44－S 002

特許情報機械検索システムの開発研究

昭和45年3月

財団法人 日本情報処理開発センター

序に代えて

社会，経済の発展にともない，各種情報の蓄積・加工・供給を有機的かつ効果的に行なう方法として，とくにコンピュータによる情報処理の役割りの重要性が認識されております。

また，最近は情報社会への指向とあいまって情報処理分野の拡大とともに，その高度化の方向が検討されつゝあり，大きな発展期を迎えているといえます。

しかし，このような情勢において，情報処理および情報処理産業の前途には，その発展の要件およびこれが他産業に与える影響といったわが国経済の動向に関連する諸問題を始め情報処理方式，ハードウェアおよびソフトウェアの技術，各種の標準化，情報処理技術者の養成等，解決を要する幾多の課題があります。

当財団は情報処理に関するこれら諸問題解決のため各種の事業を実施しておりますが，この報告書はその一環として特許情報処理に関する基礎的研究を行なうため，（株）特許デイタセンターに委託したものの一部で，コンピュータによる特許情報検索の実用性を検討するため，「自然語を使用する広域検索システム」を開発し，これに基づく実験，研究の結果をとりまとめたものであります。

なお，この事業は，日本自転車振興会の機械工業振興資金による「昭和４４年度情報処理に関する調査研究補助事業」のうち，「特許情報関係に関する調査研究」として実施したものであります。

ここに本調査，研究にご尽力ならびにご支援を賜った関係各位に心より感謝の意を表しますとともに，本報告が各方面に利用され，わが国情報処理産業発展の一助として寄与できますよう念願いたす次第であります。

昭和４５年３月

財団法人　日本情報処理開発センター
会長　難波捷吾

機械検索システム開発研究委員会
企画調査小委員会

委員長	川島　順	株式会社　特許データセンター
委　員	青木　秀実	住友電気工業株式会社
	井上　修	日本ＩＢＭ株式会社
	小川　義久	財団法人　日本情報処理開発センター
	草間　基	日本電気株式会社
	中村純之助	株式会社　日立製作所
	林　秀行	通　商　産　業　省
	宮崎　嘉夫	特　　許　　庁

（委員名は５０音順）

目　　次

Ⅰ　総　　論

Ⅰ 総　論

はじめに

技術革新，貿易自由化などを背景として，企業における国際的競争は激化の一途をたどっている。このような国際情勢下において，技術文献および権利文献の二面性を持つ特許文献は，最近，特に重視されるようになってきた。

一方，年々公表される特許件数は，ますます増加し，その累積件数は膨大なものとなっている。ちなみに，昨年（１９６９年），１年間に発生した主要国の特許件数，および過去２０年間の累積件数の概数を表Ⅰ－１に示す。

表Ⅰ－１　主要国の特許件数とその累積件数

国別＼件数		１９６９年の発行件数	過去20年間の累積件数
日本	特許	32,800	348,000
	実用新案	31,680	484,000
米国		66,000	1,130,000
英国		40,000	530,000
独国		60,000	488,000
仏国		33,000	600,000
合計		263,480	3,580,000

主要国だけでも，年間約２６万件，過去２０年にさかのぼると，実に３６０万件の特許が存在する。

特許調査は，このように膨大な特許文献を対象として行なわれているが，現在は，ほとんど人手による調査，マニュアル調査に頼っているのが実情である。

蓄積文献数の急激なる増加，特許調査の需要の増大，さらに加えて技術者の不足は，これ以上特許調査を人手によって行なうことの困難性に拍車をかけ，必然的に，機械力による特許調査の早急な実行がクローズアップされてきた。

特許文献の機械検索の必要性は，かなり前から提唱されているが，その研究，実施は遅々として進んでいない。

その一因は，特許文献は，単なる技術文献としてだけではなく，権利文献であ

るので，検索に対する要求，特に検索もれに対する要求が，他の技術文献の場合に較べて極めて高く，文献内容の解析に理想的な方法を追求し，解析そのものが複雑になり，また高度の技術的専門知識が要求されるため，蓄積に非常な労力と多数の専門技術者を要するためであろう。

したがって，比較的狭い技術領域で，しかも限られた文献数を対象とする時は，ある程度可能であるが，広い技術領域の，しかも上記のような膨大な文献数を対象として考える場合は，今迄のような精密なシステムでは，到底蓄積のためのデータ解析を行なうことが不可能となり，文献数が多いために要求される機械検索が，文献数が多いために実行できないという矛盾に遭遇する。

したがって，膨大な文献を対象として考える場合にはまず第一に蓄積に要する労力をなるべく軽減し，その軽減した結果，当然起り得るエラーは，他の機械的処理の可能な方法によって極力救済する。第二に，機械に完全な検索を行なわせることをせずに，機械で行なう部分と，人間の行なう部分に分けて，機械と人間とが共同して完全な検索を遂行するように，全体として最も経済的になるように，両者の分担を決める。という着想に基づいて，以下に述べる自然語による広域検索システムを開発した。

1. 特許情報サービスシステムの全体構想

検索システムの開発を考える場合，検索システムは，あくまでも，情報サービスの一部分を構成しているに過ぎないので，他の関連業務との関係を忘れてはならない。

特許情報サービス中，特許明細書を中心としたサービスの種類と形体は次のとおりである。

(1) 複写サービス

(2) 二次，三次資料の作成（抄録，索引類）

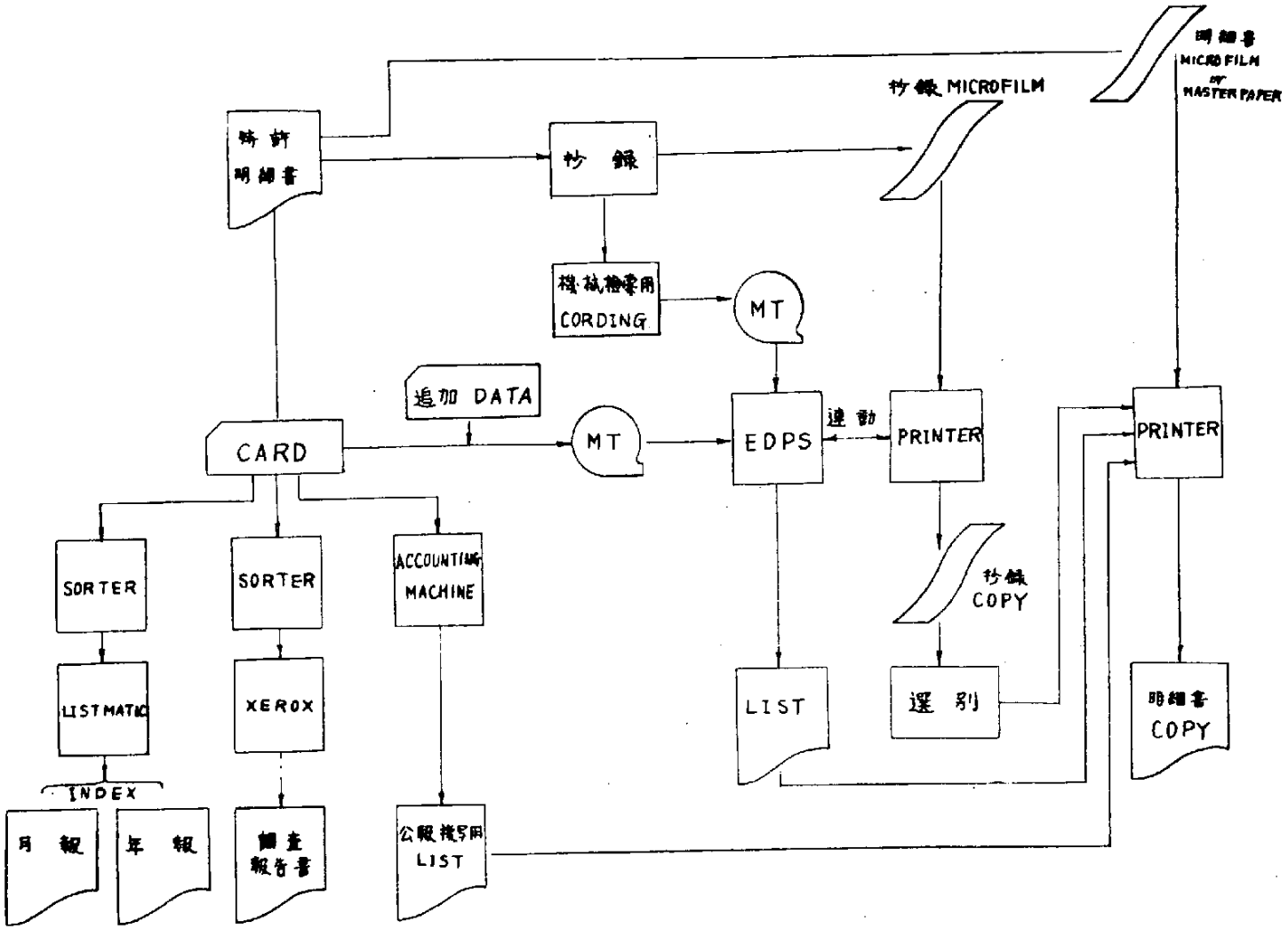
(3) 調査サービス（検索サービス）

(4) 特殊資料の作成（マイクロフィルム，磁気テープ化）

(5) 翻訳サービス

実行面において，上記各種サービスを有機的に結び付けることが大切である。

図Ⅰ－1－1は，これらの各種サービスを体形化した試案であるが，日本語はカナ文字によって磁気テープ化し，抄録は機械による自動抄録，自動編集などに



図Ⅰ－1－1　特許検索・複写サービスフローチャート

よらず，通常の方法によって行なうことを前提とした中間的なものであって，もっと機械化が進めば，当然，異ってくる。

このシステムの特徴の一つは，販売用に作成する抄録を使用して，機械検索用のデータ解析を行ない，さらに，機械で検索した回答の内容チェックにその抄録を使用することである。これによって蓄積データの解析が容易となり，また機械によって検出された多数の回答のチェックを簡単に行うことができるので，多少検索によるノイズが多くても実用上差支えないこととなる。

なお，この図では，索引作成にカードを使用しているが，これは，発明の名称や出願人名などが必要なリストを作成する場合，カナ文字では一般市販用には不適当であるので，リトマチック方式によって，カード上に原語をタイプしたものを使用している例を示している。勿論，漢字を必要としないリスト，インデックスなどは，この方法によらず，磁気テープファイルのデータを使用し，電子計算機によって行なうことができる。また将来，漢字情報を磁気テープに読込むようになれば，この工程はすべて不要となり，電子計算機によって，すべてのリスト，インデックスの作成が可能となる。また，明細書，抄録のプリントアウトはマイクロフィルムによっているが，これを明細書または抄録から，直接複写してもよく，抄録のチエックも，それをリプリントせず，抄録を直接チエックする方法に置換えてもよい。

このシステムによる製品とその利用関係は次のようになる。

製　品	加工品または使用目的
① 明細書	
全　文	→マイクロフィルム
書誌的事項	→リストマチックカード
技術内容	→抄録
② 抄　録	→販売
	→マイクロフィルム
	→調査サービス（マニュアル）
	→機械検索用データ抽出（磁気テープ）
	→機械検索回答チェック

③　リストマチックカード

→名称入り索引作成

→書法的事項による調査

④　検索用磁気テープ

→販売

→索引作成

→検索サービス用

⑤　マイクロフィルム

→販売

→複写

⑥　索引類　→販売

→調査サービス用

2. 機械検索システムの基本構想

2.1. 広域検索システム

2.1.1. 広域検索システムの必要性

特許文献の機械検索は，かなり以前から研究されているが，その大部分は狭い領域を対象とする検索システム（以後，個別検索システムと称す）である。

特許文献の場合には，さきにも述べたように，検索要求事項が，比較的細かい技術思想を対象とする場合が多いので，細かい技術思想が識別できるようなコーデングシステムを作成しなければならず，このようなコーデングシステムの作成は，狭い領域ではある程度可能であっても，広い領域では，コーデングシステム自体が，複雑かつ膨大となり，その作成は勿論，使用上も非常な困難を伴うことや，このように精密なコーデングシステムを使用する蓄積作業は，莫大な労力と多数の専門技術者を必要とするので，大量の文献を対象とする時は，ほとんど不可能になる，などの理由によって，やり易い部門，範囲を対象とした個別検索システムが発達してきた。

しかしながら，利用者の立場から見ると，そのような特定領域に対する質問は全体としてはごく稀にしか起らず，しかも各システム毎に，思想，体系，検索手段など，すべて異り，またその境界領域，収容範囲が必ずしも一定していないな

どの利用上の不便のため，実際に活用されている例は余り見受けられない。したがって，かなり広い部門（領域）を一度にカバーするような検索システム（広域検索システム）の開発が切望されている。

2.1.2　広域検索システムの問題点

広域検索システムを考える場合，まず第一に大切なことは，蓄積の労力を軽減させることである。蓄積の労力の軽減とは，蓄積のための延べ人員が少なくてすむということ以外に，特殊専門家を余り必要としないということも非常に重要なことである。

広域検索システムの場合は，必然的に，その対象とする情報量が個別検索システムに比べて桁はずれに多く，したがって個別検索システムの手法をそのまゝ取り入れたのでは，まず，内容解析を行なう技術者の確保で行きずまってしまう。

第二には，蓄積の労力，すなわち，内容解析を簡素化することは，とりもなおさず，内容の質的低下を意味するし，特に，検索もれの危険性が増える。したがって，何か別の手段で検索もれを防止することを考えねばならない。

第三には，検索もれを少なくすることに力を入れすぎると，必然的にノイズが多くなる。ノイズ防止を機械的手段で行なうか，または，ノイズの識別が簡単に人によって行なえるような方法を考えておく必要がある。

2.1.3　広域検索システムの基本方針

前項の問題点を要約すれば，蓄積の労力の軽減と，その結果として起る検索もれ，ノイズの増加に対してどのように対策を講ずるかということである。

これらの問題点を解決するために次のような方法を用いてみた。

(1)　労力の軽減と特殊技能者の節約

①　キーワードに自然語を使用する。

日本特許は日本語のカナ文字

外国特許は英語

②　抄録からキーワードを抽出する。

(2)　検索もれの防止

①　複数の原データからキーワードを抽出する。

日本特許は抄録と請求範囲

米国特許はガゼットと邦文抄録

② 機械的処理のできる方法で追加キーワードを補充する。

分類表のキーワードを機械的に追加する。

(3) ノイズの防止

① 分類とキーワードによってダブルチエックし，その結果を評価する。

② データのキーワード，質問語，マッチング方法などにウエイトをつけて，その結果を評価する。

③ 回答を抄録によって人間が二次チエックする。

2.1.4. 広域検索システムにおける各領域ファイル相互の取扱い

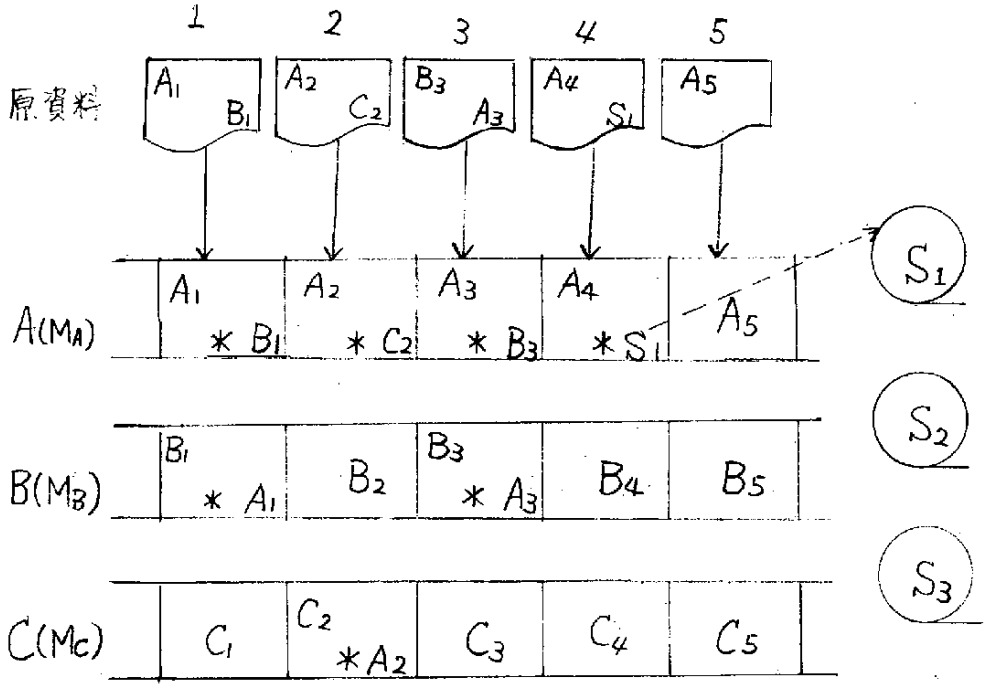
全分野を対象とした広域検索システムを開発することも可能であるが，実際に検索を行なう場合，全く関係のない部門についてまで検索を行なうことは不経済である。したがって，あらかじめ大きな技術領域毎に分けてファイルを構成するのが望ましい。しかし，このファイルの分割を余り細分化すると，境界領域の数が増え，境界領域に属する文献をどちらに所属させるかが問題となり，また検索の場合にも，関連領域のファイルについても検索しなければ脱落する可能性が生じ，結果的には細分化した意味があまりなくなる。したがって，概念上明確に区別できる程度の大きな単位で，分割するのが望ましい。

しかし，概念上明確に区分できる単位で分割しても，一文献中に複数の領域に属する内容が含まれている場合があるので，それは次のように処置したい。

図Ⅰ－2－1に示されるように，仮に，全技術領域を電気(A)，化学(B)，機械(C)，……のように分割した場合，それぞれの領域毎に検索ファイルM_A，M_B，M_C……が構成される。

実際の文献 1, 2,……，nを解析したとき

文献1にA領域の内容		A_1
	B領域の内容	B_1
文献2にA領域の内容		A_2
	C領域の内容	C_2
文献3にB領域の内容		B_3
	A領域の内容	A_3
文献4にA領域の内容		A_4
	個別検索の内容	S_1 （後述する）



図Ⅰ－２－１　各領域ファイル相互の関係

文献５にＡ領域の内容　　A_5

が含まれていたとする。

まず，Ａ領域の内容を含む文献のみを解析して，Ａ領域のファイルM_Aを構成する。Ｂ領域についても同様，ファイルM_Bを，Ｃ領域については，M_Cを，以下同様にして作成する。このようにして各領域のファイルを構成する際に，他領域の内容を含んでいるときは，その旨の表示および，他の領域における収容個所（図Ⅰ－２－１においては，＊B_1，＊C_2，＊B_3，……の表示）を附しておく。

このようにしておけば，実際の検索に当って，質問が複数の領域にまたがつて

いるときは，いづれか一方の領域，例えばA領域のファイルMAで検索し，MAに関する条件を満足した文献中，他の領域，例えばB領域に関する内容が含まれているとの表示＊Bのある文献についてのみ，B領域のファイルMBを検索すればよい。この際MAをメインルーチン，MBをサブルーチンとしてプログラムを組んでおけばよいが，それを更に能率よく行なうためには，各領域のファイル毎に，その領域に関連のある他の領域のファイル中の文献についてのみ濃縮テープを作成し，それをサプリメントテープとして，その領域のファイルに付属させておけばよい。これを図について説明すると，A領域のMAには，サプリメントテープとして（B_1，B_3，……）（C_2，……）が付属することとなる。

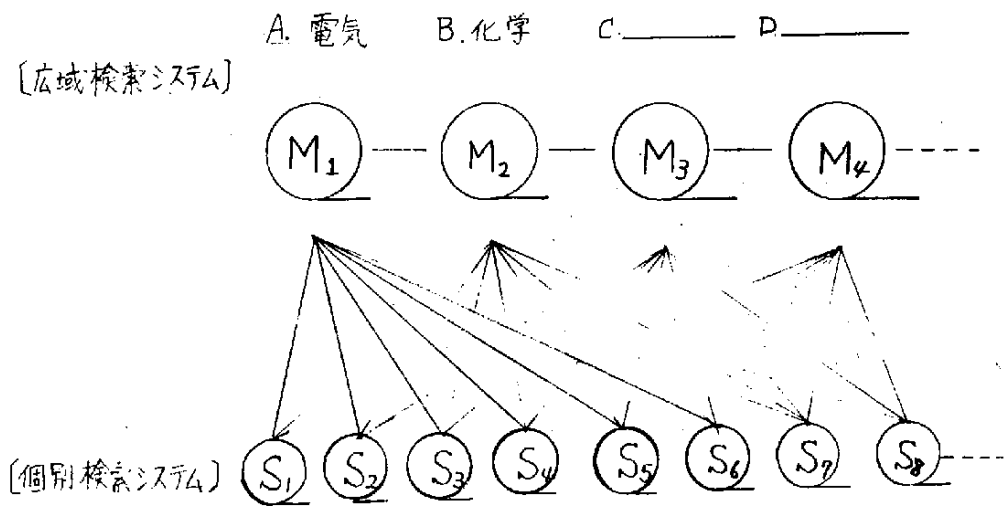
2.2. 個別検索システムとの関連

広域検索システムのファイルと個別検索システムのファイルとの関連も上記広域検索システムにおける各領域ファイル相互の関連と原則的には同じ取扱いで解決できる。（図Ⅰ－2－2参照）

即ち，個別検索システムの対象とする技術領域に属し，しかも蓄積されている文献については，広域検索システムのファイル中のその文献の蓄積データ中に，個別検索システムを代表するキーワードを設定し，さらにそのキーワードの一部に，個別検索システムにも蓄積されている旨の特別の符号を付けておけばよい。

実際の検索に当り，そのキーワードを含む質問で，しかもさらにそのキーワードが表わす概念の細かい条件についてまで要求されているときは，広域検索システムのファイルによって検索し，該当するもののうち，そのキーワードの一部に個別検索システムにも蓄積されている旨の特別の符号のあるものについてのみ，個別検索システムによって検索すればよい。

しかしながら，個別検索システムは，各システム相互，全く関係なしにシステム設計が行なわれている場合が多いので，質問の仕方の統一，検索システムの調整など，今後に残された問題が多い。



図Ⅰ－２－２　個別検索システムとの関連

2.3　実験用情報検索システム

上記のような着想に基づき，自然語による広域検索システムの開発研究を開始したが，現在，この種の広域検索システムは，国内，国外をとわず，数える程しか実施されておらず，またその詳しい内容もほとんど知られていない。

したがって，まず，自然語による広域検索の実用性を追及する基礎実験，ならびに，システム設計のための基礎データの収集をしなければならない。

本報告書では，上記の基礎実験，基礎データの作成に適した実験用検索システムを開発し，そのシステムを用いて問題点の究明および，システム設計上必要とされる基礎データを作成した経過，ならびに結論を以下に記述する。

実験用検索システムは，日本語のカタカナ，ひらがな，漢字を含む文献をカナ化したカナ文字を対象としたものと，英語を対象としたものの二種類作成した。

両システム共，基礎実験およびデータの収集が主目的であるので，ある程度実

用性を犠牲にして，ファイル構成，検索プログラムの作成を行なった。また，インプットデータにはデータやシステムの解析を行なうために必要な各種コード，（この大部分は実用化段階では不用になるものが多い）を付け，以後の解析を容易にするための配慮をした。

インプットにはカードを用い，用語の整理，シソーラスの作成のための便を考え，キーワードカードは，キーワード1語につき1枚のカードを使用した。

2.4. 汎用情報検索システム

2.4.1. 汎用情報検索システムと実験用情報検索システムとの関係

今回の研究の対象にした実験用情報検索システムは，実行段階の機能をそなえた汎用情報処理システムを開発するための予備的システムであって，汎用システムの開発に必要なシステムの解析，システム設計に必要なデータの収集を主目的としているが，特にその対象が特許文献という特殊な性格を持つ情報源であることに鑑み，検索そのものに対する利用者の要求が，一般技術文献のそれに較べて遙かに高くかつ複雑であるので，まず，これら利用者の要求を満足させるための検索理論を究明し，それを実験的に実証して後，その実績に基づいて，実行システム，すなわち，汎用情報検索システムを開発しなければならない。

では次に，その目標とする汎用システムは，どのように構成され，かつどのような機能を備えているのかを説明する。

2.4.2. 汎用情報検索システムの機能

特許情報サービスシステムを大きくわけると，①情報を必要とする要求者側の質問と②それを処理する過程と③質問に対する回答とに大別することができるように思われる。

まず，システム全体としてはどのような情報分野においてもその要求を受け入れることができなければシステムのサービス性に欠けることになる。かといってある分野についてはこちらへ，また他の分野についてはあちらへというのでは，もうすでにここで情報の要求者すなわちシステムの利用者自身が広い知識と情報源をもたねばこのシステムを効率よく利用することはできないということになる。そこでどのような技術分野に関する質問でも，すべてひとつのシステムの受け入れ窓口から実行されるようにならねばいけないのではないかと考えられる。

さらにまた，全般にわたる質問の処理が可能でも，その質問を受理し実行する

際に操作が複雑であっては何にもならない。それゆえ，システムは受け入れた質問をある程度自動的に処理し質問のテーマの中の主たる要素を表わすキーワードを設定できるようにするべきである。もちろん，このとき日常の言語ないしは技術用語から質問テーマを分解，分析し質問分野を自動的に選択決定できるシステムであることが望ましい。そのために，汎用のシステムは自然語によるシソーラス及び関連語分布を組みあわせた質問自動分析部分を有しているのが好ましい。

さらに，質問が自動的に解析された後そこからどの分野の，どのファイルによって情報処理検索を行なえばよいかが決定されることになる。

それぞれの部門，分野に適したファイル構成をとり，質問内容によって，自動的に，ないしは意識的に広域検索系の中で行なうか，または個別の情報ファイルを使って情報処理を行なうかが選択されるようになることが必要である。この際，質問と適切なファイル選択とを関連づけるためには実行上の簡便さを考慮し，簡単なパラメータ指示によって行なわれるようにすべきである。

