

## オンライン情報蓄積・検索システム“UNIQ-1”の設計と開発†

小 畑 征二郎†† 松 沢 茂†† 宮 崎 正 俊††

大学などの計算センタのユーザには、学術情報データベースの検索サービスをしてほしいという要望と、個人や研究室などの単位で学術情報の管理とその利用が手軽に行えるようにするシステムを提供してほしいという要望がある。我々は、これらの両方の要求を満足すべくオンライン情報蓄積・検索システム UNIQ-1を開発した。

このシステムのデータベースの構造はインバーテッド・ファイル型とし、インバーテッド・ファイルへのキーワードのアドレッシングはハッシュ法を用い高速化を計っている。マンマシン・インタフェースでは、幅広い利用者層をカバーできるように、機能に応じたコマンド体系を設けた。特に、検索においては2系統のコマンド体系を設け、利用者のレベルに応じた使い方ができるように配慮した。また、システムは、拡張性、デバッグ性、移行性などを考慮して、モジュール構造とし、FORTRAN 言語で作成した。

現在、UNIQ-1は東北大学大型計算機センターのACOSシステム900(OSはACOS-6)のTSSの下で稼動し、個人、研究室などの学術情報の整理と検索およびセンタの公開文献データベースの検索サービスに供されている。

### 1. はじめに

研究者にとって、自分の研究分野の動向を把握しておくことは重要なことである。このため、従来より論文の抄録や抜粋を集めた二次情報誌が出版され、その検索技術も種々考案されている。しかし、近年のように研究論文の数も増えてくると、いままでのような印刷物による検索では間に合わないのが実状である。

このような状況のもとで、情報を計算機システムを使って検索するシステム、いわゆる情報検索システムが各所で研究開発され、検索サービスが行われるようになった。わが国では日本科学技術情報センターのJOIS<sup>1)</sup>、東京大学大型計算機センターのTOOL-IR<sup>2),3)</sup>、広島大学のHUNDRED<sup>4)</sup>などがある。その他、計算機メーカーや企業でも多数開発されている。しかし、これらによってサービスされているデータベースは、上記の二次情報誌の磁気テープ版の購入によるものや大きな組織によって収集、作成されたものが大部分である。また、システムそのものもデータを一括して入力する方式をとっているものが多い。

一方、大学などでは、個人または研究室などの単位で文献データを収集し、データベースとして利用したいという要望をもつ研究者も多い。このレベルで使わ

れるシステムは、プログラミングや計算機の知識がなくとも簡単にデータベースを構築でき、利用できることはいうまでもなく、データの発生のとどリアルタイムでデータを蓄積できることが必要である。

我々は、これらのニーズに応えるとともに、計算センタなどでサービスする大量データの検索サービスもできるオンライン情報蓄積・検索システムであるUNIQ-1(Universal Information Query)を開発した。本稿では、その設計方針、システム構成、利用例などについて報告する。

### 2. システムの設計方針

#### 2.1 システムの目標

システムの設計では、機能面における目標を次のように設定した。

(1) システムで扱うデータは、主として文献データのような文字情報とする。

(2) データベースの定義は、ファイル・システム概念を知らなくとも、TSS 端末(以下端末という)から会話型処理で行えるようにする。

(3) データの蓄積は、バッチ処理による一括入力と端末からの会話型処理によるものを可能とする。会話型処理では、データの修正を可能とする。

(4) 索引ファイルへのキーワードの登録は、データの投入時に自動的に行われるようにする。

(5) 検索の方式は、キーワード指定によるQ-A方式とする。走査の方式はインデックス・サーチとス

† Design and Implementation of Online Information Retrieval System "UNIQ-1" by SEIZIRO OBATA, SHIGERU MATSUZAWA and MASATOSHI MIYAZAKI (Tohoku University Computer Center).

†† 東北大学大型計算機センター

トリング・サーチを併用し、データの集合演算を可能とする。

(6) システムに関する案内や教育活動がある程度独自にできるように、ニュースやガイダンス機能および検索を支援するためのコマンドを設ける。

(7) オンラインで逐次蓄積するデータベースと一括入力する大量データベースに対して、それぞれに対応したデータベースの保全機能を設ける。

(8) 利用記録は、単に利用状況を示すにとどまらず、利用教育、システムの開発などに対して有意義なデータを提供できるようにする。

(9) UNIQ-1は不特定多数の人が使うので、データベース単位の機密保護、データベースの管理者と検索利用の利用可能範囲の制限などきめ細かな機密保護機能を設ける。

