

# 情報検索システムへのMSS導入に関する一考察

藤 鉄雄, 高橋章二, 片岡雄二, 森原一郎  
 (電電公社 横須賀電気通信研究所)

## 1. はじめに

オンライン情報検索システムにおける記録媒体には、DASDを使用するのが一般的であるが、システムの経年化に伴いデータ量が増大する一方、アクセス特性に変化が生じる。この為アクセス頻度の低下したデータはデータの保存コスト上、MTに格納し、必要に応じてオフラインで処理したり、削除したりする傾向にある。<sup>(1)~(5)</sup> この結果、MTに退避されたデータを手りするには、日単位の時間を必要とすることとなる。

電電公社においてもDIMS-E上でDASDを記録媒体として使用したオンライン情報検索システム“DORIS-2”を提供しているが、一定の期間を経過したデータの扱いをどうするか課題となっている。

筆者らはオンライン情報検索システムにおける上記問題に対処する一方式として、経年化したデータはDASDよりMSSへ移送して保存することにより、保存コストを低減させるとともに、オンライン処理も可能となる方式について検討を行った。

本報告では、オンライン情報検索システムへのMSS導入に関し、以下の項目を中心に述べる。

- (1) 処理方式
- (2) ユーザインタフェース
- (3) データの管理方式
- (4) 評価

## 2. オンライン情報検索システムの特徴

我が国におけるオンライン情報検索システムの研究は昭和44年、科学技術会議諮問オ4号として、「科学技術情報の全国流通システム(NIST)」構想が発表されて以来活発化し、各種の公的サービスも開始されるに至った。

(表1に我が国における代表的なオンライン情報検索システム(科学技術関連)の例を示す)<sup>(5)~(10)</sup>

オンライン情報検索は

- (1) 要求が生じたその時突で膨大なデータの中から、瞬時にほしい情報が入手出来る。
- (2) 検索処理に若干の知識を

表1. 我が国における代表的なオンライン情報検索システム(科学技術関連)

提供機関	システム名	主な提供データベース	開始時期(年)
日本特許情報センタ	PATOLIS	国内特許 実用新案 米国特許	46
日本科学技術情報センタ	JOIS	MEDLARS(医学) CCondensates(化学) TOXLINE(毒物学) JICST 理工学	47
東京大学大型計算センタ	TOOL-IR	CACCondensates(化学) INSPEC(物理, 電気) XDC(X線結晶文獻データ)	48
筑波大学学術情報処理センタ	IDEAS-77	BAPreviews(生物) CASearch(化学) COMPENDEX(工学) ESI(環境) Excerpta Medica(医学) SCII(引用文献)	52
広島大学	HUNDRED	BIOSIS(生物) INSPEC(物理, 電気) Index of Mathematical Papers(数学)	52
東北大学大型計算センタ	UNIQ-1	METADEX(金属)	55

持つのみで誰でも簡単に処理出来る。

(3) 全国どこでも一律のサービスが受けられる。

等のメリットがあり、さらに情報に対する対価支払いについての社会通念が固まって来たことと相まって、広く世の中に浸透して来ている。

一方、サービス提供者はユーザに最新情報を提供するために、DBを常に最新の状態に維持してゆく必要がある。

データ量はシステムにより異なるが、一例を現在市販されている科学技術関連のデータで見ると、表2にその例を示すように、毎年勿量のデータが提供されており、これらを用いたDBの更新処理が実施されている。

電電公社においてもオンライン情報検索システム「DORIS-2」をDEMOS-E上で昭和50年より提供しているが、ここにおいてもデータ量は毎年20~30%の増加率を示している。

一般に情報の寿命は蓄積する情報の内容、並にそれを利用する利用者により影響される為、普遍的でないが、指数関数的に変化する傾向にある。<sup>(3)(4)</sup> (図1参照)