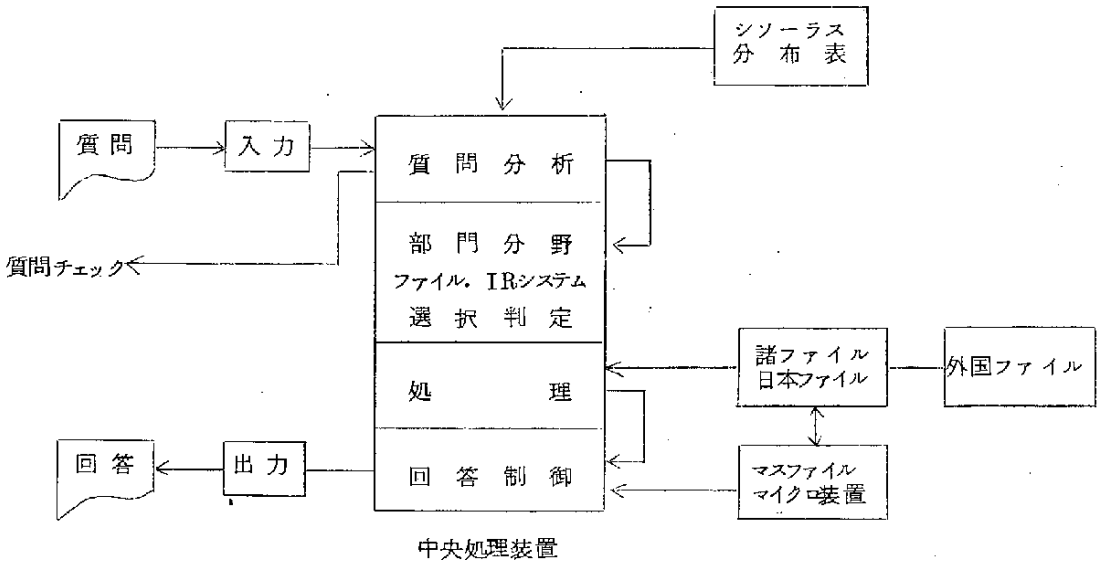
さらに，処理過程においては質問要求とその部門，分野に適した処理プロセスが用意されファイルの決定と同時に処理プロセスも自動的に選別決定され各技術分野に適した処理過程をたどるようにする。例えば電気部門と化学部門とではそれぞれ主になる技術用語が異なるので，電気部門ではできる限り言葉中心で，場合によっては数値，数式，化学記号を並用処理する。化学部門では化学記号および化学式が主体となり，場合によっては通常言語をも用いる，といったように質問と処理ファイル，処理システムを柔軟性のあるものにしようということである。

また，処理と同時に回答においても回答のレベルを自由に調節することを可能にし，要求者との対話によって制御することができるようにしなければならないように思われる。また，回答内容も単一のもののみにとどまらず，番号のみのパラメータ形式のもの，抄録カードないしは明細書に至るまで大容量のファイル管理末端機構，例えばマイクロフィルム解読プリントアウト装置などを接続して質問者の答の内容に対処できるようにすることである。したがって，システム上ではキーワードや分類などを組合わせて作成した情報領域を検索し，回答要求によっては直接，末端のファイル管理装置をアクセスして回答を出すことも行なわれるようにする。

以上のシステムは，日本語による文献のみにとどまらず，主要各国のデータに

も連絡し共通に同一操作手段によって処理ができるようにする。

以上のシステムプロセスの概要は次図（図Ⅰ－２－３）のようなものが考えられる。



図Ⅰ－２－３　システムプロセスの概要

２．４．３　汎用情報検索システムの利用方式

前節で述べたシステムを実際に利用するにはどうすればよいかについて述べよう。

現段階では，ヨーロッパ語族の文章による文献もさることながら，特に日本語による文章の特殊性から日本言語の文章のコンピュータによる自動処理が非常に難しい。それ故に，良いシソーラスなどがあったにしても原データを解析しコンピュータへの入力源データにするまでには，かなりの人手を要求する。そこで一度原データを入力可能形態にしたら一度に多数の人がそのシステムを利用することができることが望ましいのが当然である。システムの繁雑性とデータ作成の困難と経済性を考えると，情報要求源が各個別にデータを作成していたのでは分野もある程度限られ，産業，研究界総体からみると不必要な労力，経費または時間

の重複，浪費となることは明らかである。そこで，一度データを作成したらそれを一括して多くの人が利用できるシステムを構成する必要がある。

利用者は各個の手持ちのコンピュータなどの処理能力に応じて，このシステムを使用することができる。これを表Ⅰ－２－１に示す。

表Ⅰ－２－１

システムの利用度合 ／ 現在の機械化状況	広域検索用ファイル	個別検索用分野別ファイル	システムプログラム及びファイル	調査依頼質問	マイクロ装置など末端素子用ファイル
小型ないしは中型の計算機使用		○			
中型以上の計算機所有	○	○	○		
中型以上の計算機及びマイクロ装置など所有	○	○	○		○
計算機は所有しない				○	
その他	o 定例速報出版物 o T．S．S on-lineによる情報サービス				

表Ⅰ－２－１について説明を補足すると，これはこのシステムを利用する際にどどのような形式で利用するかを表わしている。

まず，小型ないしは中型のコンピュータを所有しているならば，個別検索用分野別データファイルを使用して，自分の必要とする部門または分野の情報処理を自力で行なうことができ，随時必要なときに，必要な形で検索処理が可能である。また中型以上大型機に近い機種ないしは大型のコンピュータを所有しておればすべての部門分野にわたる広域検索用ファイルを総て使うことができる。また，さらにマイクロ写真選択解読出力装置を完備すれば，オリジナルマイクロフィルムのコピーを利用することによって，出力も原データの形で得ることも不可能ではない。

なお，さらに将来，専用データ通信回線の自由割当化が実施されれば，簡単な端末装置ないしは手持ちコンピュータなどをオンラインで接続し，サービス網を完備することも可能となる。また，テーマ別調査速報のような定例出版物として

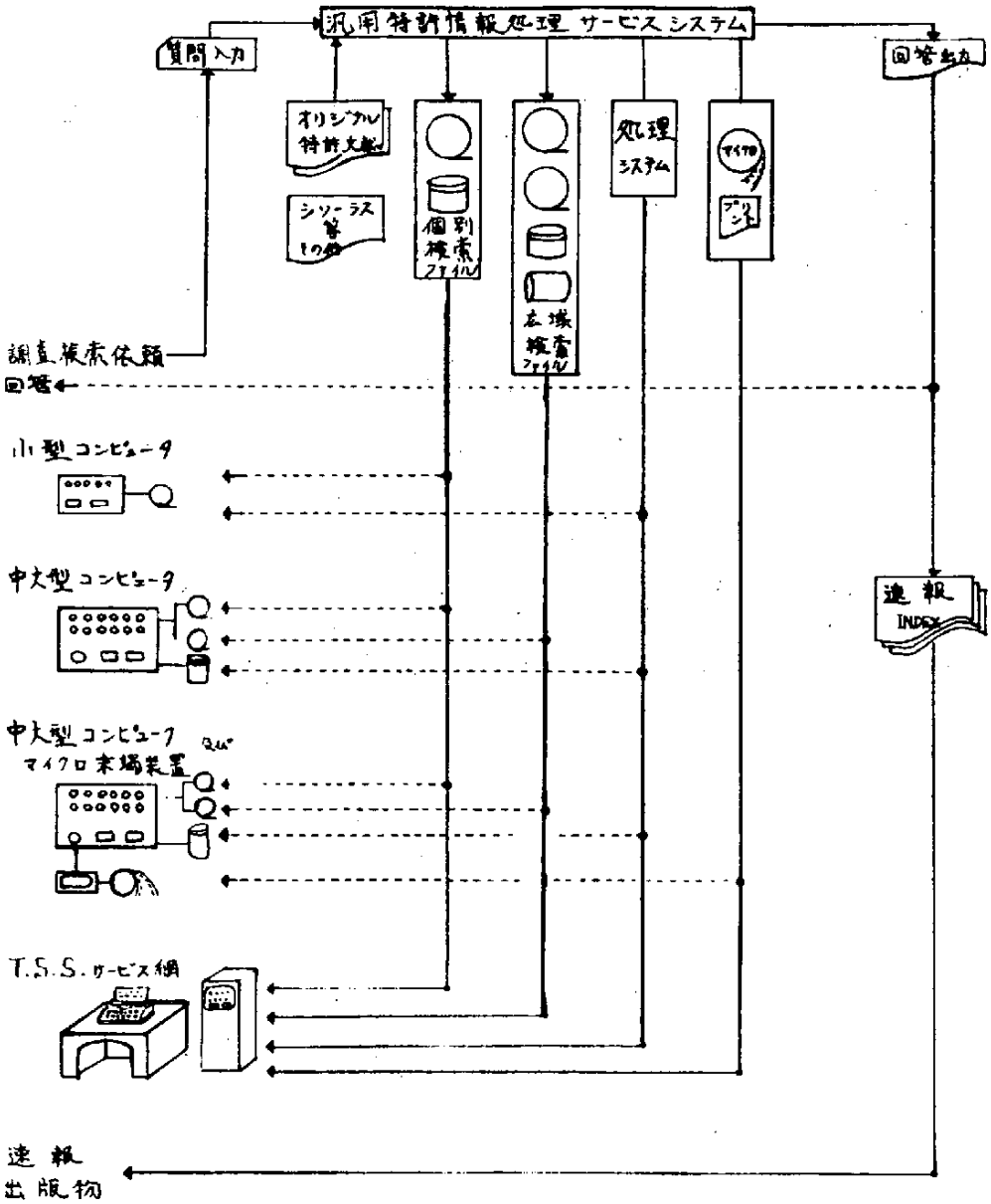
回答を定例的に得ることも可能である。

このように，原資料の収集，管理および加工，ソースデータおよびファイルの作成，コンピュータによる検索またはデータ処理，回答または資料の作成，利用者への情報または資料の提供などの一連の機能を備え，各機能を有機的に結合したシステムは，情報検索システムというよりも，むしろ，汎用情報処理サービスシステムと呼ぶべきであろう。

この汎用情報処理サービスシステムの利用経路を図解すると次図Ⅰ－2－4のようになる。

引　用　文　献

1）川島順，科学技術資料協会における特許資料のカード化，ドクメンテーション研究，18〔6〕，161～168，('68)



図Ⅰ－２－４　汎用特許情報処理サービスシステムの利用経路

Ⅱ　実験用情報検索システム

Ⅱ 実験用情報検索システム

1 システム設計の基本方針

1.1 方針

自然語による広域検索の実用性の有無を検討するための実験用検索システムの設計と，このシステムによって，汎用検索システムの設計のために必要な基礎データの収集を目的としている。

一般的には，次の項目，すなわち，

① キーワードの抽出に用いた原資料が適当であったか

② データの処理加工方法が適切であったか

③ 検索方法が，精度，時間，コストの面から見て実用性があるか

という以上3点の解析が容易にできること，および，さらに自然語の広域検索において，当然予想される欠陥をカバーするために取り入れた次の手段，すなわち

①検検索もれを防止するために採用した

(i) 複数の原資料からのキーワードの抽出

(ii) 分類表のキーワードの追加使用

② ノイズ防止のために採用した

(i) 分類とキーワードによるダブルチェックによる評価

(ii) データのキーワード，質問のキーワードおよびマッチング方法に付けたウエイトによる評価

が，有効であったか否かを判断することのできるシステム設計を基本として考えた。

1.2 対象技術領域

通常，技術情報に対する情報活動は，化学部門が最も活発であり，当然，化学部門から着手すべきであるが，化学部門は英国のダーウェント社が既に着手し，現在かなり普及しつつある。電気部門は化学部門についで需要が多いが，文献解析が化学部門より困難であるので，現在ほとんど行なわれていない。機械部門は需要も少なく，またその研究も一番遅れている。

したがって，今回は，需要の多い割合には開発の遅れている電気部門を対象と

して採用した。なお電気部門とは，いわゆる強電，弱電の両者を含む範囲で考えている。

1 3　機種選定

特許情報のコンピュータによる機械処理を行なう場合，情報量に対する処理能力がまず問題となる。この際，能力とは実際の検索処理を行なう場合の処理機能と，それ以前のデータ処理の機能とをいう。非常に多数の量の情報をコンピュータに入力してファイルを作る場合，できる限り短期間に，しかも入力後の修正の量と複雑さを防ぐ上からもできる限り正確なデータ入力ソースを要求する。それらを考え合わせた上で最近のコンピュータの動向を見るとキャラクタ機能によるものからバイト機能によるコンピュータへと切りかわりつつあり，それにともない磁気テープも7トラックから9トラックへと移行している。さらに日本語を通常言語（自然語）によってカナ表現をしている関係上，カナ文字のコードが問題となる。バイト機能のコンピュータではコード数が多くとれるので英数字，カナ文字それぞれ独立したコードで表現されているが，キャラクタ機能のコンピュータではカナ文字コードは2進数の組合わせの限界上，英数字と区別するため言葉の前後にシフトコードを入れることによって表現されているので，必然的に言葉をタイプしてカードなり紙テープなりにするときのエラーの確率は高くなるはずである。事実，データ解析の際に行なったテストデータでは，キャラクタ機能の機種によるパンチ結果はバイト機能の機種によるパンチ結果に較べて悪い結果が示された。さらに一般市場のパンチの処理能力は，シフトコードの入ったものより独立コードのものの方が能力が非常に大きく，さらに正確度も高い。またこのことはコンピュータ内での処理の際もシフトコードの始末をしなくてすむだけ単純でプロセスもそれだけ簡単となる。現在普及しているバイト機種コードは29コードに共通なものが殆んどであり，それをキャラクタ機能の機種である7トラックの磁気テープに変換することも，その逆よりは容易に行なえる状態であり，このことは29コードでソースデータを作ってもキャラクタ機能の機種にも使用できるということになり，結局29コードでデータ入力ソースとすることになった。使用機種は，将来オンラインタイムシェアリング方式で処理する場合は別として仕事のしやすさという点を重視し，中型機を選ぶことにした。

このような観点から見た場合，次の機種が適当であると思われる。

バイト機能の中型コンピュータ

富士通,	FACOM	230シリーズ (モデル25, 35)
日立製作所,	HITAC	8000シリーズ(モデル8210)
東芝,	TOSBAC	5400シリーズ(モデル10, 20)
三菱電機,	MELCOM	1530, 1600
沖電気,	OKITAC	5000, 7000
〃	OUK	9000, 9200, 9300
日本IBM,	IBM	360シリーズ (モデル20, 25)
日本ユニバック,	UNIVAC	418Ⅱ

なお, キャラクタ機能コンピュータでも, 29コードのデータをインプットする際, 特殊コードコンバータを用いて変換するか, コンピュータを介して29コードの磁気テープをキャラクタコードの磁気テープに変換すれば使用可能であり, その場合の機種としては次のものが該当する。

キャラクタ機能の中型コンピュータ

日本電気,	NEAC	2200シリーズ(モデル200, 250)
富士通,	FACOM	230シリーズ(モデル20, 30)
三菱電機,	MELCOM	3100シリーズ (モデル10／30)
日本IBM,	IBM	1600

2 機器構成

前項で述べたように, この実験に使用したコンピュータは, バイト機能をもつ中型機を対象とし, しかも使用する際の地理的条件を考慮して, 次に述べるFACOM230－25を使用することとした。

(1) 機械構成

FACOM230－25

中央演算処理装置　1台

記憶容量　65, 544B (＝Byte)

サイクルタイム　0.75μs／B

命令数　84

汎用レジスター　8

入出力チャンネル　6

磁気ドラム装置　1台

記憶容量　512KB

アクセスタイム　8.4 ms

磁気テープ装置　4台

テープ転送速度　60KB／S

トラック数　9

磁気ディスク装置　2台

記憶容量　5.2 MB

アクセスタイム　87.5 ms

カードリーダー　1台

読込速度　800枚／分

ラインプリンター　1台

印字速度　500行／分

使用文字数　109文字（カナ，英数字，特殊記号）

一行の字数　136字／行

コンソールタイプライター（F792A）　1台

紙テープリーダー（F749E）　1台

読込速度　1200桁／秒

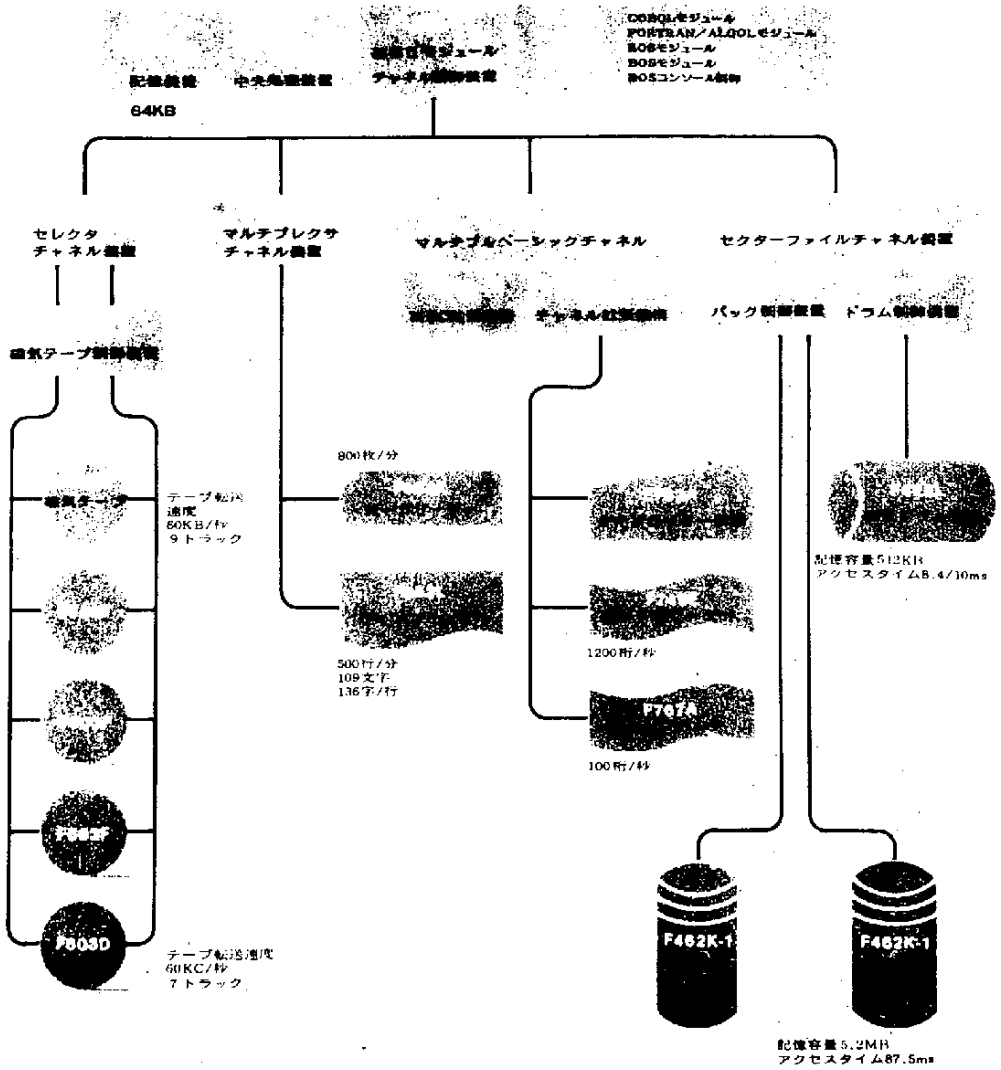
紙テープパンチ（F767A）　1台

さん孔速度　100桁／秒

他にコードコンバート用として

磁気テープ装置（F603D）　1台

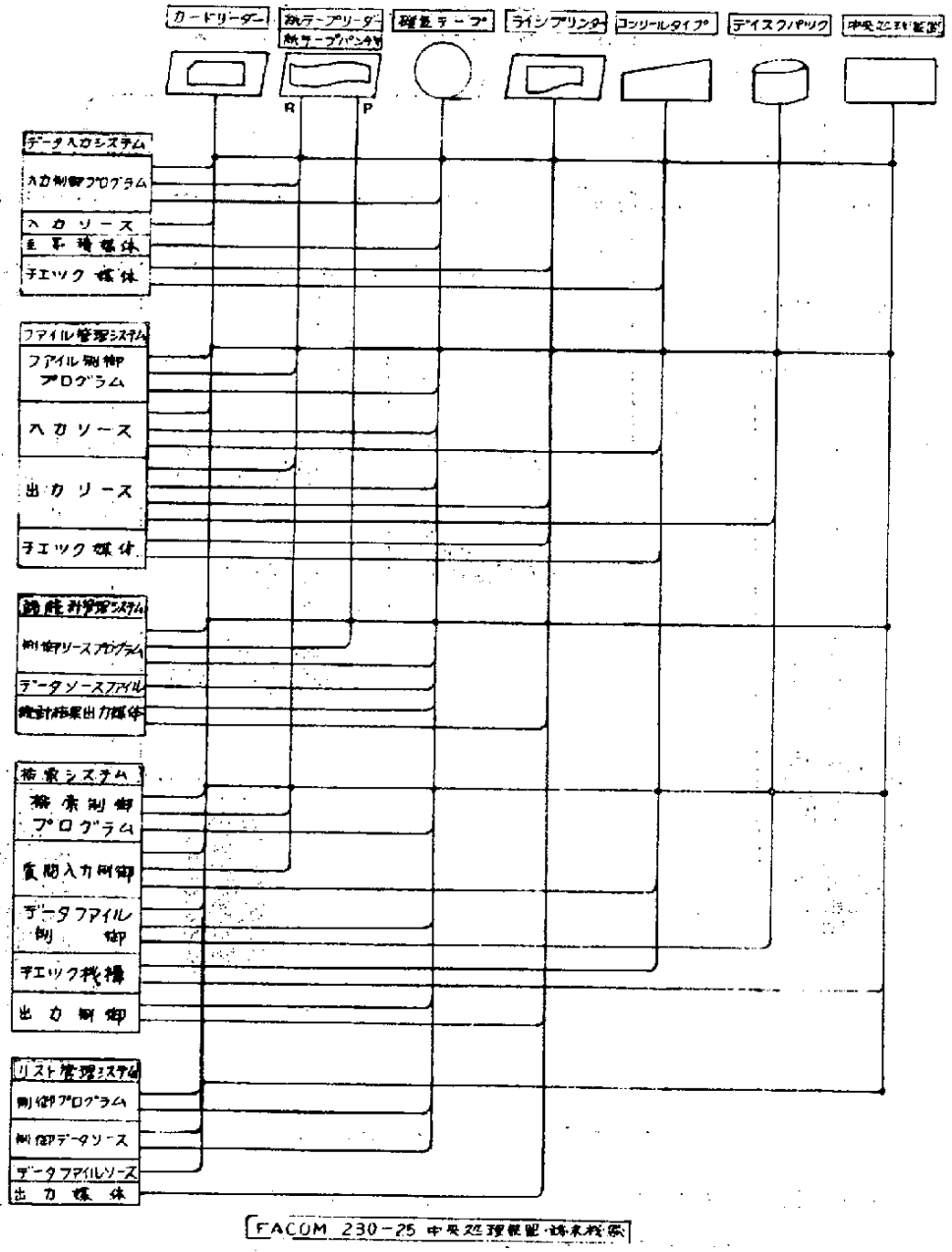
このシステムの構成図を次図Ⅱ－2－1に示す。



図Ⅱ－２－１　システムの構成図

（注）　図中，左端のテープユニットの５台中，最後の１台Ｆ６０３Ｄはキャラクタ機能による機種とバイト機能による機種とのコードおよびトラックコンバート用に専用される。

(2)　本実験による使用形態は次図Ⅱ－２－２の通りである



図Ⅱ－２－２　機器構成使用形態

3　日本語による検索システム

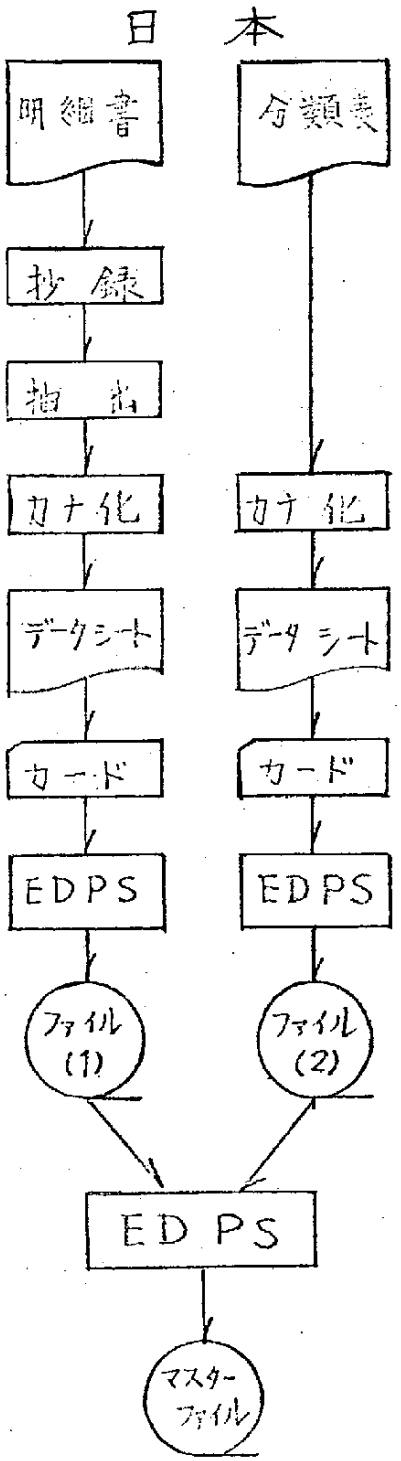
3.1　システムの概要と特徴

蓄積工程は図Ⅱ－3－1で示されるように，明細書から抽出されたキーワードと，分類表から抽出されたキーワードとが，電子計算機によって合成され，検索用のデータファイルが構成される。明細書からのキーワード抽出は，日本特許抄録を使用し，抄録上にアンダーラインを引くことによって行なわれる。アンダーラインされたキーワードは，コーディングシート上に転記し，そこでカナ文字化される。カナ化されたキーワードは，パンチカードにパンチされ，磁気テープに読込まれ，磁気テープファイル(1)を構成する。もちろんキーワード以外のデータ，すなわち，書誌的事項も，コーディングシート上に記載され，パンチされて，磁気テープに読込まれる。