### 2.2 ファイルの構造

キーワードによる文字情報の検索方式には、ストリング・サーチとインデックス・サーチがある。ストリング・サーチはデータ1つ1つを読み出して走査するため検索には時間がかかるが、各種の比較方式がとれるためきめ細かな検索が可能である。一方、インデックス・サーチは索引ファイルだけをサーチして検索するため処理は速いが、比較的大雑把な検索となる。したがって、一般にはインデックス・サーチを一次検索に使い、それによって得られたデータ集合に対する二次検索としてストリング・サーチが使われる<sup>5)</sup>。UNIQ-1でもこの方式を採用することにした。

このようなシステムの性能は索引ファイルの構造に左右される。このときの性能の評価尺度は、(1)検索に要するCPU時間(以下検索時間という)、(2)索引ファイルへのキーワードの登録に要するCPU時間(以下登録時間という)および(3)索引ファイルのスペース量である。

ところが、索引ファイルは、基本的には図1に示すようにキーワード部とデータ・ポインタ部からなり、

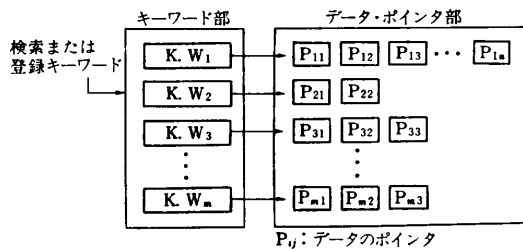


図1 索引ファイル  
Fig. 1 Index file.

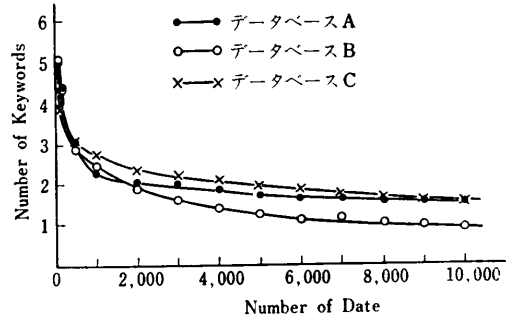


図2 データ当りのキーワード数  
Fig. 2 Keywords per Data.

その性能はキーワード部の構造に依存する。このキーワード部の構造としては、(1)シリアル・サーチ型の構造、(2)バイナリ・サーチ型の構造、(3)階層ブロック構造、(4)ランダム・サーチ型の構造などが考えられる。これらのうち、検索時間ではランダム・サーチ型が最も小さく、使用するスペース量ではシリアル・サーチ型やバイナリ・サーチ型が最も小さい<sup>6)</sup>。図2は、市販データベースに収録されているデータのタイトル項目からキーワードを切り出して、新規に索引ファイルに登録しなければならないキーワードの数を調べたものである。これによると、どのデータベースにおいてもデータ500件では約3個、5,000件では1~2個程度である。このことは、キーワードの占めるスペース量はデータの占めるそれに比較すると微々たるものであることを示している。また、データをオンラインで逐次蓄積する場合、キーワードの登録は検索と登録のためのキーワード部の操作の2つのステップからなる。ところが、シリアル・サーチ型やバイナリ・サーチ型の構造では、検索時間も大きく、キーワードの挿入のための操作も複雑である。階層ブロック構造の場合も同様である。これらはいずれもオンライン逐次蓄積向きではなく、一括蓄積向きといえる。

一方、ランダム・サーチ型の構造では、検索が高速で、操作も比較的単純であり、スペース量の問題に関してもランダム化手法の問題として独立に研究、改善の可能性がある<sup>7)-9)</sup>。このような理由によりUNIQ-1ではランダム・サーチ型の構造を採用した。

### 2.3 マンマシン・インタフェース

マンマシン・インタフェースの設計では、次のことを目標とした。

(1) 機能および使用者のレベルに応じた使い方ができること。

(2) 応答のためのキータッチ数および出力数では

きるだけ少なくすること。

(3) 分かりやすいコマンド名を付けること、また、他システムと機能およびコマンド名などである程度の互換性をもたせること。

これらを実現するために、(1)に関してはニュースやガイダンスは必ず打ち出されるものと選択によって打ち出されるものに分け、さらに詳しく知りたい場合にはコマンドで聞けるようにする。また、コマンド系には、システムを呼びだしさえすればシステムが誘導してくれ誰でも使えるシステム誘導型のもと、作業効率を重視したシステム受動型の2系列を設け、利用者が選択できるようにする。

