

# 東大オンライン情報検索システム 'TOOL-IR' の運用経験

根岸正光, 牛丸 守, 山本毅雄 (東大)

## 1. TOOL-IR の概略

学術情報システム「TOOL-IR (Tokyo university On-Line Information Retrieval system)」は、全国の国公私立大学・高専・文部省所属研究機関等の研究者に最新の文献情報を提供し、さらに、本システムを媒介とする研究者相互間の知識の共有・交換を促進すべく計画され、現在、化学関係文献 (CAS)、X線結晶学文献 (XDC)、計算機及び制御関係文献 (INSPEC) の3種のデータ・ベースを対象として、東大大型計算機センターにおいて運営され全国共同利用に供されている。<sup>(注)</sup>

利用者は、昨年12月末現在、約132名 (CAS利用者のみ) で、75年9月より音響カプラー付電話回線端末を通じて全国的なオンライン利用が進められており、月間16000件程度の文献が検索されている。稼働時間はセンターの常態時間と同じで、センターのバッチ、専用回線TSSは平日午前9:30~午後7:00、音響カプラー付端末では午後10:00まで利用できる。東大大型計算機センターの機器構成は表1のようであるが、この内、TOOL-IR関係のデータ・ベース用として、ディスク4バッチ (計400MB) が接続されている。データ・ベースの大きさは表2の通りである。

表1 東大大型計算機センターの機器構成

cpu	HiTAC 8800 X2 + 8700 X2	主記憶	4MB
ディスク	29MB X 32台	100MB X 12台	
RJE端末	14台	TSS 端末	64台

表2 TOOL-IR データ・ベースの大きさ

種別	文 献 数		ファイル容量	
	キーワードのみ	内容とも	転置ファイル	線型ファイル
CAS	(1974年1月~1975年6月) 47万8千 文献	(1975年7月分~1976年1月分及び投票分) 26万 文献	122 MB	145 MB
XDC	—	(1935年~1975年) 13830 文献	4.7 MB	49.5 MB
INSPEC	—	(1976年1月分) 2618 文献	2.0 MB	5.0 MB

TOOL-IRシステムは、主として学術文献データをオンライン (ディスク) データベースとし、これに対してTSS, リモートバッチ, オープンバッチ, クローズドバッチなど種々の方法で検索できるシステムである。

これまでの多くの文献情報システムと比較して、次のような点が特徴といえよう。

●TSS処理・バッチ処理とも共通のコマンド・サブコマンドを使う——表3にTOOL-IRの利用者コマンド (主コマンド), 表4に同じくサブコマンドを示す。利用法はTSS処理 (専用回線端末, 音響カプラー付移動端末), リモート・バッチ処理, オープン・バッチ処理, クローズド・バッチ処理 (郵送サービス) と種々あるので、これらを組合せて、たとえば検索質問の作成や、すぐに結果を得たい少量の検索はTSSで、大量の文献リストの出る検索はバッチ処理で、と使い分

(注) TOOL-IRの研究・開発は昭和48.49年後は、文部省特定研究「広域・大量情報の高次処理」の化学における情報処理班、50年後は、同開発C班 (C班代表者 藤原鏡男, 統括班代表者 島内武彦) によって行われものである。

表3 TOOL-IR 主コマンドの種類

	名称	対象データベース	年次
CAS	NEWCAS	最新之号	最近2週間分
	CAS	最近巻(84巻)	1976年前半分(累積中)
	CAS 83	83巻	1975年後半分
	CAS 82	82巻	1975年前半分
	CAS 81	81巻	1974年後半分
CAS 80	80巻	1974年前半分	
XDC	XDC	最新版	1935年~1995年
	XCALC	シ	シ
INSPEC	IEE	1976年1月号	1976年分(累積中)

表4 TOOL-IR のサブコマンド

名称	機能
SEARCH	一次検索。内容・著者名等で検索し、文献集合を作る。
AND	二次検索。文献集合間の積集合(A∩B...)を作る。
OR	二次検索。文献集合間の和集合(A∪B...)を作る。
DIF	二次検索。文献集合間の差集合(A-B)を作る。
DISPLAY	文献集合の内容表示。
QSAVE	一連の質向を、ファイルにしまう。
QUSE	一連の質向を、ファイルからとって使う。
KEEP	文献集合をファイルにしまう。
ENTER	文献集合をファイルからとって使う。
EXTRACT	文献集合の一部を抽出して別の文献集合を作る。
DUMP	文献IDを手で内容表示。
REMINO	文献集合に対する質向の内容表示。
COPY	内容を個人ファイルへコピーする(XDCのみ)。
END	検索を終了する。
KILL	サブコマンドの処理を中断する。

質向の結果えられ文献のリストを、自分のファイルにたくわえることができる。このリストは、ENTER サブコマンドによって呼び出し、内容を印刷したり、さらにくわしい検索の出発材料として使うことができる。

