

# 情報検索システムと端末装置に関する一考察

## Information Retrieval System and Terminal Equipment

福岡浩平\* *Fukuoka Kōhei*  
 石山佳夫\* *Ishiyama Yoshio*

今日のように情報過多の時代には、その目的に合った情報を整理蓄積し、必要に応じスピーディに取り出すという情報検索システムに対する要求が多い。

本稿では、情報検索システムで使われているハードウェアに注目し、発展経過、情報検索機器の分類、情報検索システムの分類、アプリケーション、情報検索システムとシステムタイプ、端末装置の順で整理してみた。

また、一つの例として、漢字ビデオ装置とマイクロフィッシュ検索装置を端末として構成したシステムを想定し、端末に対する要求条件を探ってみた。

その結果、オンライン会話形システムが検索システムにとり、充足度の高いシステムであり、端末装置の役割も重要な位置を占めているとの感を深めた。

### 1 緒言

コンピュータシステムの利用形態が大量事務処理や技術計算から進んで、更に高度な処理、データベースシステムへと進むにつれ、マン・マシン間のインタフェース手段としての端末装置の持つ役割が重要視されてきている。情報検索システムにおいても、端末装置の役割は重要な位置を占めている。

情報検索システムとは、あらかじめ整理蓄積されている情報の中から、必要に応じて適切な情報を提供するシステムである。図1に示すように、(1)情報の整理、(2)情報の蓄積、(3)質問の分析、(4)検索、(5)情報の回答・提供の五つのサブシステムより構成される。

同図中の一次情報は生データ、二次情報は加工されたデータを示す。すなわち、文献検索システムを例にとると、一次情報とは、文献そのものを示し、二次情報は、検索を行なうために作られた検索キーワード、書誌的事項(題名、著者名、発行年月日など)などを示す。

本稿では、情報検索システムの発展経過、検索機器の分類、情報検索システムの分類とアプリケーション、情報検索システムにおける端末装置の順で、情報検索システムと端末装置について述べる。

### 2 情報検索システムの発展経過

必要な情報を素早く取り出したいという要求に対しては、コンピュータが発表される以前から、いろいろな工夫がなされていた。しかし、大量のデータの中から素早い検索を行なうという課題に対しては、十分な解決策が得られていなかった。20世紀に入り、コンピュータが世に出て以来、情報検索システムの利用技術が飛躍的に進歩した。表1に、情報検索システムに関連する技術年表を示す。