このためオンライン情報検索システム提供者にとっては新規データの追加とともに、経年化したデータの処理が必要となるが、データ量の増加については、経年化したデータの削除、MTへの退避による対応の他、記録媒体(DASD)の増設や、システム更改で対応する傾向にある。<sup>(2)(3)</sup>

MTへ退避されたデータへのアクセスは必要に応じオフライン処理が実施されるが、これには以下の問題がある。

- (1) 検索式を十分考慮して作成しておかないと、結果が期待したものと反する危険が大きい。
- (2) オフライン処理に伴う運用が伴う
- (3) ターンアラウンドが遅くなる。

筆者らはこの問題を解決する方式として、経年化したデータはDASDからMSSへ移行して保管する方式について検討を行った。

### 3. MSS導入形態

DASDの他にMSSを記録媒体として使用したオンライン情報検索システムを構築する場合、記録媒体の利用は以下の分類に従って使用するのが性能、コスト等の観点から効果的である。

- (1) DASD ; ① 検索性インデックス  
                  ② アクセス頻度の高いデータ
- (2) MSS ; ① アクセス頻度の低いデータ

表2 データの発生量

種別	件数(万件/年)
CAS	30~35
INSPEC	15~16
MEDLARS	23~24
JICST文献	33~40
国内特許 実用新案	32~33

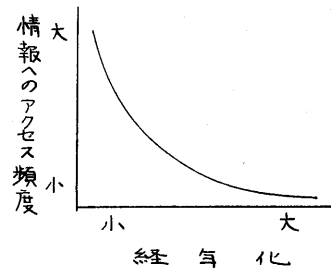


図1 データの経年化とアクセス頻度

一般に情報検索サービスはサービス提供者とその利用者、並にシステム運用者に分類して実施される。

MSSを導入してオンライン情報検索サービスを行う場合、上記担当者の対応は以下の通りである。

(1) サービス提供者

- ① 端末よりDASD上にデータベースの創成を行う。
- ② DASD上へデータの追加を定期的に行うとともに、利用状況を監視する。
- ③ DASDの空間が不足したり、データベースの利用率が低下するとDASD上のデータをMSSへ初送する。
- ④ DASD上のデータをMSS上へ初送する時期 (ex. 3年分のデータはDASD上に、それ以前のデータはMSS上) を利用者に予め通知する。

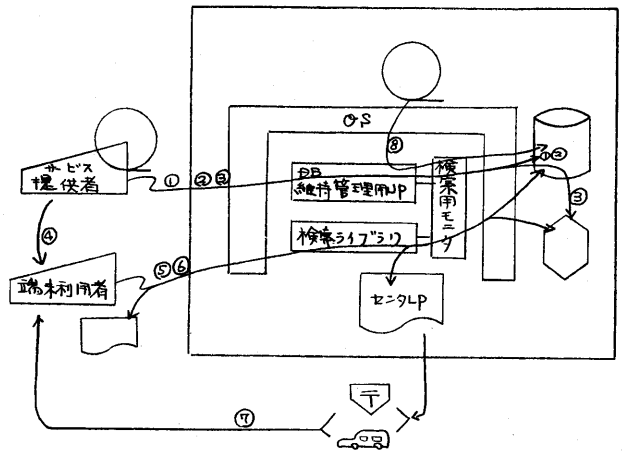


図2 MSSを用いたオンライン情報検索の処理概要

(2) 利用者

- ⑤ 回線を通じて検索処理を行なう。
- ⑥ 最終的なデータの入手は、処理時間(MSSへのアクセス時間)、出力量等を考慮し、出力先(端末/センタ)を指定する。

(3) システム運用者

- ⑦ センタ出力が指定されたデータを郵送等により利用者へ配布する。
- ⑧ データベースの維持(生成、更新)がセンタ処理の場合、データベースの維持処理を行なう。(以上の処理形態を図2に示す)