●検索質向ファイル、検索結果ファイルの貸し借り、やりとりができる——PERMIT コマンド、SHARE コマンド、TRANS コマンド、RECVF コマンドなど、システムに既に備わっているコマンドを使って、検索質向ファイルや検索結果ファイルを貸し借り、あるいはやりとりすることができる。これにより、たとえば遠隔地間の研究者同志の協力や協同作業が、ファイルを媒介にして簡単にできるようになる。

●システムの質向ライブラリを利用できる——専門外の分野を検索するとき、検索質向の模範例あるいは参考として、その分野の専門家が作ってシステムに登録した検索質向を、QUSE サブコマンドで呼び出して使うことができる。

TOOL-IR の運営経過は表5の通りであり、CAS については、試用公開から数えて1年半程の運用実績となる。

2. データ・ベースの内容

TOOL-IR の定常的な運営は、基本的に次の3局面をもっている。即ち①ソース・データ(磁気テープ)の入手、②TOOL-IR 用データ・ベースへの変換(データ・ベースの更

けることができる。

●検索質向を自分のファイルにたくわえることができる——TOOL-IR コマンドでは、会話形式で複雑なロジックを含む一連の検索質向を作成し、これをQSAVE サブコマンドで自分のファイルに貯蔵することができる。この質向は、ファイル編集コマンドを使ってTSS 処理すればバッチ処理で修正でき、QUSE サブコマンドで呼び出して使うことができる。したがって、検索質向を、何回のセッションにもわたってゆくり作成することができる。また、一連作成した検索質向でも、対象分野の進歩や自分の興味の変化に応じて、次第に作りかえてゆくことが容易である。

●検索結果を自分のファイルにたくわえることができる——KEEP サブコマンドによって、複雑な

表5 TOOL-IR 運用経過

74.8.15	CAS キーワードファイル試用公開
10.12	XDC
75.1.11	CAS 線型ファイル付加
2.13	CAS 著者名ファイル付加
4月	CAS データ・ベースの定期更新(2週間毎)開始
5.06	CAS 応用システム CODEN 公開
6.11	CAS システム一般公開・利用統計システム付加
9.16	XDC 応用システム XCALC 公開
10.01	音響カラー付端末利用開始
10.20	CAS 一回文献保存投票
11.05	XDC 検索コマンド拡張・利用統計システム付加
76.2.12	INSPEC 試用公開

新)、③ ユーザによる検索、である。ソース・データは、現在、CAS, XDC, IEE等の海外機関によって収集、編集、磁気テープ化され、そのものを扱っており、独自の文献情報の収集、ファイル化は行っていない。ここで三者の内容を説明する。

#### (1) CASデータ・ベース

Chemical Abstracts Service (CAS)は米国化学会の下部機構であり、1907年以來、化学および関連分野の国際的抄録誌 Chemical Abstractsを刊行してきた。1968年以來、ここで抄録される全文献についての情報を、Chemical Abstracts Condensates磁気テープの形で頒布するようになった。内容は、抄録された文献の書誌事項(論文・特許・レポートの表題、著者・特許出願者名およびその所属、掲載誌名及び巻・号・頁その他)、および抄録中から抽出されたキーワード・フレーズである。抄録対象は、125ヶ国、14,000種の科学・技術関係逐次刊行物、26ヶ国の特許、各国の会議議事録・学位論文・政府公報・単行本等で、これらの中から化学又は化学技術に関連した新しい情報を含むものを中広く抄録している。そのため、化学はもとより物理・薬学・医学・生物学・農学・化学工学など、多くの隣接分野で基本的な情報源となる。頻度は週1回で、年52本の磁気テープが送られてくる。抄録文献数は30~35万件/年である。

化学関係の研究においては、この抄録誌による研究動向のフォローが、従来から研究手順の一部として確立されており、TOUL-IRは、この土壌の上に計算機技術による検索の機械化を導入したということになる。このことから、ユーザである化学関係研究者の協力が、システム開発の当初より得られるなど、好条件に恵まれたといえる。

#### (2) XDCデータ・ベース

Cambridge Crystallographic Data Centreはイギリスのケンブリッジ大学化学教室に1965年に設立されたもので、科学教育省の科学技術情報オフィスOSTIより財政援助を受けている。また世界的にも結晶学データセンターとして認められており、干渉法による(X線、中性子線を用いる)結晶構造解析結果の集成、検索、評価、合成および流通を行っている。TOUL-IRで使用しているXDCデータ・ベースはこのデータセンターで作成された磁気テープによるもので次の内容を持ち、同所との契約により、日本における学術研究の目的のための使用を認められている。

