

*LISA* データベースの特色と利用上の考慮事項

——索引ファイルの観点から——

The Characteristics of Search Keys for *LISA* Databases

細 野 公 男

*Kimio Hosono*

*Résumé*

*LISA* machine-readable tapes have been produced since 1976 and become an important database for the information retrieval systems which provide "library and information fields" with document information. *LISA* tapes are the replica of the printed-versions and they also have following unique access points;

- 1) classification codes, made of CRG classification schedule.
- 2) two kinds of uncontrolled keywords sets, one is classification headings and the other is the entries to be appeared in the printed alphabetical subject index.

These access points themselves are powerful tools as search keys for machine systems, but search methods will be different between machine and manual systems. Therefore, it is not wise to use them which are devised for the printed-versions in the machine system without any modification. For example, manual searches from multiple aspects can only be done in the alphabetical subject index. Classification codes and classification headings are made only for single aspect searches.

This paper analyzes three kinds of access points from the point of recall and precision, and presents access procedure from their constituents to the access points and reasons why they are valuable.

I. 序

II. *LISA*

A. 刊行の経緯

B. 構 成

C. 収録範囲とタイムラグ

III. *LISA* テープ

A. 構 成

---

細野公男：慶應義塾大学図書館・情報学科助教授

Kimio Hosono, Associate Professor, School of Library and Information Science, Keio University.

## LISA データベースの特色と利用上の考慮事項

### B. タグの種類

## IV. LISA テープ用索引ファイル

### A. アクセス・ポイント

### B. 索引ファイル

## V. 結 語

## I. 序

過去十数年間における機械可読 データ ベースの急増は、オンライン情報検索および情報伝達技術の進歩と相まって、情報提供活動に大規模な変革を生じさせつつある。

1960年代における MEDLARS テープの普及や1971年に稼動を開始した MEDLINE は、このような変革を触発した大きな要因の1つである。以来、データベースの数の増加は著しく、1977年時点では情報サービスに定常的に使用されるデータベースは500種を越えたとの推定がなされた。<sup>1)</sup>

現在、DIALOG, ORBIT, BRS, JOIS などの商用オンライン情報検索システムが興隆しており、その他様々なオンライン情報検索システムが稼動している。このようなシステム相互間には、コマンドや操作方法などに差異があり、複数のシステムを利用するにはその使い分けが必要である。従って、個々のデータベースの構造や IR システムの特性にかかわらず一定の標準的な共通コマンドや操作方法によって検索できれば便利である。ここに CONIT (Connector for Network Information Transfer) 思想<sup>2)</sup> やシソーラスにおける中間辞書の開発などの考えが生まれる所以がある。

しかし一方ではデータベースの構造は主題領域の特徴を充分反映すべきであり、またシステムのコマンドや操作手順もデータベースの特徴に合致させるべきとも考えられる。例えば、化学分野のデータベース CACON を収録する IR システムでは、キーワードや著名名など通常の検索キーの他に、セクション・サブセクション番号や化合物の登録番号などからも検索ができれば非常に便利である。TOOL-IR では X線結晶学データベースの検索には、化合物名検索のための特別なコマンドとして、COMPOUND と FRAGMENT が用意されている。<sup>3)</sup>

さらに、図書館・情報学分野のように主題範囲が流動的で、周辺分野との境界線および周辺分野の領域の設定が困難なところでは、データベースに収録される一次情報および検索キーの形式は、安定した主題領域を対象と

したデータベースの場合とは異なるであろう。(例えば、流動的な分野では索引語は統制語いにもとづくよりも自由キーワード方式に従う方が得策であろう。)

従って、現在の段階では IR システムは、その IR システムで使用されるデータベースの特徴と密接に関連したファイル構成、検索機能、操作手順を持つことが、質の高い情報検索サービスを提供するために必要である。

本稿では、図書館・情報学分野のデータベースである *Library and Information Science Abstracts* (LISA と略称する) をとりあげ、その特色を紹介し、上記の観点に立ち利用の側面から LISA をデータベースとする IR システムの索引ファイルについての考察を行う。

## II. LISA

### A. 刊行の経緯<sup>4),5),6)</sup>

LISA は、Library Association (LA) から刊行されていた *Library Science Abstracts* (LSA, 季刊) の後身である。

1940年代の終りに R.N. Lock と C. Muriel Lock は、おもに図書館学を学ぶ学生を対象にしてタイプ印刷の小規模な抄録サービスを始めたが、このサービスの対象を図書館現場まで拡大できるようにするため、1950年に LA が Lock 夫妻を編者としこの抄録サービスを発展させた正規の抄録誌を刊行した。これが LSA である。1951年には H. Allan Whatley が編集を引継ぎ1968年までその職をつとめた。抄録の編集作業は当初から編集者や協力者の本業の片手間に行われてきたが、1960年代の後半になるとこのような処理方法では情報量の増大に対処しえなくなり、専任のスタッフが必要となってきた。

1960年代は図書館界を取り巻く環境にも大きな変化が生じた時代であった。情報科学が興隆し、旧来の図書館学分野にも経営管理技術、コンピュータ、利用者研究などが導入され、図書館学自体も変化した。そしてこのような時代の傾向を反映して新しい抄録誌が生まれた。例えば、1963年創刊の *Informatics Abstracts* (*Referativnyi Zhurnal*; *Informatika* の英語版)、1966年創刊

の *Documentation Abstracts (Information Abstracts* の前身) などである。同じ頃 Aslib が情報科学を対象とした抄録誌の刊行を検討していた。

かくして情報量の増大と主題領域の拡大は、LSA の編集・刊行作業の変更を余儀なくした。LA は 1968年に Aslib との協力関係を樹立し Aslib から情報科学分野の抄録の提供をうけるなどして対象主題領域の拡大をはかり、図書館学だけでなく情報科学ならびに複写、出版などの周辺分野の文献をも抄録の対象とするように編集方針を変更した。この変更に伴い編集責任者が Tom Edwards にかわり名称も 1969年 1/2 月号から *LISA* と改められ隔月刊で刊行されるようになった。