情報検索技術はこのような発展経過をたどり、現在に至っているが、その過程で、現在でも十分使用可能な検索機器が数多く開発された。

### 3 情報検索機器の分類

初期のころは、情報検索といえばどのように整理しておくかが問題とされ、専ら分類論が盛んであったが、整理された

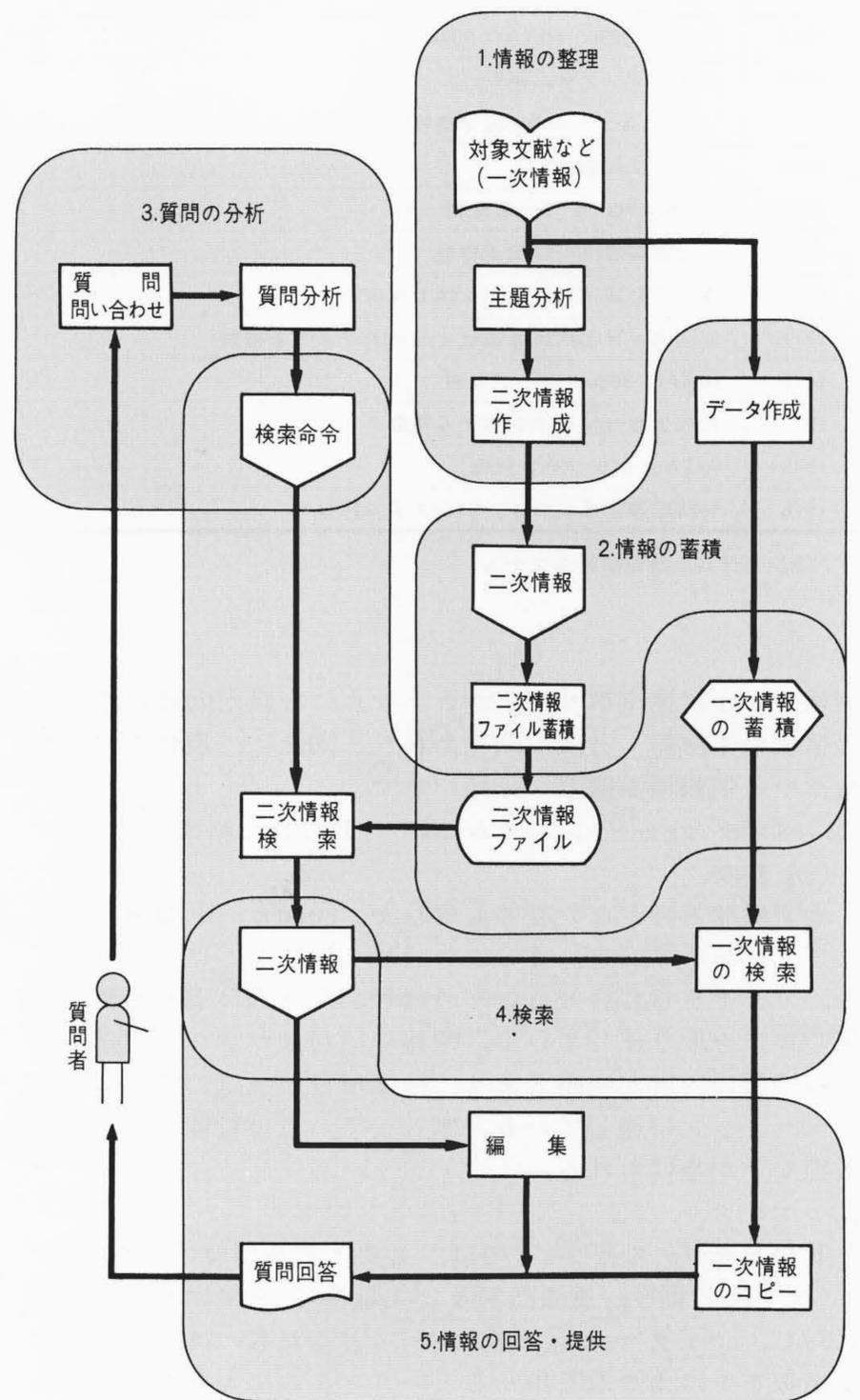


図1 情報検索システムモデル図 情報検索システムは、情報の整理同蓄積、質問の分析、検索及び情報の回答・提供の五つのサブシステムから成る(一次情報:対象文献など、二次情報:検索キーワード書誌的事項)。

\* ファコムハイタック株式会社

表1 情報検索システムの発展経過 情報検索技術、コンピュータ機器、情報検索システムの三つの観点からの発展経過を示す。

年(西歴)	発展経過
19世紀	近代図書館技術の開発
1841	大英博物館の目録法
1876	デューイの分類法
1880	パンチカード(ホラリス)開発
1945	ペンシルバニア大学“ENIAC”
1948	米農務省ラピッドセレクト
1951	国際十進法(UDC)協会設立
1953	Taubeユニタム索引法
1955	GRA記事索引
1957	日本科学技術情報センター設立
1957	日本電信電話公社電気通信研究所MUSASHINO I
1958	HIPAC I01発表
1959	Luhns, Kwickなどでシソーラスを用いるIR技法
1961	ハーバード大学“SMART” 米議会、ドキュメンテーション及び情報検索について聴取会
1962	HITAC 3030, HITAC 3010
1963	外務省IRシステム稼動 “MEDLARS”稼動(医学情報IRシステム)
1964	HITAC 5020発表
1967	HITAC 8000シリーズ発表
1969	国会図書館IRシステム稼動
1970	東急不動産株式会社IRシステム稼動
1971	財団法人日本特許情報センターIRシステム稼動
1972	HITAC 8050シリーズ発表
1973	国民生活センターIRシステム稼動
1975	HITAC Mシリーズ発表
1976	特許庁漢字オンラインIRシステム稼動

注：IRシステム＝情報検索システム

情報をいかに速く取り出すかという点にも目が向けられ種々の情報検索機器・方式が開発された。図2に、現状で使用されている情報検索機器の分類を示す。

(1) パンチカードシステム(以下、PCSと略す)をベースとしたもの

検索機器と呼べないものも含むが、80欄カードなどを利用して、パンチの内容、カードの分類の仕方などに工夫を凝らし、カードを重ね合わせるなどの物理的な方式により必要な情報のキーを取り出している。情報の絶対量が少なく、必要とする情報の内容が明確であり、検索項目が明確にできるアプリケーションに適している。例として、人事管理、個人情報管理などが挙げられる。

(2) マイクロフィルムをベースとしたもの

PCSをベースとしたものは、必要とする情報のキーワード(例：氏名番号、重要語句及び文献番号)を検索しているのに対し、マイクロフィルムをベースとしたものは、必要とする情報をそのまま取り出せるというところに大きな特長がある。アパーチャカード方式は、80欄カードにマイクロフィルムをはり付け、カード分類機により検索を行なうという点では、PCSをベースとしたものであるが、一次情報を中心に考え、こちらに分類した。超マイクロフィッシュ検索機は、現在まだ情報検索システムにおいては普及せず、地図検索など一部で使用されている。