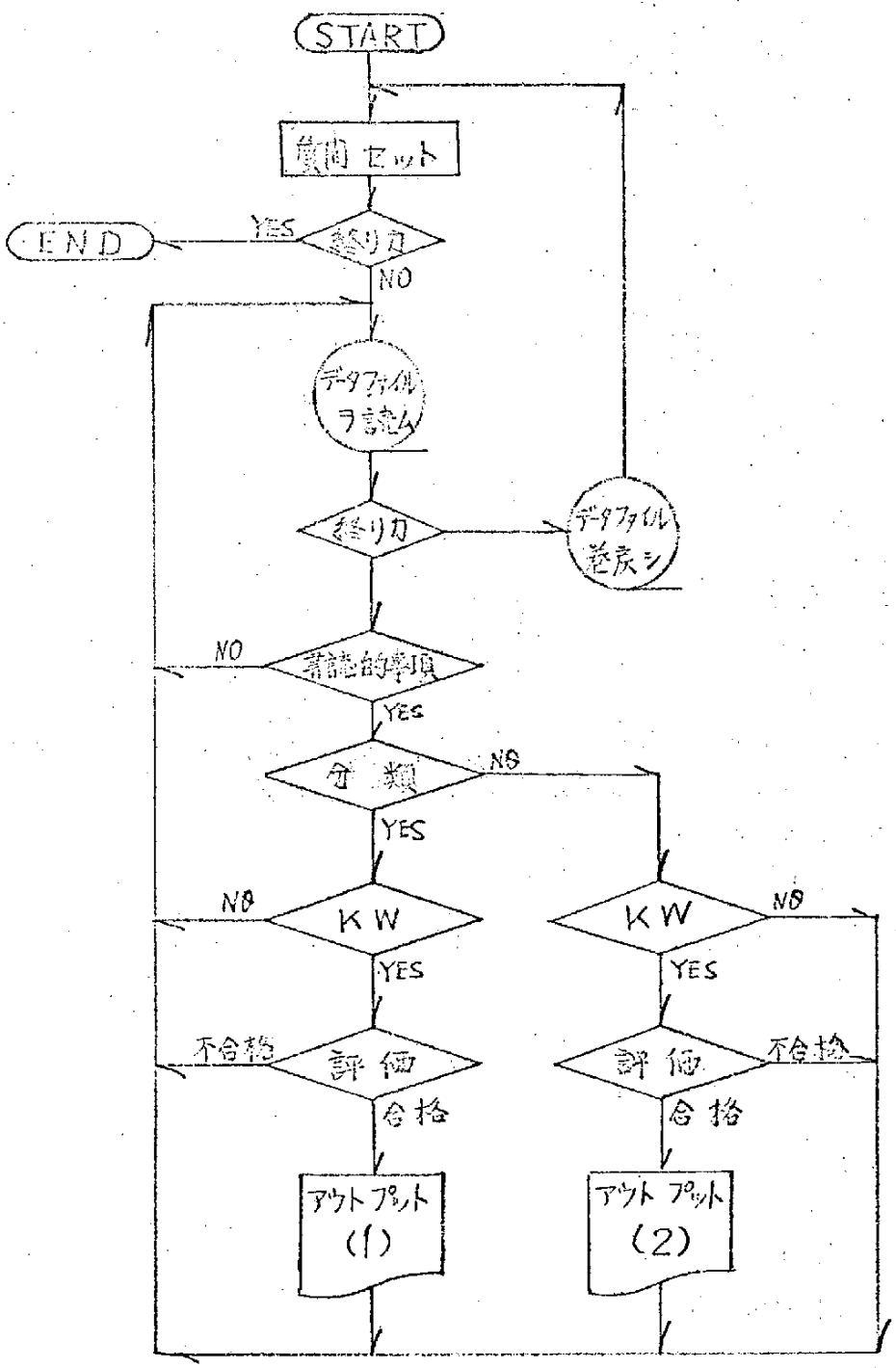
一方，分類表のキーワードは，明細書のキーワードと同じ方式で，カナ化され，分類表のまま，読込まれ，磁気テープファイル(2)を構成する。主ファイル(1)の，各特許に付けられた分類に相当する主ファイル(2)上の分類表を呼出し，各特許の分類の種目に対応する分類表上のキーワードを自動的に取り出し，ファイル(1)上の各特許のキーワードの後に追加する。

この際，すでに抽出されたキーワードと同じキーワードは，自動的に除外し，その特許のデータとして採用されていなかったキーワードのみを追記する。以上の操作をすべて，コンピュータによって行なわせ，検索用の総合ファイルを自動的に編集することができる。

検索は図Ⅱ－3－2に示すように，通常の方法と同様に，書誌的事項について検索した後，キーワードについて，検索する。検索の結果，条件を満たしたものは，さらに，その条件を満たした程度によって評価し，一定の基準に達したもののみ，回答として取り出す。この際，条件の充足度に応じて，回答に優劣をつけ，その順番に応じてアウトプットすることも可能である。



図Ⅱ－３－１　データ蓄積工程図



図Ⅱ－３－２　検索フローチャート

この評価は，さらに分類を加味して行なうこともできる。この際は，図に示すように，分類をパスしたものと，落ちたものとに分け，落ちたものについてもキーワードで検索し，キーワードで合致したものは採用する。しかし分類で，合致したものに較べて，低く評価し，その結果，合格点に適しないものは落すが，合格点に適したものは採用する。

このシステムの特徴を要約すると次の通りである。

① 自然語のカナ文字を使用する。

② 抽出材料として，明細書の特許請求範囲と抄録の2種類を使用する。

③ 検索もれ防止に日本特許分類のキーワードを追加補充する。

④ ノイズ減少のため，分類とキーワードによるダブルチエック，キーワードのマッチングの程度および質問語のウエイト付けによって，評価する。

⑤ データ解析，システムの評価がし易いように，各種コード，インジケータを設けた。

3.2 データ

3.2.1 データの収集と解析

(1) 原資料（特許公報）

・ 日本特許公報

特許出願が審査にパスすると特許公報としてその内容全部が公表される。

特許公報(図Ⅱ-3-3)に記載されている項目中，主なものは次の通りである。

① 公告番号

② 公告日

③ 出願番号

④ 出願日

×⑤ 変更出願

×⑥ 遡及日

⊗⑦ 優先権主張原特許番号

⊗⑧ 優先権主張国名

⊗⑨ 優先権主張日

○⑩ 特許分類

○⑪ 発明者名

98(3) A 0
(98(3) A 4)
(98(3) A 5)

特許庁

特許公報

特許出願公告
昭42−16201
公告 昭42. 9. 4
（全5頁）

広帯域共軛リアクタンス回路

特願 昭38−11992
出願日 昭38．3．7
発明者 鋪春男
東京都港区芝5の7の15．日本電気株式会社内
出願人 日本電気株式会社
東京都港区芝5の7の15
代表者 渡辺賞衛
代理人 弁理士 芦田坦

図面の簡単な説明

第1図は、従来公知の1対の共軛リアクタンス回路の構成図、第2図は、一般に知られている格子型全周波通過回路の構成図、第3図は、本発明の適用例となる可変位相補償回路の構成図、第4図は、本発明に係わる広帯域共軛リアクタンス回路を説明するための定位相差分波器の一例の構成図、第5図は、その動作説明図、第6図は、本発明に係わる1対の広帯域共軛リアクタンス回路の一例の構成図を示す。

発明の詳細な説明

無線周波において、1対の共軛なリアクタンス回路を得る方法としては、従来は、第1図1Aと1Bに示すように1方のリアクタンス $j.x.$ に対して、他方は、特性インピーダンスが R_0 オームで長さが使用中心周波数において $\frac{1}{4}$ 波長の長さを有する遅延線路11を介して1Aと同一のリアクタンス $j.x.$ を接続することによつて共軛なリアクタンス $Z_{12}=\frac{R_0^2}{jx}$ をえていた。

この方法は、使用中心周波数において共軛なリアクタンスとなつていても、使用周波数が中心周波数から10%以上偏ると、共軛な関係が満足されなくなる欠陥を有していた。したがつて共軛な関係において、1対のリアクタンスを忠実に運動せしめることは増々困難であつた。このような運動可能で共軛な関係にある1対のリアクタンス回路は、たとえば広帯域超多重無線中継回線において使用せられる回線位相補償回路において必要となつている。これは、広帯域超多重無線中継回線の回線の品位は、端局装置における回線の位相歪の集中等化の良否に負うところが大であるためである。

しかしながら、補償すべき位相歪は未知であるから、可変型のものが必要である。この種の目的に対して種々の提案があるがいずれも第2図(A)に示すような格子型全周波通過回路を数段縦続した構成と等価であつてかつ高周波であるために不平衡型に変形している。たとえば第2図(B)は、第2図(A)と等価な不平衡回路である。この基本回路において伝達インピーダンス R_0 オームを不変に保ちながら位相特性の形を変えるためには、つねに

$$L_p C_p = L_s C_s$$

$$\frac{L_s}{C_p} = R_0^2$$

の二つの関係が同時に保たれていなければならない。すなわち、同時に三つの変数をある一定の関係において、変化しなければならない。これは事実上困難であるため、一般にはさらに変数の中の一つを固定した形式が取られている。しかしこのように変数を二つに限定しても、必ず1つのインダクタンスと、もう一つのキャパシタンスを連動しなければならない。このように異種のリアクタンスを一定の関係で連動せしめることはとくに高周波においては容易でない。これを避ける方法として、たとえば、第3図に示すようなハイブリット回路を使用する方法が考えられる。

第3図(A)は基本型を示し、同図(B)はVHF帯で実現するための同図(A)の変形を示している。T_1 はいわゆるハイブリッド変成器 T_2 T_3 T_4 はこれを高周波において実現するための広帯域変成器で綜合して、T_1 と等価である。ここで T_2 の特性インピーダンス Z_0 は $\frac{RO}{2}$ オーム、T_3 T_4 の特性インピーダンスは R_0 オームに選ばれている。この広帯域ハイブリッド変成器については、IRE 1959 August P−1339 Some Broad Band Transformers" に述べられている。このハイブリッド回路の共軛は二端子31と32とに共軛なインピーダンス関係を有するリアクタンス Z_{33} と Z_{34} とを接続すれば、入力電圧 E_{30} を加えたとき出力はすべて負荷35に表われる。かくして得られた位相補償回路の遅延時間特性は、リアクタンス Z_{33}（または Z_{34}）の値によつてのみ決定されるからリアクタンス Z_{33} と Z_{34} とが広帯域

○⑫　発明者住所

○⑬　出願人名

○⑭　出願人住所

⊗⑮　代理人名

⑯　発明の名称

⑰　内容の説明

⑱　特許請求範囲

×⑲　図面

注１）　×　ない場合がある。

２）　○　複数個ある場合がある。

３）　⊗　ない場合もあるし，複数個ある場合もある。

・　対象分野

日本特許分類中，次の分類に該当するもの（副分類で該当するものも含めた。）を対象とした。

類	類の名称
55	発電，電動
56	変電
57	電池
58	送電，配電
59	一般的電気部品
60	電線，ケーブル，配線
61	電気絶縁
62	電気材料
93C	弧光燈
D	放電燈
E	白熱電燈
96(1)	電気通信
96(2)	電信通信
96(3)	電話通信

9 6 (4)	電話交換
9 6 (7)	伝送
9 6 (8)	多重伝送
9 7 (3)	写真電送，模写電送
9 7 (5)	テレビジョン
9 8 (3)	伝送回路，空中線
9 8 (5)	基本電子回路
9 9	電子管
9 9 (5)	半導体装置およびそれに使用する半導体
1 0 0	電気的諸装置
1 0 2 B	録音，再生一般
E	磁気的録音，再生
1 1 0	電気，磁気量の測定
1 1 4 A	計算機

・　特許抄録

キーワードの抽出材料には，発明協会発行の日本特許抄録カードを使用した。

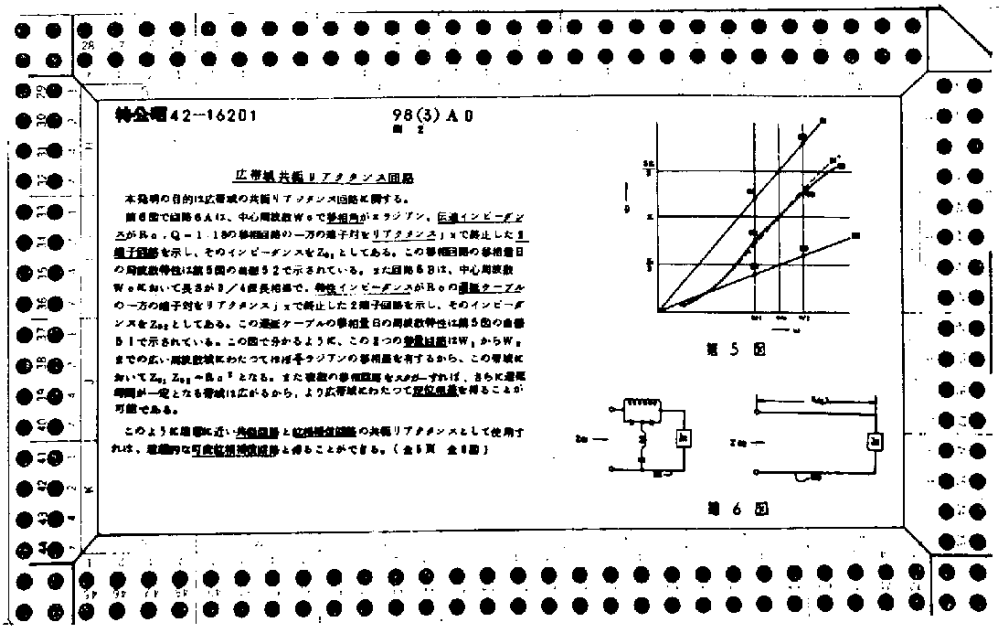
この抄録カードは，昭和３７年以来，主要部門について発行されているもので，図Ⅱ－３－４に示すように，片面に抄録および図面裏面に特許請求範囲が記載されている。なお，この抄録の作成は㈱特許デイタセンターが担当している。

今回の実験では，抄録からキーワードの抽出の可否を検討するために，キーワードの抽出材料のサンプルとしてこの抄録を使用したが，抄録と，請求範囲の両方が記載されているので，同時に両方の抽出を行なうことができて便利である。

・　データ数

昭和42年9月1日～１１月１５日の間に発行された特許公報で，前記分類に主，副いづれかが該当するもののうち，上記抄録の発行されているもの2,０００件を対象とした。

この選定に当っては，内容的に取捨選択せずに，９月１日発行のものから分類に該当するものを番号順に拾い出した。なお，該当する公報であっても抄録の作成されていないものは除外した。



特公昭42—16201　　98(3)A0

広帯域共振リアクタンス回路

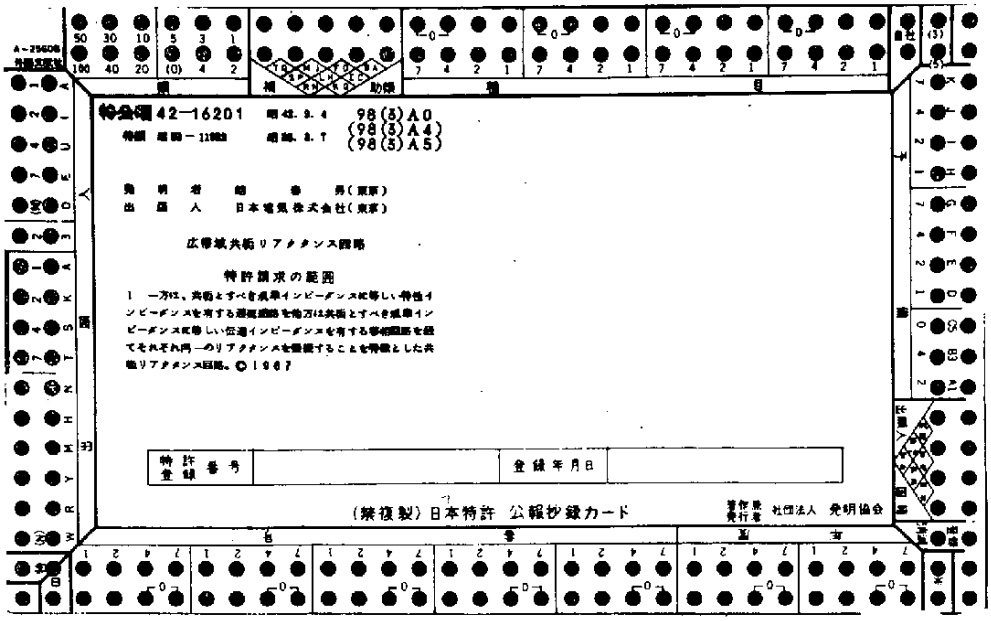
本発明の目的は広帯域の共振リアクタンス回路に関する。

第6図で回路6Aは、中心周波数W_0で等相角がπラジアン、伝達インピーダンスがR_0，$Q=1$の等相回路の一方の端子対を「リアクタンス」xで終止した2端子回路を示し、そのインピーダンスをZ_{01}としてある。この等相回路の等相量Bの周波数特性は第5図の曲線52で示されている。また回路6Bは、中心周波数W_0において長さが3／4波長相当で、特性インピーダンスがR_0の通路ケーブルの一方の端子対を「リアクタンス」xで終止した2端子回路を示し、そのインピーダンスをZ_{02}としてある。この通路ケーブルの等相量Bの周波数特性は第5図の曲線51で示されている。この図で分かるように、この2つの等量回路はW_1からW_2までの広い周波数域にわたつて相等ラジアンの等相量を有するから、この等域において$Z_{01}\,Z_{02}=R_0{}^2$となる。また複数の等相回路をカスケードすれば、さらに通過帯域が一定となる等域は広がるから、より広帯域にわたつて平坦減衰を得ることが可能である。

このように最簡に近い共振回路と位相補償回路の共振リアクタンスとして使用すれば、理論的な可能位相補償減衰器と称することができる。（全5頁　全8図）

第5図

第6図



特公昭42—16201　昭42.9.4　98(5)A0 (98(5)A4) (98(5)A5)

特願 昭33－11903　昭38.3.7

発明者　船　倉　勇（東京）
出願人　日本電気株式会社（東京）

広帯域共振リアクタンス回路

特許請求の範囲

1　一方は、共振とすべき減衰インピーダンスに等しい特性インピーダンスを有する通過回路を他方は共振とすべき減衰インピーダンスに等しい伝達インピーダンスを有する等相回路を設けてそれぞれ同一のリアクタンスを接続することを特徴とした共振リアクタンス回路。◎1967

特許登録番号		登録年月日	

著作権者　社団法人　発明協会

図Ⅱ－3－4　日本特許抄録カード

昭和44年9月1日～11月15日の発行特許件数と対象とする分類に該当するものの数は次の通りである。（表Ⅱ－3－1参照）

表Ⅱ－3－1　　データの抽出件数　　単位：件

	9月	10月	10月(1-15)	合計
特許発行件数	3600	2800	1400	7800
対象分類に該当するもの	865	765	371	2001
データとして採用したもの	865	764	371	2000

(2)　解析

キーワードの抽出は，抄録カード上にアンダーラインを引くことによって行なった。

抽出個所は，発明の名称，抄録，特許請求範囲の三個所とした。

キーワードの抽出基準は下記によった。

・　一般的基準

①　抽出の対象とするディスクロージャー

発明の対象とするディスクロジャーに焦点をあわせ，従来法や参考例に関するものは除外した。

②　技術内容による重点の置き方

電気部門に関する技術用語に重点を置き，他の部門に属する記載については，代表的用語のみにとどめた。

例えば，化学的製法についての記載があるときは原則として，その製法による最終製品に関するキーワードのみを抽出した。

注：Ⅰ－214「広域検索システムにおける各領域ファイル相互の取扱い」の原則による。

・　具体的基準

①　文献の内容を代表するような技術用語を主体とする。

②　発明の対象物,用途などに関する具体的記載中の技術用語は特に重点的に抽

出する。

③ 頻度の多い技術用語も対象とする。

④ 同義語，類似後，上位概念などはできるだけ選定する。特に上記②に関するものはもれなく拾うこと。

⑤ 実施例中の細かい条件，細部構造に関するものなどはなるべく除外する。

⑥ 一度選定した語が再度現われたときは，はじめに選定したものだけにする。

⑦ キーワードには特殊な文字，記号などはなるべく使用しない。

⑧ 次の場合は両方共キーワードとして抽出する。

イ （接頭語＋語幹） （語幹）

ロ （接頭語＋語幹） （語幹＋接尾語）

ハ （語幹1＋語幹2＋接尾語） （語幹2＋語幹1＋接尾語）

ニ （接頭語＋語幹2＋語幹1） （接頭語＋語幹2＋語幹1）

ホ 反対語

⑨ 次の場合は前者をキーワードから除外する。

（語幹） （語幹＋接尾語）

⑩ 語＋語で表現した方がよいキーワードには＋の位置に＊の記号をつける。

⑪ キーワード数は最高25個までとする。

(3) リンクの使用

抽出基準⑩にあるように，試験的にリンクを採用して見た。リンクの採用が検索上どのように効果があるかを見ることも，この実験の目的の一つである。

すなわち，独立した二語が意味的に結合していて，それを切り離して別々のキーワードにした場合，その文献の内容を十分表現できないときは，その二語の間に＊を入れ，一つのキーワードと同じ取りあつかいをした。

検索の場合は，この二語を完全なる一語と見なして検索もできるし，また，別々に切り離しても検索できるようにした。

例，電子ビームの利用率→電子ビーム＊利用率

直流レベルの検出→直流レベル＊検出

(4) ネガテブな条件の採用

特許の中には，従来使用されていた特定の手段，材料などを用いなくてもよいということが特長になっている場合がある。

この実験では，上記の条件をマイナス条件として試験的に採用して見た。

実際のコーデングに際しては，マッチングの仕方との関係で，次のように取扱った。

例，「感温制御器を使用しなくても……できる」→　－＊カンオン　セイギヨキ

3.2.2　カナ文字化

抄録上にアンダーラインしたキーワードをデータシート上でカナに直す。（図Ⅱ－3－5参照）

カナ化の場合問題になるのは，カナふりの方法，すなわち表記法と，ある単位で分けて書く方法，すなわち分かち書き法にある。これらを統一しておかなければ，検索に際し，該当するキーワードがあっても検出されなくなる。また，コンピュータを使用する場合，使用文字に自ら制限があるので，その制限範囲内の文字を使用するように規定することも必要である。