(2)に関しては、コマンドの省略形やパラメータの既定値を設けて入力量を減らせるようにする。また、コマンドのパラメータは一括入力を可能とし、応答メッセージを省略できるようにする。

(3)に関しては、コマンド名は無意味な記号化をさけ、分かりやすい名前を付ける。また、他システムとの互換の問題では、類似な機能には同じような名前を付ける。その結果1つの機能に対して複数のコマンド名が付く場合もありえる。

### 3. システムの構成

#### 3.1 データベースの構成

UNIQ-1 で作成する個々のデータベースは、データ

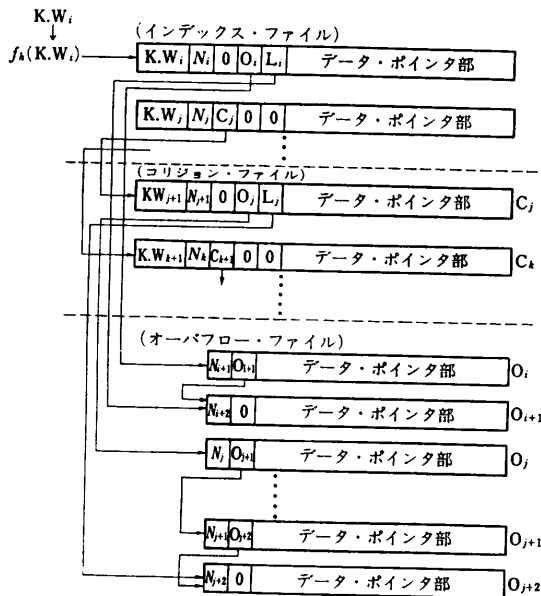


図3 各レコードの論理的关系  
Fig. 3 Logical relation between each record.

ベース定義ファイル、索引ファイルおよびデータ・ファイルで構成される。

データベース定義ファイルは、シーケンシャル・ファイルで、ここにはデータベースを構成するファイルの名前とそれらの大きさ、データの形式とその許容数、キーワードの切り出し項目、キーワードを切り出すための区切文字、不要語、データベースの管理者名など、データベースを特徴づける種々の情報が書き込まれている。そして、このデータベース定義ファイルの名前がデータベースの名前となる。

索引ファイルにおけるキーワードのランダムイズはハッシュ法により、衝突の処理は連鎖法 (chaining 法) による。索引ファイルは物理的には、ランダムイズ・ファイル、コリジョン・ファイルおよびオーバフロー・ファイルの3つのファイルからなる。これらのファイルのレコードの関係を示すと図3のようになる。Nはそのレコードでもつデータのポインタ数、Cはそのキーワードと衝突したキーワードの格納アドレス、Oはデータ・ポインタ部のあふれ部分を格納するオーバフロー・レコードのアドレス、Lはその最後のレコードのアドレスである。各ファイルの大きさやレコード長は、データベースごとに任意に設定できる。ただし、キーワード部の大きさは固定長12文字である。\*

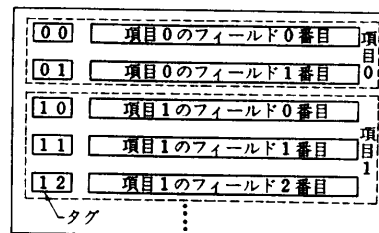


図4 データの形式  
Fig. 4 Data format.

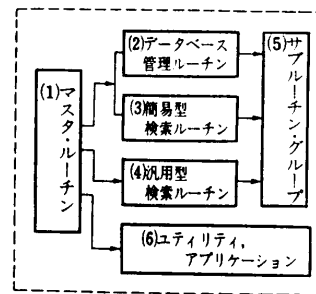


図5 UNIQ-1の構成  
Fig. 5 The configuration of UNIQ-1.