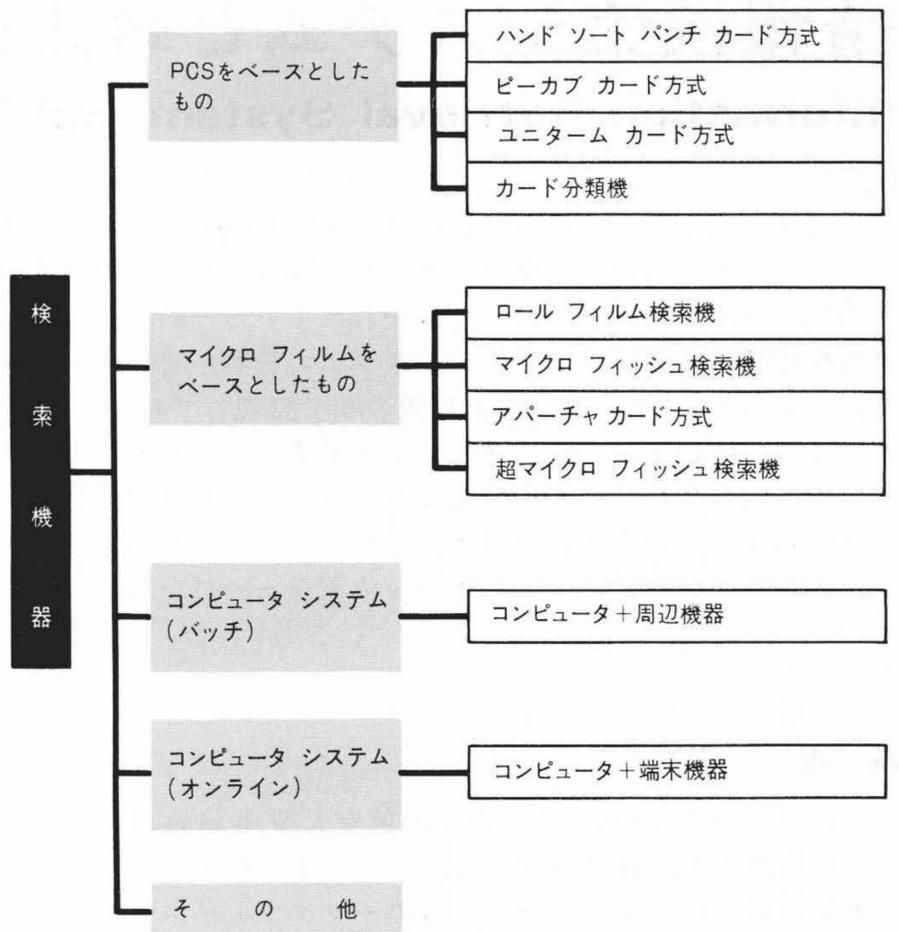


図2 検索機器の分類 情報検索に使用される機器、方式をパンチカードシステム、マイクロフィルムシステム、コンピュータシステムの三つの分類にして示す。

(3) コンピュータシステムをベースとしたもの

現在、大規模な情報検索システムの大部分が、コンピュータシステムをベースとしたものである。情報検索システムは、一般に単独でシステム建設を行なった場合、情報の蓄積に対する投入コストが大きく問題となっており、国立国会図書館における国会会議録検索システムのように、会議録索引編集システムのような他のシステムのデータベースを利用しているものが多い。また大別すると、バッチ形とオンライン形の二つがある。

ここでは、コンピュータシステム以外の機器についても述べたが、次にコンピュータシステムをベースとしたものについて述べる。

#### 4 情報検索システムの分類とアプリケーション

##### 4.1 情報検索システムの分類

情報検索システムの分類基準としては、検索方式、検索する情報の量・レベル・分野などのいろいろな基準があるが、ここでは、質問パターンと回答件数及びその精度の二つの観点の組み合わせにより分類し図3に示す。

情報検索システムを、質問パターンの観点によりみると、定型的な単一項目、定型的な複数項目及び多次元の複数項目の三つの形式に分類できる。

定型的な単一項目とは、既に決まっている1種類の検索キーだけにより質問されるパターンの場合である。例えば、預金システムのように口座番号という一つのキーによるだけで検索する場合である。すなわち最も簡単な質問パターンである。

定型的な複数項目とは、既に決まっている、幾つかの検索キーに基づき検索を行ない、その結果に論理演算を行ない検索する場合である。代表的なものとして、人事情報検索におけるスキルインベントリなどが挙げられる。