(a) 文献ファイル: 1935年より現在までの有機および有機金属化合物の結晶構造解析のすべての文献を次の内容で記載してある。内容は6週間毎に更新され、絶えず最新の状態に保たれている。件数は1973年4月には8000に、1976年4月には12000になると推測される。(6週間毎に約150件が追加される)。内容は化合物名、同義語、分子式、著者名、掲載ジャーナル巻、頁、年、化合物分類コードからなる。

(b) 数値データ・ファイル: 1960年以後現在までに発表されたすべての有機化合物・有機金属化合物の結晶構造データを記載してあり、6ヶ月毎に更新される。1973年4月の件数は5000以下であり、1976年4月には10000件に達すると推測される。内容は結晶の単位セルの大きさ(およびその標準偏差)、スペースグループ、非標準の対称操作、単位セル中の分子数(分子式数)、密度の計算値、密度の測定値(およびその標準偏差)、測定温度、融点、原子座標、結合の長さ、結合データ等である。

XDCでは配布されるテープが、新しい文献・数値だけでなく、既収録のものも訂正も含めた全文献・数値を再録した新版の発行という形で行われる点、CASの

場合とは考え方が異なり、データ・ベースの更新方法にも差が生ずることになる。なお、XDCは文献情報だけでなく、数値データも持っているため、これに対応する応用システムが、野上法正氏（東大・理）によって作成され、50年9月から公開されている。

### (3) INSPEC データ・ベース

INSPEC (International information services for the physics and engineering communities) 磁気テープは、英国 IEE (The Institution of Electrical Engineers) より刊行されているもので、物理、電気・電子工学、計算機・制御の3分野にわかれている。TOOL-IRでは、この内の Computer and Control セクション (INSPEC-1 C) を導入しているが、これは月刊で、年間収録文献数は約2400件である。以上3種のソース・テープの内容をよとめると表6の通りである。

表6 ソース・テープの比較

種別	フォーマット	刊行	収録数	分野	内容	1件当りの収録
CAS	SDF <sup>(1)</sup>	週刊	35万件/年	化学	書誌事項・キーワード	572B
XDC	カード・インジ	6週間毎	14,000件	結晶学	書誌事項・結晶構造データ	3580
INSPEC	MARC方式 <sup>(2)</sup>	月刊	24,000件/年	計算機・制御	書誌事項・キーワード・要旨	785

(1) Standard Distribution Format — CAS独自の形式 (2) 米国 Library of Congress の標準形式

### 図1 INSPEC の出力例

```

SER#: 847026          TYPE:06          ABS#:C7601281
AUTH:DIJKSTRA, E.W.
TITL:CORRECTNESS CONCERNS AND, AMONG OTHER THINGS, WHY THEY ARE RESENTED
HTTL:SIGPLAN NOT. (USA) CODN:SINODQ
VISS:VOL.10, NO.6   PAGE:546-50      DPBL:JUNE 1975
AAFL:BURROUGHS CORP., NUENEN, NETHERLANDS NREF:0
TEXT:PROVIDES AN HISTORICAL PERSPECTIVE OF SOFTWARE RELIABILITY AND PROGRAM
CORRECTNESS
CIDX:FAULT-TOLERANT COMPUTINGPROGRAMMING
FIDX:HISTORICAL PERSPECTIVEPROGRAM RELIABILITYPROGRAM CORRECTNESS
SCOD:C8120          UCOD:WCAAAF      TRMT:G P
CTTL:INTERNATIONAL CONFERENCE ON RELIABLE SOFTWARE
LCNF:LOS ANGELES, CALIF., USA          DCNF:21-23 APRIL 1975
SPSR:IEEY ACMY ET AL.
  
```

```

SER#: 847816          TYPE:06          ABS#:C7602101
AUTH:DIJKSTRA, E.W.
TITL:GUARDED COMMANDS, NON-DETERMINACY AND A CALCULUS FOR THE DERIVATION OF
PROGRAMS
HTTL:SIGPLAN NOT. (USA) CODN:SINODQ
VISS:VOL.10, NO.6   PAGE:2-14      DPBL:JUNE 1975
AAFL:BURROUGHS, NUENEN, NETHERLANDS     NREF:2
TEXT:SO-CALLED 'GUARDED COMMANDS' ARE INTRODUCED AS A BUILDING BLOCK FOR
ALTERNATIVE AND REPETITIVE CONSTRUCTS THAT ALLOW NONDETERMINISTIC PROGRAM
COMPONENTS FOR WHICH AT LEAST THE ACTIVITY EVOKED, BUT POSSIBLY EVEN THE
FINAL STATE, IS NOT NECESSARILY UNIQUELY DETERMINED BY THE INITIAL STATE.
FOR THE FORMAL DERIVATION OF PROGRAMS EXPRESSED IN TERMS OF THESE
CONSTRUCTS, A CALCULUS IS SHOWN
CIDX:PROGRAMMINGPROGRAMMING THEORYFAULT-TOLERANT COMPUTING
FIDX:CALCULUS FOR THE DERIVATION OF PROGRAMSGUARDED COMMANDSREPETITIVE
CONSTRUCTSYNONDETERMINISTIC PROGRAM COMPONENTSYFORMAL DERIVATION
ALTERNATIVE CONSTRUCTSYPROGRAM SEMANTICS
  