\* 12文字をこえるキーワードは、登録の際も、検索の際も自動的に12文字にカットされる。つまり、12文字をこえるキーワードの検索では自動的に12文字の前方一致を行うことになる。

UNIQ-1 で扱うデータの大きさは、最大99フィールド(1フィールド=78文字)と限定している。その形式は図4に示す通りで、各フィールドに付いている2桁のタグによってフィールドを識別する。つまり、上の桁によって項目を、下の桁によって項目内のフィールド番号を示している。このタグと項目の内容の対応はデータベース定義ファイルで行う。このようなデータを格納するデータ・ファイルは、固定長レコードのランダム・アクセス型のファイルで、1データが1レコードに入り切らない場合には、オーバーフロー領域にリンクする。なお、実際にUNIQ-1が動作するときには、検索のためのワーク・ファイル、ログ・ファイル、ニュース用ファイルなどもアクセスされる。

### 3.2 UNIQ-1の構造と機能

UNIQ-1は図5に示すように6つの部分からなり、それぞれの機能は次の通りである。

#### (1) マスタ・ルーチン

このルーチンは、UNIQ-1が呼び出されると最初に起動され全体を制御するとともに、UNIQ-1に関するニュースの出力、データベースのオープン・クローズ、利用記録などの処理を行う。

#### (2) データベース管理ルーチン

このルーチンは、データベースの管理者として登録されている人だけが使用でき、データベースの作成、データの入力と修正、キーワードの切り出しと索引ファイルへの登録、データベースの保全のためのデータのセーブやリスト・アップ、データベースを構成するファイルの使用状況に関するデータの収集などの処理を行う。

#### (3) 簡易型検索ルーチン

このルーチンは、簡易型の検索をサポートする部分で、ここでは「索引検索」、 「直接検索」および「索引・直接検索」の3つの方式の検索をサポートする。索引検索は、いわゆるインデックス・サーチを行うもので、ここでは簡易型と記述式型の検索ができる。直接検索は、データを直接読み出して指定された項目に対し

てストリング・サーチを行うものである。索引・直接検索は、索引検索によって得られたデータ集合に対して直接検索を可能にしたものである。

#### (4) 汎用型検索ルーチン

このルーチンは、汎用型の検索をサポートする部分で、大別すると次のようなコマンドを提供する。

- 1) Q-A方式の検索用コマンド
- 2) プロファイル検索に関するコマンド

```
SYSTEM ?UNIQ
  (( NEWS-1 1981,8,25 ))
** METADEX IS UPDATED ON 1981,8,25.
  DATABASE/NEWMETA CORRESPONDS TO THE JUNE ISSUE OF 1981

FUNCTION(APPLICATION,MANAGEMENT OR RETRIEVAL) ? M
DATABASE ? *
*** CREATION OF DATABASE DEFINITION FILE ***
NAME OF DATABASE ?
=IDATA
KADAIBANGO OF DATA-MANAGER
=J20000020
NUMBER OF DATA
=500
KANJI DATA (YES/NO)
=NO
NAME OF ITEM
01=TIT:
02=AUT:
03=JOU:
04=KEY:
05=
TAG NO. OF KEY-ITEM
=2
=4
① ** DELIMITER OF KEYWORDS **
  BLANK( ), SEMI COLON(;), COMMA(,)
  OTHER CHARACTERS
=
** USELESS WORDS **
  THE,IS,A,OF,AN,AND,AS,AT,BY
  FOR,FROM,IT,ITS,OR,ON,TO,WITH,IN
  ノ,ニ,シイテ,カ,チ,オ,アリ,ト,スル,ヨリ,チ,ス,オケル,ニテ
  OTHER WORDS
=
PASSWORD OF FILE-INITIALIZATION
=QDATA

REQUEST SIZE 1096(LLINKS) CREATE(YES/NO) ? Y
PASSWORD ? QDATA

DATABASE ? IDATA

* * * * YOUR DATA-TAG * * * *
* 1-TIT: 2-AUT: 3-JOU: 4-KEY:

KIND OF WORK ? DM
INPUT DATA FROM KB OR FILE ? KB
MAX SIZE = 499 CURRENT SIZE = 0
NEXT DATA NO. = 1 INPUT DATA !
=1
=11 A STOCHASTIC EVALUATION MODEL FOR DATABASE ORGANIZATIONS
=12 IN DATA RETRIEVAL SYSTEMS
② =21 KENNETH F. SILER
=31 COMM. ACM, VOL.19 NO.2 PP.84-95(1976)
=41 DATABASE ARCHITECTURE, PERFORMANCE AND EVALUATION;
=42 INVERTED, THREDED AND CELLULAR LIST, INFORMATION
=43 RETRIEVAL, DATABASE ORGANIZATIONS, QUERY COMPLEXITY;
=44 STOCHASTIC MODEL, MONTE CARLO SIMULATION
=99
NEXT DATA NO. = 2 INPUT DATA !
```

図6 データベースの定義とデータの入力

Fig. 6 Definition of database and input of data.

