標準規格ならびにプロトコルの重要性、データベースの共同構築、そして学術目的の技術創造的応用などに移行させる上で、NACSIS は主体的な役割を果たすことができるだろう。

新しい情報システムを構築する過程では、伝統的な学術情報システムすなわち図書館の機能を理解すること、そしてこの理解を将来設計に統合することが重要である。図書館の伝統的任務としては、知識の組織構造に関する理解、教育上の基礎的素養としてこの方面への関心が一般に欠けている社会科学、歴史学、人文科学ならびに科学研究者の情報要求に関する理解がある。新しい情報システムの設計においては、全ての学問分野の必要性を反映することが決定的に重要である。図書館界が伝統的に果たしてきた機能は、新しい情報システムの将来における成功にとって必須のものとなろう。目標は、私見では、電子化された大学における図書館職の概念を無視することではなく再定義することである。新しい図書館情報大学の教科課程はこの新しく発生しつつある課題を反映している。

新しいアプリケーションが開発され、ローカル・エリア・ネットワークが設置され、パーソナル・コンピュータの利用が増加するにつれて財政的問題が表面化する。技術の特性とりわけハードウェアとソフトウェアの両面における継続的な陳腐化と組み合わされた急速かつ継続的な変化は、伝統的な

大学の予算制度に対する新しい課題となっている。現在の能力が倍加し成熟するにつれて新しいデータベース、新しい技術的能力そして国際的ネットワークへのアクセスなどに対する利用者の要求の爆発的増大は、当然、発生することになる。NACSIS の目的を全うする過程で、避けることのできない財政に関する諸課題に、今、取り組み始めるのは決して時期早尚ではない。(文責：内藤衛亮)

付) 3

大学図書館における情報サービスの在り方—公開講演記録—*

パトリアバットン (Patricia M. Battin) 女史は、コロンビア大学の副学長であり図書館長である。昭和62年5月16日より6月2日まで日本学術振興会外国人研究員招へいにより、学術情報センターに初の外国人研究員として滞在し、大学図書館等の訪問調査と3回の公開講演をした。

女史は、8月より、コロンビア大学を離れ、米国図書館振興財団と関係の深い National Commission on Preservation and Access の代表として、新たな分野で活躍する予定であり、多忙なさなかに学術情報センターの招きに応じたのである。

東京では、センター学術情報サービスならびにネットワークの研究開発の現状と方向、コロンビア大学における図書館と計算機センターとの統合という、米国でも先端的な改革の意義などについて討議し、慶応義塾大学、東京大学、筑波大学、図書館情報大学などの図書館、富士通を訪問調査した。また、国立国会図書館では、講演と訪問調査、文部省国立大学附属図書館事務部課長会議、センター主催の公開講演会 (日本学術会議講堂) で講演した。

京都では、天理図書館、国立民族博物館、京都大学図書館を訪問調査したほか、センター主催の公開講演会 (京都大学薬学部講堂) で講演した。

講演の詳細は、いずれセンターの手によって改めて公開の予定であるが、以下あらましを紹介する。

大学における学術研究過程と情報—機能、形態、そして運営

大学では、情報技術の進展に伴って、研究者も学生も通信回線を経由した計算機利用から情報検索そして情報交換へと進化しつつあるが、その一方で、知識とその記録形態の違い、あるいは本質である内容と単なる手段の区別が、曖昧にされつつある。

大学は、図書館を通じて、社会全体の中で知識の生産・流過程に貢献するという基本的機能を果たしてきた。情報技術の進展のさ中であって、単なる知識やデータと知恵に裏付けられた情報とを、教育過程においては明確に区別すべきである。図書館が果たしてきた機能、大学内での役割は、変化する情報技術とは、別の次元にあり、情報を求めている研究者・学生との関係と情報内容こそが、重要なのである。情報の流れる過程にあって、技術と直結した物理的形態は、確かに激しい変貌を遂げつつあるが、この変化にふさわしい程に、情報の役割についての認識は深められてきたらうか。