```

### 3. データ・ベースの保守

TOOL-IRのデータ・ベースは、新しいテープの到着に従って更新される。ここでは、容量の大きいCASデータ・ベースの場合を中心として、データ・ベースの更新手順について説明する。更新処理には2種類ある。すなわち、テープの到着の後にこれを交換し、データ・ベースに追加してゆく処理で、CASの場合では、ほぼ2週間毎に行われる。(週刊のテープ之本をよとめて更新する。)これを「通常更新処理」と呼ぶことにする。通常更新によつて、データ・ベースは逐次拡大して

ゆくが、これを無限に継続することはオンライン・ディスク容量の制限から不可能であり、一定間隔で古くなる文献を追出し、データ・ベースを再編成する操作が必要になる。これがオスの更新処理で、これを「再配置処理」と呼ぶことにする。CASの場合、この配置を約半年ごとに行っている。

このような更新処理の方法にあわせて、検索の場合の利用者用コマンドも、対象とするデータ・ベースを指定するように設計されている。(表3) NEWCAS コマンドは、最近着のテープに含まれる文献のみを対象とする検索を行うもので、新規のユーザ以外は、このコマンドによって自己の研究分野についての検索を更新のたびに実行しておれば、いわゆる *Current awareness* を確保できる。CAS コマンドは、NEWCAS の文献を逐次累積したものを検索するもので、CAS の場合半年分を1巻としているので(テープ1本を号と呼ぶ)、巻の切替時期にいったんカラになり、半年間内で満杯となる。CAS 83、82 等々のコマンドは以前の巻(バックナンバー)に該当するもので、遡及検索の際に使われる。

通常更新処理とは、従って、NEWCAS のデータ・ベースの入替と、CAS データ・ベースへの最新号を追加することであり、再配置処理とは、半年毎の巻の切替時期に、それまでの CAS データ・ベースを CAS 72 (表3の例では CAS 84) として分離し、新しい CAS データ・ベースの蓄積を開始することである。

#### (1) 通常更新処理

通常更新処理は、次の3つの作業からなる。

① キーワード・着者名の抽出——テープ中の文献からキーワードと着者名を抽出して、②の転置ファイル用のデータを作る。キーワードは、CAS や IEE によって各文献に付加されている、いわゆるキーワード、あるいはキーフレーズを使うだけでは不十分であるので表題及び、論文要旨等からも新たな抽出処理を行っている。

② 転置ファイルの作成——抽出されたキーワード、着者名をソートし、転置ファイルを作る。転置ファイルは、NEWCAS 用に新しいものが作成され、また CAS 用については、併合処理が行われる。

③ 文献レコードの線型ファイルへの追加——テープの内容は、TOOL-IR のための簡単な変換をほどこした後(文献 ID のくくり出し)、CAS 用線型ファイルに追加(ファイル拡張)され、また NEWCAS 用に新しいものが作成される。

④ Welcome メッセージの更新——TOOL-IR では主コマンド投入直後に出力される Welcome メッセージの中に、ユーザへの伝達事項を挿入しているが、こゝにデータ・ベースの更新状況を表示する。

#### (2) 再配置処理

再配置処理は定型的な作業となるはずであるが、オンライン・ディスク容量の制約から、次のようなややめんどうな手順を踏んで行っている。即ち、以前の巻(CAS 83 等々)すべての線型ファイルをそのままオンラインで保持しておくことは容量上不可能なので、文献情報の本体である線型ファイルは、実際には CAS 用のもの及び一つ以前の巻(現時点では CAS 83 用)が保持されている。従って、巻の切替時期にも、少なくとも半年分のデータは保存されることになるが、それ以前のものを単純に消去してしまうのでは、利用者へのサービスが低下する。そこで、次に述べる投票方式によって、利用者の必要とするデータだけは保存しておくことにした。

投票方式——巻の切替に先立って、ユーザに必要と思われる文献のリストアップをしてもらう(投票依頼)。TOOL-IR 管理者は、スペースの許す範囲内で、投票

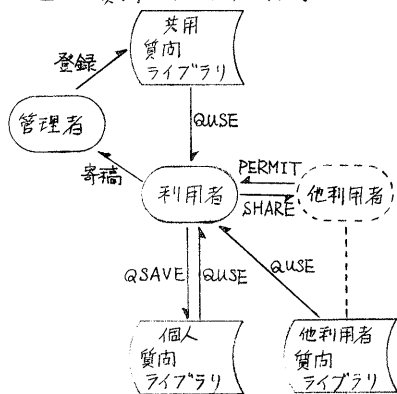
の多から、文献を線型ファイルに保存しておく方式である。但し、検索の依頼及び、必要文献のリストアップともに手作業による文書授受で行うとすれば、管理者、ユーザ各々の負担が大きくなり、円滑なシステム運営が阻害されることになる。そこで、TOOL-IRの検索サブコマンド、及びオペレーティング、システムのファイル共用機能とメモリー機能とを組合せて投票方法を考案した。これによれば、ユーザは、保存してほしい文献リストを通常の検索セッションの中で、個人ファイルとして作成した後、管理者にこのファイルのアクセス許可を与えるだけでよく、管理者側では、これらのファイルを併合して、全体としての保存要求文献リストのファイルを作成する。