手作業で刊行されていた *LISA* は、収録される抄録数の増大 (例えば、1973年は約3000件、1974年は約3700件、現在は約4000件) につれ、年間索引を中心に刊行の遅れが顕著になってきた。これは1974年から国際色を深めたため (抄録対象の地域的範囲を拡大)、データ処理量が増大したことによる。この変更は購読価格にも影響を与え、1950年以來継承されていた個人でも購入可能な価格政策が放棄された。また、国際化を反映し英国外からの購読が増加し、例えば1950年中期の *LSA* では購読者の45%であったのに対し、1970年代中期の *LISA* では80%となっている。

LA は収録方針の変更により生じた情報提供の遅れに対処するために、同時期に *LISA* 刊行の機械化を検討した。その結果 INSPEC の協力のもと1976年1/2月号から機械編さんが実施され、現在に到っている。刊行の機械化に伴い副産物として生まれる機械可読版の *LISA* 磁気テープは、Learned Information から頒布されており、現在印刷版に対応する隔月刊の新規テープと、1969年までさかのぼる累積テープとが用意されている。

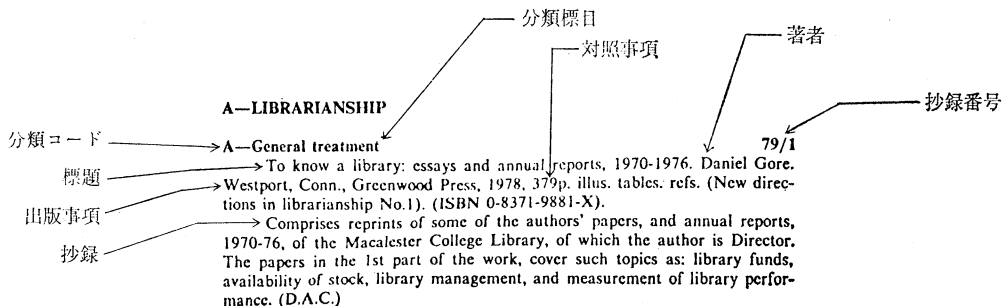
## B. 構成<sup>7),8)</sup>

*LISA* の印刷版は、隔月に刊行される抄録誌本体と年間索引とから構成され、抄録誌本体は、記事、著者索引、件名索引から構成されている。各記事は Classification Research Group (CRG) で開発された図書館・情報学およびその周辺分野を対象としたファセット分類表、*A Classification of Library and Information Science* の分類記号にもとづいて配列されている。各記事には第1図で示すように、CRG 分類表にもとづく分類コード、分類標目、抄録番号、標題、著者、出版事項 (あるいは掲載誌名・巻号表示)、対照事項、抄録が記載されている。もし記事が会報のような合集である場合には、記事全体としての抄録の後に各章の抄録番号が表示されている。

*LISA* を使用した文献検索では CRG 分類表にもとづき該当する分類記号の項で記事を逐次通覧する方法があるが、このような一面的な検索の非能率を補うために件名索引 (alphabetical subject index) と著者索引があり、どちらも年間索引に累積される。

*LISA* が使用している CRG 分類表は、未公開の1971年版であり、概略は印刷版各号の冒頭に示されている。この分類表では、図書館・情報学分野の重要な主題をアルファベットの大小文字で、周辺分野の主題を数字で表現している。しかし図書館・情報学固有の分野でも場所の共通細目は DC の地理記号、利用者の専門分野あるいは図書資料の主題分野は UDC 番号を使用している。

分類コード、分類標目、件名の付与は、抄録の主題分析にもとづいて行われる。主題分析の結果は例えば、"Planning the building of the teacher college library at Cardiff College of Education" のように句形式で表現される。次に主題構成要素に CRG 分類表中の



第1図 *LISA* 印刷版におけるレコード(記事)の構成

適切な記号が割り当てられる, 上例では, Qi (buildings), Nbm (planning), Gk (teacher college library), D 29 (場所の共通細目, Cardiff) となる。これら記号を一定の順序 (citation order, 結合順序) で結合して分類コードが作成されることになるが, LISA が採用している結合順序は以下の様である。

- (1) S/Z 処理, 作業
- (2) N/R 組織, 管理, サービス, 資源
- (3) J/M 資料, 資料・図書館の利用
- (4) E/H 図書館, 利用者
- (5) Ab/D 共通細目 (場所, 時, 主題, 形式)

従って上の例では分類コードは QiNbmGkD27 となる。なお1969~70年とそれ以後では使用する CRG の分類表に違いがあるので, 分類コードから検索する場合には, 1969~70年の LISA はそれ以後のものとは異なった出版物と考えた方がよい。

一方分類標目は, 主題から抽出された重要語の集合であり, 上の例では Buildings. Planning. Teacher college libraries. Cardiff College of Education (UK) となる。

件名索引に収録される件名も抄録の主題分析の結果得られたキーワード句から作られる。従って表現される概念は分類標目とほぼ同様であるが, chain indexing 手法にのって作成される。

一般に chain indexing では, まず本文の見出し (分類コード) と逆の結合順序で重要語を結合させて件名を作成し, 以下順次最左翼の重要語を落してさらに件名を作成していく方法をとる。しかし LISA ではこの方法で機械的に作成するのではなく, 作成された件名が検索キーとしての効果を高めるように, また重要語の結合順序がなるべく自然の順序に従うように修正がほどこされている。また, 補足的な重要語をつけ加え, より多くの検索キー (件名) を提供する工夫がなされている。上例での件名は以下のようになる。

Cardiff College of Education (UK): Teacher college libraries: Planning: Buildings  
 Teacher college libraries: Planning: Buildings  
 Planning: Buildings: Libraries  
 Buildings: Libraries  
 Library buildings  
 Architecture: Library buildings

なお, 件名索引には件名の他に, 必要に応じて“を見よ参照”, “をも見よ参照”などがなされている。

[例] Bookmobiles. See Mobile libraries  
 A.L.A. See American Library Association  
 Indexing: Subject See Subject indexing  
 Public libraries See also County libraries,  
 District libraries, Municipal libraries,  
 Rural libraries