研究者や学生の生活様式、研究活動の態様は、情報技術の進歩によって大きく変貌しつつあり、情報を提供するものにとっては、大きな挑戦となっている。この変化に対応するためには、新しいインフラストラクチャが必要なのである。これは、新しい交通手段として、例えば鉄道が導入された際に、

* 学術情報センターニュース No. 5 (1987.9) (再録)

生活様式が変化し、発展し、また、運営経費の捻出の問題が発生したことも明らかのように、情報の場合も、その基幹・基盤の整備と運営方式すなわち情報のインフラストラクチャが、今や大きな課題となっている。この問題は端末あるいはワークステーションの前にいる利用者の立場で、情報を形態に関わりなく、機能として捉えることが重要なのである。

米国における研究情報ネットワーク

米国では、議会図書館 (LC: Library of Congress) を中心に、研究図書館、大学図書館が全国的ネットワークとして研究図書館連合 (RLG: Research Libraries Group)、そして公共図書館、大学図書館を対象とする OCLC などが発展してきた。

図書館ネットワークで結び、目録データベースを構築することによって ILL サービスが実現した。分散したコレクションが、これらの機能によって全国的に調整された、ひとつのまとまりとしての意義を持つことになった。

図書館振興財団 (CLR: Council on Library Resources) は、これまでも図書館ネットワークの発展に主導性を発揮してきたが、最近では典拠コントロールの導入、徹底とネットワーク相互接続に力をそそいでいる。

研究者の生活様式の変化と図書館の変貌

研究者・学生がワークステーションから研究・学習活動をする時代になりつつある。学内のシステムが、これに対応し、そして学外の全世界的なネットワークとの接続を実現・充実していく過程で、図書館も進化しなければならない。問題は CO-ROM、オンラインの発達はめざましいものの、学術資料の大部分が、依然として伝統的な印刷形態であることである。

新情報インフラストラクチャ

研究活動が情報技術の進歩に支えられて変化しつつある。研究成果の発表が、機械可読形式となり、オンラインネットワークを利用する方式も充実しつつある。文書編集と簡便な印刷能力の高度化は学術成果の発表方法に新たな側面を付け加えた。

研究者・学生は自分のワークステーションから学内外、国内外の情報源と自由に円滑にコミュニケーションしたいのである。

学術研究過程の理解の重要性

学術コミュニケーション過程と研究者・学生の情報要求を理解することが、情報提供にたずさわるものにとっては重要である。情報伝達手段として書籍以外の形態 (format) が出現しているので、これらを統合するサービスが必要である。

図書館や計算センターの専門的能力とされていたものが、ワークステーションを手に入れた利用者

の側でも機能しつつある。情報資源と情報技術の新しい枠組を理解することは緊急の課題である。

SIS—コロンビア大学学術情報システム

SIS: Scholarly Information System は、コロンビア大学の図書館（26 館）と計算機センターを統合したもので、全学にサービス・ポイントのある図書館と一ヶ所にあった計算機センターとが、利用者に対するコンサルティング・サービスをいっそう緊密な形で提供している。

高等教育における情報課題

高等教育機関における情報サービスの課題として、当面、次のようなものがある。

- 1) フォーマットに関わりなく学術情報の蓄積・保存。
- 2) 先端技術の導入。
- 3) 運営資金の確保と支出管理戦略の確立。
- 4) 情報提供における平等の確保。
- 5) すべてのサービスの統合提供。
- 6) コレクションに基づくサービスから利用（アクセス）を基本とするサービスへの移行。
- 7) 適及変換資金の確保と労働集約的な努力の注入。
- 8) 劣化資料の保存と新技術による高度利用の提供。
- 9) 効率のよい電子出版システムの完成。
- 10) 情報技術の学術的応用にむけての全学的努力。

日本における訪問調査で得た印象

図書館のアプリケーションが、全国的な目録作成システムの開発を指向した米国とは対照的に、日本における開発は、組織内に限定された縦形であった。これら個々独立の、一般に非互換性のシステム群を水平的に連結するという NACSIS の目標は、研究者の情報要求にとっては、望ましく、また理想的（visionary）に見える程の対応策である。