## 4 記憶構造

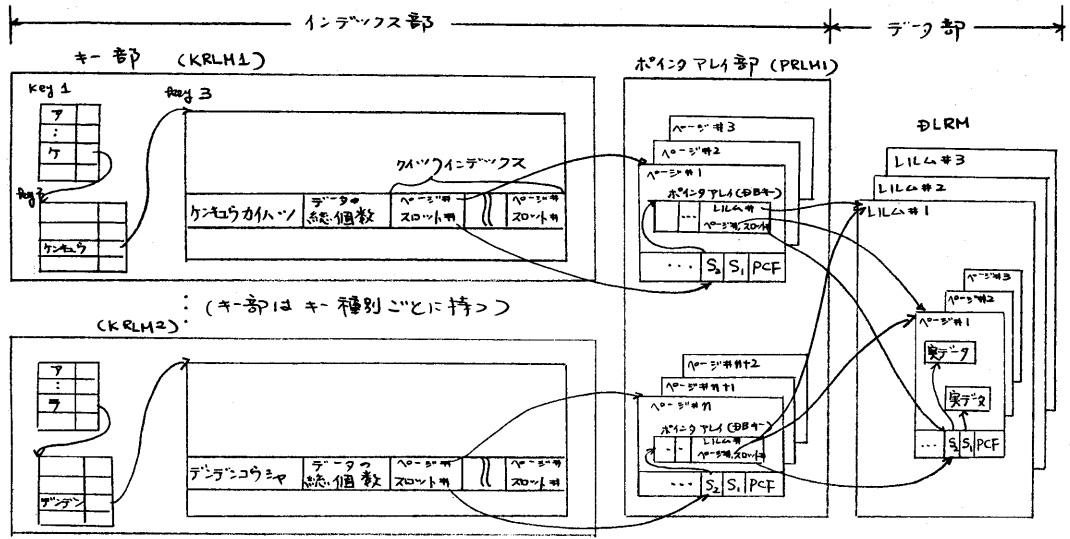
### 4.1 記憶構造の概要

情報検索システムにおける記憶構造はデータの増加への柔軟な対応、ファイルエリアの効率等を重視し、インバーテッド・インデックスをベースとした記憶構造であり、MSSの導入においても本構造とする。

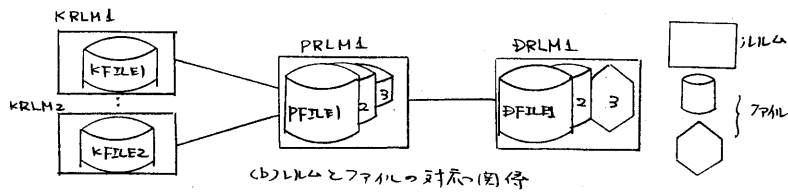
なおキー・インデックスとデータ本体は3. で述べた分析に従って配置する。

インデックス部はキー部とポインタアレイ部より構成され、キー部は各キー項目ごとに該キーに属するデータの総数とクイックインデックスで構成する。ポインタアレイ部はキー項目と実データとのマッピングをDBキーとして保持する。

図3に記憶構造の概念図を示す。



(a) インデックス部とデータ部の構成



(b) LILMとファイルの対応関係

図3 記憶構造

#### 4.2 データベースの再編成処理(テープのDASD → MSSへの移行)

DASDからMSSへのデータの移行はサービス提供者がデータの利用状況、ファイルの空状況等を考慮して実施する。この場合、移行単位としては①ファイル、②データ、③キーが単位となり得るが、データの利用特性は図1に示した如く、経年化に伴って低下すること、データの追加処理は一定期間のものをまとめて行なわれる傾向にあること、移行処理の容易性等を考慮し、ファイル単位とする。

移行に伴う管理情報の変更としては、論理情報(DBMSで管理する情報)、物理情報(OSで管理する情報)がある。

論理情報の変更としては、事前検索時に使用するデータの所在情報(DASD上 or MSS上)を、物理情報の変更としては、要求データへアクセスする時使用する物理媒体上のアドレスをそれぞれ変更する。