この実験で試験的に採用したカナ化法の規定の要約したものを次に示す。

①　表記法

文部省新カナ使い50音表を使用し，用法はカナ文字会の用字法を原則として用いた。

特異点

(イ)　長音は通常ウを用いるが，長音記号「－」を用いる。

(ロ)　ヲ→オ，ヂ→ジ，ヅ→ズ　に統一

②　分かち書き法

原則，有意の2漢字単位で分かち書きする。

例，自動電信方式→ジドー△デンシン△ホーシキ

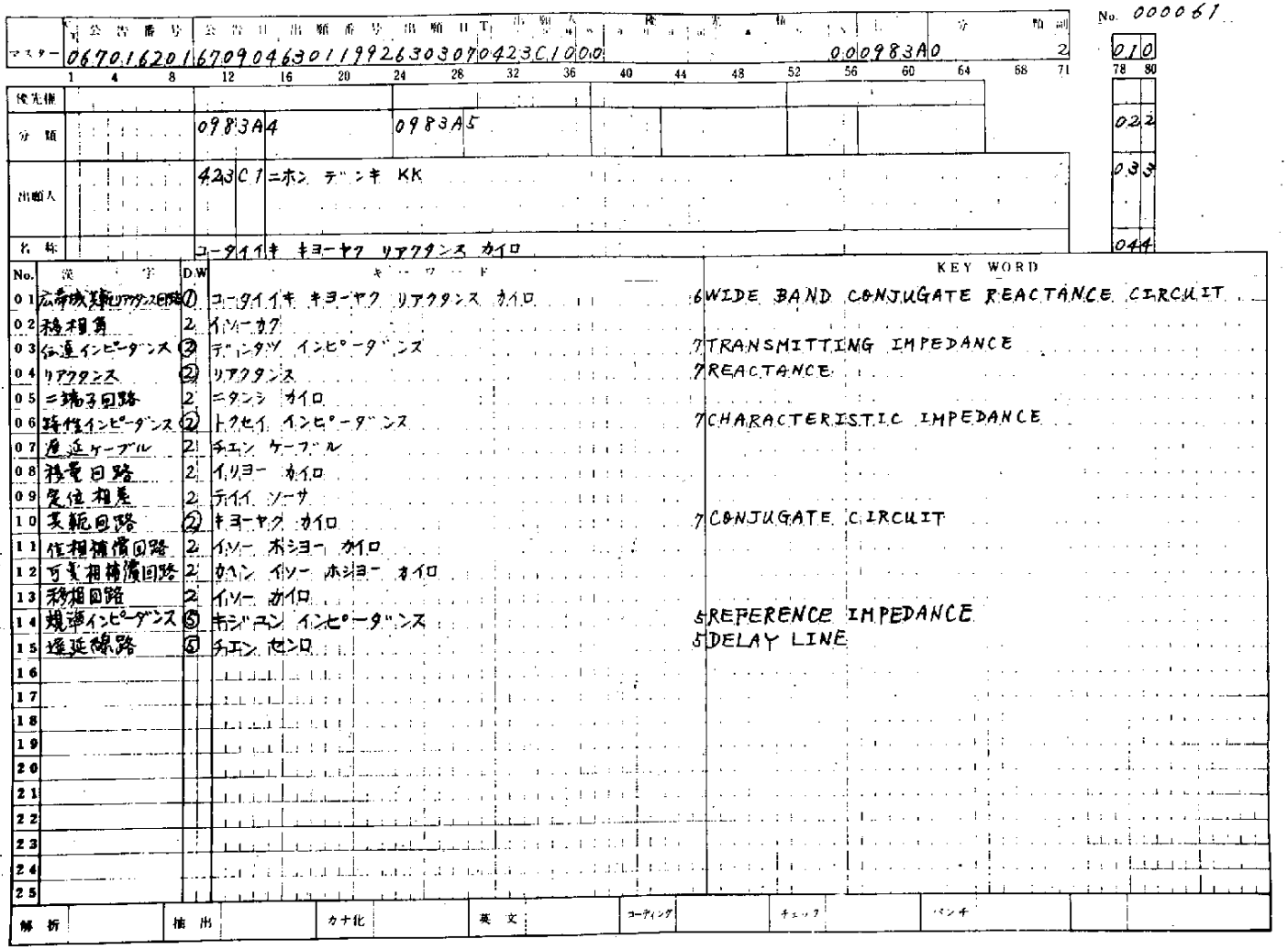
例外規定

(イ)　有意の2漢字に附属する1漢字は，2漢字に付ける。

例，光電管→コーデンカン

極超短波→ゴクチョータンパ

(ロ)　次に示す能動型接尾語は単独で切り離す。「用，型，形，式，的，性，状，


No. 000061
マスター 06701620167090463011992630307042 3C1000 0.0 0983A0 2
010
022
033
044
0983A4
0983A5
423C1 ニホン デンキ KK
ユータイイキ キョーヤク リアクタンス カイロ
No. 漢字 DW キーワード KEY WORD
01 広帯域共役リアクタンス回路 ① コータイイキ キョーヤク リアクタンス カイロ 6WIDE BAND CONJUGATE REACTANCE CIRCUIT
02 結合量 2 インピーカク
03 伝達インピーダンス ③ デンタツ インピーダンス 7TRANSMITTING IMPEDANCE
04 リアクタンス ② リアクタンス 7REACTANCE
05 二端子回路 2 ニタンシ カイロ
06 特性インピーダンス ② トクセイ インピーダンス 7CHARACTERISTIC IMPEDANCE
07 同軸ケーブル 2 ドウジク ケーブル
08 移電回路 2 イデンカイロ
09 受信相差 2 ジュシンソーサ
10 共役回路 ② キョーヤク カイロ 7CONJUGATE CIRCUIT
11 位相補償回路 2 イソー ホショー カイロ
12 可変相補償回路 2 カヘン イソー ホショー カイロ
13 移相回路 2 イソー カイロ
14 規準インピーダンス ⑤ キジュン インピーダンス 5REFERENCE IMPEDANCE
15 遅延線路 ⑤ チエン センロ 5DELAY LINE
16
17
18
19
20
21
22
23
24
25
解析 抽出 カナ化 英文 コーディング チェック パンチ

図Ⅱ－３－５　データシート

化」

（説明）これらの接尾語は，用途，性質，形式などを表わすので，これらの語をキーワードの性質を判別するための手段として利用できないかということを検討するために試験的に切り離して見た。

また，これらの接尾語は，かならず付ける場合と，省略する場合，または付けたり，付けなかったりする場合など，色々な慣用的用法があるので，この不規則性が，どの程度検索の邪魔になるかを見ることもその目的とした。

例，電信用継電器→デンシンヨーケイデンキ

(ｲ) 反語中「非」「不」のみ切り離し，その後に＊を挿入する。

例，非結晶 → ヒ＊ケツショウ

（説明）反語の場合は，語幹のみで検索する可能性もあるので，否定語と語幹とを切り離した。

全部の反語について，このような規定を設けなかったのは，その代表的な非，不について試験的に行なって，その利害得失を検討し，検討の結果，このような取り扱いを将来するかしないかを決めるためである。

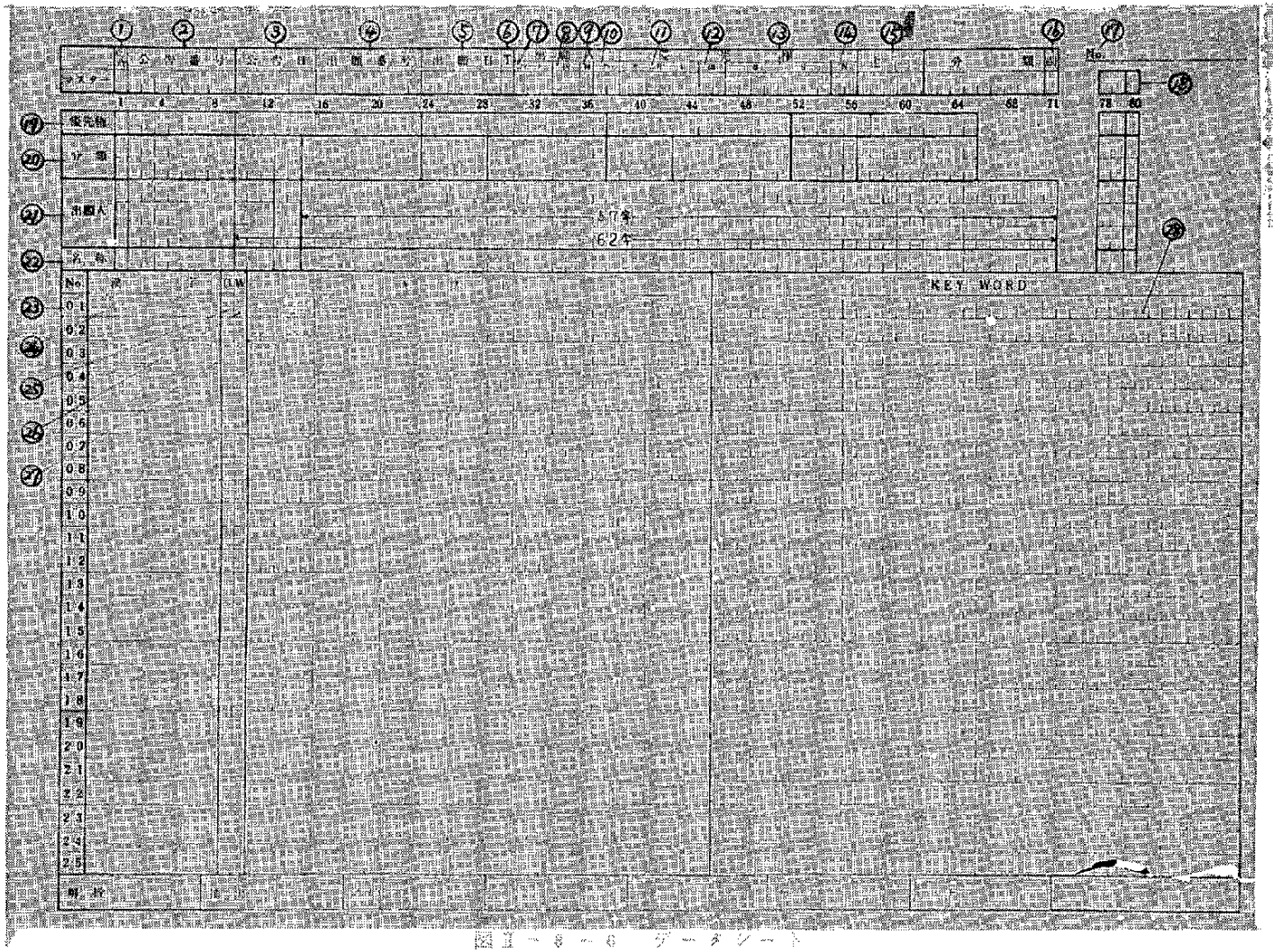
(ﾊ) カナ書きの外来語，英，数字，化学名，化学式，化学記号，数式，計算機用語にない特殊文字などの取り扱いは別途定めた。

323 データシートおよびインプットカード

(1) データシート

図Ⅱ－3－6に示すデータシートを使用した。

インプット項目および，その内容は次の通りである。



図Ⅱ－３－６　データシート

図Ⅱ－３－６

イ　書誌的事項（表Ⅱ－3－2参照）

表Ⅱ－3－2　　書誌的事項一覧表

番号	内容	カラム数	説明
1	C／T	1	
2	公告番号	8	年度は西歴末尾2桁
3	公告日	6	〃
4	出願番号	8	〃
5	出願日	6	〃
6	T	1	追加，分割，変更等特許の種類
7	出願人コード	3	
8	〃　国籍	2	
9	〃　種類	1	法人，個人の別
10	〃　外	2	複数出願人の数
11	優先権主張日	6	年度は西歴末尾2桁
12	〃　国	2	
13	〃　番号	8	
14	〃　N	2	複数優先権の数
15	特許分類	14	主分類
16	副分類数	1	
17	整理番号	6	
18	カードコード	2	
19	優先権	14	複数優先権4個迄
20	副分類	14	副分類8個迄
21	出願人コード 出願人名	5 57	出願人3人迄
22	発明の名称	62	

ロ　キーワード（表Ⅱ－3－3参照）

表Ⅱ－3－3　　キーワード項目一覧表

番号	内　　容	カラム数	説　　明
23	カードコード	2	
24	和文キーワード	－	原データから転記する
25	D	1	キーワードの出所　注1）
26	W	1	キーワードのウエイト　注2）
27	カナ化キーワード	35	磁気テープファイルでは46字分取ってあるので，書ききれないときは続けて記入する
28	英文キーワード	40	〃

注1）　キーワードの出所

キーワードの抽出された場所，すなわち，発明の名称，抄録，請求範囲などをコード化し，検索実験の結果の解析を行なうために付けたものであって，実用段階には不要となるものである。しかし，分類表からのキーワードも，このコードで区別してあるので，キーワードの出所によって，回答を評価することも可能である。

なお，出所のコードを次に示す。（表Ⅱ－3－4参照）

表Ⅱ－3－4　　キーワードの出所コード

内容	コード
分類表	0
名称	1
抄録	2
簡易抄録	3
明細書	4
請求範囲	5
名称＋請求範囲	6
抄録＋請求範囲	7

注２）　キーワードのウエイト

キーワードそのものにもウエイトを付ける予定で研究したが，ウエイトの付け方の思想的統一が困難なことと，質問またはマッチングにおけるウエイトとの関係の調整が困難であったため，今回は見送った。

なお，参考までにウエイト付けの基準の試案を次に示す。

案１　キーワードのウエイト付けを次の７段階にわける。

1.　重要主題

発明の全内容を表現するキーワード

2.　主要主題

発明の大部分を表現するキーワード

3.　上位主題

発明の内容を表現するキーワード

4.　下位主題

発明の内容と関係しているが，表現されることがあまりにもせますぎるキーワード

5.　副次的主題

発明を構成する中間的方法，可能な用途などに関するキーワード

6.　関連主題

発明に関するキーワード

7.　その他の主題

ある利用者には有効であるが，他の利用者には必要がないキーワード

案２　キーワードのウエイト付けを次の４段階に分ける。

1.　発明の対象物

主題そのもの，または主題の具体的な説明に用いられている主要語

例　（名称）　小型電動機用ブラシ保持器

（説明）　この発明は例えば真空掃除機に使用する小型電動機

2.　発明の主要構成要件

発明のポイントを説明するために使われている主要語，即ち，主要部品，主要主段,などに関連するキーワード

3.　発明の作用効果

発明の利点，効果等を説明する主要語

この場合2つの語の連合したものが使われる場合が多い。

例　電子ビームの利用率，ひいては再生映像の明るさが低下するのを防止できる。

電子ビーム＊利用率

再生映像＊明るさ

4　その他

（説明）　案1は，キーワードの重要度をそのまゝ表現したものであるが，実際に抽出する場合その判定が困難であり，かえって混乱をまねくおそれがあるので実行上，不向きである。

案2は，キーワードの重要度よりも，むしろ，キーワードの性格，即ち，ファセット的要素で分類されているので，その性質を検索に利用すれば，ノイズ軽減に役に立つのではないかと思われる。

(2)　インプットカード

インプットにはＩＢＭ８０欄カードを使用した。データの内容に応じ次の4種のカードを作製した。（図Ⅱ－３－７）

①　マスターカード

②　複数優先権，副分類併用カード

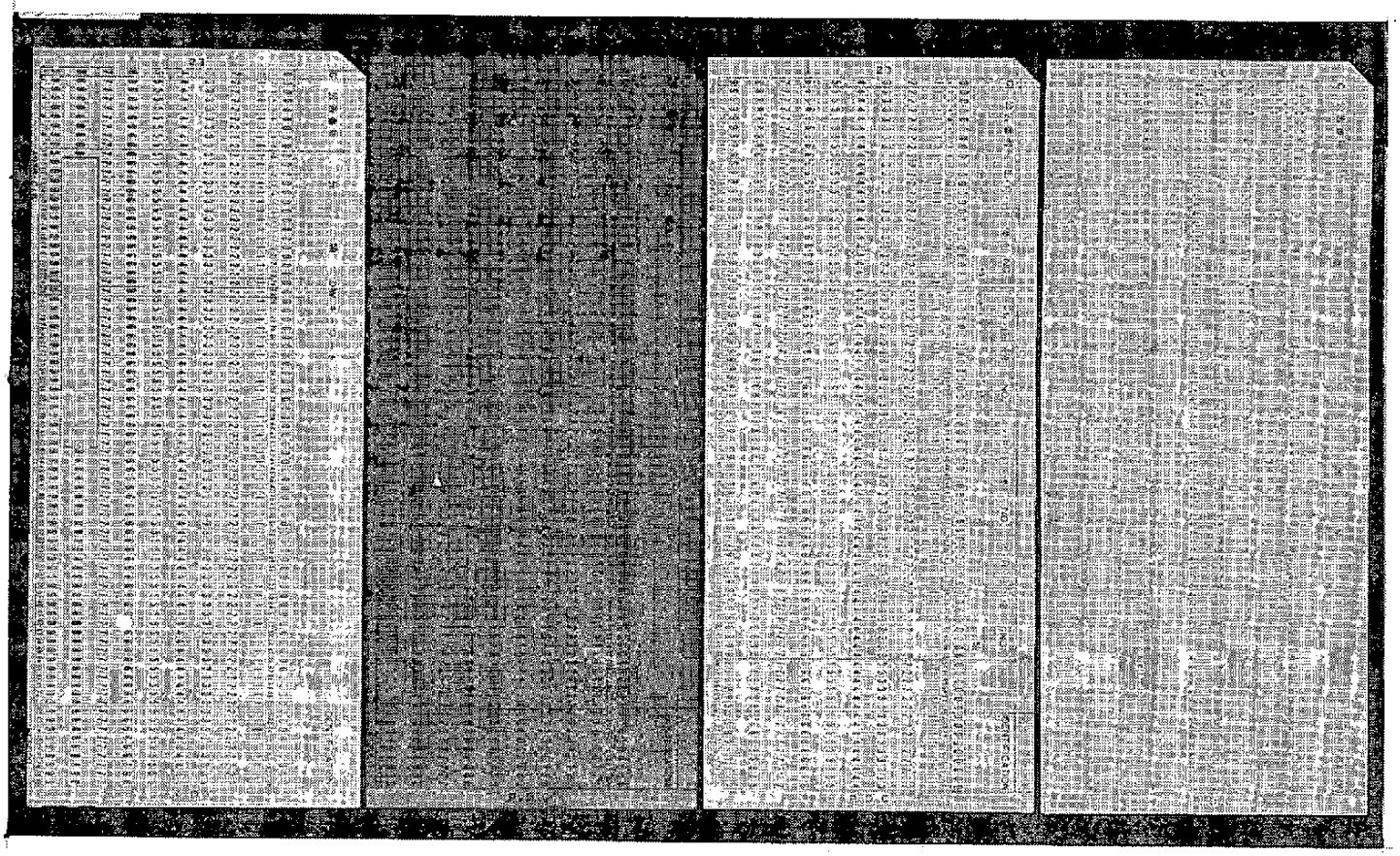
③　出願人，発明の名称併用カード

④　キーワードカード

マスターカードには書誌的事項の主要項目が全部入るように設計されている。マスターカードに入らなかった複数優先権，副分類，出願人名，および，発明の名称に関する項目は，それぞれ別カードとした。

複数優先権，副分類は，1枚のカードに4個迄，出願人，名称は1枚のカード1つ宛パンチするようにした。出願人が多数いる時は，その数だけカードを使用する。

キーワードカードは，キーワード1語につき1枚のカードを使用する。このカードには，キーワード，出所，ウエイトの外，シソーラス作成のための用語整理の便を考えて，キーワードの所属文献番号，主分類の補助類迄も記入してある。さらに，カードによって人間が，用語整理を行う場合を考慮に入れて，漢字キー



図Ⅱ－3－7　インプットカード

ワードを記入する欄も設けてある。

3.2.4 分類表

自然語による検索の場合には，原データから抽出されたキーワードだけに頼ると，文献の記載方法が偏っているときに，検索もれを生ずる危険性がある。

この実験では，標準的用語を補助キーワードとして機械的に追加することにより，その欠陥を除くこととし，その手段として，分類表のキーワードを使用した。

この実験に使用した日本特許分類表は図Ⅱ－3－8に示すように10進分類である。

この分類表中，3.2.1「データ収集と解析」の項で述べた，対象分類について，3.2.2「カナ文字化」の項で述べたカナ化法の規則に従ってカナ化し，キーワード1語につきキーワードカード1枚にパンチし，磁気テープに読み込み，3.3「検索システムとプログラム」の項で述べる方法によって，機械的に分類キーワードを補助キーワードとして各データファイルのキーワードの後に追加する。

分類表をカナ化したデータシートを図Ⅱ－3－9に記載する。

3.2.5 ソース データ ファイル

解析したデータをコンピュータで処理するために，カードにパンチした状態のデータを各分献単位で指定されたフォーマットに従ってコンピュータに読み込み，ファイルを作成する。このとき，ソースデータは，初めから後の仕事に適した形態をもつファイルとはせず，データ蓄積登録の意味を持つ主ファイルと，実際のデータ処理を行なう仕事用のサブファイルとに分け，主ファイルを汎用性のある形を持ったファイルにしておく方が後のデータ処理に都合がよい。データシート上のデータは，カードにそれを転移するとき，パンチミスや転移作業の繁雑性および複雑性を避けるために，各データをブロックに分け，ブロック単位でカード化するのが好ましい。

また，データの種類によっては，カード枚数およびカード中に収容されているデータの数が不定であるので，各カードには，データの種類，順序，ブロック内のデータ数，カード枚数などを制御するコントロール コードを附し，入力時にチェックおよびコントロールができるようにした。

ソースデータをコンピュータにインプットしてファイルを作る時点では，分割集団化されているデータを各項目別集団にまとめながら，あまり手を入れて加工

第98(3)類　伝送回路，空中線

この類は，昭和42年7月1日付分類改正によって新設された類である。

この類は，第96類，第98類に分類されていた伝送回路，空中線に関する事項を再編成して分類したものである。

したがって，この類の補助類および種目に属する事項について，改正前どこに分類されていたかを知るには，廃止された第96類，第98類を参照されたい。

A　集中定数回路
B　分布定数回路
C　立体回路
D　空中線

A　集中定数回路

0　集中定数回路（←集中定数回路網に関する測定）
01　素回路網
02　擬似回路網
1　結合回路，分岐回路（←サーキュレータ）
2　変換回路，整合回路
3　共振器，ろ波器
31　共振器
32　ろ波器
321　普通ろ波器
322　圧電ろ波器
323　磁わいろ波器
324　機械的ろ波器
325　集積回路ろ波器（製造法，集積構造→99(5)H）
4　分波器，合波器
5　伝送特性補償回路
51　周波数領域等化器
52　時間領域等化器
6　減衰器（←アイソレータ）
7　移相器（←ジャイレータ）
8　遅延回路，遅延線

B　分布定数回路

0　分布定数回路
1　結合回路，分岐回路
2　変換回路，整合回路
3　共振器，ろ波器
4　分波器，合波器

C　立体回路

0　立体回路
01　回路素子
1　結合回路，分岐回路（←サーキュレータ）
2　変換回路，整合回路
3　共振器，ろ波器
31　共振器
32　ろ波器
4　分波器，合波器
5　伝送特性補償回路
6　減衰器（←アイソレータ）
7　移相器（←ジャイレータ）

D　空中線

0　空中線
01　放射素子の形状，構造
011　可とう性をもつもの
012　伸縮自在なもの
013　折りたたみ自在なもの
1　放射パターンの形成
11　放射素子と他の装置との組み合わせ
12　空中線配列
2　共振形空中線
3　非共振形空中線

図Ⅱ－3－8　日本特許分類表

分類 98(3)A		
番号	項目	キーワード
001	0	シュー チュー テイスー カイロ
002	01	ソカイロモー
003	02	キ゛シ゛ カイロモー
004	1	ケツコ゛ー カイロ」
005		フ゛ンキ カイロ」
006		サーキュレータ
007	2	ヘンカン カイロ」
008		セイコ゛ー カイロ
009	3	キョーシンキ」
010		ロハキ
011	31	キョーシンキ
012	32	ロハキ
013	321	フツー ロハキ
014	322	アツテ゛ン ロハキ
015	323	シ゛ワイ ロハキ
016	324	キカイ テキ ロハキ
017	325	シュー セキ カイロ ロハキ
018	4	フ゛ンパ゜キ」
019		コ゛ーハキ
020	5	テ゛ントー トクセイ ホショー カイロ
		ー カイロ

図Ⅱ－３－９　分類表カナ化データシート

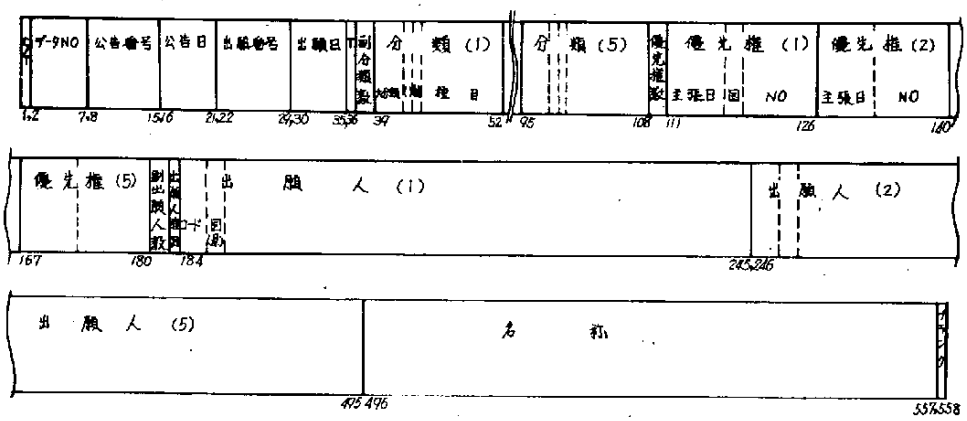
せずに主ファイルを作成した。そして，データ管理システムによって，この主ファイルから仕事の内容に適した形をもつファイルに作成し直すことができるようにした。その方が仕事もし易く，データのファイルとしても整ったものが作成されることになる。また，定期的にデータの修正を行なったり改定したりする際にも主ファイルのみを行なって，その日付けをコードにして各ファイルの制御を行なうことができ，繁雑さを避けられて有利である。また，主ファイルは文献の書誌的事項とキーワードとを区別して別ファイルとした。

次に，この実験用情報検索システムのために作成したファイルについて述べる。

(1) 日本特許主ファイル

a 書誌的事項主ファイル

このファイルには，特許の明細書に記載されている書誌的事項が常用言語（以下自然語と書く）により，別図Ⅱ－3－10のフォーマットに従って蓄積されている（バイト数558バイト）

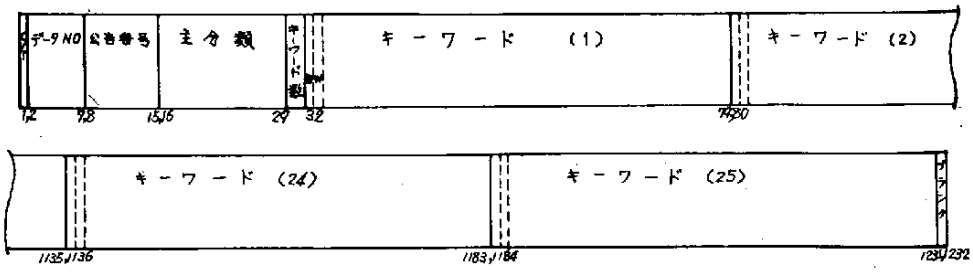


図Ⅱ－3－10 書誌的事項主ファイル

b キーワード主ファイル

各文献から解析されたキーワードが書誌的事項のファイルに関連づけされて，別の主ファイルとして別図Ⅱ－3－11のフォーマットに従って蓄積さ

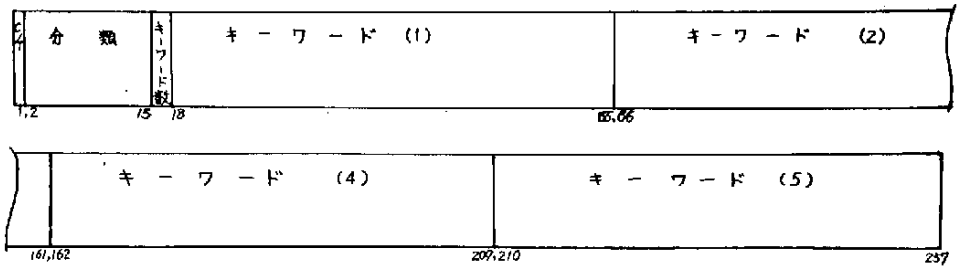
されている。（バイト数１２３２バイト）



図Ⅱ－３－１１　キーワード主ファイル

c　日本特許分類キーワード主ファイル

このファイルは，日本分類の分類表から抽出したキーワードを分類標数と共に別図Ⅱ－３－１２のフォーマットで磁気テープ化したものである。このデータは各文献の書誌的事項の中の分類のコードと関連づけて使えるようにしたものである。（バイト数は２５７バイト）