ところで件名では複合主題は基本的な主題のコロン結合で示されることが多いため, 意味のあいまいな件名が生じる。例えば, Staff: Libraries は library staff なのか staff library なのか判断できない。このようなあいまいさを避けるため LISA では一種のロールを採用して, 次のように処理している。

[例] Staff: libraries: (Type of library)

Staff: Library staff

#### C. 収録範囲とタイムラグ

図書館・情報学分野をカバーする二次資料として LISA とよく比較されるものに,

- (i) *Informatin Science Abstracts (ISA)*
- (ii) *Bulletin Signalétique ; Science de l'information documentation (BS)*
- (iii) *Referativnyi Zhurnal Informatika (RZI)*
- (iv) *Computer and Control Abstracts (CCA)*
- (v) *Library Literature (LL)*

があり, Gilchrist 等 (1971)<sup>9)</sup>, Dansey (1973)<sup>10)</sup>, Edwards (1976)<sup>11)</sup> などによってこれらの二次資料の収録範囲, タイムラグの調査が行われている。

#### 1. 収録範囲

LISA が対象とする主題領域は, 図書館・情報学全般および出版, 複写, 書籍販売, 書誌, 印刷, 読書などの周辺分野であるが, 索引誌とは異なり収録雑誌の記事を網羅的に採録するわけではない。例えば, ニュース, 特定主題の書誌, 1頁以下の記事, 編集者への手紙などは収録されない。<sup>12), 13)</sup> とくに創刊当時は採録数が2567件<sup>14)</sup>と現在に比べ少なかったため, 採録もれに対して具体的な指摘がなされている。<sup>15)</sup>

LISA, ISA, BS, RZI, CCA の1971年版を対象に Dansey が行った主題領域の分析では, LISA は他の二次資料に比べ図書館学関係の雑誌からの記事の採扱が情報科学・ドキュメンテーション関係のものよりも多く図書館学分野が強い反面, コンピュータ科学, 複写, 経営科学などの周辺分野が相対的に弱いことが指摘されている。Edwards の調査は上記の5誌に LL を加えた6誌の1973年版を対象に Dansey と同様な分析を行った

ものだが、その調査でも *LISA* は *LL* と共に図書館学が強いことが示されている。

*LISA* を含む二次資料間での収録雑誌の重複量についても調査がなされている。Gilchrist 等が行なった *LISA*, *ISA*, *RZI* (英語版) の1969年版を対象にした調査では、3誌全てに採録された雑誌数は51誌であった。一方 Dansey の調査 (4誌を対象) では40誌, Edwards の調査 (6誌を対象) では13誌が共通であった。

資料の形態別では、*LISA* は逐次刊行物からの記事の収録はよいが、会議資料、レポートなどの非逐次刊行物の収録状況は充分でないとの結果が Dansey, Edwards の調査で示されている。なお、Edwards<sup>16)</sup> によると1973年および74年版での非逐次刊行物の割合は、それぞれ13%, 17%であり創刊以来あまり変化はない。

2. タイムラグ

Turtle<sup>17)</sup> は文献が公刊された期日とそれが *LISA* あるいは *LL* に収録された期日との時間的ずれを月単位で調査した。調査対象は1972年版から抽出された318文献 (*LISA*) と1607文献 (*LL*) である。*LISA* は抄録誌であるから *LL* よりも1文献あたりの処理時間は長くなるであろうが、これは *LL* の収録範囲の広さ、収録数の多さで相殺されると仮定している。結果は *LISA* の場合タイムラグの平均、モード、中央値はそれぞれ3.8ヶ月、2.0ヶ月、2.8ヶ月であり、*LL* では6.7ヶ月、4.0ヶ月、4.7ヶ月であり、*LISA* の方が短くなっている。また地域別

第1表 *LISA* および *LL* 地域別タイムラグ (Turtle と Robinson の調査による) 単位: 月

	北米	英国	オーストラリア・ ニュージーランド	ヨーロッパ	アジア	アフリカ	(A)
<i>LISA</i>	3.6	2.4	1.9	4.4	8.1	3.6	3.8
<i>LL</i>	4.5	5.7	6.6	19.0	8.2	11.8	6.7
(B)	4.4	4.9	5.7	14.7	8.1	10.7	(C) 6.2

注 (A)地域差を考慮しない場合のタイムラグ (B) *LISA* と *LL* をまとめた地域別タイムラグ (C)誌別、地域差を考慮しない場合のタイムラグ

の平均タイムラグも第1表のように *LISA* が短い。

一方、Gilchrist 等は、1969年の *LISA*, *ISA*, *RZI* に収録された16誌とこの3種の二次資料が Aslib library に受け入れられた期日を調べ、原文献とその抄録とのタイムラグを求めた。それによると *LISA* 99日 (対象は115文献), *ISA* 427日 (332文献), *RZI* 165日 (182文献) となっている。同じ調査方法で Dansey, Edwards もタイムラグを求めている。これらの結果を Turtleの

第2表 図書館・情報学関係二次資料のタイムラグ 単位: 日

二次資料名	1969 (Gilchrist)	1971 (Dansey)	1972 (Turtle)	1973 (Edwards)
<i>LISA</i>	99	99	114	144
<i>LL</i>	—	—	201	282
<i>ISA</i>	427	382	—	282
<i>RZI</i>	165	194	—	223
<i>BS</i>	—	124	—	182
<i>CCA</i>	—	86	—	97

調査によるものも含めて示したのが第2表であり、*LISA* の速報性は他の二次資料よりは高いといえよう。

III. *LISA* テープ

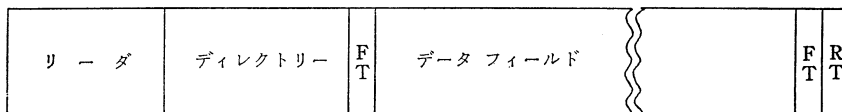
A. 構成

*LISA* 磁気テープは印刷版の機械編さんに伴う副産物として1976年に生まれたもので、印刷版の各号に対応し年6本のテープが頒布されている。

*LISA* テープでは各記事に対応する各論理的レコードは、1ブロック2000バイト (文字) の物理的レコードで

第3表 レコード長の分布 単位はバイト (文字)