多次元の複数項目とは、今まであげた二つの形式と異なり

検索精度	質問パターン		定型的なパターン		一義的に質問パターンが定まらないもの
	質問項目数	回答件数	(1)単一項目	(2)複数項目の組合せ	
			(3)多次元の複数項目の組合せ		
検索結果が一意に決まるもの*	(A) 1件	預金システム 自動車保険照会システム	座席予約システム CAIシステム***		
	(B) n件		人事情報検索システム(スキル インベントリ) 不動産情報検索システム	技術情報検索システム 特許情報検索システム	
検索結果確率的に決まるもの**	(C) n件		文献情報検索システム 判例検索システム	会議録検索システム 文献情報検索システム	

注：\* 完全検索：技術的に100%の再現率、適合率の回答が、どの質問についても得られるもの。  
 \*\* 確率検索：技術的に、ある確率での再現率、適合率の回答しか得られず、質問によってもその精度が変わるもの。  
 \*\*\*CAIシステム：教育システム

図3 情報検索システムの分類とそのアプリケーション 質問と回答結果の精度という点から情報検索システムを分類した。

質問パターンが、一義的に定まらず広範囲の分野に対し、いろいろな角度からの質問が考えられるパターンである。一般に、データの構造は、メッシュ構造になっていたり、複数のトリー構造になっている。

以上が質問パターンからみた場合であるが、次にシステムの検索精度の点から分類してみる。

一般に検索精度は、検索ノイズを適合率、検索漏れを再現率で評価する。検索された文献の中で、要求に合ったものをA、合わなかったものをB、また、要求した文献の中で検索できなかったものをCで表わすと、各々次式で表わされる。

$$\text{適合率}(\alpha) = \frac{A}{A+B} \dots\dots\dots(1)$$

$$\text{検索ノイズ} = 1 - \alpha \dots\dots\dots(2)$$

$$\text{再現率}(\beta) = \frac{A}{A+C} \dots\dots\dots(3)$$

$$\text{検索モレ} = 1 - \beta \dots\dots\dots(4)$$

理想の検索システムとは  $\alpha = \beta = 1$  のシステムである。

検索システムを、この検索精度と回答件数から分類すると、 $\alpha$ 、 $\beta$ が共に近似的に1になり、回答件数が1件のタイプと複数件のタイプ、それに $\alpha$ 、 $\beta$ がある数値以上は1に近づかないものがある。

前者の回答件数が1件に定まるものの代表例としては、先ほどあげた口座番号から一つの預金者ファイルを検索する場合、複数件に定まるものとしては、人事情報検索の場合のように複数人の情報が代表される場合がある。また後者の場合は、蓄積時の分類の仕方が質問者とは異なるというような分類ミス、質問者のレベル、必要とする内容のレベル、質問の不明確さなどにより同じ結果が出ても、要求する人により適合文献数Aが変わり、確率的にしか検索できないものである。例としては、会議録検索などが挙げられる。

4.2 情報検索システムのアプリケーション

表2に、HITACユーザーにおける代表的な情報検索システムの一例として、官公庁の場合を示す。

漢字を使用したシステムが多いという点が目立っている。

表2 HITACユーザーにおける代表的な情報検索システム 官公庁におけるHITACユーザーの代表例を示す。

項番	ユーザー名	アプリケーション	出力装置	概要
1	外務省	国会会議録調書	ビデオ端末 漢字ラインプリンタ	自然語をキーワードとしてオンライン会話処理で検索を行なっている。
2	国民生活センタ	生活相談	ラインプリンタ	商品に対するクレームなどに対する過去の相談事例について検索を行なっている。
3	国会図書館	国会会議録	漢字ラインプリンタ	国会会議録を検索対象とし、発言者、事項などにより検索を行なっている。
4	特許庁	特許情報	漢字ビデオ端末	特許分類番号などをキーとし、出願書類などの検索を行なっている。
5	日本科学技術情報センター	会話型文献検索	ビデオ端末 データタイプライタ端末	理工学文献、医学文献、化学文献計250万を対象としてサービスを行なっている。
6	財団法人日本特許情報センター	特許情報	ラインプリンタ 漢字ラインプリンタ	不特定ユーザーを対象にし、特許情報検索サービスを行なっている。
7	防衛庁技術研究本部	文献検索	ラインプリンタ データタイプライタ端末 ビデオ端末	技術文献約20万件を対象として、TSS、リモートバッチ、バッチ処理による検索を行なっている。

注：ユーザー名はアイウエオ順  
 TSS=タイムシェアリングシステム

表3 検索システムとコンピュータ システムのタイプ 検索システムに要求される条件に対するバッチ、オンライン問合せ、オンライン会話型システムの各々の充足度比較を示す。