なお、再配置処理については、特にこれを設けず、通常更新処理の中でデータ・ベース内容の先入先出の管理を行う方法なども検討したが、処理時間、ワーク・ファイル容量等々に作業上及び手続上種々の難点が指摘され、現在の方法に落ち着いたものである。

### (3) 質問ライブラリの保守

TOOL-IRにおける検索は、比較的単純な機能の検索コマンドを組合せて行うように設計されているが、<sup>(4)</sup> 実際問題として機能検索には、それなりのコツ、要領といえるものがあって、これは特にキーワードの選択と、それらの間の論理演算の施し方の面で顕著である。そして、適切な文献リストを得るには、結局、当該分野に対する専門知識が必要となることが多い。しかしこれでは、多少とも自分の専門とすれば分野に關した文献を捜そうとする場合、必ずしも好結果が得られないことになる。そのためTOOL-IRでは、個人用の質問ライブラリとは別に、ユーザ共用の質問ライブラリを備えている。<sup>(3)</sup> これは各ユーザが、各々の分野での検索の経験(検索された結果と要求するものとの異同をチェックすることは、かなり難しい作業で、専門知識も必要である。)の積み重ねから、良好な一連の検索コマンド群(以下質問という)が得られた場合に、これを個人用ライブラリに登録しておくことととどめず、TOOL-IRユーザ全体に公開してもらう場である。現在までのところ、公用文献の検索等20件程が寄稿され、登録されている。これらの質問ライブラリの利用状況は、後述の利用設計システムで把握している。

図2. 質問ライブラリの保守



新しい質問の登録、即ち、質問ライブラリの更新は、作業量も少ないので、寄稿の毎に随時行っている。そこで問題は、むしろ、いかにして新しい質問の寄稿を促進するかであり、この点、後の問題点の項でもふれる。

## 4. 利用統計からみた利用状況

TOOL-IRに組込まれている利用統計システムの概要及び、75年6,7,8月の統計は、既に報告したが、<sup>(2)</sup> その後の経過もあわせてここにまとめてみる。

利用統計システムは、TOOL-IRを1年程の試用期間を経て、東大大型計算機センターの全ユーザへ公開するに伴って稼働させられたものである。75年6月分以降の利用状況が蓄積されている。XDCについては、75年11月分から採取を始め、76年、INSPECに關しては、データベース公開と同時に76年2月分より統計採取を開始した。統計採取の目的は、オービシステムの利用

状況を正確に把握し、システムの改善のための資料を得ることになり、オニに、ソースデータ供給機関への集計報告の作成があげられる。採取項目は、ユーザのプライバシー保護に留意して下記の項目とし、月間の集計結果をNewsletter等でユーザへ報告している。

採取項目	
job番号, job-id, user番号, 使用端末番号, 月日, 開始時刻, 経過時間, CPU時間, 使用検索コマンド種別, 出力文献id, 出力モード, 主コマンド種別(対象データベース), 質問ライブラリ質問名	

(1) 統計表

統計結果のリストは、現在下記①②の二種が月次に作成されており、③以下のリスト作成のプログラムを準備中である。

① 統計項目ダンプリスト——これは統計システムで得られた事項に必要な変換を加え、その子系出力しているもので、主としてチェック用に使っている。

② ヒット・カウント・レポート——これは統計項目の中の一般的なものについて、セッション毎の情報を日別時間順に出力し、日計と月計をとるもので、Tool-IR全体をとりまとめたものと、CAS, XDC, INSPEC各々のセッションだけを抜き出したものの計4種類が作成されている。以下の統計結果の分析は、主として、このレポートに基づくものである。

③ 地域別、ユーザ別ヒット・カウント・レポート——これは②とほぼ同一の内容のものを、分類キーをユーザ番号にしてソートし、集計を施して出力するもので、新ユーザの増え方や、各種の値の一人当りをみるのに便利なものである。

④ 質問ライブラリ利用表——これは、前述の公開されている質問ライブラリに含まれる各質問の利用頻度などを把握するもので、質問ライブラリ充実のための基礎的資料として使っている。