日本の大学図書館では、関心の多くが、研究者、学生による技術の利用や、この変化しつつある利用が情報に対する要求条件に及ぼす影響などよりは、図書館業務の自動化に置かれている模様である。印刷資料の迅速な整理業務と貸出業務を支援するために、効果的、効率的な自動化システムが開発されてきた。次の段階の重点目標として、学術研究への新技術の有効な応用に、いっそう関心を払うべきである。互換性のある標準規格ならびにプロトコルの重要性、データベースの共同構築、そして学術的な目的で技術を創造的な応用などに移動させる上で、NACSIS は主体的な役割を果たすことができるだろう。

新しい情報システムを構築する過程では、伝統的な学術情報システムすなわち図書館の機能を理解すること、そしてこの理解を将来設計に統合することが重要である。図書館の伝統的任務としては、

知識の組織構造に関する理解、教育上の基礎的素養としてこの方面への関心が一般に欠けている社会科学、歴史学、人文科学ならびに科学研究者の情報要求に関する理解がある。新しい情報システムの設計においては、全ての学問分野の必要性を反映することが決定的に要要である。図書館界が伝統的に果たしてきた機能は、新しい情報システムの将来における成功にとって必須のものとなろう。目標は、私見では、電子化された大学における図書館職の概念を無視することではなく、再定義することである。新しい図書館情報大学の教科課程は、この新しく発生しつつある問題を反映している。

以上がパトリア・バッチェン女史の講演要旨と印象の骨子である。今回、初の外国人研究員として女史をセンターに向かえるに当たっては、日本学術振興会、文部省の後援のほか、国立国会図書館とアメリカ大使館のご好意が大きな支えとなったこと、貴重なサバティカルを女史の予定に合わせて来日されたコロンビア大学図書館人文歴史部アクセス課長外山良子氏の公開講演会における流麗な通訳が成功の大きな要因であったことを付記して謝意に代える。

学術情報センター日誌

昭和61年

- | | | | |
|--------|---------------|--------|------------|
| 12. 12 | 学術会議第5常置委員会見学 | 12. 20 | 課金委員会 |
| 12. 13 | ネットワーク委員会 | 12. 23 | 総合目録小委員会 |
| | | 12. 25 | ネットワーク小委員会 |

昭和62年

- | | | | |
|-------|--|--------------|--|
| 1. 28 | 総合目録小委員会 | 3. 24 | 橋爪宏達助教授帰国 |
| 2. 3 | 運営協議員会議 | 3. 28~ 4. 12 | 猪瀬所長、海外出張(アメリカ合衆国、フランス) |
| 2. 6 | 評議員会議 | 3. 29~ 4. 10 | 浅野教授、海外出張(中華人民共和国) |
| 2. 26 | 第2回シスポジウム各古屋大学にて開催 | 5. 10~ 5. 17 | 猪瀬所長、海外出張(アメリカ合衆国) |
| 3. 1 | 文部省科学研究費補助金・海外学術調査(タイ・インドネシア・ニュージーランド・オーストラリア)安達淳助教授出発 | 5. 18~ 6. 1 | 米国コロンビア大学副学長・図書館長バッチェン女史招聘 |
| 3. 2 | 医学関係データベース懇談会 | 5. 23 | 米国AT&Tベル研究所副所長 Penzias氏一行、来訪 |
| 3. 2 | 接続図書館説明会 | 6. 2 | 第3回運営協議員会議開催 |
| 3. 3 | 文部省科学研究費補助金・海外学術調査(シンガポール・中国)内藤衛亮助教授出発 | 6. 3 | 第3回評議員会議開催 |
| 3. 3 | 総合目録小委員会 | 6. 8 | 英国大使 Whitehead氏、英国公使 B. Hitch氏、Dr. Bradley氏、来訪 |
| 3. 5 | 橋爪宏達助教授、中国科学技術大学出張 | 6. 13~ 6. 20 | 猪瀬所長、海外出張(フランス) |
| 3. 11 | ネットワーク小委員会 | 6. 14~ 6. 21 | 内藤教授、海外出張(タイ、中華民国) |
| 3. 14 | ネットワーク委員会 | 6. 19 | 塩川文部大臣、来訪 |
| 3. 15 | 文部省科学研究費補助金・海外学術調査安達淳助教授、内藤衛亮助教授帰国 | 6. 25 | カナダ公使 Armand Blum氏ほか、来訪 |
| 3. 17 | データベース委員会 | 7. 1~ 7. 11 | 浅野教授、海外出張(アメリカ合衆国) |
| 3. 19 | 昭和61年度後期セミナー閉講式 | 7. 9 | カールスルーエ大学視察団来訪 |
| 3. 23 | 課金委員会 | | |
| 3. 24 | 総合目録委員会 | | |