充足度	システムのタイプ	バッチ	オンライン	
			問合せ応答	会話
総合評価		14	26	33
評価項目	1. ユーザーの思考速度についていけるか(人間の連想)	0	1	2
	2. 検索回答が即時に入手できるか	0	2	2
	3. 事前検索ができるか	0	1	2
	4. 検索式、検索戦略の変更が容易か	1	1	2
	5. 検索過程を確認できるか	1	2	2
	6. 検索精度(再現率、適合率)をユーザーごとの要求に合わせられるか	1	1	2
	7. 検索回答による検索へのフィードバックはできるか	0	1	2
	8. いつでも必要なときに検索できるか	0	2	2
	9. データ ファイルの拾い読みができるか	0	1	2
	10. 問題発見のための新しい検索アルゴリズムを作り出せるか	0	1	2
	11. 二次情報検索システム特有のインデックス時のひずみは救えるか	0	1	2
	12. マン マシン インタラクションは容易か	0	1	2
	13. 遠隔地からでも容易にできるか	1	2	2
	14. 他の情報センタへリンクできる可能性	0	2	2
	15. 情報センタの事務処理、検索処理を削減できるか	0	2	2
	16. システムの稼働効率を高くできるか	2	1	1
	17. 大量情報処理は容易か	2	1	1
	18. システムのランニング コストは安い	2	1	1
	19. システムの建設コストは安い	2	1	0
	20. 単位当たりの情報コストは安い	2	1	0

注：評点基準 0 = 悪い, 1 = やや良い, 2 = 良い

更新→探索処理→検索文献の表示→質問者自身の適合性の判断の反復処理が可能となり、検索精度を高めることができる。

オンライン会話型、問合せ応答型及びバッチ型の三つについて、検索システムに要求される条件に対する充足度を比較すると表3のようになる。

### 6 情報検索システムと端末装置

前述のように、情報検索システムにおいては、オンライン会話型のシステムにより高度なシステムが実現できる。その場合、システムの顔としての端末装置は重要な役割を果たす。

情報検索システムにおける端末装置として要求される条件を以下に述べる。

(1) 素人にでも扱えるよう操作性の良いこと。

端末のオペレータは、質問者自身が行なうことが望ましく、その場合、次の質問項目の指示などを行なうオペレータ ガイダンス、入力したデータを表示する入力モニタなどの機能があることなどが挙げられる。

(2) 入出力速度が速いこと

シソーラスの表示、質問式の入力、回答の書誌的情報の表示、大量の原本の出力などがあるため、伝送速度、表示速度が速いことが望ましい。

(3) 回答結果の見やすいこと

フィードバックを、数多く行なうためには、スピードとともに、出力結果の見やすさが重要なポイントとなる。漢字での出力、原本と同じフォーマットでの出力、編集出力などは、回答結果を見やすくすることができる。

(4) 運用コストを低くできること

端末のコストなども重要なことであるが、特に蓄積時のコスト、回答出力コストなどの運用コストを低くできることが挙げられる。ビデオ端末のように、検索過程での出力はソフト コピーで、必要とする情報の出力はハード コピーでという端末は一つの有効な方法である。

(5) その他

複数の情報センタに、同一の端末でサービスを受けられるような回線インタフェースであること、質問式の更新が容易であること、保守性の良いことなどが挙げられる。

表4に、オンライン端末のタイプとその特長を示す。

### 5 情報検索システムとシステム タイプ

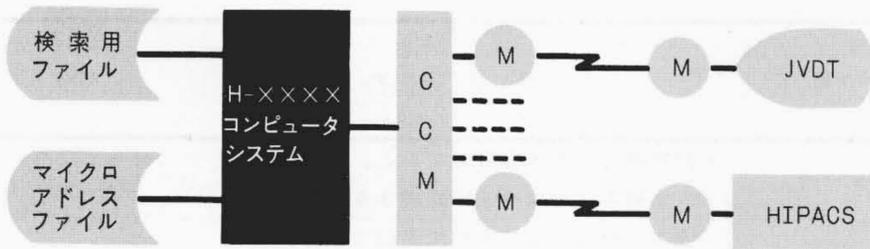
コンピュータ システムのタイプとして、バッチ型、オンライン問合せ応答型、オンライン会話型の三つがある。