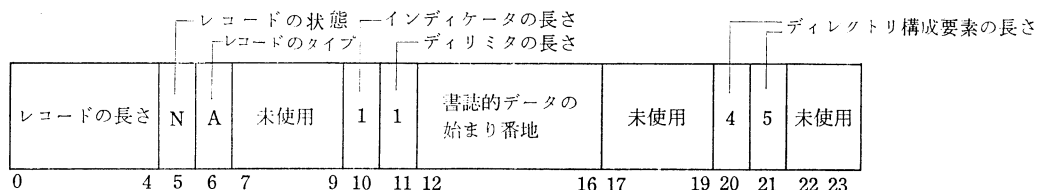
レコード長	頻度	レコード長	頻度	レコード長	頻度	レコード長	頻度
—399	1	800—999	53	1400—1599	116	2000—2199	21
400—599	22	1000—1199	79	1600—1799	60	2200—	12
600—799	22	1200—1399	134	1800—1999	31		
最短長	373	最大長	2912	平均長	1353		



注: FT はフィールドターミネータ, RT はレコードターミネータを表す

第2図 *LISA* レコードの構造

LISA データベースの特色と利用上の考慮事項



第3図 リーダの構造

表現されている。論理的レコードの長さが2000バイトを越える場合は、複数個のブロックで表わされるが、論理的レコードの最大長は6200バイトに制限されている。第3表は1979年1/2月号におけるレコードのレコード長を示したものである。

記録形式は MARC フォーマットとほぼ同様であり、ISO 2709「書誌情報交換用記録形式」に従っている。第2図はレコードの全体、第3図はリーダーの構造を示したものである。

リーダー部分のレコードの状態とタイプは当分の間常に N (新規レコード) と A (印刷物) である。またインディケータおよびデリミタの長さは MARC (いずれも2字) と異なりいずれも1字である。ディレクトリは各レコードを構成するフィールドの種類 (タグ)、長さ、始まり番地を示すもので、MARC と同様にそれぞれ3桁、4桁、5桁の数字で表現されている。しかし使用されているタグ番号は MARC とは異なっている。

LISA テープで使用されている文字セットには、大文字と基本的な記号だけを含む交換用文字セットと、大小文字、イタリック、ボールドフェイスなどの識別記号、さらに特殊記号をも含む完全文字セットとがある。完全文字セットで作られるテープは印刷版作成用であるため、機械検索に使用する際は識別記号の削除などデータの修正が必要となる場合がある。

**B. タグの種類**

タグの種類とその内容は、1976年に Learned Information から発行されたマニュアル<sup>18)</sup> に示されているが、以後変更がなされた箇所もあるので、1979年1/2月号のプリントアウトを参考に以下に言及する。

- 001 コントロール番号: タグ 300 と同じ内容で印刷版の抄録番号に相当する。西暦の後2桁と一連番号で表わす。
- 100 標題 (英語): 標題が英文でない場合は英訳標題。
- 101 標題 (原語): 非アルファベットの場合は作成

されない。

- 110 抄録本体: 当レコードが副出である場合には抄録はなく、このレコードの抄録が収録されているレコードへの参照が示される。

例: FOR ABSTRACT SEE ENTRY AT CLASSIFICATION NUMBER HdAw

- 111 抄録の補足: 学会報や論文集のような合集の場合は、合集全体だけでなく個々の論文も抄録の対象となるが、その際の補足的なデータを収録する。

例: THE FOLLOWING ARTICLES HAVE NOT ABSTRACTED .....

- 120 分類コード: CRG 1971年版にもとづく分類記号。

- 121 分類標目: 分類コードにほぼ対応し各記事の見出しとなるキーワード。印刷版では分類コードの後にダッシュをはさんで表示されている。

- 130 件名: 印刷版の件名索引に収録されるキーワード。補足として使用されるキーワードはカッコで示される。

例: WRITING (OF HISTORY: OF LIBRARIANSHIP) HISTORY (OF LIBRARIANSHIP) HISTORIOGRAPHY (OF LIBRARIANSHIP)

- 133 副出指示: 当レコードが副出である場合には1が与えられている。

- 150 掲載誌名, 合集名, 収録書名:

- 160 言語:

- 200 著者:

- 202 団体著者, 補足: タグ 200 に収録された著者名の変形もここで示される。

例: CORRY, E. LOWRY— (200ではCORRY, ELIZABETH となっている)

- 210 編者: 合集の場合はここに収録される。

- 300 抄録番号: コントロール番号と同じ。

- 301 サブレコード指示: 合集の場合個々の文献の抄

録も作成されるが、抄録された文献の一連番号  
(コントロール番号の西暦を落したものを)を示す。

例: 102/190/201/272

400 書誌事項: 雑誌論文では巻号, 年, 頁など, 単  
行書や合集などでは出版事項・対照事項を示す。

#### IV. LISA テープ用索引ファイル

機械検索では普通, 検索時間を短縮するために検索キー  
ーとしての利用頻度が高いと思われる項目をレコードか  
らとり出し, 検索用の索引ファイルが作られる。その際,  
効果的にデータファイルへアクセスできるように, デー  
タベースあるいはそのレコードの特色をよく考慮して索  
引ファイルを作ることが望ましい。

##### A. アクセス・ポイント

LISA テープをデータベースとする IR システムでは  
索引ファイルの種類として考えられるものとして, 著者  
(200, 202, 210), 分類コード (120), 分類標目 (121),  
件名 (130) がある。

著者には, 個人著者, 団体著者, 編者, 会議名などが  
含まれ, 重要なアクセス・ポイントである。

分類コードは, LISA では唯一の統制された検索キー  
である。ファセット分類表にもとづいて作成されるた  
め, コードは組合せによって作られた最終の複合概念だ  
けでなく, その構成概念をも明瞭に示している。例えば  
分類コード JaBge の Ja は図書館利用, Bge はリサー  
チを意味する。従って分類コードではその構成要素もま  
たアクセス・ポイントとすることができる。

分類標目は分類コードの自然語訳ともいえるが, 必ず  
しも対応しているわけではない。例えば, Bvr D437—  
Czechoslovakia のように Bvr (図書館学教育) に関す  
るキーワードがない場合がある。しかし一方では比較的  
多くのキーワードを付与して分類コードよりも詳細に主  
題を表現している場合もある。

〔例〕 GfD485—Technological University Libraries.  
Sweden. Chalmers Tekniska Hoegskala. Lib-  
rary.