データの移送単位をファイルとすることで、上記2項目の管理情報の変更のみで、処理プログラムに影響を与えることなく、対応可能としている。

図4にファイルの移送と、それに伴う管理情報の変更状況を示す。

### 5. 端末利用者インタフェース

システムで使用する記録媒体は仮想化し、端末利用者は意識しないのが通常で

あるが、MSSを記録媒体として使用する場合、アクセスに数十秒以上を必要とする。この為MSSへのアクセス中、端末は無応答状態となり、端末利用者に無用の誤解を与える要因となる。

この問題に対処する為端末コマンドの延長上でMSSへアクセスする必要のある場合には、端末利用者が該処理を認識した上で実行する方式とした。

表3にDORIS-2が提供するサブコマンド一覧を示す。

これらサブコマンドにおいて端末利用者が記録媒体を意識する必要のあるサブコマンドとして、検索条件式を投入する「/LOGIC」サブコマンド、検索結果を出力する「/KPRINT」サブコマンドを例にとりユーザインタフェースの変更状況について述べる。

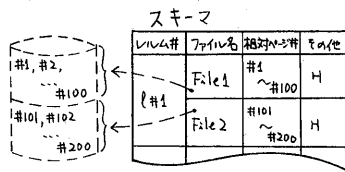
### 5.1 /LOGIC サブコマンド

検索条件式の投入は/LOGIC サブコマンドを用いて行う。

本サブコマンドは指定された条件式をもとにキー部をサーチし、条件に合致したデータ件数をお出力する。

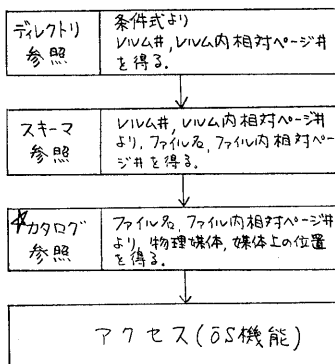
この場合、該当するデータの総件数とともに、データの所在(DASD上 or MSS上)も分類して出力する。(図5参照)

### 論理的情報の管理(DBMS)



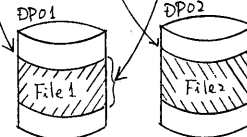
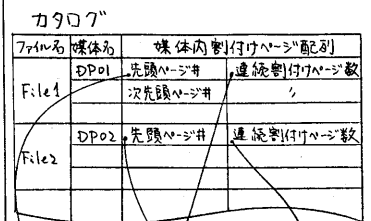
File 1 をMSSに転送した場合は H→L に修正。

### (DBMSの処理)



※ 本処理においてカタログがMSV用に修正されていることにより自動的にMSSにアクセスする。

### 物理的情報の管理(OS)



File 1 をMSSに転送

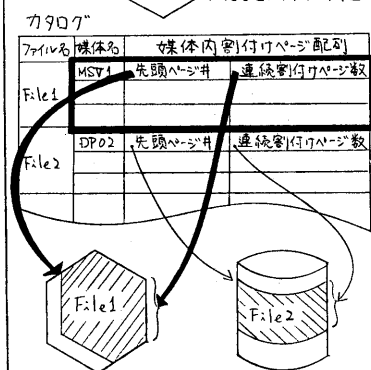


図4. ファイルの移送に伴う管理情報の変更

表3 DORIS-2 サブコマンド一覧

番号	サブコマンド名 正式名	略称	機能 概要
1	/RANGE	/RAN	カビス内の区間を指定し検索の範囲を設定する。
2	/EXPAND	/XEP	指定された語幹を持つ索引語とカビス初階に通知する。
3	/THESAURUS	/THE	指定された索引語のシノニムとカビス利用者へ通知する。
4*	/LOGIC	/LOG	論理条件式による検索を行う。
5	/WEIGHT	/WEI	ウェイト条件式による検索を行う。
6	/SEARCH	/SEA	サーチ条件式による検索を行う。
7	/PRINT	/PRI	検索結果を編集し出力する。
8	/DELETE	/DEL	部分集合の削除を行う。
9	/GUIDE	/GUI	部分集合の案内を行う。
10	/SAVE	/SAV	質問式をPFUファイルに保存する。
11	/RATE	/RAT	使用量とカビス利用者へ通知する。
12	/INCLUDE	/INC	包摂条件式による検索を行う。
13	/TSEARCH	/TSE	TFサーチ条件式による検索を行う。
14*	/KPRINT	/KPR	正式による検索結果の高度出力を行う。
15	/INFORM	/INF	カビス案内情報を通知する。