図3において、質問項目数が多次元の組み合わせにより行なわれ、検索結果が確率的にしか定まらないアプリケーションほどオンライン会話型が適すると言える。

オンライン会話型システムにより、シソーラスの表示→質問

表4 オンライン端末タイプとその特長 代表的なオンライン端末タイプの機能一覧を示す。◎, ○, △, ×の順で機能が優れていることを示す。

端末タイプ	機能	入出力文字種					直接入力	直接出力	入力モニタ	オペレータガイダンス	伝送速度	表示速度	表示文字数	入力内容更新	編集出力	確認性	原本と同じ出力
		英字	数字	片仮名	漢字	単語											
プリンタ		○	○	○	-	-	-	○	×	×	中△	△	◎	×	○	×	×
キーボード プリンタ		○	○	○	○	-	○	○	○	×	△	△	◎	×	○	×	×
データ タイプライタ		○	○	○	○	-	○	○	○	×	△	△	◎	×	○	×	×
ライン プリンタ		○	○	○	-	-	-	○	-	-	速○	○	◎	×	○	×	×
紙テープ読取せん孔機		○	○	○	-	-	-	×	×	△	×	◎	×	×	×	×	×
キー セット		△	△	△	△	△	-	-	×	×	△	×	×	×	×	×	×
カード読取機		○	○	○	-	-	-	×	×	△	×	×	×	×	×	×	×
光学文字読取機		○	○	△	-	-	-	×	×	△	×	×	×	×	×	×	×
ビデオ データ ターミナル装置		○	○	○	-	-	○	○	○	○	○	○	○	○	○	○	×
漢字キーボード プリンタ		○	○	○	○	-	○	○	○	×	△	△	◎	×	×	×	×
漢字ビデオ端末装置		○	○	○	○	-	○	○	○	○	○	○	○	○	○	○	×
マイクロ フィッシュ検索装置		○	○	○	○	○	○	×	×	○	◎	◎	◎	×	×	◎	○



注：JVDT=漢字ビデオ端末装置  
HIPACS=マイクロフィッシュ検索装置  
CCM=コミュニケーションコントロールマルチチャンネル  
M=マルチチャンネル

図4 モデルシステム構成図 漢字ビデオ装置により質問，検索を行ない，マイクロフィッシュ検索装置により会議録本文を検索する。



図5 HITAC T-560/40漢字ビデオデータシステム 表示文字数：480文字，16×18ドットマトリックス，表示文字種：4,096文字(標準)，磁気カード読取可能な漢字ビデオデータシステムである。

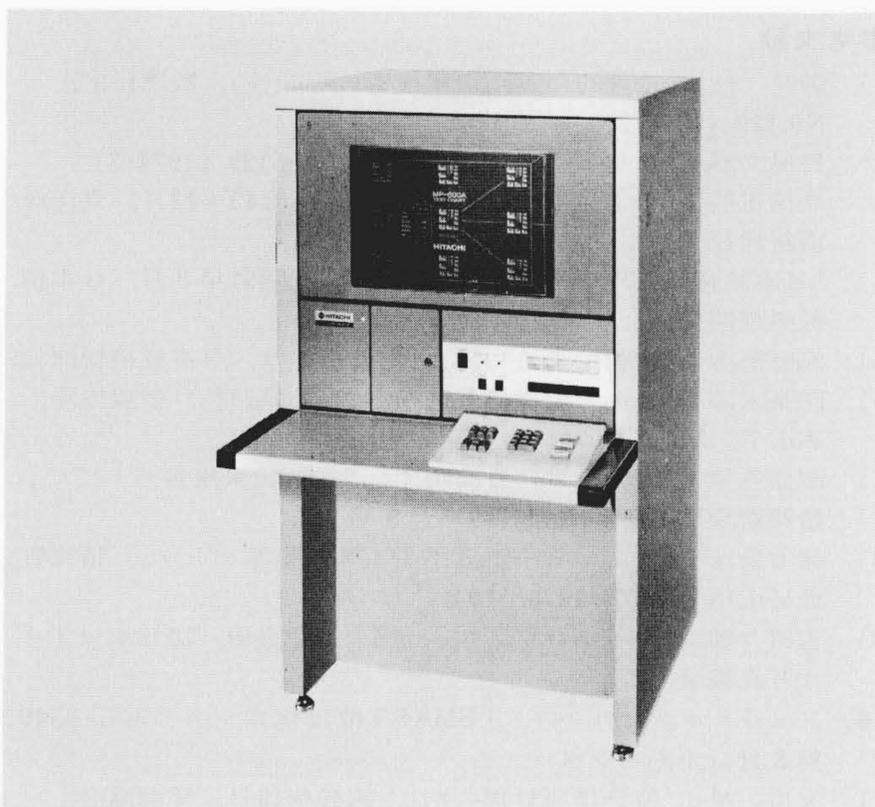


図6 HIPACSマイクロフィッシュ検索システム 概略仕様(MP-600Aの場合) COMフィッシュ，一般マイクロフィッシュ600枚を収納でき，平均アクセスタイム4秒のマイクロ検索装置である。

## 7 オンライン会話型検索システムの事例

オンライン会話型検索システムの事例として，国会会議録検索システムについて述べる。

このシステムは，国会会議録の内容について検索を行なうシステムである。図4に示すように，端末装置として，漢字ビデオ装置，マイクロフィッシュ検索装置が使用されている。HITAC T-560/40漢字ビデオデータシステム，HIPACSマイクロフィッシュ検索システムの仕様を図5に，外観を図6に示す。