分類標目には個人名, 機関名など固有名も多く使用され  
ており, 特定の検索に有効である。分類コードにも  
Bko (LA), Bji (STE) における LA (Library Associa-  
tion), STE (Steinberg, Heinz 人名) のように固有名が  
使用されるが, その数は分類標目に比較して少ない。な  
おこれらの分類コードに対応してその完全名が分類標目  
に表示されている。第4図は1979年1/2月号における分

FOLKERBIBLIOTEKERNES AUTOMATION SYSTEM (FAUST)  
FOOD TECHNOLOGY  
FOR PATENTS  
FORMATS  
FORMS  
FORTHCOMING EVENTS BULLETINS  
FRANCE  
FRANKFURT BOOK FAIR (1978)  
FREEDOM OF INFORMATION ACT AND PRIVACY ACT  
FREQUENCY DISTRIBUTIONS  
FUNEN  
FURTHER EDUCATION  
FUTURE DEVELOPMENTS  
FUTURE EVENTS  
GEDANKEN INDEXING  
GENERAL TREATMENT  
GEODARCHIVE  
GEOGRAPHY  
GIDEP (GOVERNMENT-INDUSTRY DATA EXCHANGE PROGRAM)  
GIMSØYA  
GLADSAXE  
GLASGOW UNIVERSITY  
GOETEBORGS UNIVERSITET  
GOFFMAN'S TRANSITION OF WORK OCCURRENCES  
GOTHENBURG  
GOVERNMENT DEPARTMENT LIBRARIES  
GOVERNMENT PLANNING  
GOVERNMENT PUBLICATIONS  
GOVERNMENT STATISTICAL SERVICE  
GOVERNMENT STATISTICS  
GRADUATE LIBRARY SCHOOLS  
GRADUATE SCHOOL OF SOCIAL WORK INVESTIGATION  
GRADUATES  
GRÖNNINGEN  
GUIDES AND HANDBOOKS  
HEALTH AND SAFETY EXECUTIVE  
HEALTH AND SAFETY LABORATORIES  
HEALTH EDUCATION  
HEERENLOO INSTITUTE  
HEILOO  
HIGHER EDUCATION LIBRARIES  
HISTORICAL ABSTRACTS AND AMERICA: HISTORY AND LIFE  
HISTORY  
HOECHST AG  
HOLOTHENE CONCEPT  
HOSPITAL LIBRARIES  
HOSPITAL PATIENT LIBRARIES  
HUMAN RIGHTS  
HUMBERSIDE AUDIO VISUAL CATALOGUE  
HYBRID ACCESS METHOD  
IBM STAIRS  
ICELAND  
ILLINOIS UNIVERSITY  
IMPACT MEASUREMENT  
IMPLICATIONS FOR DERWENT PUBLICATIONS LTD  
IMPLICATIONS OF PUBLISHING AND BOOKSELLING  
IMPLICATIONS OF LAW  
IMPLICATIONS OF READING  
IN-LINE INFORMATION RETRIEVAL  
IN-SERVICE TRAINING

第4図 分類標目の例

類標目の一例である。

件名は印刷版の件名索引に収録されるキーワードであ  
るため, LISA 件名索引の大きな特色である chain  
indexing 手法によって作られている。従ってその数は  
分類標目中のキーワード(句)よりもはるかに多く, キ  
ーワードの形, 主題の表現方法は, 分類標目の場合とは  
異なっている。

〔例〕 分類標目 History. Writing. USA

件名 Writing (of History: of Librarianship)  
History (of Librarianship)  
Historiography (of Librarianship)

## LISA データベースの特色と利用上の考慮事項

この例から明らかなように、分類標目が主題を表わす構成概念の単なる列挙である場合が多いのに対し、件名では前置詞やカッコを使用して構成要素にロール機能を持たせ主題の文脈がわかるようになっている。これはLISA テープが印刷版のレプリカであるためである。つまり印刷版ではレコードがファセット分類にもとづくカテゴリ順に配列されており、分類標目は主題の文脈が明示されている分類コードに引続いて表示されている。(第1図参照)従って、分類標目にはカテゴリを示すキーワードを必要としないし、また主題の文脈を考慮する必要もない。前例の分類標目に Librarianship がない

のはこの理由による。一方件名はレコードの配列とは全く独立であるから、構成概念を網羅しかつ主題の文脈を表示することになる。

索引ファイルはレコードの配列順序と全く独立に作成されるため、機械検索では手検索の場合と異なり、レコードの上述した配列順序は検索時に何んらの情報も与えない。この点から“ことば”を使用したアクセス・ポイントとして件名も有効である。