\*: 本稿で詳述するサブコマンド



## 5. 評価

### 5.1 性能評価

#### (1) 評価モデル

性能を評価する上で  
の評価モデルを図8に  
、システム定数を表4  
に示す。

#### (2) ステージング回数

データへのアクセス  
分布が一様と仮定して

##### (i) マウント済確率

$$= 16/200 = 0.08 \dots \textcircled{1}$$

##### (ii) データヒット率

(データのステージング不確率)

$$= 0.8/20 = 0.04 \dots \textcircled{2}$$

MSS上データの総出力  
件数を工件/時とすると

##### (iii) VTOCのステージング回数

$$= x(1-\textcircled{1})$$

$$= 0.92x \text{ (回/時)} \dots \textcircled{3}$$

##### (iv) データのステージング回数

$$= x(1-\textcircled{2})$$

$$= 0.96x \text{ (回/時)} \dots \textcircled{4}$$

よってステージング回数は $\textcircled{4}$

$$1.88x \text{ (回/時)} \text{ となる。} \dots \textcircled{5}$$

#### (3) ステージングに要する時間の算出

DRD 2台、ステージ  
ング処理時間分布/ステ  
ージング要求発生分布  
が指数分布とすると、  
ステージング処理をM/M  
/2( $\infty$ )の待ち行列と見  
ることが出来る。

(図9参照)

上記の仮定より図9  
における到着率 $\lambda$ ;入  
サービス率 $\mu$ は以下  
で表わすことが出  
来る。

$$\lambda = \textcircled{5}/3600$$

$$= 1.88x/3600 = 0.000522x \dots \textcircled{6}$$

$$\mu = 1/10 \text{ (単独での1回のステ  
ージング処理時間を10秒とする)} \dots \textcircled{7}$$

```

/000/ ON
          92-02-10, 13:40
/001/ LIB RETRIEVE

*** RETRIEVE カイシ
AUTHOR: DENDENKOSYA
SERVICE: JOHOKENSAKU

***JOHOKENSAKU カイシ
%:/RAN 0001,0012

%/LOG
[01]: テンテンコウシ
***100ケンテマス (H:60ケン,L:40ケン)
[02]: マジックユリイソツ
ケケ601ケンテマス (H:250ケン,L:351ケン)
[03]: ケンキョウカイソツ
***161ケンテマス (H:61ケン,L:100ケン)
[04]: 01*(02+03)
***14ケンテマス (H:5ケン,L:9ケン)
[05]: /EOD
%:/KPR 51,,H
    
```

← センタとの接続  
← ライブツリ起動

← 検索期間の指定

← 検索条件式投入開始指示

H) DASD上のデータ件数  
L) MSS "