図7に示すように，検索過程では，漢字ビデオ装置が使用

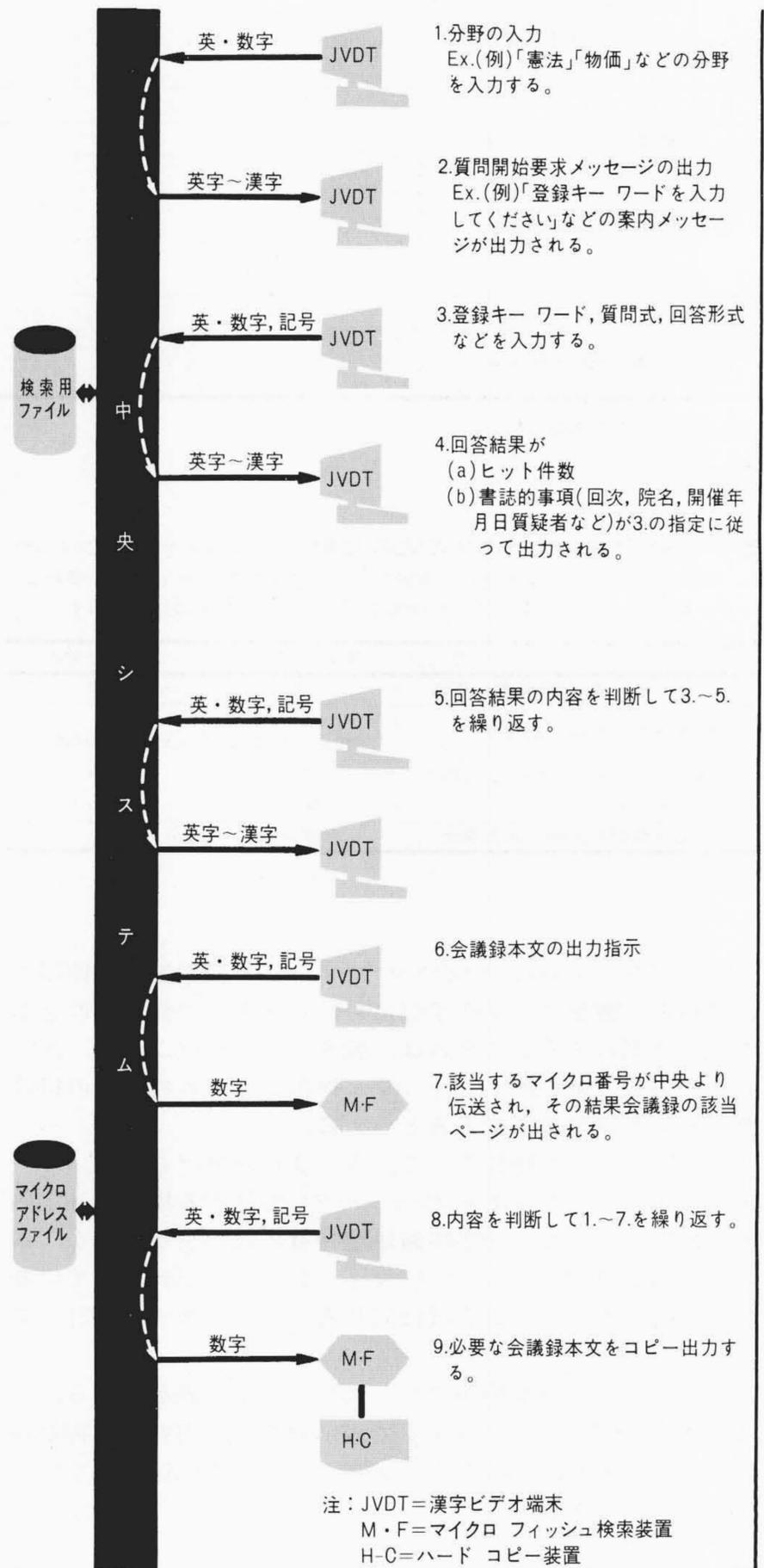


図7 会話型システムにおける質問，検索フロー例 会議録本文を検索するまでの端末の操作フローの一例を示す。

表5 端末に要求される条件(手順別) 表4の各手順における端末への要求条件を示す。

項番	手 順	入出力文字種					データ量(字)	要 求 さ れ る 条 件	
		英字	数字	記号	片仮名	漢字			
1	分野の入力	○	○				5~10	・非熟練者でも入力可能なこと。 ・入力結果がモニターで確認できること。	
2	質問開始メッセージ出力	○	○	○	○	△	50~100	・漢字, 片仮名交じり文で表示できること。 ・表示時間が1~2秒程度であること。 ・出力内容の変更が可能なこと。	
3	質問式などの入力	○	○	○			5~10	・質問入力のガイダンスができること ・質問入力の修正が可能なこと。 ・質問入力キーが表示可能なこと。 ・入力結果がモニターで確認できること。	
4	回答結果の出力 (ヒット件数書誌的事項)	○	○	○	○	△	100~500	・回答部分がプリンク機能などで明確に表示できること。 ・漢字, かな交じり文で表示できること。 ・回答内容の編集が可能なこと。 ・出力内容がシステムの変更により変更できること。	
5	省 略								
6	会議録本文の出力指示	○	○	○			5~10	項番3に同じ	
7	会議録本文の出力	○	○	○	○	○	2,000~4,000	・会議録1ページ分が表示できること。 ・原本と同じ内容で表示できること。 ・拾い読みが可能なこと。 ・応答時間が3~5秒以下のこと。	
8	省 略								
9	会議録本文のコピー	○	○	○	○	○	2,000~4,000	・A4 1枚3~5秒以下でコピーできること。 ・コピー コストが安価であること。	

注：項番は図7の手順と同じ

表6 漢字方式とマイクロ方式の比較 情報をすべてデジタル化し、磁気ディスク装置に入力した漢字方式と、マイクロ フィッシュに撮影し、マイクロ フィッシュ検索機により検索するマイクロ方式との比較を示す。

項 目	方 式	漢 字 方 式	マ イ ク ロ 方 式
1ビット単位の処理(編集, 更新)		可	不可
蓄積インプット時の正確性		× (校正が必要)	○ (校正が不要)
蓄積インプット コスト(単位当り)		高い	安い
検索スピード		遅い	速い
漢字文字情報の取扱いの容易性		悪い	良い

され、回答の出力に、マイクロ フィッシュ検索装置が使用されている。表5に、操作手順における端末への要求条件を示す。この例にみるシステムは、表6に見られるような、漢字ビデオ方式とマイクロ フィッシュ検索システムの両者の特質を生かしたシステムであるといえる。