図Ⅱ－３－１２　日本特許分類キーワード主ファイル

d　英訳キーワード主ファイル

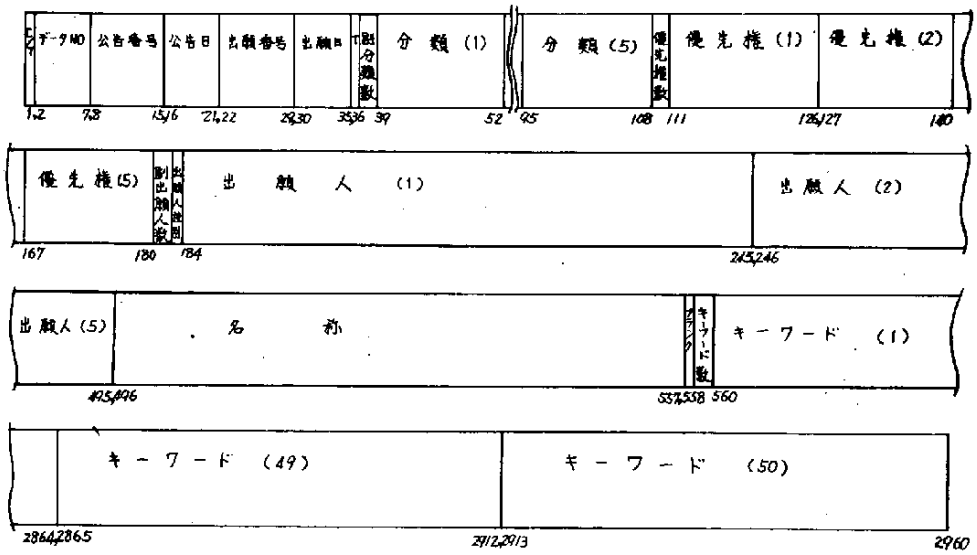
このファイルは，ｂキーワード主ファイルの和文キーワードに対応する英

文抄録より抽出して，ｂ主ファイルと同様に配列したものである。したがって，ｂにあるキーワードでも英文抄録になければ，対応部分が空白になっている（バイト数は１２０７バイト）

(2) 日本特許検索用ファイル

a 検索用総合ファイル

このファイルは，(1)ａの書誌的事項主ファイル，(1)ｂのキーワード主ファイル，(1)ｃの日本特許分類キーワード主ファイルの内容を総合し，図Ⅱ－３－１３のフォーマットに従って編集しなおしたもので，主として検索用に使用する。（バイト数は２７６１バイト）



図Ⅱ－３－１３ 検索用総合ファイル

b 検索用サブファイル

上記の(2)ａの総合ファイル中，日本特許分類キーワードのデータを除いたもので，分類キーワードを使用しない場合の検索に用いる補助的なファイル

である。（バイト数は１７６１バイト）

c　特定領域検索用サブファイル

このファイルは，検索能率を向上させるために特定の領域の文献のみを，(2) aの検索用総合ファイルから抽出して構成したファイルで，磁気テープとディスクパックとの両方がある。

(3)　日本特許統計用ファイル

諸統計，リスト作成のために便利な各種サブファイルを作成した。その主なものは，次に示す通りである。

a　キーワード処理用サブファイル

このファイルは，(1) bのキーワード主ファイルからキーワードに関する諸統計をとるために作成したファイルで，文献番号順に配列されている。

b　キーワードリスト作成用サブファイル

このファイルは，上記のキーワード処理用サブファイルを，キーワードの文字の５０音順に配列しなおしたもので，キーワードの頻度の統計やキーワードリストの作成に用いられる。

c　和→英キーワードファイル

このファイルは，和文キーワードと，それに対応する英文キーワードを，和文キーワードの文字の５０音順に配列したファイルで，和→英キーワードリスト作成用に使用される。

d　英→和キーワードファイル

このファイルは，上記，和→英キーワードリストの逆引きで，英文キーワードの文字のアルファベット順に配列したもので，英→和キーワードリスト作成用に用いられる。

上記の各ファイルの関係を図Ⅱ－３－１４で示す。

図中

１－ａ：書誌的事項主ファイル

１－ｂ：キーワード主ファイル

１－ｃ：日本特許分類キーワード主ファイル

１－ｄ：英訳キーワード主ファイル

２－ａ：検索用総合ファイル

2－b：検索用サブファイル

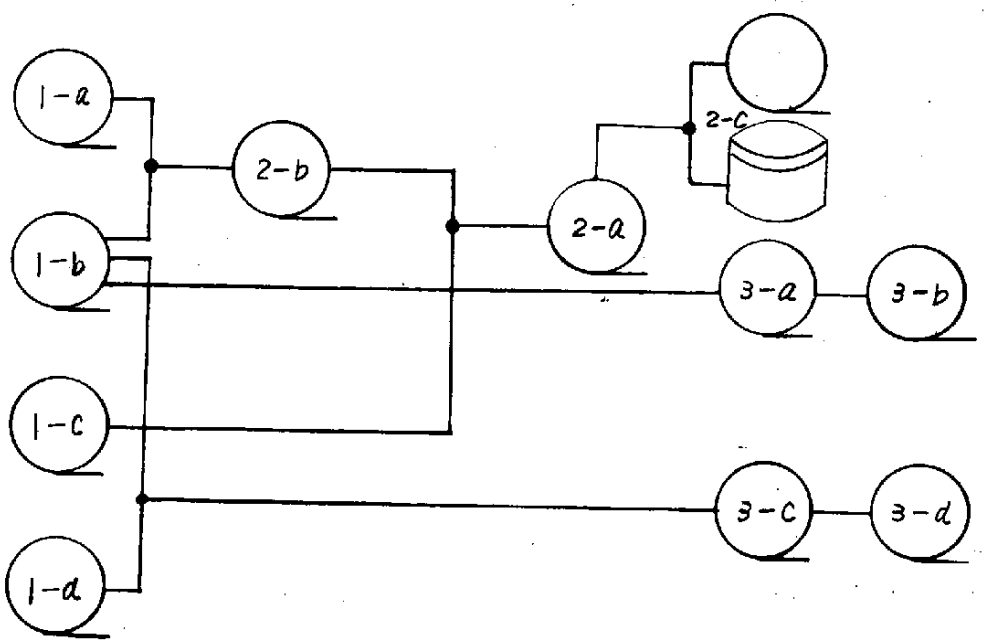
2－c：特定領域検索用サブファイル

3－a：キーワード処理用サブファイル

3－b：キーワードリスト作成用サブファイル

3－c：和→英キーワードファイル

3－d：英→和キーワードファイル



図Ⅱ－3－14　日本特許用ファイル系統図

(4)　米国特許主ファイル

a　書誌的事項主ファイル

このファイルは，日本特許の主ファイルと同様に，米国特許の書誌的事項を蓄積したファイルである。主なる相違点は，主分類コードのエリヤに米国特許分類を，副分類コードのエリヤに日本特許分類を記録した。バイト数は

５５８バイトである。

b　キーワード主ファイル

このファイルは，日本特許のキーワード主ファイルと同様に米国特許のキーワードを蓄積したファイルで，バイト数は１２３２バイトである。

(5)　米国特許検索用総合ファイル

このファイルは，上記(4) a，bの書誌的事項と，キーワードとをマージしたもので，主に検索用に使用するために作成したファイルである。

(6)　米国特許統計用サブファイル

a　キーワード処理用サブファイル

米国特許のキーワードに関する統計をとるために作成したサブファイルで，文献番号順に配列したもの。

b　キーワードリスト作成用サブファイル

米国特許のキーワードに関する統計および，キーワードリストを作成するために用いるファイルで，キーワードの文字のアルファベット順に配列したファイルである。

米国特許に関する各ファイルの関係を分け易く示すと次図Ⅱ－３－１５のようになる。

なお，図中

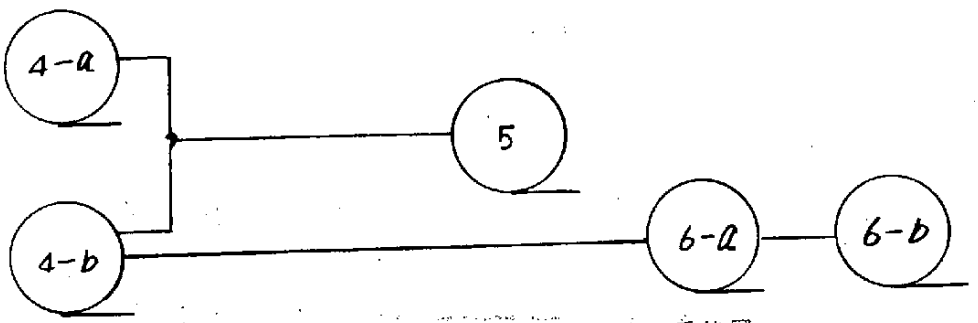
４ａ：書誌的事項主ファイル

４ｂ：キーワード主ファイル

５　：検索用総合ファイル

６ａ：キーワード処理用サブファイル

６ｂ：キーワードリスト作成用サブファイル



図Ⅱ－３－１５　米国特許用ファイル系統図

表Ⅱ－３－６　(1) 主分類分布

分類	A	B	C	D	E	F	G	H	J	K	L	M	合計
55	30	1	25	3	—	—	—	—	—	—	—	—	59
56	7	72	48	11	23	—	—	—	—	—	—	—	161
57	16	15	51	4	—	—	—	—	—	—	—	—	86
58	1	5	1	26	2	5	7	21	1	0	—	—	69
59	80	4	33	18	36	7	22	—	—	—	—	—	200
60	13	35	24	33	19	—	—	—	—	—	—	—	124
61	0	0	3	2	1	0	—	—	—	—	—	—	6
62	13	12	10	0	—	—	—	—	—	—	—	—	35
93	—	—	0	27	4	—	—	—	—	—	—	—	31
96(1)	5	7	15	0	0	—	—	—	—	—	—	—	27
96(2)	3	15	4	3	1	—	—	—	—	—	—	—	26
96(3)	8	42	—	—	—	—	—	—	—	—	—	—	50
96(4)	0	0	24	35	6	—	—	—	—	—	—	—	65
96(7)	1	3	18	0	10	3	7	0	—	—	—	—	42
96(8)	0	6	2	0	—	—	—	—	—	—	—	—	8
97(3)	9	0	4	—	—	—	—	—	—	—	—	—	13
97(5)	6	3	5	9	8	7	1	0	0	10	4	—	53
98(3)	9	9	12	8	—	—	—	—	—	—	—	—	38
98(5)	23	20	18	6	7	6	8	0	10	—	—	—	98
99	23	10	3	7	—	32	4	—	—	—	—	—	79
99(5)	11	24	50	5	11	18	8	6	45	24	—	—	202
100	2	0	19	36	—	—	—	—	—	—	—	—	57
102	—	2	—	—	110	—	—	—	—	—	—	—	112
110	3	10	4	2	0	0	14	11	0	3	1	12	60
114	162	—	—	—	—	—	—	—	—	—	—	—	162
													1,863
その他													137
合計													2,000

3.2.6　データ作成とその問題点

(1)　データ作成

2,000件の日本特許の解析，コーデング，カードパンチに要した人員，時間，ミスの発生率などを次に述べる。

なお書誌的事項は，原文から直接コーデングシート上に内容を転記し，カナ化およびコード化をも同時に行なった。キーワードは，抄録カード上で解析を行ない，その内容をコーデングシート上に転記した上で，カナ化した。

原データの蓄積件数2,000件における分類別件数は表Ⅱ－3－6(1)，(2)に示すとおりであり，これを技術分野ごとに大別すると次表Ⅱ－3－5のようになる。

表Ⅱ－3－5　技術分野の分類件数

技術分野	分類	内容	件数	%
強電	55～62	発変電	771	38.55%
	93(C)(D)(E)	電燈		
弱電	96(1)～100	電気通信	870	43.5%
	102(B)(E)	音響装置		
電子計算機関係	110	磁気量の測定	222	11.1%
	114A	電子計算機		
その他		金属化学など	137	6.85%
計	－	－	2,000	100%

表Ⅱ－３－６　(2) 主分類分布

分類	A	B	C	D	E	F	G	H	J	K	L	その他	合計
10	3			1				7	3		3	2 (R)	19
12	8	2	2										12
13	1												1
15						3						1 (P)	4
20	3	1	2										6
21	1	4											5
24									1				1
25								2				2 (N)	4
26	1	1		2									4
39				2									2
53	1				2								3
54	1	10											11
72	1												1
74	1		1										2
76		2											2
78					1								1
79	2												2
85						1							1
86						1							1
92	1												1
93						3							3
94	1												1
95					1								1
101					9	2							11
103			2	1						6			9
104	3												3
105	1												1
106			1										1
107				2									2
108					1								1
109		3											3
111	2		4		1				1				8
113	1												1
115				2		4							6
116				1									1
136		1		1									2

＊解　　析

電気部門の日本特許２０００件を解析するために次表Ⅱ－３－７に示すようなマンパワーが必要であった。

表Ⅱ－３－７　解析者の内容

専　門	キーワード抽出の経験		合　計
	あ　り	な　し	
電　気	5	5	１０
物　理	0	4	4
化　学	1	1	2
合　計	6	１０	１６

なお，解析作業は，技術内容毎に細分化して専門家に割当てるのではなく，番号順に一率配布した。

結果から判断すると，抄録からキーワードを抽出する場合は，原文献の内容がかなり濃縮されているので，技術の専門の人が行なったものと，専門外の人が行なったものとの優劣の差はほとんどない。むしろ，その人の個人差，特に，自然語のキーワードが検索にどのような働きをするかというシステムの内容を認識しているか否かの方が，より多く影響しているように思われた。

解析に要した時間数は，他の仕事の合間に行ったので，正確には握めなかったが，内容の説明から全件数完了迄４０日間要した。この期間に作成した件数と人員の関係は次表Ⅱ－３－８に示す通りである。

表Ⅱ－３－８　解析件数表

所定期間における作成件数	該当人員
２０	1
６０ － ９０	6
１１０ － １７０	7
２７０ － ３００	2

実際に作業した経験から，１人半日に２０～３０件消化するのは容易であるが，

それ以上作業を継続すると内容が乱雑になるおそれがあるので，充分余裕のある解析者を確保しておく必要がある。

また上記の数字は，抄録上のキーワードにアンダーラインをし，さらにその内容をコーデングシート上に転記する作業迄含めている。

単にアンダーラインを引くだけならば，作業能率は倍近く上るが，それは次に述べる理由によって好ましくない。

当初の予定では，抄録上にアンダーラインを引く作業と，そのキーワードを漢字でコーデングシート上に転記する作業とは分けて行なう予定であった。事実，漢字の転記には意外と時間がかゝり，技術者にそのような単純作業を行なわせるのは不経済と思われるが，その反面，その作業を技術内容の分らない別の人に行なわせた場合は，内容を理解できないために，当然防げる記載上の不統一などに基くミスが発生し，カナ化段階を機械的に行なうことが困難となる。

したがって，この実験では，技術者によってキーワードの解析，抽出，コーデングシート上への転記を同時に行ない，その際，さらに記載方法の統一を行ない以後のカナ化工程をなるべく機械的に行なえるようにした。

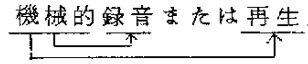
漢字キーワードの転記の際，記載方法の統一を行なった点は次の通りである。

① 原文のミスプリントの修正（抽出キーワードについてのみ）

② 外来語，化合物名などの表記法の統一． 例，トランジスター→トランジスタ

③ 二重修飾語の整理

例 機械的録音または再生



→ 機械的録音
　 機械的再生

④ 不要語の消去

例 光による通信→光通信

⑤ リンクの使用

電子ビームの利用率が… →

電子ビーム＊利用率

⑥ 重複語の除去，脱落の防止

アンダーラインだけ引いてあると重複しているか否か分りにくいが，書き出

すことによって，重複と脱落が容易にわかる。特に請求範囲は裏面に記載されているので，その必要性が大きい。

書き出し段階で以上の点を統一しておけば，カナ化は，内容の全く分らない人に行わせることも可能である。実際に一部について，転記を技術系以外の人に行なわせて見た結果，①，③，⑥に対する誤りが多く，その他，内容を理解していないためか，書き出しそのもののミス（誤字，脱字等）もかなり多かった。

したがって，内容の分る人によって書き出しを行なうことは，このシステムの実行上，非常に重要なことである。

・解析結果のチエック

解析した結果を技術者によってチエックした。その結果，抽出すべきキーワードの脱落，不用キーワードの採用を発見したので，それを補正した。２０００件の特許から抽出されたキーワード数およびチエックにより追加，削除した件数は次表Ⅱ－３－９のとおりである。

表Ⅱ－３－９　抽出キーワード数

a	当初抽出キーワード数	２９７７９　件
b	追加キーワード数	３６０　〃
c	削除キーワード数	２２　〃
	差引最終キーワード数	３０１００　〃
	抽出ミス（ｂ＋ｃ）／ａ	１.３　％

なお，解析によるミスは以後のコーデングの工程でもある程度発見された。

・書誌的事項のコーデング

書誌的事項は，作業内容から見れば，次の３種類に大別される。

① 単純に書き写すもの

公告番号，公告日

出願番号，出願日

出願人の数

優先権主張日，番号，数

主副分類，副分類の数

② コード化するもの

出願の種類

出願人コード，国，種類

優先権主張国

カードコントロールコード

③ カナ化するもの

出願人名（株式会社，Co，Corp などの記載方法は統一した）。

発明の名称

上記の作業を短大卒以上の女子7名，男子2名で行った。この作業は，他の仕事の合間に行い，集中的に行なわなかったので，正確な作業時間は算出されなかったが，日本特許2000件を行なうのに，約1ケ月要した。この間の最高は1人741件で，実働20日として1日平均40件となる。

しかし，実際にこの作業を専門に行った場合は，1人1日平均50～60件処理できるものと思われる。

・チエック

短大卒以上の女子3人が，コーデング内容についてチエックを行なった。この作業は，上記と同様，他の仕事の合間に行なったので，正確な作業時間は算出さ

表Ⅱ－3－10　書誌的事項コーデングミス

欄	ミスの内容	件数	割合
マスターカード	T項目書き忘れ	2	3.9
	優先権N書き忘れ	1	2.0
	副分類数書き忘れ	27	52.9
優先権	二重書き	3	5.9
出願人	コードミス	1	2.0
	社名カナ化ミス	1	2.0
	社名規則違反	4	7.8
名称	フリガナ間違い	5	9.8
	表記法間違い	4	7.8
	未完成	1	2.0
その他	コントロールコード	2	3.9
合計		51	100

れなかったが，全数チエックを行なうのに１０日間を要した。実際に作業した経験から，１人，１日平均１００～１５０件処理できるものと思われる。

なお，上記書誌的事項のコーデング中，未経験者の行なった５８７件についてチエックを行なった結果を示すと次表Ⅱ－３－１０の通りである。

全体に対するミスの発生率は５１÷５８７＝８７％で，このミスの内容を分析して見ると，

① T（特許の種類），N（優先権の数），副（副分類の数）など，特殊欄の書き忘れが目立った。この大部分は，記入方法の指示が徹底せず，後で入れる積りで空白としていて，そのまま忘れたものによるものと思われる。

② 出願人の社名についての規則に従っていないものが多かった。

③ カナ化に関しては，フリガナの間違いが最も多かった。これは作業に従事した人の大部分が技術系以外の人で，しかも特許にはほとんどなじみのない人であったため，特許用語に不慣れなためによるものと思われた。

しかし，後述するキーワードのカナ化，特に分ち書きによるミスがほとんどなかったのは，発明の名称そのものは，検索には使用しないので，キーワードの項の規定を厳密には適用しなかった（もっとも，キーワードの分ち書きの規定は，単語を対象としているので，通常の文章に適用することはできない）ので，キーワードのカナ化のミス発生率に較べて遙かに少いのはそのためである。

・キーワードのコーデング

キーワードのコーデング，すなわちカナ化には，１３人が従事した。その大部分は女性で，しかも，この規定を作成して始めての作業であるので，すべて未経験者といえる。作業は，他の仕事の合間に行い，完成迄約１ケ月を要した。この作業を専従者によって行なえば，１人１日８０件前後は可能と思われるが，長時間行うと急激にミスの発生が多くなる。

キーワードは直接検索に使用するので，そのチエックは特に入念に行ない，最も経験のある３人の女性によって，全数２回重ねて行なった。この作業は比較的集中的に行なったが，１回のチエックに約１週間，計２週間を要した。専従する時は１人１日１００～１５０件可能と思われる。

このチエックによって発生したミスの種類と割合は表Ⅱ－３－１１の通りである。

表Ⅱ－3－11　キーワードのカナ化によるミスの種類と割合

ミス		1回		2回		合計	
種類	内容	件数	割合%	件数	割合%	件数	割合%
抽出ミス	重複	1	0.6	0	－	1	0.5
	転記ミス	0	－	1	2.0	1	0.5
	原文ミスプリント	0	－	1	2.0	1	0.5
	削除	0	－	1	2.0	1	0.5
	不適当な切り方	5	3.1	0	－	5	2.4
	追加	6	3.6	1	2.0	7	3.3
	化学記号の特例	17	10.4	1	2.0	18	8.5
	1行1語	1	0.6	0	－	1	0.5
カナ化ミス	フリガナ（単純）	17	10.4	1	2.0	18	8.5
	〃（解釈）	23	14.0	11	22.0	34	15.9
	分ち書（単純）	10	6.1	4	8.0	14	6.6
	〃（解釈）	28	17.1	2	4.0	30	14.1
	表記法	11	6.7	0	－	11	5.2
	書き忘れ	0	－	1	2.0	1	0.5
	用・型・式等の特例	10	6.1	10	20.0	20	13.6
	アルファベットの特例	2	1.3	0	－	2	1.0
	化学名の特例	18	11.0	4	8.0	22	10.3
	外来語の特例	15	9.2	4	8.0	19	8.9
	不明瞭	0	－	8	16.0	8	3.8
合計		164	100%	50	100%	214	100%

注1）この統計は，最初に作業を行なった200件（キーワード総数3015語）についてチエックした結果を示すものである。

2）ミスの件数はキーワード1語単位で算定した。

3）抽出ミスとは，本来，抽出者がコーデングシート上に記載する段階におけるミスであって，カナ化段階のミスではないが，カナ化のチエック段階で発見されたので，ここに記載する。なお，カナ化およびそのチエックは，ほとん

ど技術内容の分らない人によって行なっているので，この工程によって発見された抽出段階のミスは，その大部分偶発的に発見されたものである。この工程の後で，カナ化のキーワードを英訳した際，抽出段階のミス，特に抽出すべきキーワードの脱落が，かなり発見された。これについては表Ⅱ－3－12を参照されたい。

4）カナ化における「単純」とは，規約上の明らかな誤りによるもの，「解釈」とは規約上明確でないか，またはその後設定した規約によって誤りとされたものを指す。

5）特例とは，それぞれの項について設けた規約に従わなかったものを指す。

表Ⅱ－3－12　キーワードのカナ化によるミスの発生率

ミスの種類	ミスの内容	件数	発生率%
抽出ミス		35	1.2
カナ化ミス	フリガナ（単純）	18	0.6
	〃　（解釈）	34	1.2
	分ち書（単純）	14	0.5
	〃　（解釈）	30	1.0
	表記法	11	0.4
	書き忘れ	1	0.1
	用・型・式等の特例	20	0.7
	アルファベットの特例	2	0.1
	化学名の特例	22	0.8
	外来語の特例	19	0.7
	不明瞭	8	0.3
	合計	214	7.1
全キーワード数		3015	

注1）この表は，データシート200件の総キーワード数3015語に対するミスの発生率を示す。

2）件数は，前表1回，2回のチエックの合計数による。

3）抽出ミスは合計数のみを示した。

この表から分るように，単純なミスは比較的少く，規約上不明瞭だったり，特例を設けた点にミスが集中されている。したがって，このような規約を作る時はなるべく，単純な法則で総てが処理できるようにすべきである。また1回だけのチエックではまだかなりのミスが残っていることも注目すべき点である。

・カードパンチ

日本特許2,000件について作成して，インプットカードの種類と枚数は次表Ⅱ－3－13の通りである。

表Ⅱ－3－13　インプットカードの種類と枚数

	種　　類	枚　　数	特許1件当りの枚数
1	マスターカード	2,000	1.00
2	副分類カード	1,089	0.54
3	副優先権カード	483	0.02
4	出願人カード	2,116	1.06
5	名称カード	2,000	1.00
6	キーワードカード	30,100	15.05
	小　　計	37,788	18.89
7	分類表カード	2,533	
合　　計		40,321	

上記のカード作成には，1，2，3については，英数字のキーパンチマシン，4，5，6，7についてはカナ，英数字のキーパンチマシンを使用した。

パンチはすべて外注した。外注は数社に分けて行なった。発注から納品迄の期間，ならびに，納品をチエックして発見されたミスの件数を次表Ⅱ－3－14に示す。

表Ⅱ－3－14　パンチ所要日数，パンチミス件数

カードの種類	件数	発注先	期間日	ミスの件数
マスターカード	1,000	A社	12	1
	1,000	B社	14	2
その他書誌カード	5664	B社	13	31
		C社	12	
キーワードカード	2211	C社	4	13
	27568	D社	24	37
合計	37443		79	84