← 検索式投入終了

← DASD上のデータのみに  
近未来へ出つ

電電探検研、延長1線状導光路から成る透明なタブレット2種を試作。

電電公社武蔵野電気通信研究所、文部省標準91名の会話型音声認識システムを試作。

電電公社武蔵野探研、超LSI世代のメモリーの256kビットRAMの試作完了。

電電武蔵野研究所、加入者ごとに専用プロセッサを置いた分散制御交換システム試作。

電電公社、光損失を大きく減らした新ファイバーを開発。

検  
索  
結  
果

#### 図7. 検索結果例

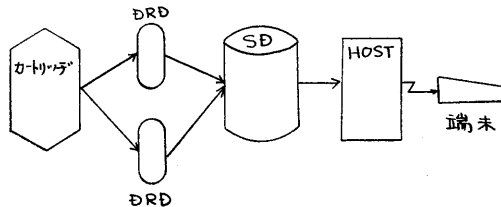


図8. 性能評価モデル

表4. システム定数

項番	項目	定数
1	MSS容量	20GB(100MBx200)
2	SD容量	0.8GB
3	VTOC数	16台
4	DRD	2台

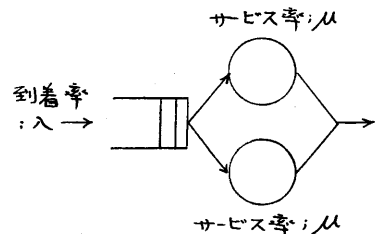


図9 待行列モデル

1回のステージングに要する時間 $T_s$ は待行列の公式より次で表せる。

$$T_s = \frac{4\mu}{(2\mu + \lambda)(2\mu - \lambda)} \dots \dots \textcircled{8}$$

式⑧へ⑥, ⑦を代入すると式⑧は④となる。

$$T_s = \frac{0.4}{(0.2 + 0.000522\lambda)(0.2 - 0.000522\lambda)} \text{ (秒)} \dots \dots \textcircled{9}$$

式④より1時間当りのMSS上のデータの総出力件数 $\lambda$ と1回のステージングに要する時間 $T_s$ の関係を図10に示す。

(4) 検索時間の算出

几件のデータを出力する場合

(i) ステージングに要する時間

ステージング回数は⑤と同様にして1.88 $n$ となるため、ステージングに要する時間は1.88 $nT_s$  …… ⑩となる。

(ii) SSD上のデータの読み込み、端末へデータを出力するに要する時間は1データ3秒とすると $n$ 件では3 $n$  (秒) となる… ⑪

(iii) 検索時間

上記(i), (ii)を並列処理すると、検索時間 $T$ は⑫となる

$$T = \textcircled{10} + \alpha \text{ (秒)} \dots \dots \textcircled{12}$$

⑩>⑪ということから、上記並列処理が理想的に行なわれたと考えると $\alpha = 3$ となる。  
よって式⑫は次の通り表せる。

$$T = \frac{0.744n}{(0.2 + 0.000522\lambda)(0.2 - 0.000522\lambda)} \dots \dots \textcircled{13}$$

式⑬をもとに出力件数 $n$ と処理時間 $T$ の関係を図11に示す。  
尚且DASD上のデータ検索時間は⑪であり、合せて図11に示す。

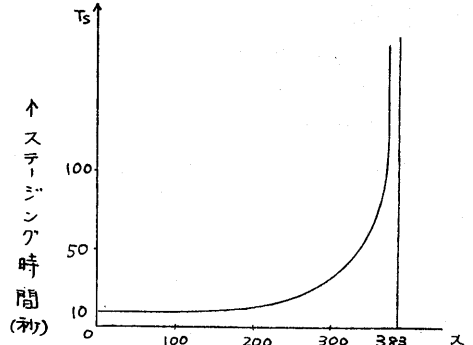


図10 MSS上データの総出力件数(件)とステージング時間

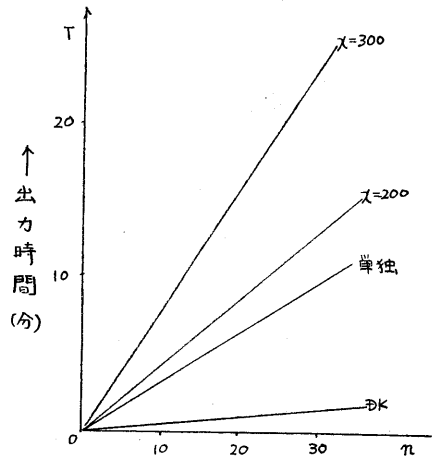


図11 出力件数(件)と処理時間

5.2 ファイル保存コストの比較

ファイル保存コストをDASD, MSS, MTで比較する。

この場合、MTには使用回数に依り、MTのハンドリングコストが必要である。MSSは装置全体のコストが高い等の要因があり一概に決定しがねるが、大略以下と想定する。(4)(5)

$$DASD : MSS : MT = 1 ; \frac{1}{50} \sim \frac{1}{100} ; \frac{1}{50} \sim \frac{1}{200}$$