⑤ 検索文献リスト——検索された文献のIDを統計項目として採集しているもので、これらをもとめて再びデータベースを検索して、文献リストを得、これに種々の観点から分析を加えようとするものがこれである。当面、雑誌論文に関しては、掲載誌別のカウントをとり、多く参照される雑誌は何かを明らかにすることが考えられる。これは、図書館等で購入すべき雑誌を決定する際に参考になろう。また、文献リストに含まれるキーワードを抽出して、頻度表を得ることによって、研究の全体的な傾向、研究者集団の関心の動向を判別できると思われ、これは、研究者全体へなんらかのヒントを与えることになろう。

(2) 利用状況の概括

Tool-IRの現在までの利用状況を示す主なデータを表7, 8にまとめる。

表7 CAS データベース利用実績 (1975.7.1~1975.12.25)

実利用者数	132名							
運転日数	136日							
利用回件数	1542件							
主コマンド数	2635個							
経過時間	延318時間33分							
CPU時間	延34844秒							
コマンド 使用回数	種別	SEARCH	AND	OR	DIF	EXTRACT	ENTER	計
	総数	16372	6542	1881	351	20	145	25311
	コマンド当り	6.21	2.48	0.71	0.13	0.01	0.06	9.61
出力文献 レコード数	種別	書誌情報	内容情報	両方	全情報なし			計
	総数	2527	7293	82495	172			92487
	コマンド当り	0.96	2.77	31.31	0.07			35.10
	利用者当り	19.14	55.25	624.96	1.30			700.66

表8 XDC データベース利用実績 (1975.11.1~1975.12.25)

実利用者数	33名							
運転日数	55日							
利用コマンド数	99件 3.00件/利用者, 1.80件/日, バッチ15件, TSS 84件							
コマンド総数	118個 3.58個/利用者, 1.19個/日, バッチ18個, TSS 100個							
経過時間	延14時間46分 0分43秒/バッチコマンド, 10分25秒/TSSコマンド							
CPU時間	延521秒 7.39秒/バッチコマンド, 5.11秒/TSSコマンド							
サブコマンド 使用回数	種別	SEARCH	AND	OR	DIF	ENTER	計	
	総数	321	68	6	3	1		399
	コマンド当り	2.72	0.58	0.05	0.03	0.01		3.38
出力コマンド数	種別	化合物名	文献情報	化学構造情報	全情報		計	
	総数	3479	1276	80	63			4898
	コマンド当り	29.48	10.81	0.68	0.53			41.51
	利用者当り	105.42	38.67	2.42	1.91		148.42	

表9

1975年12月末現在の利用者所属

所属	CAS	XDC
東北大	1	
宮城教育大	2	
群馬大	2	
筑波大		1
千葉大	2	
東大(本郷)	78	14
(駒場)	7	3
(センター内)	7	4
電通大	4	2
玉京大	3	1
東工大	1	2
立教大	1	
早大	1	
国文学研究資料館	1	
横浜国大	2	
横浜帝大	1	
信州大	2	
名大	2	1
富山高専	1	1
京大	1	1
阪大	5	2
南西学院大	1	
姫路工大	4	
広島大	1	
九大	1	1
計	132名	33名

公衆回線端末, リモートバッチ端末等からの利用を含めたユーザの全国的分布状況は, 表9の通りである。遠隔地ユーザは, オンライン検索の他, センターで行っているクロード・バッチ処理(郵送サービス)による利用が可能である。今後, 大学間コンピュータネットワークの充実, 公衆回線端末の普及等による利用者の拡大を期待したい。

利用統計からみて平均的な利用形態は次のようである(CASの場合)。1回の検索に, 9~10個のサブコマンドを使用し, 35文献を出力する。コマンドの種別をみると, SEARCH 6.2回, AND 2.5回, OR 0.7回のように使われている。CPU時間は, バッチの場合45秒, TSSでは15秒であるから, バッチの方が複雑な検索と, 大量の出力のために使われていることになろう。TSSでの検索は, 1回, 15分程度(経過時間)で行われている。また利用者当りの数字では, ジョブ件数が6ヶ月間で11.7件であるから, 月2回程利用していることになり, データベース更新サイクルの2週間とほぼ一致するが, 途中からの新規利用者も多いので, 実際には, 月3~4回は利用していると思われる。同じく6ヶ月間で利用者当り700文献(月間120件)という出力文献数もやや多いような気がするが, これは, 新規利用者による検索出力が反映しているように思われる。

(3) 判明した傾向, 現象

6ヶ月間の統計の蓄積によって判明した傾向, 現象について述べる。但し, この間, 公衆回線端末の付加等によって, 新規利用者の加入が多く, 統計の変動も大きいので, 一概に結論できないかもしれない。

① 利用量の変動 —— 月別にみた場合, 利用量は大学という環境を反映してか, 夏休み等休業期間にははかなりおろこんでいる。(システムは休業期間中も稼働している。)そこで, この期間はデータベースの更新間隔を上げることがよいといえるかもしれない。