- | | |
|--|--|
| <p>7. 15～ 7. 16 韓国学術振興財団 Mookun-Lee 氏ほか、来訪</p> <p>7. 15～ 7. 19 浅野教授、海外出張（カナダ）</p> <p>7. 26 英国キングストン・ポリテクニク John Lindsay 氏、来訪</p> <p>8. 1～ 8. 19 市川研究開発部長、海外研修旅行（ハンガリー、オーストリア、スイス、フランス）</p> <p>8. 25 総合目録委員会開催</p> <p>8. 25 韓国中央国立図書館 Won-Ho Jo 氏ほか来訪</p> <p>8. 25 米国科学技術参事官 Richard W. Getzinger 氏ほか来訪</p> <p>8. 29～ 9. 7 井上研究主幹、根岸教授、安達助教授、海外出張（イギリス）</p> <p>8. 30～ 9. 6 猪瀬所長、海外出張（イギリス・フランス）</p> <p>9. 21～ 9. 26 浅野教授、海外研修旅行（アメリカ合衆国）</p> <p>9. 26～10. 4 濱田教授、内藤教授、海外出張（西ドイツ）</p> <p>10. 3～10. 17 浅野教授、海外出張（フランス、西ドイツ、スペイン、イギリス、アメリカ合衆国）</p> | <p>10. 3～10. 18 根岸教授、小山助教授、森岡データベース課長、海外出張（アメリカ合衆国）</p> <p>10. 13 米国 AT&T ベル研究所所長 Powers 氏一行来訪</p> <p>10. 16 ワシントン大学教授 Mr. Karl Lo 氏ほか来訪</p> <p>10. 23～11. 5 猪瀬所長、海外出張（ユーゴスラビア、フランス、アメリカ合衆国）</p> <p>11. 9 プレハブ（別館）竣工式</p> <p>11. 18 米国 AT&T ベル研究所社長 Ross 氏来訪</p> <p>11. 20 ネットワーク委員会開催</p> <p>11. 26 データベース //</p> <p>11. 30～12. 6 猪瀬所長、海外出張（フランス）</p> <p>12. 7 欧州会議エネルギー・技術委員会委員長一行来訪</p> <p>12. 2～12. 17 浅野教授、大山教授、向井システム管理課長、海外出張（フランス、西ドイツ、イギリス、アメリカ合衆国）</p> <p>12. 8～12. 10 第1回学術情報ネットワーク国際会議開催</p> <p>12. 11 第1回 NACSIS 国際フォーラム開催</p> <p>12. 16～12. 18 第2回公開シンポジウム開催</p> |
|--|--|

昭和63年

- | | |
|--|---|
| <p>1. 14 課金委員会開催</p> <p>2. 3 ネットワーク委員会開催</p> <p>2. 9 国立国会図書館長来訪</p> <p>2. 24 評議員会議開催</p> <p>3. 1～ 3. 13 浅野教授、海外出張（アメリカ合衆国）</p> <p>3. 8～ 3. 18 宮澤助教授、海外出張（アメリ</p> | <p>カ合衆国）</p> <p>3. 22 猪瀬博日本放送協会放送文化賞受賞</p> <p>3. 22 総合目録委員会開催</p> <p>3. 22～ 3. 26 猪瀬所長、海外出張（フランス）</p> <p>3. 24 課金委員会開催</p> <p>3. 28 カナダ大使館 Broodfoot 氏ほか、来訪</p> |
|--|---|