件名の第2の特色は、国名・地名以外の固有名が使用されていない点であり、また分類標目として固有名が使用されている場合には、件名が付与されないことがある。

FINANCE (INFLUENCE ON CULTURAL USE OF MATERIALS; LIBRARIES)  
FINGERPRINT TECHNIQUE (PRE-1801 BOOKS; COMPUTERISED PRINTED UNION CATALOGUES)  
FIRE PREVENTION (LIBRARY BUILDINGS)  
FO-50 SERIES (INTEGRATIVE LEVELS THEORY; PRINCIPLES; SYSTEMATIC SUBJECT COLLIGATION)  
FOOD TECHNOLOGY (COVERAGE BY ABSTRACTING AND INDEXING SERVICES)  
FOOD TECHNOLOGY (COVERAGE BY INFORMATION SERVICES)  
FOOD TECHNOLOGY (INFORMATION WORK)  
FOOD TECHNOLOGY (LIBRARY MATERIALS)  
FOREIGN LANGUAGE TRANSLATIONS (ARTICLES; PERIODICALS; SCIENCE AND TECHNOLOGY; LIBRARY)  
FORM OF HEADING (CLASSIFIED CATALOGUE CODE)  
FORMATS (COMPUTERISED BIBLIOGRAPHIC RECORDS)  
FORMATS (MICROFORMS (FRINGE SUBJECT))  
FORMATS (OUTPUT; SEARCHING; ON-LINE INFORMATION RETRIEVAL)  
FORMS (STRATEGIES; SEARCHING; ON-LINE INFORMATION RETRIEVAL)  
FORTHCOMING EVENTS (INFORMATION BULLETINS ON)  
FRANCE (LIBRARY LAW; AND REGULATIONS; LIBRARIANSHIP)  
FRANCE (TEXTILES; COMPUTERISED MULTILINGUAL INFORMATION SERVICES)  
FRANKFURT (WEST GERMANY) (INTERNATIONAL BOOK FAIRS; READING PROMOTION; CHILDREN)  
FREQUENCY DISTRIBUTIONS (WORDS; STATISTICAL TECHNIQUES; AUTOMATIC SUBJECT INDEXING)  
FUEL INDUSTRY (INFORMATION WORK)  
FUND ALLOCATION (BOOK FUNDS)  
FUNDS (BOOK FUNDS)  
FUNEN (DENMARK) (COUNTY RESOURCE CENTRES; ROLE IN SCHOOL RESOURCE CENTRES)  
FURTHER EDUCATION (LIBRARIANSHIP)  
FUTURE DEVELOPMENTS (ADMINISTRATION; OF INFORMATION WORK)  
FUTURE DEVELOPMENTS (SCOTLAND; LIBRARIANSHIP)  
FUTURE DEVELOPMENTS (SLOVAKIA; LIBRARIANSHIP)  
FUTURE DEVELOPMENTS (WEST GERMANY; ON-LINE UNION CATALOGUES)  
GEDANKEN INDEXING (UTILITY-BASED CRITERIA; TERM SELECTION; SUBJECT INDEXING)  
GENERAL PUBLIC (ROLE IN MUNICIPAL LIBRARIES; PARTICIPATIVE MANAGEMENT; DECISION)  
GEOGRAPHIC LOCATION (SITES; BUILDINGS; LIBRARIES)  
GEOGRAPHIC LOCATION (SITES; NORWEGIAN BIBLIOTEKHOOGSKOLE; NORWAY; GRADUATE LIBRARY SERVICES)  
GEOGRAPHICAL FACTORS (INFLUENCE ON VARIATIONS; CITATION ANALYSIS; ARTICLES; PERIODICALS)  
GEOGRAPHY (ON-LINE INFORMATION RETRIEVAL)  
GIMSØYA (NORWAY) (MOBILE LIBRARIES; COUNTY LIBRARIES)  
GLADSAXE (DENMARK) (MUNICIPAL LIBRARIES; BLIND AND PARTIALLY SIGHTED; USERS; TALKING)  
GOFFMAN'S TRANSITION OF WORK OCCURRENCES (STATISTICAL TECHNIQUES; AUTOMATIC SUBJECT INDEXING)  
GOTHENBURG (SWEDEN) (ADVICE CENTRES; SOCIAL SERVICES; CONSUMER PROTECTION AND INFORMATION)  
GOTHENBURG (SWEDEN) (LIBRARIES)  
GOVERNMENT DEPARTMENT LIBRARIES (CLASSIFICATION AND CATALOGUING)  
GOVERNMENT PLANNING (LIBRARIANSHIP)  
GOVERNMENT PUBLICATIONS (DOCUMENT DELIVERY)  
GOVERNMENT PUBLICATIONS (LIBRARY STOCK)  
GOVERNMENT STATISTICS (MAGNETIC TAPE; COMPUTERISED INFORMATION SERVICES; USE FOR INFORMATION)  
GRADUATES (STUDENTS; USERS; LIBRARIES)  
GRONINGEN (NETHERLANDS) (NURSERY SCHOOL LIBRARIES; ROLE IN LANGUAGE DEVELOPMENT; GUIDES AND HANDBOOKS (USE PROMOTION; LIBRARIES)  
HANDBOOKS AND GUIDES (USE PROMOTION; LIBRARIES)  
HASH FUNCTIONS (USE FOR CHEMISTRY; STRATEGIES; SEARCHING; ON-LINE INFORMATION RETRIEVAL)  
HEADINGS (CHOICE; CLASSIFIED CATALOGUE CODE)  
HEADINGS (FORM; CLASSIFIED CATALOGUE CODE)  
HEALTH (OCCUPATIONAL HEALTH LIBRARIES; GOVERNMENT DEPARTMENT LIBRARIES)  
HEILOO (NETHERLANDS) (BRANCH LIBRARIES; MUNICIPAL LIBRARIES; BUILDINGS)  
HERZEGOVINA (YUGOSLAVIA) (LIBRARIANSHIP)  
HIGHER EDUCATION LIBRARIES (GOVERNMENT PLANNING)  
HIGHER EDUCATION LIBRARIES (GUIDES AND HANDBOOKS; USE PROMOTION)  
HIGHER EDUCATION LIBRARIES (NEWSPAPERS; STOCK)  
HIGHER EDUCATION LIBRARIES (PLANNING; BUDGETS)  
HIGHER EDUCATION LIBRARIES (PLANNING; COOPERATION)

第5図 件名の一例



従って、件名は固有名からの検索には適しているとはいえない。第5図は1979年1/2月号における件名の一例である。

**B. 索引ファイル**

IR システムの評価基準には、応答時間、コスト、データファイルや索引ファイルの更新の容易さ、再現率、精度などがあるが、ここでは再現率と精度の面から LISA テープの索引ファイルについて言及する。

1. 分類コードファイル

分類コードには複合主題を構成する要素が記号的に明示されているため、コードだけでなくその構成記号からでもアクセスできるように索引ファイルを作成することが可能であり、それがファセット分類の利点を活用することにもなる。