② KEEP/ENTERサブコマンドの利用 —— KEEP/ENTERサブコマンドは, 検索結果をユーザの個人ファイルに保存し, また, これを後の検索時に使用するためのコマ



ンドで、検索時間の節約ができ、また、データ・ベースの内容の推移をみるのに便利であるが、このサブコマンドの利用回数は逐次増加している。このサブコマンドは、検索自体に不可欠のサブコマンドではないから、当初からこれを使うユーザーは少なく、ある程度システムとなじんでから使い始めると考えられ、これが統計データにも現われているのであろう。

③ 主コマンド当りサブコマンド使用回数の変化——主コマンド当りサブコマンド使用回数が増加していることは、平均して質問がより複雑になってきていることを示す。これは新規利用者がいるが、なにより経験者が次第に増加しているところから当然のことといえる。

しかし、ここで興味があることは、SEARCH, AND, OR の3つのサブコマンドの中で、ORサブコマンドの増加傾向がみられることである。SEARCHサブコマンドは、ORの機能を含むので、単一の概念についての検索には必ずしもORサブコマンドを使う必要はない。しかし、機械と対話しながらオンライン検索していく場合、慣れるに従って、複雑な論理式を作って一発回答を期待するよりも、小出しに質問を出してその結果をみながら検索を組み立ててゆく方が、よりよい結果が得られることかあかってくる。<sup>(4)</sup> ORサブコマンドがより多く使われるようになってきているのはこのような理由かと思われる。

#### (4) 検討したい仮説

次に、今後検討してみたい仮説を挙げてみる。

① データ・ベース更新と利用量の関係——current awareness を重視するユーザーの立場からすれば、更新の都度、自己の関心分野について検索してみるという利用形態が考えられ、利用量の変動の一部が、更新の回数と関係することは当然ありうることで、この点に注目して、統計値の整理をしてみたい。なお、current awareness に関しては、現在、TOOL-IRの一部として、SDI を試行中であり、近々実用に移す予定である。

② 新規ユーザーの動向——新しく TOOL-IR の利用者になった人が、どんな経過で、システムとなじみ、これによって利用形態が変化してゆくかを知ることは、システムを運営する側にとり、システムとその運用方法の改善、利用者のための講習のしかたなどに非常に参考になるものと思われる。そこで、このようなデータを集める方法を考えてゆきたい。

③ データ・ベース別の利用者の動向——データ・ベースの内容の差異、即ち利用者の研究分野の違いは、システムの利用形態にも反映されると思われる。そこで、今後稼働させたい INSPEC の利用状況を CAS と対照しながら追うことで、利用者集団の特徴といったものが引出せるのではなからうか。

### 5. 運営上の留意点と今後の課題

情報システム運営の基礎は、システム管理者と、利用者間及び利用者相互間の意思疎通の円滑化にある。情報システム自体の機能として、本システムの場合であれば、本来の情報検索機能、情報提供機能に加えて、システム管理者と利用者というシステムをとりまく人々の意思伝達の手段としての機能が極めて重要で、これはシステムの長期的な成長・発展の成否を握るものである。本節は、このような運営上の留意点、問題点と今後の課題・計画等について述べる。

#### (1) 利用者の反応とシステムの改良

CAS データ・ベースは、74年夏に試用が始められた。当時、DISPLAY サブコマンドには文献番号だけが表示される暫定版を使っており、物理化学関係の利用者が

新か、 $\pi$ 。その後コマンドが完備するとともに、有機化学関係の利用が増加した。さらに毎年秋に公衆回線端末が利用可能になると同時に、利用者数は急増した。この間、熱心な利用者が次第に多くなり、ましておし、問合せ、苦情、要望、利用状況等々にフリエ、システム管理者との通信が頻繁となっている。

XDCデータベースの利用者はシステム開発の当初から比較的熱心な人々が多く、現在に至っている。

システムの使用開始後、利用者の要望をとり入れて、下記のような改良をシステムに施した。

①ストップワードの削減——CASにフリエ、キーワードを抽出する際に、無視する語(ストップワード)のリストとして、当初、Chemical Tablesのもの参考し、1233語を用いてシステムの運用を始めるところ、STRUCTURE、HUMAN、DATA等の重要な語がこれに含まれており、検索上不便であることがわかり、利用者の協力によってこのリストの見直しを行い、これを345語までに削減した。

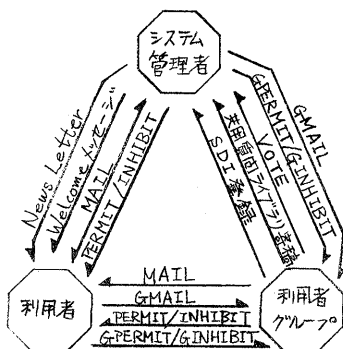
②化合物構造の表示その他——XDCにフリエ、DISPLAY DUMPコマンドによる内容表示方法に化合物構造の線型表示をとり入れるなど、利用者の要望に合わせて改良を加えた。同時に、検索したデータを、FORTRANプログラムで扱ったり形で、個人ファイルへ転写するCOPYコマンドも付加した。