5. 18 OECD、Bennett 氏、来訪
5. 24～ 5. 30 浅野教授、海外出張（イタリア）
5. 25～ 5. 29 山田研究開発部長、海外出張（シンガポール）
6. 1 ネットワーク委員会開催
6. 5～ 6. 11 小山助教授、海外出張（フィンランド）
6. 14 運営協議員会議開催
6. 16 評議員会議開催
6. 16 スウェーデン大使館 Hornmark 氏ほか、来訪
7. 12 米国国際貿易委員会（ITC）A. Hred E. Eckes 氏ほか、来訪
7. 18～ 7. 27 井上研究主幹、海外出張（アメリカ合衆国）
7. 22 参与会開催
7. 29 第3回公開シスポジウム開催
7. 30～ 8. 7 飯田助教授、大山助教授、済賀システム管理課課長補佐、海外出張、（オランダ、西ドイツ、連合王国）
8. 28～ 9. 25 猪瀬所長、海外出張（フィンランド、スウェーデン、デンマーク、連合王国）
8. 29～ 9. 9 山田研究開発部長、井上研究主幹、内藤教授、宮澤助教授、雨森総務課長、海外出張（オーストラリア、シンガポール）
9. 13 UCLA 図書館情報学部長 Hayes 氏来訪
9. 13～ 9. 19 浅野教授、海外出張（アメリカ）
9. 13～ 9. 22 橋爪助教授、海外出張（アメリカ）
9. 24～10. 9 田中事業部長、海外出張（アメリカ）
9. 25～10. 1 内藤教授、海外出張（タイ、中国）
9. 26 英国図書館日本情報サービスキング氏来訪
9. 26 英国スターリング大学副学長 Arthur 氏来訪
10. 1～10. 9 安達助教授、海外出張（アメリカ、カナダ）
10. 7～10. 23 根岸教授、海外出張（アメリカ）
10. 9～10. 16 猪瀬所長、海外出張（フランス）
10. 9～10. 15 山田研究開発部長、井上研究主幹、橋爪助教授、海外出張（韓国）
10. 23～11. 3 大山助教授、海外出張（インド、スリランカ、モルジブ）
10. 26～11. 3 桂助手、海外出張（マレーシア、インドネシア）
10. 31～11. 6 浅野教授、海外出張（アメリカ）
11. 5～11. 14 根岸教授、安達助教授、海外出張（イギリス、フランス）
11. 11 米国議会下院予算委員会 Lombard 氏来訪
11. 14 台湾 Academia SINICA 電子計算機センター長 Ching 氏来訪
11. 24 課金委員会開催
11. 27～12. 2 猪瀬所長、海外出張（西ドイツ）
11. 29～12. 4 根岸教授、海外出張（西ドイツ）
11. 30～12. 10 浅野教授、海外出張（オーストラリア）

11. 30 昭和 63 年度大学図書館職員講習生来訪

1. 7～ 1. 16 浅野教授、海外出張（アメリカ）

平成元年

1. 8～ 1. 13 猪瀬所長、海外出張（アメリカ）

1. 29～ 2. 4 浅野教授、海外出張（フィリピン）

1. 8～ 1. 15 小山助教授、安達助教授、桂助手、海外出張（アメリカ）

2. 7 米国議会図書館日本課長 Matumoto 氏来訪

1. 8～ 1. 19 橋爪助教授、海外出張（アメリカ）

2. 20 昭和 63 年度行政情報システム（GIS）に関する国際研修団来訪

1. 20 ロンドン大学図書館文書館情報学大学院 McIlwaine 氏来訪

2. 20 第 6 回学術情報センター・シンポジウム開催

1. 23 総合目録委員会開催

2. 22 データベース委員会開催

1. 28～ 2. 17 内藤教授、海外出張（イギリス、東ドイツ、イタリア）

2. 23 アジア経済研究所保延理事他来訪