LISA におけるレコードの重出件数は、1979年1/2月号では21件であるが、これは主題に関してのみである。例えば、学校図書館と公共図書館の関連を主として扱う文献には、HykGpAwFv と FvAwHykGp をそれぞれ分類コードとする2つのレコードが作成される。(Awは combination, relationships など関連を表わす。) しかし個々の分類コードに関しては構成要素の組合せの順序がきめられているので重出はない。LISA 印刷版では多面的な検索は、件名索引でのみ可能なのである。

従って、分類コードを使用して機械検索を行なう場合には、再現率を上げるため構成要素から分類コードにアクセスできることが必要になる。

IR システムでは普通どのようなキーワード(句)、分類記号が検索キーとして利用できるかを示すため、そのシステムですでに利用できる検索キーのリストを出力するコマンドが用意されており、利用者はこのリストにもとづいて適切なキーワード(句)、分類記号などを選択し検索を行う。LISA をデータベースとする IR システムでは、構成要素を記事検索の手がかりとすることがで

入力: “通覧” コマンド B Aw←構成要素  
 出力: 1 EgrAw←分類コード 2←頻度  
 2 FvAwHykGp 1  
 3 HdAw 1  
 4 HykGpAwFv 1  
 … … …

第7図 分類コードの通覧

きるように、分類コードや構成要素を対象にした同様の“通覧”コマンドが考えられる。例えば第6図、第7図のようなリストが出力されることになる。第6図は記号列 Bk で始まる構成要素のリストであり、第7図は構成要素 Aw を含む分類コードのリストである。なお記号列は LISA データからシフトコードにもとづいて切り出された構成要素そのものあるいはその部分である。また分類コードと構成要素とが一致する場合もある。第6図、第7図で示されたコマンドの利用により、構成要素から適当な分類コードにアクセスし、その分類コードにもとづいて記事が検索できることになる。

構成要素から分類コードへのアクセスを可能にするための補助ファイル作成には注意すべき点がある。

まず、CRG のファセット分類記号は大文字と小文字の組合せになっており、大文字がファセットを表わしている。このため分類コードから構成要素を抽出するには、大小文字の識別が必要になる。LISA の完全文字セットのテープでは、特殊文字(シフトコード)を用いて大小文字の区別がなされているので、シフトコードを手がかりにして構成要素を切り出すことになる。従って、LISA テープは完全文字セットのものを使用しなくてはならない。

ところで分類コードには Bko (ASLIB), Bji (DEW) のような機関名や人名の省略形が使用されているが、これらの分類コードをシフトコードにもとづいて機械的に切断していくと、構成要素として A, S, L, I, B および D, E, W が作られることになる。これを回避するにはカッコ内の要素で大文字が2回以上続くものは無視すればよい。Aslib や Dewey などの固有名は分類標目としても使用されているから、固有名からのアクセスは分類標目で行えるからである。カッコ内に含まれる要素には Bw (Rm) D73, M (664) のように、分類記号や UDC 番号もある。このうち分類記号は切り出し、UDC 番号は無視するのが妥当であろう。(M (664) の M は資

入力: “通覧” コマンド A Bk←記号列  
 出力: 1 Bk  
 2 Bkbcg  
 3 Bkbp  
 4 Bkbq  
 5 Bkbqgg  
 6 Bker  
 … … …

第6図 構成要素の通覧

料を, 664 は食品製造技術を示す)

## 2. 分類標目ファイル

分類標目は主題のクラスからではなく, “ことば” から文献を検索するキーであり, “REPORT OF THE WORKING PARTY”, “MUSEUMS OF LIBRARIANSHIP”, “ABSTRACTING AND INDEXING SERVICES” など句形式をとることが多い。検索精度を考えると句形式をとるキーワードをそのまま使用して索引ファイルを作成する方がよいが, その結果再現率が低下するのは避けられない。例えば, 部分一致・後方一致機能を持つシステム以外では, 分類標目中のキーワードとして “ABSTRACTING AND INDEXING SERVICES” のみが付与されているレコードを, “INDEXING SERVICES” からは検索出来ないからである。

再現率の低下を回避するために, 分類コードに関して採用された方法がここでも考えられる。つまりキーワード句中の重要語からそのキーワード句にアクセスできる機能を持つことである。これは分類標目に使用されているキーワード句が無統制であることを考えれば非常に有効であろう。

ところで, 重要語からキーワード句にアクセスするには, システムに採録されている重要語を示すコマンドが必要となる。第8図はその出力例を示したものである。分類標目の中には句形式をとらないものもあるため, 重要語と分類標目とが一致することもある。第8図ではそれを \* 印で示している。なお重要語の抽出は, キーワード句中のことばが無統制であるだけでなく誤字などもあるため, ストップワード・リストにもとづいてなされることになろう。

```

入力: “通覧” コマンド A      COMPUT←文字列
出力:   1  COMPUTER
        2  COMPUTERS *
        3  COMPUTING
        4  COMPUTER
        5  COMPUTER
        ⋮
        ⋮
        ⋮   ←重要語

```

第8図 重要語の通覧

第9図はある特定の重要語を含むキーワード(句)をリストするコマンドの出力例である。キーワード(句)は重要語からしか探せないため, 分類標目を検索キーとする検索では, あらかじめ正確にキーワード(句)を知らないかぎり, 第8図, 第9図の段階を経ることになる。従って, この方法では, “data bank” のように複

数の常用語で重要概念が表現される場合は, 検索もれが生じやすくなる。もし “data” および “bank” が常用語とみなされれば, 両語共第8図には出現せず, 従って第9図のコマンドも意味をなさないからである。

```

入力: “通覧” コマンド B      LIBRARY ←重要語
出力:  1  AMERICAN THEOLOGICAL LIB-
        2  RARY ASSOCIATION          1
        3  BALLOTS AND OHIO COLLEGE
        4  LIBRARY CENTER            1
        5  BALLOTS, OCLC, WASHINGTON
        6  LIBRARY NETWORK           1
        7  BRITISH LIBRARY           3
        8  BRITISH LIBRARY RESEARCH AND
        9  DEVELOPMENT DEPARTMENT   1
        ⋮
        ⋮
        ⋮
          ↑
        キーワード句