- 3) 回答出力用コマンド
- 4) 検索を補助するコマンド
- (5) サブルーチン・グループ

これは、前述の(2), (3)および(4)で共通に使えるサブルーチンを集めたものであり、これには文字列の分解、データの読出し、ストリング・サーチ、データの集合演算などを行うものやハッシュ関数がある。

- (6) ユティリティおよびアプリケーション

これらはデータベースの作成と管理をサポートし、次のような処理を行う。ユティリティの一部はオフラインで動作する。

- 1) 大量データを一括入力する
- 2) キーワード・リストを出力する
- 3) 著者索引リストを出力する
- 4) コード変換をする
- 5) ログファイルの編集を行う

#### 4. コマンド体系とその特徴

UNIQ-1の目的は、1章で述べたように素人でも簡単にデータベースを構築でき、利用できることにすることにある。つまり、データベース管理ルーチンは誰でも使えるようにする必要があるし、データベースを管理する特定の人が使うにしても頻繁に使うものではないので、作業効率よりもむしろ使いやすさを重視する必要がある。また、誰でも容易に使えるようにするという点では、検索についても同じことがいえる。ところが、検索では個人や研究室利用のデータベースだけではなく、計算センタなどでサービスする広範囲利用の比較的大規模なデータベースのサービスにも堪え得る必要がある。すなわち、検索においては幅広い利用層をカバーする必要がある。このような理由で、コマンド体系には2系統のものを設けた。データベース管理ルーチンと簡易型検索ルーチンはシステム誘導型である。汎用型検索ルーチンでは、初めはシステム誘導型であるが、実際の検索作業の部分はシステム受動型を主体としている。これは慣れた人はそれなりの効率のよい検索ができ、慣れない人は

初めのシステム誘導型の部分で学習をしてから検索作業に入れるように配慮したものである。また、受動型の部分でもコマンドの使い方が聞けたり、パラメータの指定を忘れた場合にはシステムが誘導するなど初心者にも楽に使えるように工夫をしている。また、汎用型検索では、大量データの検索に便利のように検索の範囲を制限したり、プロフィールによる検索などを可能にしている。

表 1 主なコマンド  
Table 1 Principal Commands.

コマンドの形式	機能
SEARCH = [[ {Keyword} ] [AND] [OR] [DIF] ] [ {String} ] [ {Set no} ]	キーワードによるインデックス・サーチ (一次検索)
SSEARCH [Set no] [[ [ {Tag/String} ] [AND] [OR] [DIF] ] [ {Tag/String} ] [ {Set no} ] [ {Tag/RO.n} ] ]	ストリング・サーチ (二次検索)
QLIST [Range]	質問式とデータ集合の大きさの出力
PRINT [Set no] [, Data no]	検索結果の端末出力
QSAVE [Profile] or PMAKE [Profile]	質問式の保存 (プロフィールの作成)
PUSE [Profile]	プロフィールによる検索
TEACH [command]	コマンドの使い方の教授
OFF or END	検索の終了

```

FUNCTION(APPLICATION,MANAGEMENT OR RETRIEVAL) ? M
DATABASE ? DATABASE/NEWMETA

* * * * YOUR DATA-TAG * * * *
* 1=NO.: 2=TIT: 3=ABS: 4=AUT: 5=SOU: 6=IND: 7=AID:

KIND OF WORK ? ID

METHOD OF RETRIEVAL (SIMP OR EXPR) ? SIMP

① [KEYWORDS ?
HYDROGEN NIOBIUM

ANSWER COUNT = 4
PRINT (YES OR NO) ? N

② [TAG AND KEY-EXPRESSION (EX. 2,SYSTEM^*ABORT^) ?
=4,H. MULLER^
=>

ANSWER COUNT = 1
PRINT (YES OR NO) ? Y
OUTPUT-LINE(TAG NO. S OR F) ? 2,4,5

( ** ANSWER LIST ** )

( 1 ) * 20
TIT:INVESTIGATION OF THE TERNARY SYSTEM NB -- TA -- H.
AUT:H. MULLER,K. WEYMANN,P. HARTWIG
SOU:$GJ. LESS-COMMON MET.$11 OCT. 1980$J74, (1)$L17-22$MENGLISH

TAG AND KEY-EXPRESSION (EX. 2,SYSTEM^*ABORT^) ?
    
```

図 7 索引・直接検索

Fig. 7 Indexed and direct retrieval.