注1）発注後発見されたコーデングミスなどによりカードを修正または追加した分は算入していない。

(2)　データ作成上の問題点

データ作成上特記すべき問題点を次に列記する。

①　データ解析

抄録からキーワードを抽出する場合，技術内容の理解という点はあまり問題にならなかったが，むしろ，システムの内容を理解して，そのシステムにあうキーワードを選定するか否かの方が影響が大きかった。したがって，システムに興味のない人，偏った見方をする人は抽出作業から除外すべきである。

②　カナ化

カナ化の場合の主な特徴は，表記法における長音の取扱い，および分ち書きにおける有意の二漢字単位で切る原則である。

前者の長音については，通常，「ウ」を使用しているが，この実験では「－」を使用した。「－」の使用は，コンピュータによりタイプアウトされた場合，「ウ」よりも読み易い利点があるが，将来，漢字を主体とするシステムが開発された場合，原文中に，ひらがなで記載されている長音と，その取扱いが異なるので，「－」は使用されなくなるかも知れない。

次に，分ち書きの問題であるが，実際に作業して見ると，特殊な例がでてきて，その調整に苦慮した。

その主な点をあげると次の通りである。

外来語と日本語とが結合しているもの

例　マイクロ波帯，エックス線像

英字，数字と外来語または日本語が結合しているもの

例　３６ミリフイルム，ＳＮ比

例外規定として設けた反語の前に１字漢字の付く場合

例　光不透過

例外規定として設けた，用，型などに１字漢字が付く場合

例　金属化紙，鋸歯状波

これらの特殊な例については，その都度規約を決めて処理したが，今後は，一括してその取扱いを定める必要がある。

③　カナ文字のキーパンチ

カナ文字のキーパンチは，英数字のパンチマシンに較べて，ⓐその保有台数が非常に少い。ⓑカナ文字のキーパンチを打つ熟練者が少い。ⓒ作業速度が遅いなど，多くの問題点をかかえている。さらに，コンピュータの使用機種との関連もあるが，データの性質によってどのようなコード体系を使用するかは，重大な問題である。

すなわち，今回のようにキーワード中にカナ文字と英数字が混在しているようなデータを扱う場合には，カナを表現するためにシフトコードを含んだコード系を使用すると，シフトコードの検出判定，ないしはダブルキャラクタ表現にするためのテーブルコンバートを行なわなければならず，繁雑性をまぬがれない。さらに，データエリアを可変にしても固定長にしても，シフトコードのために本来のデータエリアよりも長くなり，情報量の冗長性をまぬがれない。

また，データ作成の観点から見ると，通常の文字タイピングに加えて常にシフトという上下２段の２つの動作が加重され，そのファクターの増加はひいてはミスパンチの増加を招くことになろう。

④　書誌的事項のインプット

書誌的事項は，インプットのための文献解析を必要とせず，キーワードに較べてその作業は容易なものと思われるが，実際に作業して見ると，キーワードのように技術専門者を必要としないとはいえ，データシートの作成，カードパンチに大変な労力を要する。特に分類，各種コードなどは，１字間違えただけ

でそのデータの価値がなくなるので，考え方によっては，キーワードよりもさらに慎重にしなければならない。万一間違えた場合でも，数字の羅列が多いのでキーワードよりもミスの発見は困難である。

特許のように，対象文献が明確に定まっているものについては，どこかでまとめて，正確なオリジナルファイルを作成しておくことは，国家的に見ても非常に有益なことと思う。

3.3 検索システムとプログラム

この実験では，汎用特許情報処理サービスシステムを開発する前提にたって，実験的に自然語を用いた情報検索システムの研究を行なった。

特許文献は，技術文献としての性格の外，権利文献という特殊な性格を兼ね備えているので，両目的に使用可能な検索システムでなければならない。しかしながら，各方面で行なわれている検索システムの研究においても，なかなか実用に耐え得る好結果を得るのは困難とされ，正確さ，および再現率を向上させるためには結局，原データの精密な解析とデータの緻密さが要求され，膨大なデータ数を扱うのには適さない場合が多い。

さらに，システムを使用する上で，簡便さを持たせるために常用言語による検索システムを作ることは，日本語という特殊性からみてもなかなか大変なことである。

まず第一に，ヨーロッパ語族に比較して文字数が非常に多いということ。カタカナ，ひらがな，漢字，漢数字，アラビア数字および英字，ローマ数字などヨーロッパ語族に使用されている文字，ギリシア文字に至るまで含んでいるので，現在のコンピュータシステムで扱う場合に，ハード面でかなりの制約を受ける。したがって，この実験では日本語の常用言語を，比較的処理しやすいカナ文字，英数字で表現することにした。

また，データ作成の項でも述べたように，日本語の完全な自動処理がむずかしい現在，データを作成する際に大きな労力と経費および時間がかかったのでは実用にならない。そこで，本システムでは，すでに市販されている抄録からデータを作成することにした。しかし，原資料である明細書から抄録へ情報量を圧縮した際に起るデータの歪みによる変換誤差，さらに，それを解析して最大数25のキーワードに圧縮してソースデータとするとき，意味の歪みをいかに最小に止め

るかという問題，ひいてはこれらの変換誤差に起因する検索の再現率の低下，つまりもれをいかに防止するかという課題を解決せねばならない。さらに，質問を構成する概念または文章から選ばれた少数のキーワード（通常３～７語）によって探索が行なわれるが，その時の質問のキーワードとの内容を代表する蓄積データのキーワードとは独立に設定されているので，再び，ここで歪みが生じてくることになる。それを検索して結果を評定する際に，データ交換時に含まれた歪みをいかに調整するかということも課題となる。

いま，ここに明細書中のテーマを表現している言葉を

w_1 ， w_2 ， w_3 ， …… ，w_i ， …… ， w_m

とすれば，明細書中のテーマＰは次のように表わされる。

$$P = F^P \left(\sum_{i=1}^{m} w_i \right)$$

抄録は明細書を圧縮したものであり，抄録によって表わされているテーマＡは次のようになる。

$$A = F^A \left(\sum_{i=1}^{m} w_i - \sum_{k=0}^{n} w_k \right)$$

$$m > n \geqq 0$$

$$k \geqq 0$$

つまり，明細書Ｐから作られた抄録Ａは，明細書の言葉の集団 Σw_i の函数F^Pから脱落した言葉の集団 Σw_k を差し引いて表わされている言葉の集団の函数F^Aで表わされる。このとき，ＰとＡをいかにして近似ないしは等しくさせるかということは，言いかえれば抄録をいかに原明細書に近づけるかということであり，またＰはＡの函数ｇ(A)として表わされるので，ＰをＡに近づける，ないしは同等とするということは，とりもなおさずｇが常に１であるような要素をみつけることである。

さらに，この抄録から最大２５語までのキーワードを抽出して，抄録の内容を圧縮したデータシート上のデータをＤとすると，明細書から見れば，

$$D = F^D \left(\sum_{i=1}^{m} w_i - \sum_{k=0}^{m} w_k - \sum_{j=0}^{r} w_j \right.$$

抄録から見れば

$$D = f\left(F^{A}\left(\sum_{i=1}^{m} w_i - \sum_{k=0}^{n} w_k\right) - \sum_{j=0}^{r} w_j\right)$$

となる。

つまり，データとしての入力ソースDは，抄録で表現するのに使われていた言葉の集団の関数 $F^{A}\left(\sum_{i=1}^{m} w_i - \sum_{k=0}^{n} w_k\right)$ から，さらに，キーワード化したために $\sum_{j=0}^{r} w_j$ なる脱落した言葉の集団をさし引いて構成された言葉の集団の持つ意味集合体である。

つまり，Dとして有するキーワード数DKWSUは

$$DKWSU = m - n - r \leqq 25$$

である。

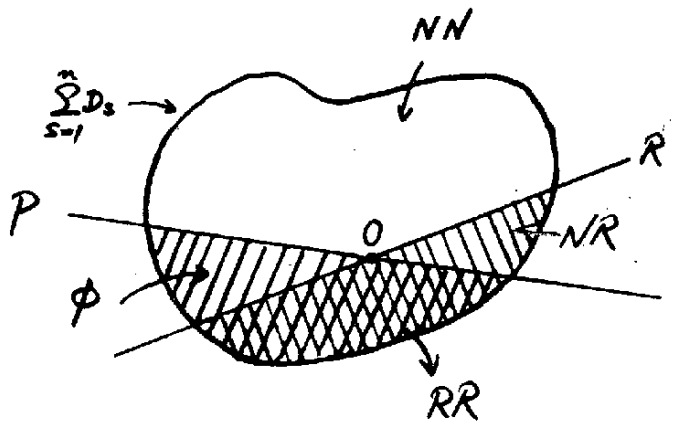
そこで，圧縮されたデータ集合 ΣD_S の中のあるデータ D_S を明細書集合 ΣP_S の中の P_S に対して $D_S = P_S$ であるとして，質問のテーマによって構成される最大限20語，通常3から7語程度の質問キーワード集合 Σ_Q によってテーマに等しいかまたは類似の該当文献は何かを，データ集合 Σ_{Ds} の中から探索する。しかしながら，Ps と Ds では変換過程における情報量の次元の低下に附随する歪みがあるので，それをさらに低次元の質問キーワード集合 ΣQ によって探索するということは，Ps，Ds 間の歪みを増大することになる。特に，特許文献の権利性という特殊な事情に対しては，再現率を常に一定に，それもできる限り1に近く，理想的には常に1に保ちたい，のが実情である。この非線形な歪みをいかにして矯正するかが本実験の大きな課題のひとつである。しかし，再現率を高めるために矯正した結果が適合率を低下させたのでは，あまり意味のないことである。そこで適合率を高く保って，すなわちノイズの量を増加させずに再現率を高くする必要があるのは当然である。

これを図Ⅱ－3－16について説明しよう。

図Ⅱ－3－16において，再現率が高いということは

$$\phi \propto 0$$

ということ，つまりもれた文献が0に近いかないしは0であることである。また適合率が高いということは



$\sum_{s=1}^{n} Ds$ ：データ文献集合

RR：検索された適合文献集合

NR：検索された不適合文献集合

φ ：検索されなかった適合文献集合

図Ⅱ－3－16

$$\frac{NR}{RR} \propto 0$$

ということ，つまりノイズの量が0に近づけば近づくほど適合率はよくなる。

現在，一般に行なわれている意味論的な情報検索における再現率の向上および適合率の増加を同時に実行することは，非常にむつかしいとされている。

すなわち，再現率と適合率の関係を数式で示すと次のようになる。

$$\frac{RR}{RR+NR} = \gamma \cdot \frac{RR}{\phi+RR}$$

一般には，係数γは反比例関係となるような結果となっている。

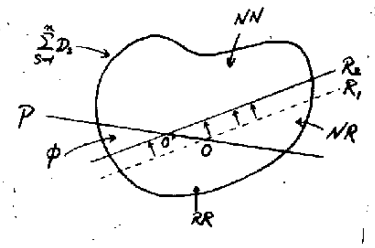
これをさらに図Ⅱ－3－17について述べると領域区分線Rが並行移動ないしは斜行移動するような結果に終ることが多い。

R_1 を改良前の領域区分線，R_2 を改良後の領域区分線とすれば，両システムの検索される適合文献集合の比

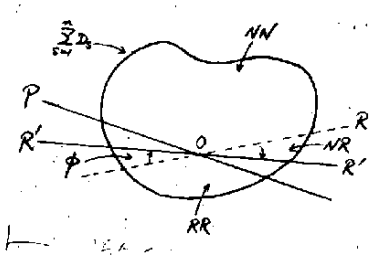
$$\frac{RR_2}{RR_1}$$

は増加し，その結果として，再現率の比

$$\left(\frac{RR}{\phi+RR}\right)_1 \Big/ \left(\frac{RR}{\phi+RR}\right)_2$$



図Ⅱ－３－１７



図Ⅱ－３－１８

は増加し，さらに適合率の比

$$\left(\frac{RR}{RR+NR}\right)_1 \Big/ \left(\frac{RR}{\phi+RR}\right)_2$$

も同様，増加する。しかしながら，適合率１と２との比が増加するということは R_1 から R_2 に移行した結果，ノイズが増加して適合率が低下することを意味する。

そこでこの情報検索システムの実験においては，結果的にみて，なんとか再現率を増加させ，φの量を０ないしは０に近づけるようにし，それを保ったままでＮＲの増加をおさえ，ないしは低下させる方法を考えようとしたものである。システム的には，上図Ⅱ－３－１７におけるＲを並行ないしは斜行移動させず，つまり現行手法におけるＰとＲとの交点を移動してφの領域をカバーするのではなくて，ＰとＲとの交点を保持したまま，原点として回転させたような結果をもたらす手法を見つけることができないものか，ということに重きを置いた。これを図Ⅱ－３－１８によって説明すると，検索領域区分線Ｒを正の時計方向に回転させることによりＲＲの領域を増加させるとともにＮＲの検索された不適合の文献集合領域を減少させることになる。これを実際の文献とデータ文献との歪みを矯正するという観点から見ると，データの集約としての情報量を増加させるという

こと，つまり

$$\sum_{i=1}^{m} w_i \quad - \quad \sum_{k=0}^{n} w_k$$

の量を低下させる，つまり抄録の意味内容を増加させるとか，$\sum_{j=0}^{r} w_j$ の量を低下させる方法，つまりソースデータキーワードの数を増加させるような方法をとることになる。しかし，これは結局，検索データとしてはなんら元の文献Pの意味論的解析能力の増加とはならず，検索の情報量のみを増加させることとなる。そこで，意味論的に歪曲されたままのソースデータDから元の文献Pに対する対応能力を増加させる別の手法を見つける必要がある。

その試みとして，この実験ではキーワードを分ち書きにし，その言葉の単位が構成する意味論的構造を，語順との関連において検索に生かそうとした。日本語という言語の持つ語順と，その意味を構成する関係とを画一的にきめてしまうのは危険ではあるが，試験的に実行した。

検索を行なう際に質問のキーワード集合 ΣQ，つまり質問テーマを表現するキーワード群 q_1，q_2，q_3，……，q_{20} の各語 qt とソースデータ文献Dsのキーワード集合 $\sum_{j=1}^{r} wj$ の各語 wj とを対照するとき，wj と qt との全体を対比するのみならず qt の部分構成要素 $qt = \sum_{x=1}^{y} q'x$ の $q'x$ についても対比を行なった。したがって1キーワードの構成部分の要素についても論理式の要素として独立に設定できるようにした。また，対比の際には対照されるソースデータのキーワード wj と対比する質問のテーマのキーワード qt との対照状態を分類してパラメータ表示するようにした。このパラメータは希望条件として質問の際に質問の条件として設定することができ，それと検索時において内部で検索対比状態を自動的に判定し表示されるパラメータと比較判定させるようになっている。さらに，このパラメータの集合を評価判断の基準値として質問時に外部から与えられる希望値と比較し，採否を決定するようにした。このパラメータの値から関数計算によって固有値を求めて採否を決定することも可能となっている。したがってこのシステムのアルゴリズムとしては論理方式，代数方式，関数方式の三方式の併用となっている。

また，もうひとつは，先に述べたデータの情報の絶対量の増加を計ることであるが，この場合も単にDの情報量を単純にA又はPの情報量に近づけるのではなく，

質的に圧縮された少ない情報で次元を増加させるようにすることが肝要である。

本システムでは，特にその点に留意し，原データの明細書を一定の基準にしたがってある分類形態の中に分布させ，その区分を定めている。その分類を利用して分類表の区分表記に使われている概念を表わすキーワードを分類コードにしたがって選択し，Dのキーワードに追加してデータのキーワードとした。その選択法は次のようにする。

分類表のキーワードの抽出方法を，具体例について説明する。

表Ⅱ－3－15　97(5)C分類表抜粋

分類種目	キーワード
0	「テレビジョン　ソーサ」
	テレビジョン　ドーキ」
1	テレビジョン　ソーサ」
10	……
⋮	
13	セイデン　ヘンコー」
14	デンシ　ビーム」
	デンジ　ヘンコー」
141	ヘンコー　ハケイ　チョーセイ」
1411	チョクセンセイ」チョーセイ」
1412	シンプク　チョーセイ」
1413	ダイケイ　ヒズミ　チョーセイ」
142	

表Ⅱ－3－15は97(5)C　テレビジョンの走査同期に関する分類表の一部をカナ化した状態を示していて，磁気テープには，このような状態で分類表が読込まれている。

いま蓄積データの特許の分類が，97(5)C1412であったとすると，この分類のキーワード，すなわち「シンプク，チョーセイ」のみを拾うのではなくて，その上位概念の分類，即ち，141，14；1；も拾ってくる。なお，0はすべての場合拾うようにした。したがって，実際には，まず，97(5)Cのファイルを

見付けだし，まず0に相当するキーワードを拾った後で，種目の頭から1字づつ読んで，それに相当するキーワードを抽出している。この結果，97(5)C141.2の場合は，次のキーワードが抽出されることとなる。

0	テレビジョン　ソーサ
1	テレビジョン　ドーキ
1	テレビジョン　ソーサ
14	デンジ　ビーム
	デンジ　ヘンコー
141	ヘンコー　ハケイ　チョーセイ
141.2	シンプク　チョーセイ

このうち，同一のキーワードは，照合することにより除去し，さらに，予め抽出された明細書からのキーワードとも照合させ，重複分を除いた状態でデータのキーワードの後に追加する。

なお，副分類があれば，副分類についても同様に，キーワードを抽出するが，蓄積対象部門以外の分類については，予め，分類表のキーワードが設定されていないので，自動的に記入されないこととなる。

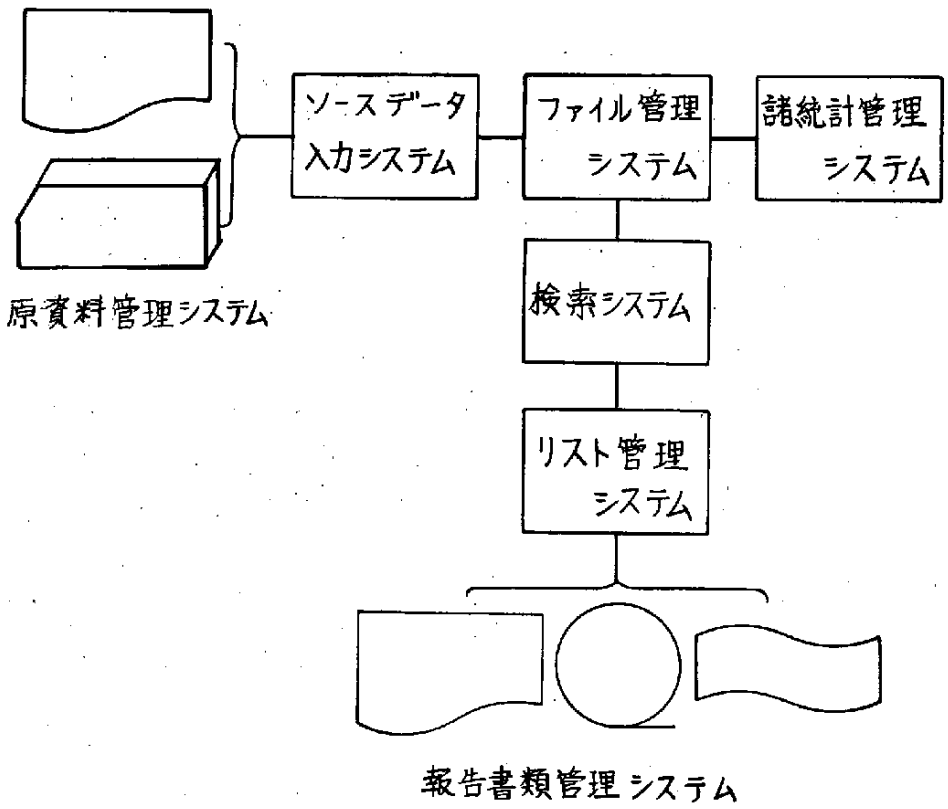
本システムのシステムブロックチャートを図Ⅱ－3－19に示す。
この実験用情報処理システムは，(1)原資料管理システム，(2)ソースデータ入力システム，(3)ファイル管理システム，(4)諸統計管理システム，(5)検索システム，(6)リスト管理システム，(7)報告書類管理システムの7ブロックからなっている。
使用したプログラム言語はＢＯＳコンパイラーによるコボル，フォートランを主としており，ソースデータ入力システム，ファイル管理システムにおいてはコボルを，諸統計管理システム，リスト管理システムではコボル，フォートラン併用，検索システムはフォートランをそれぞれ使用した。

また，質問テーマ数は単独質問，同時多重質問の両方が同一のシステムで行えるようになっており，回答は検索しながら検索された順に，各質問別のマークが付されたリストとしてアウトプットすることもできるし，また検索終了後，各質問別にまとめてリストアップすることもできる。

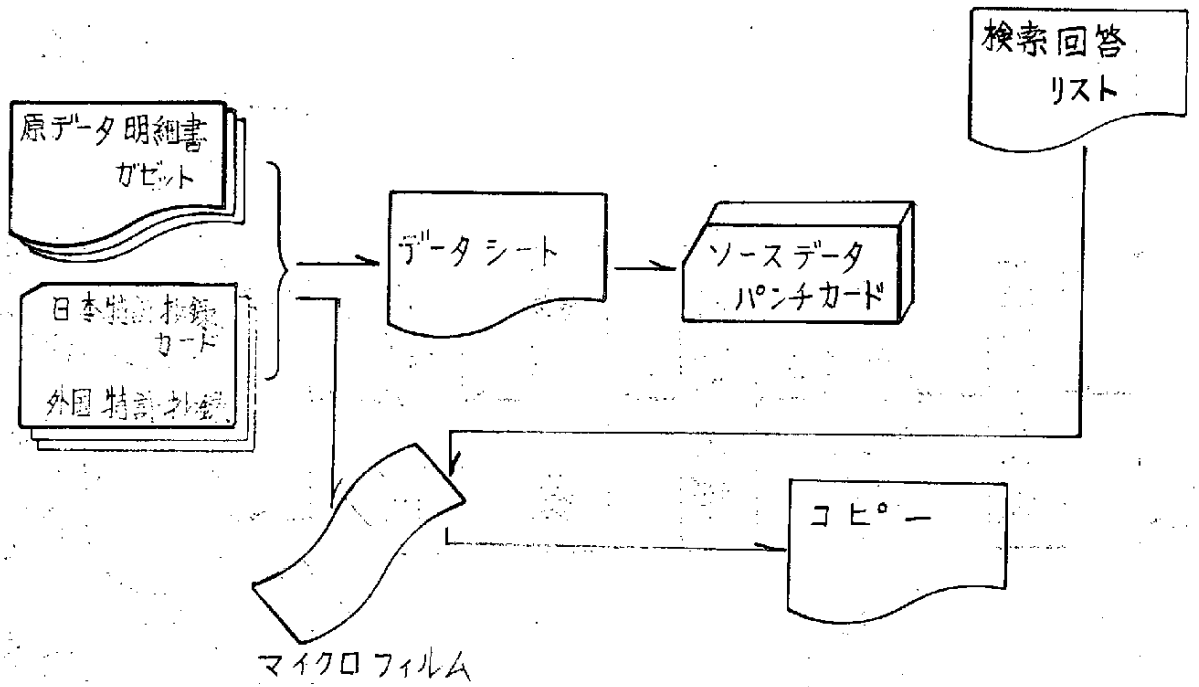
次に，このシステムの各ブロックについてプロセスを述べる。

3.3.1　原資料管理システム.