③逆引キーワードの導入——TOOL-IRでは、開発に際して、CASテープを検討の結果、キーワードは8文字超過分は切捨てとし、また、検索では、右側切捨検索(語尾変化を許す)を可能なものとする設計を行った。<sup>(1)</sup>その後XDCにフリエ、8文字をこえる化合物名の一致検索の要求があり、さらに今回INSPECの導入に当たって、MICROPROCESSORとMICROPROGRAMのように従来のままでは、正列しにくい重要な語の例が散見されるなどの事情から、9文字以上の語に關して、語尾から8文字の逆引キーワードの抽出と、左側切捨検索(語幹変化を許す)の導入を行った。

## (2)利用者・管理者間の関係強化

①利用者・管理者間の連絡——TOOL-IRでは、ユーザとシステム管理者間の意思伝達機能を重視し、種々の対策を施しているが、計算機システムを通じた連絡経路として、図3に示すようなコマンドを作成している。<sup>(3)</sup>この内、WelcomeメッセージとGMAILコマンドが管理者からユーザへのメッセージ伝達に多用されている。Welcomeメッセージは、前述の通りTOOL-IR主コマンド投入直後に表示されるもので、データベース更新の都合をさしかえて更新の状況をこれにて通知する。その他システムの変更拡張等もこれにて通知しているが、システムの事故など緊急事項は、次のGMAILコマンドをユーザに通知する。

図3 利用者・管理者間の連絡



GMAILコマンドは、OSのMAILコマンドを元にして作られたマクロコマンドであって、TOOL-IRユーザ登録簿(SYSUIDファイルと称している)を参照して、全員もしくは特定のユーザ群にメッセージを送り出すコマンドである。Welcomeメッセージと異なる点は、ユーザが、TOOL-IRに限らず東大大型センターにJOBを投入するとMailの配達がおSから表示される点であり、あまり頻繁でなくユーザへも通知が届きやすいことである。

ユーザから管理者への苦情・問合せ等も、MAILコマンドを利用できるが、これは計算機システムになれどユーザに限られるようで、一般には、電話・手紙による連絡が多い。

Newsletterの発行——管理者からユーザへの伝達事項をのせたTOOL-IR Newsletterをほぼ毎月1回の割合で作成し発送している(現在130部程度)。これには投票の依頼、質問ライブラリの追加、その他各種の連絡事項をのせており、先述の統計データもこの場で報告している。さらにユーザの寄稿もあって、総合的な機関誌の役割を担ってきている。

②ユーザ登録簿の整備——ユーザ登録簿ファイルは、現時点では、グループ別ユーザIDを収録しているのみで、またその登録もオフライン的に行われている。今後ユーザ数の増加に鑑みて、この登録簿を中心とするシステムを作成し、本格的なユーザ台帳システムにする必要があろうと思われる。

③投票システムの運用——先述の投票システムは、昨年秋に一発実施したのみで、まだ経験が浅く、投票の仕方も計算機システムに不慣れなユーザにとっては、やや繁雑であると感じられる点もあるようである。今後これらの技術的な改善を行って、より活発な投票(棄権防止)を得られるようにしたい。

### (3)応用システムの拡充

TOOL-IRを利用したシステムとしては、既にXCALCシステムがある。これは、前述の通り、TOOL-IR利用者の一人である、上野上氏によって開発されたものである。現在、このような形での応用システムの開発が、いくつか計画されており、これらの推進に努めている。

謝辞——TOOL-IRの開発にあたって討論をいただいた多くの方々、特に東大理学部・戸沢義夫氏、日製製作所・吉田郁三氏、金沢恵作氏、電通大・山崎昶助教授に感謝します。

### 関連文献

- (1) 山本, 根岸, 岡田, 山田, 牛丸, 戸沢, 吉田, 金沢, 山崎, 藤原: TOOL-IRシステムI~V情報処理学会  
オ15回大会講演予稿集, 1974
- (2) 山本, 山崎, 藤原, 根岸, 長町, 島内, 牛丸: TOOL-IRシステムVI~IX情報処理学会  
オ16回大会講演予稿集, 1975
- (3) 山本, 根岸, 牛丸, 戸沢, 岡部, 藤原: TOOL-IR: An On-line Information Retrieval System  
At An Inter-university Computer Center, 2nd USA-Japan Computer Conference Proceedings, 1975
- (4) 根岸, 山本: オンライン文献情報検索システム: TOOL-IRにおけるマン・マシン・インターフェース, 情報処理, 掲載予定