\* [ ] は省略可を示す。

```

FUNCTION(APPLICATION,MANAGEMENT OR RETRIEVAL)? R
DO YOU WANT THE LIST OF COMMANDS (YES OR NO)? Y
①
COMMAND (ALTERNATIVE)      FUNCTION
(1) SEARCH                  INDEX SEARCH--PRIMARY SEARCH
(2) SSEARCH                  STRING SEARCH--SECONDARY SEARCH
    }
(14) CHANGE                  CHANGE DATABASE
(15) OFF(END)                END OF RETRIEVAL

** YOU MAY USE INITIAL TWO LETTERS AS ABBREVIATED FORMS !!
( EX. SE FOR SEARCH )
** IF YOU WANT MORE INFORMATION ABOUT COMMANDS THEN YOU MUST USE
A "HELP(TEACH)" COMMAND !!

DO YOU WANT THE LIST OF DATABASE (YES OR NO)? Y
②
** LIST OF AVAILABLE DATABASE 1981.8.25 **
(1) DATABASE/META70 --- METADEX 1970(09264)
(2) DATABASE/META74 --- METADEX 1974(26159)
    }
(9) DATABASE/METAB1 --- METADEX 1981(19355)
(10) DATABASE/NEWMETA --- METADEX 1981.6(3180)

③ [DATABASE (IF CR THEN END)? NEW

*** DATA-TAG OF DATABASE/NEWMETA: ***
* 1=NO.: 2=TIT: 3=ABS: 4=AUT: 5=SOU: 6=IND: 7=AID:

ENTER COMMAND
④ [(- SE
  (=HYDROGEN
    #1 HYDROGEN 101
  (=NIOBIUM
    #2 NIOBIUM 60
  ⑤ (=OR TANTAA^
    #3 OR TANTAA^ 73
  (=AND #1
    #4 AND #1 5
  (=
  (=
  ENTER COMMAND
  ⑥ [(- SS 4
    (=AU/H. MULLER
      #5 AU/H. MULLER 1
  (=
  ENTER COMMAND
  [(- PR 5,1
  TAG NO. (IF CR THEN SAME AS BEFORE)? S

  [ (** ANSWER LIST (SET NO. 5) **)
  ⑦ ( 1) # 20
  TIT:INVESTIGATION OF THE TERNARY SYSTEM NB -- TA -- H.
  AUT:H. MULLER,K. WEYMANN,P. HARTWIG
  SOU:$GJ. LESS-COMMON MET.$11 OCT. 1980$J74, (1)$L17-22$MENGLISH

  ENTER COMMAND
  [(- OFF
  ONL-T: 13.958 - 14.007 CPU-T: 1.9705 HIT-T: 1 HIT-F: 0
  ** RETRIEVAL END **

```

図 8 汎用型検索

Fig. 8 Universal type retrieval.

UNIQ-1の呼びだしは、コマンド入力可能時に  
UNIQ [ Database name]\*  
と指定する。Database nameはアクセスしようとするデータベースの名前である。  
ここで指定しないと、後で公開のデータベースのリストが出力され、何をを使うかの問い合わせがあるので、そこで指定してもよい。

図6はデータベースを作成し、データをオンラインで蓄積する例である。①の部分はデータベースを定義し、ファイルを割り当てる部分であり、②の部分はオンラインでデータを蓄積する部分である。

図7は簡易型検索における索引・直接検索の例である。①の部分はHYDROGENとNIOBIUMの両方を含むデータの検索で、②の部分が①で得られたデータ集合に対して二次検索を行う部分である。二次検索では著者項目(タグ番号4)に対して、H. MULLERをストリング・サーチしている。

図8は汎用型検索の例である。①の部分がコマンド・リストで、②の部分が公開のデータベースのリストである。③にこれから使おうとするデータベース名を指定する。このときフルネームの他に、データベース・リストの番号やNEWなどと省略形で指定することも可能である。また、自分専用のデータベースを指定してもよい。④はコマンドの入力促進記号であり、これに対してコマンドを入力する。コマンドは初めの2文字だけでもよい。また、これに対して?を入力すると①に示したコマンド・リストが得られる。ここで入れられるコマンドの形式は表1の通りである。表中、[ ]は省略可能なことを示し、{ }はその中から選択することを示す。また、SSEARCHのRO(Relational

Operater)としては、GT, GE, EQ, NE, LT および LEが使える。⑤の部分はインデックス・サーチの部分で、#3のキーワードの‘^’記号は前方一致を示す。⑥の部分はストリング・サーチによる二次検索の部分で、著者項目に対して‘H. MULLER’を捜すものである。⑦の部分はヒット・データの出力の部分である。TAG NO.?に対しては、出力する項目をタグ番