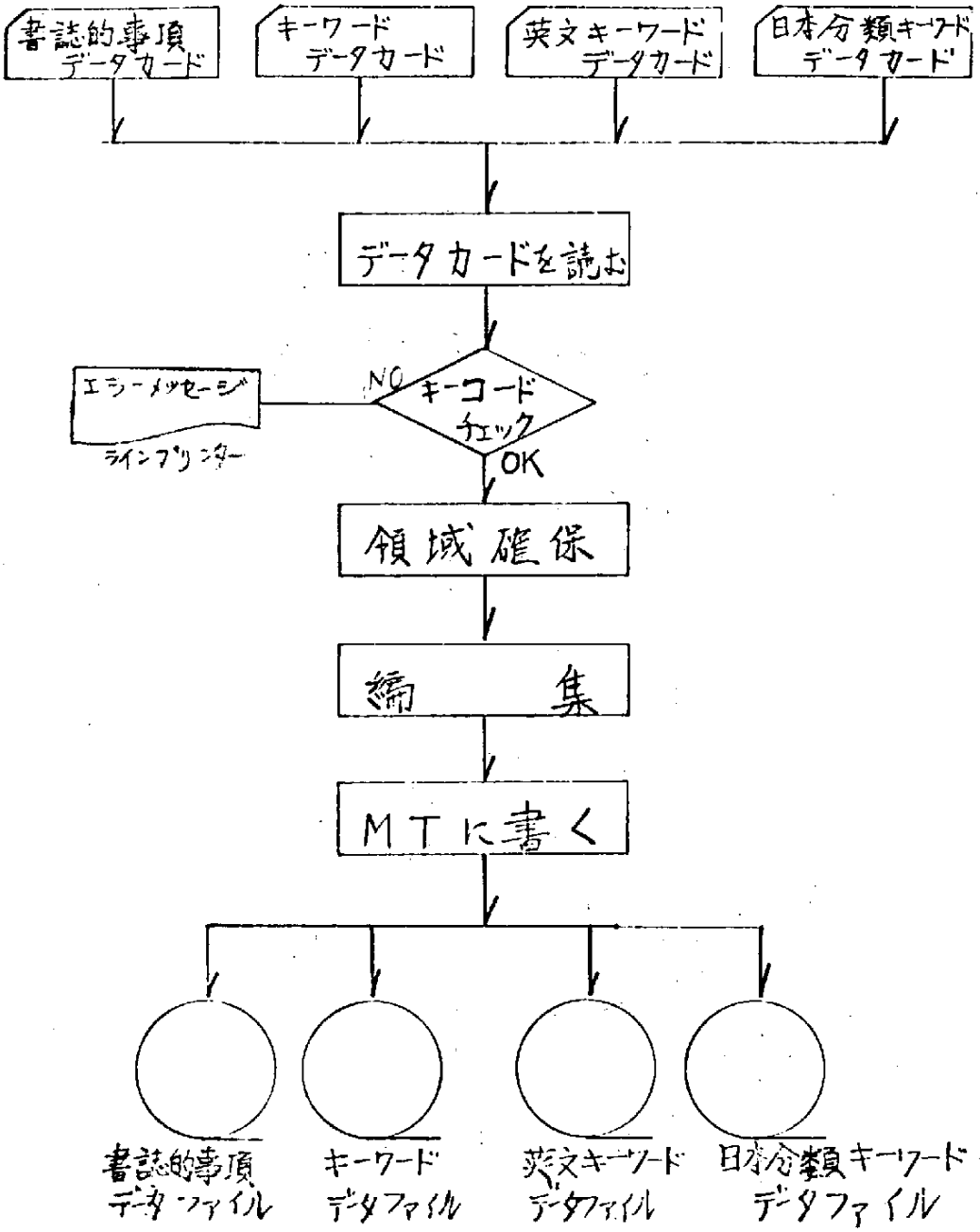
このシステムブロックは，この実験の段階では特許情報処理サービスシステムにおける目標の一つである原データの自動キーワード化が開発されていないので，このセクションは主として，人間が主体となって原データを管理し，データシートを作成し，80桁カードにパンチされる（ソースデータの作成）さらにそのデータカードの管理と同時に，最終システムでは，マイクロフイルムなどに収容された原データそのものを質問要求に応じ機械検索結果に基いて自動的に複製化することが目標とされているが，本実験システムでは，その機能の代行として，マニュアルで半自動的に複製する機能をもつXEROXなどを採用している。



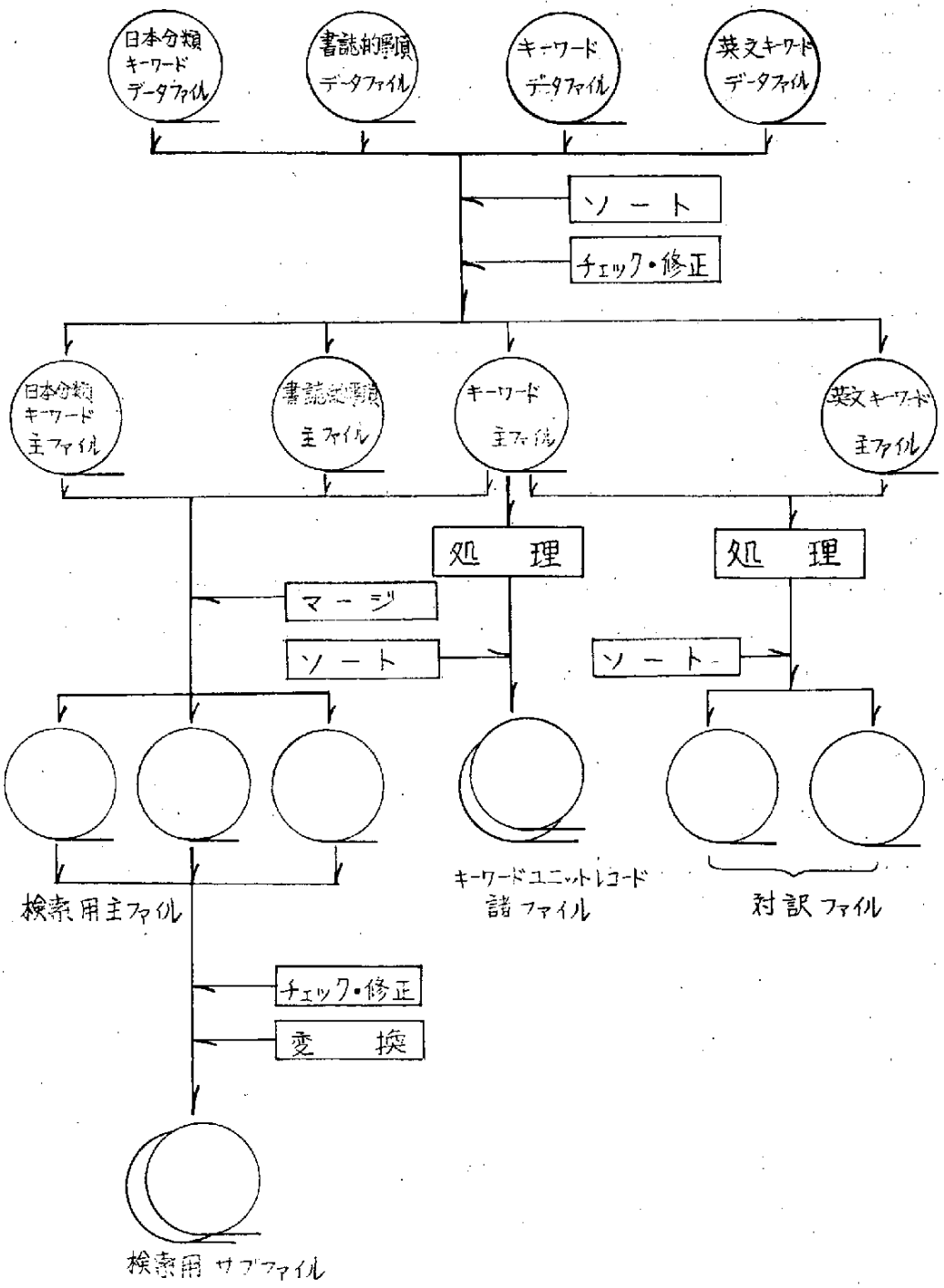
図Ⅱ－3－19　システムブロックチャート



図Ⅱ－３－２０　原資料管理システム



図Ⅱ－3－21　ソースデータ入力システム



図Ⅱ－３－２２　ファイル管理システム

3 3 2　ソースデータ入力システム

原データ処理システムで作成されたソースデータは，このシステムブロックの段階においてコンピュータに入力され，それぞれ所定のフォームにより磁気テープに変換記録される（主ファイルの作成）。このプロセスを示せば図Ⅱ－3－21の通りである。

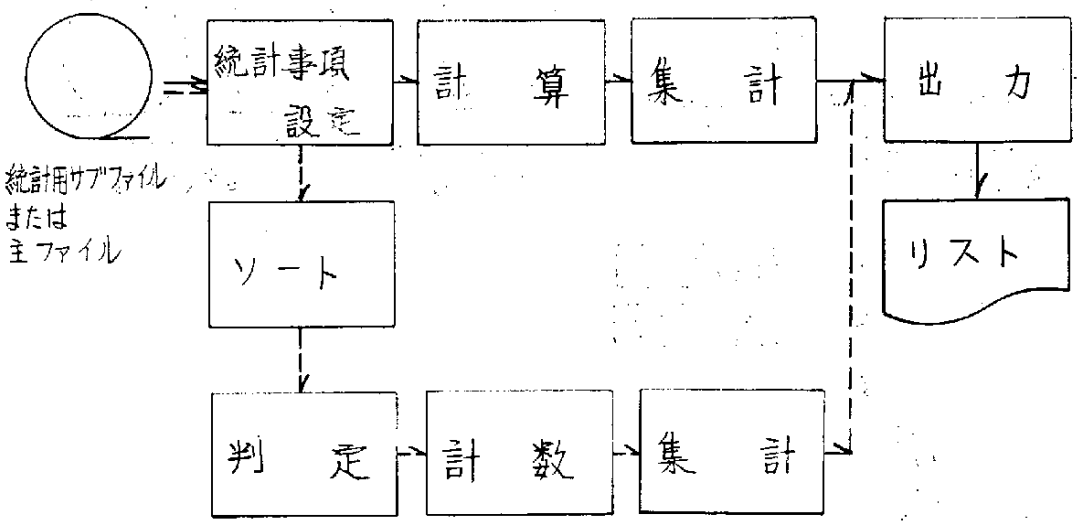
3 3 3　ファイル管理システム

このシステムブロックはソースデータ入力システムで作られた主ファイルをもとにして，それぞれの仕事に適した形態を持つサブファイルに変換したり，入力されたデータの主ファイルに対しデータの正誤チェック，データの修正，改定等を行なう機能を持つ機構である。そのプロセスは図Ⅱ－3－22に示すとおりである。

使用コンパイラシステムプログラムはＢＯＳコボルを主体とした。

3 3 4　諸統計管理システム

このシステムブロックは，蓄積データ自体の諸統計，たとえばキーワード数の分布，キーワードの文字数分布，キーワードの頻度数，特許分類の数，優先権の数，出願人の数，発明の名称の字数などを計算集計し，定められたフォーマット



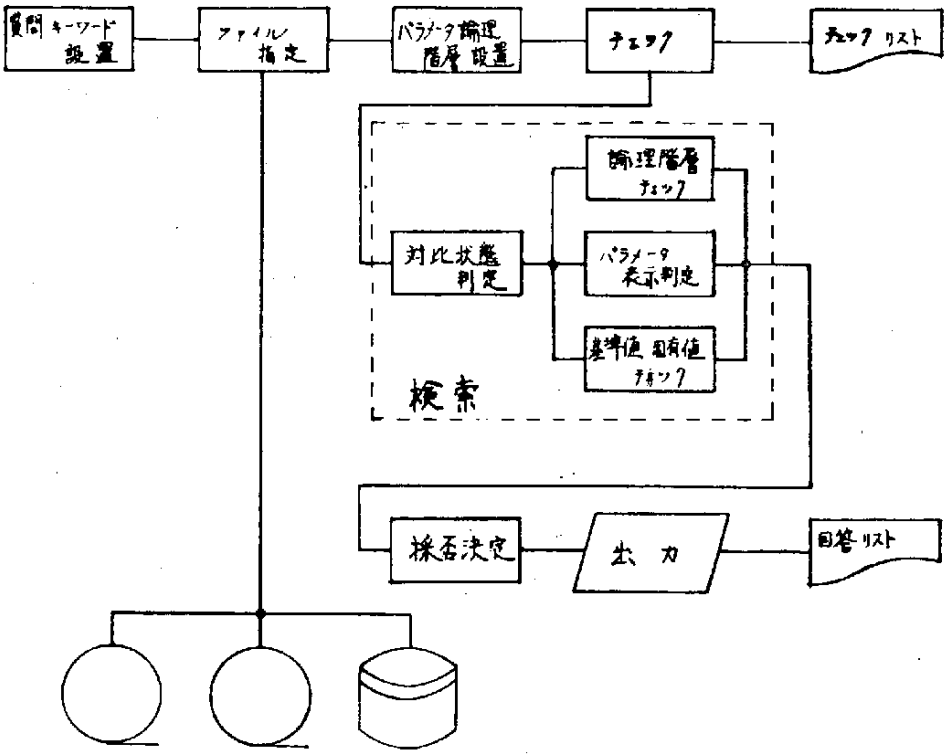
図Ⅱ－3－23　諸統計管理システム

に従ってアウトプットする機能を持つ。フォートランおよびコボルＢＯＳコンパイラーシステムプログラムを用いてソースプログラムを作成した。

３３５　検索システム

このシステムは原資料管理システム，ソースデータ入力システム，ファイル管理システムなどのシステムブロックによって作成されたデータファイルを基にして，質問テーマに従い情報検索を行う機能を果す。

質問は単独質問，多重質問同時処理の二つの方式がひとつのシステムでできるようになっている。



図Ⅱ－３－２４　検索システム

検索手順を図Ⅱ－３－２４によって説明しよう。

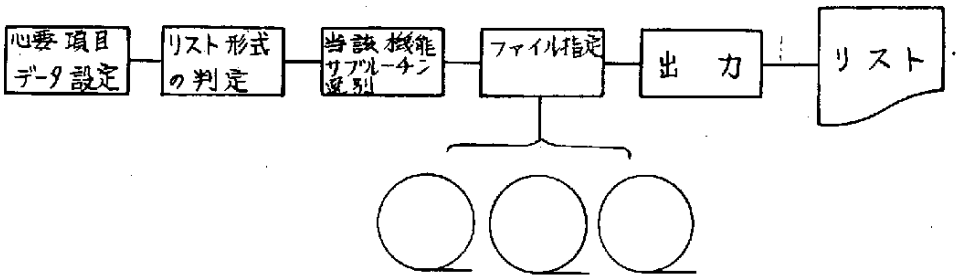
まず，質問テーマをキーワード化する。この際シソーラスが完備されておれば，シソーラスによって質問キーワードをチェックして修正する。さらに質問のテーマの趣旨にそって質問のキーワードの論理条件を設定する。論理条件の設定はその都度質問の条件として読込まなくとも，プログラム中に組込まれた既製の論理条件をコードによって指定すればよいようになっている。特別な論理条件の場合はその都度インプットすることもできる。次に各キーワードのマッチング状態により附与される数値の積算値と比較して回答の採否を決定する基準値を，論理条件，各キーワードの長短およびその重要度と勘案して決める。

以上の質問事項およびその条件をコンピュータにインプットし，正しく解読されたか否かをチェックするために，一旦アウトプットしてその内容を確認した後，検索を実行する。使用したファイルが日本特許分類のキーワードを含む場合で，分類のキーワードによって一致した場合にはマッチング状態によって附与される数値にファクターを附加し，そのレベルを制御することも可能である。データ文献が質問テーマに該当しているか否かの最終的な採否の判定は，論理条件，各キーワードのマッチング状態，および各キーワードのマッチング状態によって附与された数値の積算値と基準値との比較により決定する。

さらに，必要ならばキーワードのマッチング状態によって附与された数値とを関数評価して採否を決定することもできる。とくに，この函数評価は，文献の適合度の判別函数として固有値レベルの状態を用いることにより，文献の適合度を推定することができるので，固有値レベルの状態を函数評価することにより，文献の優先順位を決定することも可能であると思われる。選出された文献について，回答の様式に従って必要事項をプリントアウトする。多重質問の場合は実行中各質問テーマ毎にマークを附し検索順にプリントアウトすることもできるし，また，最後に各質問別にまとめてプリントアウトすることもできる。

３３６　リスト管理システム

このブロックはファイルの管理，データの管理，報告書類の作成，リスト打出しなどの要求をまとめて行なうもので，リスト形式に必要な項目およびデータを命令ソースとしてそれを解説し，該当するサブルーチンを選んで実行するようになっている。このプロセスを図Ⅱ－３－２５で示す。



図Ⅱ－３－２５　リスト管理システム

３３７　報告書類管理システム

このシステムは質問に対する回答の整理全般を扱い，速報印刷物を管理したり質問者に対する回答の受け渡し，配布の機能を有する機構である。

3.4 検索実験

3.4.1 実験と結果

作成された蓄積データに対して質問テーマによって検索実験を行なった。この検索実験では，特に改めて質問テーマを作ることをせず，既に一般の要求によって過去にマニュアルによって行なわれてあった調査テーマを与え，このシステムによる回答が得られるまでマニュアル調査によって行なわれた検索の答はふせておき，第一回テストラン以後に両者の結果を対比させた。実験手法としては数種の質問テーマに対して試行錯誤によるシステムチェックを含めてデータを収集した。

その実行例の代表的手法を次に述べよう。

① 同一質問テーマに対して設定する質問キーワードの数を変える。

② 同一質問テーマに対して設定する質問キーワード集合の意味内容を変えず，質問キーワードの分ち書きのセグメントの長さを変え，短く分解したものと長いままのものとを対比させる。

③ 質問キーワードの長短，数の加減に対応して論理階層を変化させる。

④ 質問テーマのキーワード集合，論理階層を変えずに，条件パラメータを変えて行なう。

⑤ 同一質問テーマを原データのみのファイルと補助データとして情報量を増加させた分類キーワードを含むファイルとの両者について比較する。

⑥ 各質問テーマ間の回答内容の安定性を調べる。

⑦ 単独質問処理，多重質問同時処理の実行。

以上のような観点で行なった実験の結果を，代表的な実例によって説明する。

(1) 例1

磁気録音ヘッドについて，検索実験を行なった。この実験では，キーワード抽出に使用する原資料と検索の精度の関係を追及することを主目的として行なった。

この質問内容と，回答のアウトプットの一部を参考までに次図Ⅱ－3－26に示す。

図Ⅱ－３－２６　質問および回答のアウトプット

上記の図Ⅱ－３－２６に記載されているように，この質問に用いたキーワードは，下記の６語である。

1. ヘッド
2. ジキ
3. ロクオン
4. ジセイ
5. サイセイ
6. キロク

また，各キーワードの論理式は

１×２×（３＋４＋５＋６）

この設定によって，データのキーワードのみを蓄積したファイルの場合と，データのキーワードに日本分類からのキーワードを追加したファイルを使用した場合について比較実験を行なった。

データのキーワードのみのファイルで検索した場合は４２件，分類表のキーワードを追加したファイルでは５６件検索された。

この結果を，さらに，キーワードの抽出に抄録を使用しなかった場合を仮定し，名称と特許請求範囲のみから抽出されたキーワードのみを対象とした場合，検索されたか否かをデータ上で解析した。

この結果を表Ⅱ－３－１６検索結果の内容分析表に示す。

表Ⅱ－３－１６　検索結果の内容分析表

No	公告番号 特公昭４２－	分類 Ⅰ	D－KW Ⅱ－a	D－KW Ⅱ－b	D-KW+CL-KW Ⅲ－a	D-KW+CL-KW Ⅲ－b	評価
1	１６２３６	△		A		A	○
2	１６２３７	△			A	A	○
3	１６２５１			A		A	△
4	１６８０５	△		B		A	○
5	１７３０７	△		A		A	○
6	１７３０８	△	A	A	A	A	○
7	１７３０９	◎			A	A	○

8	1 7 3 1 0	△		A		A	○
9	1 7 3 1 1	△		A		A	○
10	1 7 4 5 1	△			B	A	○
11	1 7 4 7 3	△	A	A	A	A	○
12	1 8 0 3 4	△	A	A	A	A	○
13	1 8 0 3 5	◎	B	B	A	A	○
14	1 8 0 3 6	◎	A	A	A	A	○
15	1 8 1 8 3	△			A	A	○
16	1 8 1 8 7	△			A	A	○
17	1 8 1 8 8	△			A	A	○
18	1 8 1 8 9	△			A	A	○
19	1 8 1 9 0	△			A	A	○
20	1 8 1 9 3	△	A	A	A	A	○
21	1 8 1 9 6	○			A	A	△
22	1 8 1 9 7	◎	A	A	A	A	○
23	1 8 1 9 8	◎	A	A	A	A	○
24	1 8 8 6 6	△	A	A	A	A	○
25	1 8 8 7 0	△	A	A	A	A	○
26	1 8 8 7 3	△			A	A	○
27	1 8 8 7 4	△	A	A	A	A	○
28	1 9 6 0 8			B		B	○
29	2 0 1 0 6	○	A	A	A	A	○
30	2 0 1 0 7	○	A	A	A	A	○
31	2 0 1 1 6	△	A	A	A	A	○
32	2 0 1 1 8	○	A	A	A	A	○
33	2 1 1 5 0	△	B	B	A	A	○
34	2 1 1 5 1	△	B	B	A	A	○
35	2 1 1 5 3	△	B	B	B	B	○
36	2 1 1 5 6	△			A	A	○
37	2 1 1 5 7	◎	A	A	A	A	○

38	2 1 2 6 4			A		A	○
39	2 1 2 7 9			B		B	○
40	2 1 4 6 7		A	A	A	A	○
41	2 1 5 9 4			A		A	○
42	2 1 5 9 6	△	A	A	A	A	○
43	2 2 3 5 0	△			A	A	○
44	2 2 3 5 2	△	A	A	A	A	○
45	2 2 3 5 3	△	A	A	A	A	○
46	2 2 3 5 6	△	B	A	B	A	○
47	2 2 3 5 7	△	B	A	B	A	○
48	2 2 3 5 8	△	B	B	A	A	○
49	2 2 6 6 5	△	A	A	A	A	○
50	2 2 6 6 6	○	A	A	A	A	○
51	2 2 9 8 9	△		A		A	△
52	2 2 9 9 0	△	A	A	A	A	○
53	2 2 9 9 1	△			A	A	△
54	2 2 9 9 2	△	A	A	A	A	○
55	2 2 9 9 3	△			A	A	○
56	2 2 9 9 4	△			A	A	○
57	2 2 9 9 5	◎	B	B	A	A	○

注1） 分類中，◎印は１０２Ｅ５が主分類のもの，○印は１０２Ｅ５が副分類のもの，△印は１０２Ｅ中に分類があるもの，無印は１０２Ｅに分類が付けられていないもの。

2） Ｄ－ＫＷとは，各特許分献から抽出されたキーワードのみのファイルを使用した検索結果を示す。

3） Ｄ－ＫＷ＋ＣＬ－ＫＷとは上記2）のキーワードに，分類からのキーワードを追加したファイルによる検索結果を示す。

4） Ⅱ－ａ，Ⅲ－ａとは，発明の名称と，特許請求範囲からのみキーワードを抽出した場合。

5） Ⅱ－b，Ⅲ－bとは，上記4）のキーワードの外，抄録からもキーワードを抽出した場合。

6） 評価の欄中，Aは比較的内容のよいもの，△は関連の薄いものを示す。

7） キーワードの欄中，Aは，コンピュータで検索結果を評価した結果，上位の基準値以上と判定されたもの，Bは同じく下位の基準値以上として判定されたもの，無印は検索されなかったものを示す。

これを，さらに検索件数でまとめると，次表Ⅱ－3－17の如くなる。

表Ⅱ－3－17　内容別，検索件数表

		Ⅰ	Ⅱ－a	Ⅱ－b	Ⅲ－a	Ⅲ－b
回答数	A（主）	7	23	33	42	54
	B（副）	5	8	9	4	3
	（その他）	39	－	－	－	－
	小計	51	31	42	46	57
検索もれ		6	26	15	11	0
合計		57	57	57	57	57

注1） A，Bとは前記表Ⅱ－3－16における基準値を示す。

2） （主），（副）とはⅠ（分類）における主分類，副分類が102E5につけられているもの，（その他）は，上記以外の102E類中に分類が付けられているものを示す。

この表にも示してあるように，この質問によって検索された内容について検討したところ，完全なるノイズはなく，評価の悪いものでも，明細書中に，少なくとも磁気ヘッドについての用途，利用状態などが記載されていた。

したがって，検索もれという立場から見ると，Ⅰの102E5に主副のかかったものを，分類で人間が検索した場合の検索件数12件がもっとも少く，次に，Ⅱ－aの名称と請求範囲のキーワードのみで検索した31件，Ⅱ－bの名称と請求範囲と抄録のキーワードで検索した42件，Ⅲ－aの名称，請求範囲，分類のキーワードで検索した46件，最も結果のよかったのが，Ⅲ－bの名称，請求範

囲，抄録，分類のキーワードで検索した５７件となっていて，抄録のキーワード，分類のキーワードが検索もれ防止に役立っていることが分る。

さらに，これを細かく解析して見ると，No.７の文献は，１０２Ｅ５１に主分類がかかり，内容も磁気ヘッドの磁心材料に関し，重要な文献と思われるが，データのキーワードのみの場合（Ⅱ－ａ，Ⅱ－ｂ）では，いずれも脱落している。これは，文献に「磁心」，「ヘッド」，「トランスジューサ」などのキーワードがあるが，「磁気」というキーワードが存在しなかったため，脱落したものである。

分類表のキーワードを追加した場合（Ⅲ－ａ，Ⅲ－ｂ）は，いずれも検索され，その評価も非常に高い価を示している。

No.１５の文献の場合は，やはり，Ⅱ－ａ，Ⅱ－ｂでは検索されず，Ⅲ－ａ，Ⅲ－ｂで検索されている。この原因は，原文献中に，「ビデオテープレコーダ」，「ヘッド」などのキーワードがあっても，「磁気」というキーワードがなかったためである。

同様に，No.１６，１７，１８，１９は，Ⅱ－ａ，Ⅱ－ｂでは検索されていない。この原因は，データのキーワードとして，「磁気」が抽出されなかったためであるが，原文献中には，「磁気記憶再生装置」という非常によいキーワードが存在したが，たまたま，解析者が，このキーワードを抽出しなかったためによるものである。

このように，これは解析者によるエラーもある程度救済できた例である。

No.２１の場合は，「磁気」というキーワードが原文献中に存在せず，「リードヘッド」という特殊な用語が使用されていたために，Ⅱ－ａ，Ⅱ－ｂで検索されなかったが，Ⅲ－ａ，Ⅲ－ｂでは，分類からのキーワードとの組合せで検索されている。

(2) 例２

カートリッジ型磁気テープについて，主としてキーワードの組合せと論理式の関係および分類キーワードの追加の影響を見るために検索実験を行なった。

質問は次の三つを作成し，その比較をして見た。

① 1. ジキ テープ
　 2. カートリッジ
　 3. ジキ キロク テープ

4. ジキ テープカートリッジ

5. カートリッジ シキ テープレコーダ

6. ジキ キロク バイタイ

7. ジセイ ハクマク

8. テープ

論理式 1×(2＋3＋4＋5＋6＋7＋8)

② 1. カートリッジ

2. ジキ テープ

3. ジキ キロク テープ

4. ジキ テープカートリッジ

5. カートリッジ シキ テープレコーダ

6. ジキ キロク バイタイ

7. ジセイ ハクマク

8. テープ

論理式 1×(2＋3＋4＋5＋6＋7＋8)