```

第9図 キーワード句の通覧

## 3. 件名ファイル

件名は印刷版の件名索引を作成するためのものであるから, これをそのまま機械検索用に使用するのは無意味である。件名は件名索引の中で他の件名との関連でとらえられてこそ, chain indexing 手法と相まって検索キーとして有効になるからである。手検索を意図して作られた検索キーをそのまま使用することは, 機械検索の効力を殺すことにもなりかねない。

従って, 件名をそのままの形で検索キーとせず, カッコやセミコロンを区切り記号と考え, 第10図のように構成要素に分割し整理することが考えられる。

```

MISINFORMATION (THEORIES OF INFORMATION AND KNOWLEDGE)
INFORMATION (THEORIES OF)
KNOWLEDGE (THEORIES OF)
INFORMATION SCIENCE (PROFESSION)
PROFESSION (INFORMATION SCIENCE)

```

↓

```

MISINFORMATION, THEORIES OF INFORMATION AND KNOWLEDGE, INFORMATION, THEORIES OF, KNOWLEDGE, INFORMATION SCIENCE, PROFESSION

```

第10図 件名の分割・整理

件名を分割・整理して得られるキーワード(句)の特色には前節 A. で言及した以外に次のものがある。

- i) 分類標目中のキーワード(句)に比べ索引語としての特定性が低い。

- ii) 一般にキーワード(句)の数は分類標目よりも多くなるが、分類標目中のキーワード(句)はその構成要素である重要語からアクセスできるため、検索の多面性において両者間に大きな差異はない。
- iii) キーワード句には分類標目、件名のどちらにも共通なものが多い。

以上の点から分類標目と別個に件名の索引ファイルを作成する必要はなく、分類標目として存在しないキーワード(句)のみを抽出し、分類標目の索引ファイルに加えればよい。

#### 4. 著者ファイル

著者ファイルでは、個人名、団体名などの変形(表現形式の種類)をどれだけ網羅するかが再現率に影響する。*LISA* ではタグ 202 にその変形が収録されているので、それも著者ファイルに含めなければならない。

[例] 200: DE SOLLA PRICE, DEREK

202: PRICE, D. J. DE SOLLA  
PRICE, D. J. DE  
DE SOLLA PRICE, D. J.

## V. 結 語

IR システムの設計では、検索機能の充実、効率の良い処理アルゴリズムの開発など検索面に力点が置かれており、データベースの特色に合致したファイルの作成といった蓄積面からのアプローチは、あまりなされていない様に思われる。しかし再現率と精度はデータベースの特色をいかにうまく IR システムに反映させるかに大きく依存する。特に *LISA* のように、統制語いではなく自由キーワードを採用しているデータベースでは顕著である。

社会科学・人文科学分野でのデータベース作成がさらに盛んになると、標準化の促進への努力が一方でなされているにもかかわらず、主題分野あるいは収録されるデータの種類の特色から様々の異なったタイプのデータベースが生まれうる。

古代ギリシャのプロクラスティーズではないが、現在の IR システムではデータベースを IR システムに合わせる傾向があるように思える。しかし蓄積面にもっと力点をおき、それぞれのデータベースに適した IR システムという考え方も増々必要にならう。

そのためにはデータベース生産者と利用者(IR シス

テム)との間の情報交換・コミュニケーションをより円滑にする必要があり、データベースの形式や内容の変更も利用者との協議のもとになされるべきであろう。機械可読版(テープ)における変更は、印刷版の変更よりもはるかに大きな影響を利用者に与えるからである。

- 1) Lancaster, F. Wilfrid. *Information retrieval systems*. 2nd ed. New York, John Wiley, 1979. p. 78.
- 2) 齊藤孝. 異なる情報検索システムを相互利用するシステム; CONIT 思想の提案. 第26回日本図書館学会研究大会, 三田図書館・情報学会昭和53年度研究大会合同研究大会発表要綱. 1978. p. 76-81.
- 3) 山崎昶. 情報システムのための大規模データベースの利用法. 東京, 近代科学社, 1979. p. 52.
- 4) Edwards, T. "LISA: A traditional abstracting service?," *International forum on information and documentation*, vol. 1, no. 2, 1976, p. 25-34.
- 5) Edwards, T. "Indexing LISA; Chains, KISS and the bold approach," *Indexer*, vol. 9, no. 4, 1975, p. 133-46.
- 6) *Library and information science abstracts*, (1) Jan. - Feb., 1976, 序文.
- 7) Edwards, T. "LISA: A traditional abstracting service?," *op. cit.*,
- 8) Edwards, T. "Indexing LISA; Chains, KISS and the bold approach." *op. cit.*,
- 9) Gilchrist, Alan and Alexandra Presanis. "Library and information science abstracts: The first two years," *Aslib proceedings*, vol. 23, no. 5, 1971, p. 252.
- 10) Dansey, P. "A bibliometric survey of primary and secondary information science literature," *Aslib proceedings*, vol. 25, no. 7, 1973, p. 252-63.
- 11) Edwards, Tom. "A comparative analysis of the major abstracting and indexing services for library and information science," *Unesco bulletin for libraries*, vol. 30, no. 1, 1976, p. 18-25.
- 12) Edwards, Tom. "LISA and comprehensive control of library and information science literature," *International library review*, 4, 1972, p. 68.
- 13) *LISA* Nov.-Dec., 1978, p. 350.
- 14) Gilchrist, Alan and Alexandra Presanis. *op. cit.*,
- 15) Heinz kill, Richard. "An American looks at LISA," *International library review*, 3, 1971, p. 251-6.
- 16) Edwards, Tom. "LISA: A traditional abstrac-

LISA データベースの特色と利用上の考慮事項

- ting service?," *op. cit.*, p. 28.
- 17) Turtle, Mary R. and William C. Robinson.  
"The relationship between time lag and place  
of publication in *Library and information sci-*  
*ence abstracts and Library literature*," *RQ*,  
Fall 1974, p. 28-31.
- 18) *LISearch tape services manual*. Oxford, Lear-  
ned Information, 1976. 42p.