号と記号で指定できる。例の S は標準的な出力を行うもので、その他 I (インデックス), A (抄録), F (全項目) などが指定できる。これらは任意に組み合わせて指定することもできる。

### 5. UNIQ-1 の使用例

UNIQ-1 は、主に個人、研究室または研究所などが単位での文献データの管理や検索のために使われているが、UNIQ-1 を使って当センターの一般利用者に公開サービスをしているデータベースもある。そのうちで最大のもは金属学関係の文献データベースである METADEX である<sup>10)</sup>。METADEX は米国金属学会 (American Society Metals, ASM) と英国金属学会 (The Metals Society, London) が協力して収集しているもので毎月約 3,200 件ずつ発行されている。これを東北大学金属材料研究所が 1974 年にさかのぼって購入し、当センターの一般利用者に公開している。図 7 と図 8 で検索しているのは METADEX の新着月分のものに対してである。

ところで、METADEX は磁気テープで納入されるが、その仕様は独特なものである<sup>13)</sup>。これを UNIQ-1 のデータベースとしてサービスに供するためには、まず、データのコードと UNIQ-1 のデータ形式への変

換が必要である。これはバッチ処理で行い、その結果は普通のシーケンシャル・ファイルに出力される。これをもとにデータのチェックを行う。データの内容のチェックは適当にサンプリングして金属材料研究所の専門家が行い、件数や形式のチェックは UNIQ-1 のユティリティ・プログラムで行う。次に、UNIQ-1 のデータベースとしてロードする。これはシーケンシャル・ファイルから行い、このときの索引語の切り出しは、インデックス・タームとアロイ・インデックス・タームから行っている。1 カ月分のデータを蓄積するのに要する CPU 時間は約 9 分で、ファイル容量は約 7.5 MB である。この容量は索引ファイルなどの容量も含んでいる。ロードが終ると簡単な検索のチェックを行い、ニュースなどのアップデートを行ってサービスを開始することになる。

この種のデータは月々何千件のペースで増加し、その全体の容量も非常に大きい。したがって、何年次までさかのぼってサービスするかが問題となる。当センターではスペースが許すならサービスをしようという方針なので、現在は 74 年次から全部をサービスしているが、一般の利用者のファイル・スペースを圧迫するようになった場合には、古い年次のもははずすことになるであろう。また、障害対策としては、デー

```

SYSTEM ?ERME
ト*ンナ メッセ-シ* カ* テ*マシタ カ (オフリ ナラ <フ*キ>) ?
INCORRECT CAT/FILE
( 1 )
F000 INCORRECT CAT/FILE DESCRIPTION AT <ナマエ>
( 2 )
J007 INCORRECT CAT/FILE DESCRIPTION -AT- <ファイル>
ウエ ノ メッセ-シ* ノ ウチ カ*イトウ スル モノ NO. ヲ イレテ クタ*サイ !
モシ カ*イトウ スル メッセ-シ* カ* ナイナラ <フ*キ> ヲ イレテ クタ*サイ !
↓
< メッセ-シ* >
F000 INCORRECT CAT/FILE DESCRIPTION AT <ナマエ>
< ケ*ンイン >
シタイ サレタ ショウジョウ カタログ/ファイル メイ N ソンサ*イ シナイ。 イッチ シナイ サイショ ノ ナマエ カ
インツツサレル。 マタハ、 ファイル メイ ノ ヒツヨウナ トコ ニ カタログ* メイ カ* シタイ サレタイル カ、
ソノ キ*トク ノ シタイ テ* アル。
(1) $ FILSYS ファン テ* "F" カ* COL.9 カラ Nシ*マツテ イタ。
(2) マタ* トウロク シタイナイ ファイル ヲ FILE PURGE シタ。
(3) $ FILSYS カ* $ FILSIS ニ ナツタイタ タメニ、 コノ ウシロ ノ $ PRMFL ファン テ*
コノ エラー ニ ナツタ。
(4) オナシ* シ*ユフ* ノ ナカ テ* FILE CREATE シタイル ファイル ヲ $ SELECT テ* ツカッタ
(5) $ PROGRAM SCED カ* ヌケテ イタ タメニ *CIテ*イタ ノ タメノ ファイル テイキ* カ*
マエ ノ アクティビ*ティ ノ モノト ミナサレタ。
(6) $ PROGRAM ファン ニ ツツリ ミス カ* アッタ タメニ、 ソレ イコウ ノ ファイル テイキ* ファン
カ* マエ ノ アクティビ*ティ(FILSYS) ノ モノト ミナサレタ。
ト*ンナ メッセ-シ* カ* テ*マシタ カ (オフリ ナラ <フ*キ>) ? →
** END **

```

図 9 ERME の使用  
Fig. 9 Use of ERME.