システムの主な特長として、次の3点が挙げられる。

- (1) 漢字仮名交じり文をコンピュータで処理する場合、外字処理、漢字ソートなど漢字情報処理特有の問題点があるが、本事例では、原本がマイクロ フィッシュ上に記録されているため、従来のシステムとほぼ同じ考え方でシステム設計ができる。
- (2) 安価で正確な蓄積ができ、スピーディに表示できる。
- (3) 本文と同じフォーマットで表示ができ、内容の把握が容易であるとともに、中央のコンピュータとの伝送時間を短くすることができる。

## 8 結 言

以上、情報検索システム的一端につき、ハードウェア面から述べてきたが、情報検索システムは、まだまだ未解決な点

の多いコンピュータのアプリケーション分野の一つである。その中で、端末に対する未解決な点として、画像伝送、大量伝送、検索速度、コストなど多くの点が残されている。特に今後、漢字・図形を含めた情報の検索要求が増えるとともに、今回事例として挙げた大量の静的な漢字情報の入出力が必要とされるケースが考えられるが、そのような場合、漢字ビデオ装置とマイクロ フィッシュ検索装置を端末装置として使用することが一つの有効な解決策として考えられる。

## 参考文献

- 1) 細井 勉：「論理的にみた定理検索の可能性」, 数理科学, No.129 (1974-3)
- 2) 長尾 真：「連想と検索」, 数理科学, No.129 (1974-3)
- 3) 高橋正明：「科学技術情報の検索方法」, 昭45年12月, 東洋経済新報社
- 4) 「日本語情報処理システムの研究開発」, 昭51年3月, 日本情報処理開発協会
- 5) 河野徳吉：「情報検索の知識」, 昭43年7月, 日本経済新聞社
- 6) 印東太郎：「人間の情報処理過程の数量的解析」, 情報処理, Vol. 7, No. 3 (1976, 3)
- 7) 諏訪秀策：「米国における商用オンライン情報検索サービス」, 情報管理Vol.18, No. 5 (1975, 8)
- 8) 緒方良彦：「ニューヨーク・タイムスの情報バンク」, 情報管理Vol.18, No. 7 (1975, 10)
- 9) 吉田 勉：「マイクロフィルムガイド」, 昭49, 10日本マイクロ写真協会
- 10) ジェラルド・サルトル：「SMART情報検索システム」, 昭49年8月, 企画センター
- 11) 大川 晃：「特許情報管理入門」, 昭49年10月, 発明協会
- 12) 西岡英也：「オンライン端末装置」, 昭45年4月, 産業図書
- 13) 行政情報システム研究所：「マイクロフィルム連動検索システムに関する調査研究報告書(昭51年3月)」