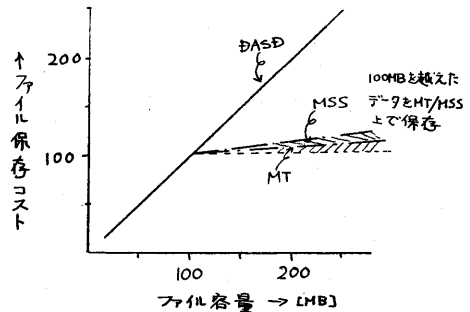


図12 ファイル保存コストの比較



この値をもとに、今100MBを越えたデータはDASDからMSS、MTへ移行して保存する場合のコストを比較すると図12の通りである。

## 6. おわりに

本稿ではDORIS-2をベースにオンライン情報検索システムへのMSS導入法についての検討結果について述べた。

本方式の特徴を要約すると以下の通りである。

- (1) 端末利用者にデータの所在(DASD/MSS)を知らせることにより、利用者がデータの出力特性を認識した上で、出力先を指定することを可能とした。
- (2) この結果、DASDに比べアクセス速度の劣るMSSをオンラインで使用することを可能とした。
- (3) この結果、ファイルコストはMTとほぼ等価で、かつMTに比べ情報入手のスピード化を図った。
- (4) DASDからMSSへのデータの移行単位をファイルとすることで移行に伴う変更処理を局所化した。

今後は本方式の有効性についてさらに分析してゆく予定である。

おわりに、本研究の機会を与えられ、また有益な御指導を賜った横須賀通研高村テ処部長、橋本統括役、山本室長に深く感謝の意を表します。また本研究に対し御討論いただいた研究室の諸氏に感謝いたします。

## 参考文献

- (1) 江里口；「JOISサービスの現状」情報管理 Vol.22, No.2, '79
- (2) 根岸, 牛丸他；「東大オンライン情報検索システム「TOOL-IR」の運用経験」情報DB研究会 DB27-2 '76-3
- (3) 岡田；「TIRS(トヨタ情報検索システム)の現状と将来」情報管理 Vol.20, No.2 '77
- (4) 情報管理研究委員会編；「総合情報管理事典」P271~P272
- (5) 山口；「我国における科学技術情報政策」情報管理 Vol.22, No.3 '79
- (6) 片岡；「JICST」情報管理 Vol.24, No.3 '81
- (7) 金平；「JAPATICのオンライン特許情報検索システムについて」情報管理 Vol.21, No.9 '78
- (8) 宮本, 池田他；「汎用文献検索システムHUNDRED」情報オ18回全国大会 S.52
- (9) 中山, 野上他；「IDEAS/77でのCA SEARCH データベース」情報オ20回全国大会 S.54
- (10) 小畑, 松沢他；「UNIQ-1でサービスしている金属学関係文献データベースMETADExの利用分析」情報オ23回全国大会, S56
- (11) 長峯, 北村他；「情報検索プログラム-DORIS-2」情報DB研究会 17-2, '80

- (12) 岡本, 北村他; 「検索形データベース管理システムの記憶構造の-実現法」  
情報オ19回全国大会 s.53
- (13) 高橋 片岡他; 「大容量ファイル空間の管理方式について」  
情報オ23回全国大会 s.56
- (14) 大塚, 廣野他; 「MDFシステムの概要」 日電技報 No.136 P56~59 '80
- (15) 伊藤 ; 「超大容量記憶技術の動向と研究課題」 研実報 Vol.29, No.6 '80