データベースは年次ごとのサブファイル形式とし、更新中のものは月ごとに変換後のデータの保存を、更新終了のものはトータル・ダンプの保存を行っている。なお、METADEX は昭和55年6月からサービスをし、月約170回の利用がある。

UNIQ-1は他のプログラムからも呼びだして使用することができる。その例としては、ERME (Error Message 検索システム) の試作を挙げることができる<sup>11)</sup>。現在の計算機システムのエラー・メッセージは、現象だけをとらえて表示するにとどまっている。したがって、原因の究明は現象を起すと思われる事柄の羅列の中から該当のものを選ぶことになる。つまり、経験に頼ることで、経験のない人は経験者かマニュアルに頼ることになる。しかし、これは現在使っている計算機に聞けることが最も望ましい。そして、その方法は得られたメッセージからキーワードを選んでAND型の検索を行えばよいし、検索手続きもできる限り単純な方がよい。ERMEでは、UNIQ-1の簡易型の検索部分をサブルーチンとして使用し、主プログラムでは端末に対する入出力とヒットしなかった質問文の記録だけを行っている。図9はこのシステムを使って、エラー・メッセージ「INCORRECT CAT/FILE DESCRIPTION ××××」について、その原因を調べた例である。簡単な手続きで目的の情報が得られることが分る。この種のデータベースでは、よりよいデータをいかにして作成するかが問題である。また、すべてのデータを最初から準備することは困難なので、ヒットしなかった質問文の記録をもとにデータを逐次増加できることが必要である。ERMEではこれも可能である。

## 6. おわりに

UNIQ-1は、以上のように種々の使われ方をしているが、今後はシソーラスを使った検索、SDIサービスについても実施の方向で検討する予定である。またUNIQ-1はFORTRANで作成し、特別なDBMSを使っていないので、他システムへの移行も比較的簡単

に行えるものと思っている。

最後に、本システムの開発にあたり、特に検索用コマンドに関して、種々ご助言いただいた東北大学金属材料研究所の小岩昌宏教授に深謝する。

## 参 考 文 献

- 1) 日本科学技術情報センター：JOIS オンライン情報検索サービス・マニュアル。
- 2) 東京大学大型計算機センター，TOOL-IR プロジェクト：TOOL-IR 利用者マニュアル(上,下)。
- 3) 根岸正光，山本毅雄：オンライン文献検索システム TOOL-IR におけるマン・マシン・インタフェース，情報処理，Vol. 17, No. 5 (1976)。
- 4) 鳴海元他：HUNDRED オペレーション・ガイド。
- 5) 時実象一：オンライン検索システムの比較，情報管理，Vol. 20, No. 12. pp. 929-944 (1978)。
- 6) SEVERANCE, D.G. & CARIS, J.V.: A Practical Approach to Selecting Record Access Paths, Computer Surv. Vol. 19, No. 4. pp. 259-272 (1977)。
- 7) 小畑，松沢，宮崎：キー・ランダマイズ型情報検索システムの性能評価，情報処理学会第20回全国大会論文集。(1979)。
- 8) 小畑，松沢，宮崎：キー・ランダマイズ型インバーテッド・ファイルの評価，電気関係学会東北支部連合大会論文集 (1979)。
- 9) 小畑，松沢，宮崎：キー・ランダマイズ型情報検索システムの最適化に関する一考察，情報処理学会第21回全国大会論文集 (1980)。
- 10) 小岩，小畑：METADEX オンライン検索サービスの利用法，東北大学大型計算機センター広報，SENAC, Vol. 13, No. 4 (1980)。
- 11) 小畑，松沢，宮崎：簡易型オンライン情報検索システム“UNIQ-1”について，NEAC-SP研究会論文集 (1978)。
- 12) 小畑，松沢，宮崎：計算機の運用管理のためのオンライン情報蓄積・検索システム，情報処理学会第19回全国大会論文集 (1978)。
- 13) 小畑，松沢，宮崎，中西，小岩：METADEXの検索サービスとその運用管理，全国共同利用大型計算機センター研究開発論文集，No. 2 (1980)。  
(昭和56年9月8日受付)  
(昭和56年11月18日採録)