③ 1. カートリッジ

2. マガジン

3. ジキ テープ

4. テープ

5. ジキ キロク テープ

6. ジキ テープ カートリッジ

7. カートリッジ シキ テープレコーダ

論理式 (1＋2)×(3＋4＋5＋6＋7)

この結果を表Ⅱ－3－18質問形式による解答の変化，によって示す。

表Ⅱ－3－18 質問形式による解答の変化

No.	公告番号 特公昭42－	オ1質問 検出	オ1質問 評価	オ2質問 検出	オ2質問 評価	オ3質問 検出	オ3質問 評価	適否
1	16231		B	×		×		×
2	17286	×			B	×		×

3	1 7 3 0 4	×		×			A	○
4	1 7 3 0 5	×		×			A	×
5	1 7 3 0 7		B	×		×		×
6	1 7 3 1 2	×		×			A	×
7	1 8 1 8 0	×			B		B	○
8	1 8 1 8 2	×		×			B	○
9	1 8 1 8 4	×		×			B	○
10	1 8 1 8 5	×		×			B	○
11	1 8 1 8 6	×		×			B	○
12	1 8 1 9 2	×		×			A	○
13	1 8 1 9 5	×		×			B	○
14	1 8 1 9 9	×			A		A	○
15	1 8 6 5 3	×		×			B	×
16	1 8 8 6 7	×			B		A	○
17	1 8 8 6 8	×			B		A	○
18	1 8 8 7 2	×		×			A	○
19	1 8 8 7 4		A		A		A	○
20	2 0 1 1 0		B	×		×		×
21	2 0 1 1 1	×		×			A	○
22	2 0 1 1 3		B	×		×		×
23	2 0 1 1 7		B	×		×		×
24	2 1 1 4 7		B	×		×		×
25	2 1 1 4 8	×		×			B	○
26	2 1 1 5 2	×			A		B	○
27	2 1 1 5 5		B	×			A	○
28	2 1 1 5 6	×			A		A	○
29	2 2 3 5 1		B	×			A	○
30	2 2 3 5 6	×		×			A	×
31	2 2 3 5 7	×		×			A	○
32	2 2 7 4 0	×		×			B	○

33	2 2 9 8 5		B	×		×		×
34	2 2 9 8 6	×		×			A	○
35	2 2 9 8 7		B	×		×		×

注１）　検出の項で，×印は検索されなかったものを示す。

２）　評価の項でA，Bは，機械で評価した点数で，Aは上位基準値以上のもの，Bは下位基準以上のもの，無印は，検索されなかったものを示す。

３）　適否の項で，○印は正解，×印はノイズを示す。

次に，上記の結果を，各質問について，正解とノイズの件数を累計して見ると，次表Ⅱ－３－１９質問形式による解答件数の変化，のようになる。

図Ⅱ－３－１９　質問形式による解答件数の変化

	第１質問	第２質問	第３質問
正　　解	3	7	2 3
ノ　イ　ズ	8	1	4
検出件数	1 1	8	2 7

第３質問における検索件数２３件を全蓄積データにおける正解数とすると，第１質問での正解は３件，ノイズは８件，再現率は，３／２３＝１３０％，適合率は，３／１１＝２７３％，第２質問の再現率は７／２３＝３０４％，適合率7／8＝８７５％，第３質問では，再現率１００％，適合率は８５０％となる。

第１質問の結果が悪かったのは「ジキ　テープ」を必須条件としたためで，これを，第２質問の如く，「カートリッジ」を必須条件にすれば，正解率は高くなる。しかし第２質問でも，まだ，かなり脱落しているので，「カートリッジ」と「マガジン」をＯＲ条件で引いて見ると，ほとんど全件数が抽出された。しかし，第３質問では，分類のキーワードを加味した結果であって，それを使用しない場合は，やはり，かなりの脱落が生じた。

この関係を次表Ⅱ－３－２０第３質問の内容分析表で述べる。

表Ⅱ－３－２０　オ３質問の内容分析表

No.	公告番号	分類	D－KW		D-KW+CL-KW		適否	備考
	特公昭42－	Ⅰ	Ⅱ－a	Ⅱ－b	Ⅲ－a	Ⅲ－b		
3	17304	◎			A	A	○	
4	17305	◎			B	A	×	マガジンではない
6	17312	◎			A	A	×	〃
7	18180	◎		B		B	○	
8	18182	◎	B	B	B	B	○	
9	18184	◎	B	B	B	B	○	
10	18185	◎	B	B	B	B	○	
11	18186	◎	B	B	B	B	○	
12	18192	◎	A	A	A	A	○	
13	18195	△		B		B	○	
14	18199	◎	B	A	A	A	○	
15	18653			B		B	×	写真フィルム用
16	18867	◎	B	B	A	A	○	
17	18868	◎	A	A	A	A	○	
18	18872	◎			B	A	○	
19	18874	◎			A	A	○	
21	20111	◎			A	A	×	
25	21148	◎	B	B	B	B	○	
26	21152	◎		A		A	○	
27	21155	◎			A	A	×	
28	21156	○	B	B	A	A	○	
29	22351	◎			A	A	○	
30	22356	◎			A	A	○	
31	22357	◎			A	A	○	
32	22740		B	B	B	B	×	
34	22986	◎	B	B	A	A	○	

注１）　分類(I)とは分類によって人間が検索した結果を示し，◎印は，１０２Ｅ９１，または１０２Ｅ２１が主分類であったもの，○印は上記類に副分類があったもの，△印は，１０２Ｅ全類中にあったもの，無印は，１０２Ｅには主副共分類が付けられていないものを示す。

２）　Ｄ－ＫＷとは，各特許分献から抽出されたキーワードのみのファイルを使用した検索結果を示す。

３）　Ｄ－ＫＷ＋ＣＬ－ＫＷとは上記２）のキーワードに，分類からのキーワードを追加したファイルによる検索結果を示す。

４）　Ⅱ－ａ，Ⅲ－ａとは，発明の名称と特許請求範囲からのみキーワードを抽出した場合。

５）　Ⅱ－ｂ，Ⅲ－ｂとは，上記４）のキーワードの外，抄録からも抽出した場合。

６）　キーワードの欄中のＡは，コンピュータで検索結果を評価した結果，上位の基準値以上で検出されたもの，Ｂは，同じく下位の基準以上として検出されたもの，無印は検索されなかったものを示す。

７）　適否の欄中，○は適合，×はノイズを示す。

この表を要約すると，次表Ⅱ－３－２１内容別検索件数表の如くなる。

表Ⅱ－３－２１　内容別検索件数表

		Ⅰ	Ⅱ－ａ	Ⅱ－ｂ	Ⅲ－ａ	Ⅲ－ｂ
回答数	Ａ（主）	１８	２	４	１１	１３
	Ｂ（副）	１	９	１０	６	７
	（その他）	１	－	－	－	－
	小計	２０	１１	１４	１７	２０
	ノイズ	４	１	２	５	６
検索もれ		０	９	６	３	０
合計		２４	２１	２２	２５	２６

注１）　Ａ，Ｂとは前記表Ⅱ－３－２０における基準値を示す。

２）　（主），（副）とは，Ⅰ（分類）における主分類，副分類が１０２

E21または102E91につけられているもの，（その他）は上記以外の102E類中に分類がつけられているもの，検索もれは102Eには分類がつけられていないものを指す。

すなわち，人間が分類によって調査した場合(I)は，102E全類の主副について調査すれば，20件全部を抽出することが可能であるが，このためには102E全類に主副のかかっているものが2000件中121件あるのでこの全数を調査しなければならないことになる。したがって，この件数から正解20件を差引いた残り101件はノイズということもできる。また，これを102E21また102E91の主副に限定して調査した場合でも，その対象件数は64件となり，そのうち正解は19件，検索もれ1件，ノイズ45件となる。これを機械で行なったときは次のようになる。

分類表のキーワードを使用せず，しかも，抄録からのキーワードも使用しなかった場合

（Ⅱ－a）では

検索件数	12件
ノイズ	1件
検索もれ	9件

分類表のキーワードを使用せず，抄録からのキーワードを使用した場合

（Ⅱ－b）では

検索件数	16件
ノイズ	2件
検索もれ	6件

分類表のキーワードは使用するが，抄録からのキーワードは使用しなかった場合

（Ⅲ－a）では

検索件数	22件
ノイズ	5件
検索もれ	3件

分類表，抄録共，そのキーワードを使用した場合

（Ⅲ－b）では

検索件数	26件
ノイズ	6件
検索もれ	0件

となり，分類表のキーワード，抄録のキーワードが共に，検索もれ防止上有効であるが，ノイズは若干増加する。この実験では，分類表のキーワードと，データのキーワードを同等にあつかったからであり，両者にウエイトを付ければ，分類表のキーワードによるノイズ増加をある程度，抑えることができる。

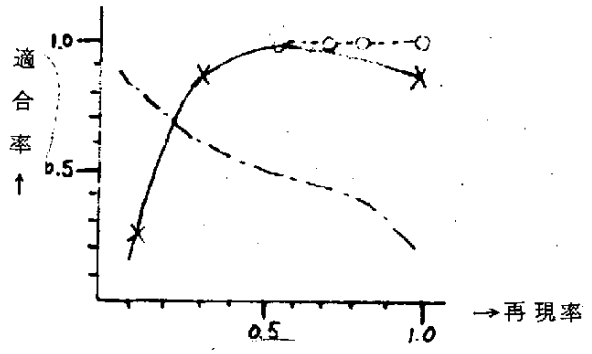
上記，例1，例2の実験結果に基いて，適合率と再現率の相関関係を計算して見ると表Ⅱ－3－22のごとくなる。なお，これを，コーネル大学で，一般文献について行った結果と比較して見ると図Ⅱ－3－27のごとくなる。

表Ⅱ－3－22　適合率と再現率の相関関係

質問	例1		例2	
	適合率	再現率	適合率	再現率
1	100	554	273	130
2	100	737	875	304
3	100	807	850	100
4	100	100	－	－

注1）　例1の質問テーマは，磁気記録ヘッド，質問形式は変更せず，蓄積データのファイルを変更した。

2）　例2の質問テーマは，カートリッジ型磁気テープで，蓄積データファイルは同一のものを使用し，質問形式を変化させた。



図Ⅱ－3－27　適合率と再現率の相関関係

図において，適合率を縦軸，再現率を横軸として表わし，点線は例１，実線は例２，一点鎖線はコーネル大学の実験例を示したものである。この結果からのみ判断すると，非常によい結果が示されているが，実験回数が少ないため，これで結論を出すことは危険であるが，さらに実験を重ねて上図のような結果が得られるならば当システムのアルゴリズムによる検索にも意義があると考えられる。

(3) 例３

多重質問について行なった実験結果を次に述べる

この検索用プログラムでは，さきにも述べたように，現在；１０問迄同時処理を行なうことができる。

今回は，５質問を同時に行なった。その質問内容のアウトプットの一部を次図Ⅱ－３－２８に示す。

回答は，検索された順序でアウトプットすることも，また，質問ごとにソートしてアウトプットすることもできる。

３.４.２　名称と分類による検索実験

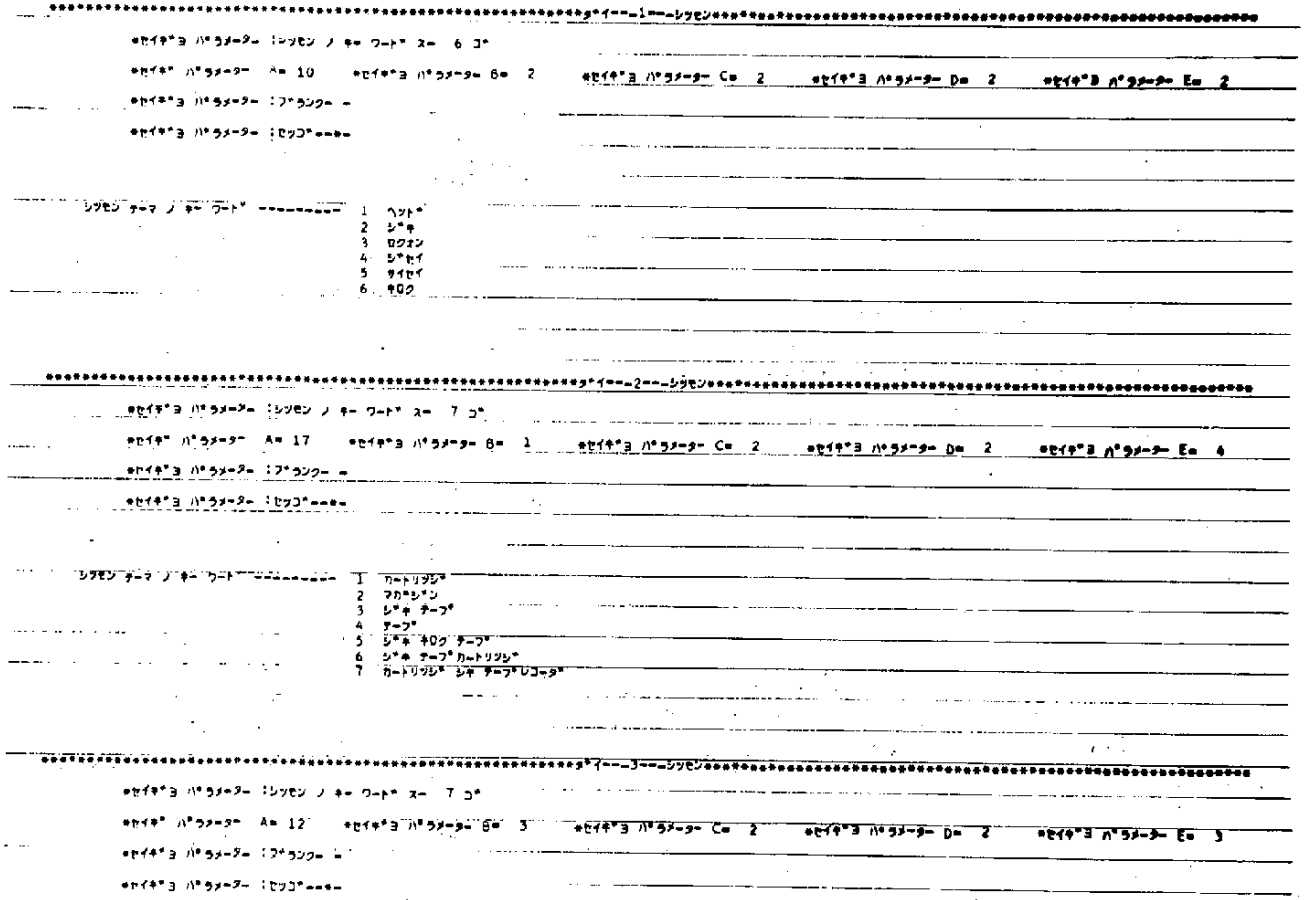
この実験用検索システムでは，発明の名称は勿論，特許請求範囲，抄録，分類表のキーワードを使用したが，発明の名称からのキーワードと，特許分類だけを用いた場合，どの程度の結果が得られるかを比較検討するために，特許庁出願マスターテープを利用して，小規模ではあるが検索実験を行なった。本実験では，明細書全体を分析してキーワードを取出したのではなく，発明の名称中のキーワードと，特許分類の併用による検索がどの程度，効果があるかということを追求した。その結果，以下のデータを得ることができたので報告する。

(1) 実験の目的

出願マスターテープには，種々の事項がインプットされているが，その中で発明の内容をプロフィルとしてつかめる事項には，発明の名称と特許分類がある。この実験では先にあげた２項目を主検索対象とするほか，他の事項をも検索対象として機械検索を行ない，特許出願の際の先行技術調査などに利用可能か否かを調査した。すなわち，当実験では発明の名称中のキーワードと分類により，特許明細書中に示される発明にどの程度まで，接近できるかを追求した。

(2) 検索指定事項

出願マスターテープにインプットされている事項のうち，次の９項目を指定



図Ⅱ－３－２８　多重質問の内容アウトプット

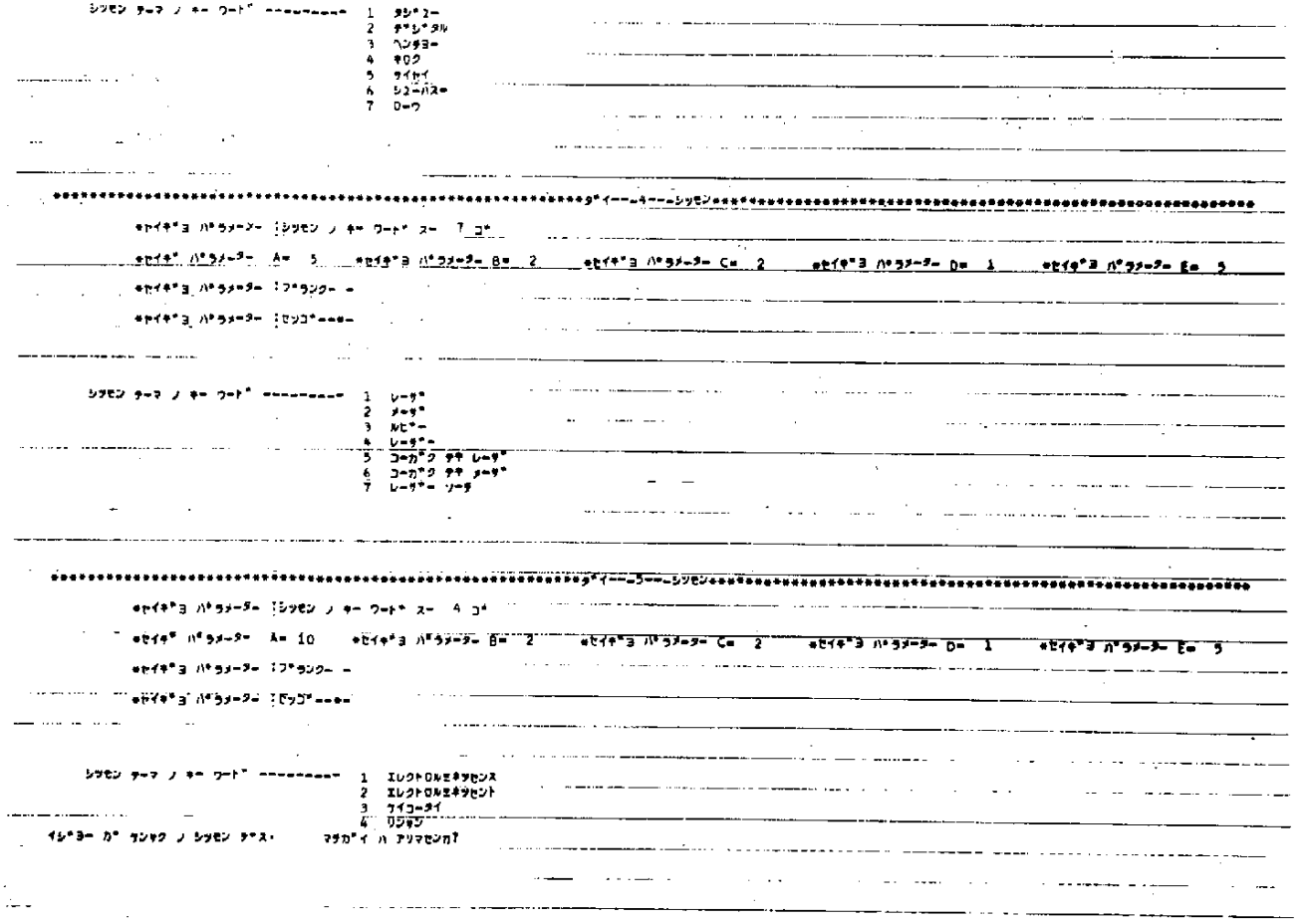


図 II - 3 - 2 8 多重質問の内容アウトプット

した。

① 出願人

② 出願日

③ 出願番号

④ 公告日

⑤ 公告番号

⑥ 優先権主張国

⑦ 優先権主張日

⑧ 発明の名称

⑨ 特許分類

(3) 検索実験

検索実験は，次の4項目を検索の基準とした。

① 出願人

② 出願日

③ 特許分類

④ キーワード

キーワードは発明の名称を分析してとり出した。

(3)－1 質問の作成

日本特許分類の全分野から，31分野を取り出し，質問作成の資料とするため，

イ 発明の名称の分析（キーワードとなりうるような用語の頻度調査）

ロ 分類調査（主分類，副分類の頻度調査）

を行なった。その結果から，発明の名称および分類を考慮して仮想の命題を考え，質問式を作成した。

なお，出願マスターテープには昭和39年4月からのデータが入ってくるので，それ以降のデータを対象とした。

(3)－2 システムの特徴

イ インプットされているデータには，明細書本文を解析して得たキーワードは含まれていないから，あまり精緻な質問はできない。

ロ 分類を用いて主題の分野を指定し，用語の意味のあいまいさを除くこと

ができる。

ハ　キーワードとなりうる用語の語幹を，検索語として使用することにより，用語全体をすべて指定しなくても，類義語および同義語に対して，検索の範囲を拡大できる。（部分マッチ）

(3)－3　真空開閉器についての検索例

特許分類の59類（一般的電気部品）のA18（真空開閉器）を選んで実験を行なった結果を報告する。

仮想命題　真空開閉器について

論理式　真空AND（開閉OR遮断）

対象　特許（55件）

公告期日　昭和39年4月1日～昭和44年3月31日

この実験ではANDを9個まで，ORはセパレータ，論理記号まで含めて70桁まで使用できるようにした。なお，否定（NOT）は検索の条件で考慮することにした。

上記論理式により電子計算機機械処理をすると，59A18のもの13件，123E9のもの1件が回答としてえられた。このように，発明の名称中に部分的に「シンクー」，「カイヘイ」，「シヤダン」が含まれていれば，そのデータをひろうことができる。回答中，特公昭43－6195のようにまったく異なるものが入っている。図Ⅱ－3－29

実用新案について上記の論理式で検索を行なえば，59A18に分類されているもの20件，そのほか58D25，59A132，59A41，66A01，66A171に分類されるものが，それぞれ1件ずつあった。

特許55件中には真空開閉器に関するもの38件であったので，特許，実用新案を合計した上での適合率は846％，呼出率は709％である。

上記の結果を得たが，適合率，呼出率は発明の名称中からどのような用語を採用するかによってかなり異なる。したがって，発明の名称を基準として検索を行なうには，発明の名称全部を分析して，一種のシソーラスを作成しなければならない。

なお，上記の質問で分類を59A18に指定すれば，当然ながら出てくる回答は59A18に含まれるものだけとなる。出願人についても，同様に指定さ

ﾀｲﾜﾝ ｼﾝｾﾝ ﾄ ｶｲﾄｳ ﾉ ｹﾝｽｳ　　S XX/XX/XX P.00001

ｼﾗｾﾝ ﾊﾞﾝｺﾞｳ　069

ｹﾝｻｸ ﾒｲｼｷﾄ　SHIRAGASHI

ｼﾗﾍﾞﾝﾋﾞ　S44.12.11

4 ﾎﾟｲﾝﾄ　ﾄｳﾛｸ　　ｹﾝｽｳ (ｳﾁ ｼﾘｮｳｶｲ)

ｼﾘｮｳﾎﾟｲﾝﾄ ﾎﾘｳ　S39.04.01 ｶﾗ S44.03.31 ﾏﾃﾞ　431771 ｹﾝ
AND
ｷｰﾜｰﾄﾞ ｸﾐｱｲ　077637 ｹﾝ
AND
ﾀｰﾑ　ｼﾝｸｳ　000227 ｹﾝ
AND
ﾀｰﾑ　ｶｲﾍｲ OR ｼｱﾀｰﾝ　000014 ｹﾝ

ｺｳｶｸﾊﾝｺﾞｳ	ﾒｲｼｮｳ	ｼｭﾘｭｳﾎﾝ	ｺｳｾｲｼﾝ	ｺｰﾄﾞﾒｲ	ｼｭﾂｶﾝﾝｼ
JPP S43-013568 S43.06.08.	ｼﾝｸｳ ｶｲﾍｲｷ	JP S39-020892 S39.04.14.	59	A18	ﾐﾂﾋﾞｼﾃﾞﾝｷ ｶﾌﾞｼｷｶｲｼｬ JP
JPP S42-017524 S42.09.14.	ｼﾝｸｳｶｲﾍｲｷ	JP S39-023848 S39.04.28. GB	59	A18	ｱｿｼｴｲﾃｯﾄﾞ ｴﾚｸﾄﾘｶﾙ ｲﾝﾀﾞｽﾄﾘｰｽ ﾘﾐﾃｯﾄﾞ
JPP S43-007328 S43.04.17.	ｼﾝｸｳｼｬﾀﾞﾝｷ ﾉ ｼﾝｸｳﾄﾞ ﾘｮｳｲｳｿｳ	JP S39-055780 S39. 9.30.	59	A18	ﾄｳｷｮｳｼﾊﾞｳﾗｲﾃﾞﾝｷ ｶﾌﾞ ｼｷｶﾞｲｼｬ
JPP S42-024924 S42.11.29.	ｼﾝｸｳｼｬﾀﾞﾝｷ	JP S39-066609 S39.11.26.	59	A18	ﾐﾂﾋﾞｼﾃﾞﾝｷ ｶﾌﾞｼｷｶｲｼｬ JP
JPP S42-013045 S42.07.26.	ﾋﾞﾃﾞｵﾃｰﾌﾟ ﾉ ｺｳｽｲ ｼﾝｸｳｼｬﾀﾞﾝｷ	JP S39-066408 S39.12. 7. US	59	A18	ﾋﾞﾃｵ& ｴﾚｸﾄﾘｯｸ ｶﾝﾊﾟﾆｰ
JPP S43-013570 S43.06.08.	ｼﾝｸｳｼｬﾀﾞﾝｷ	JP S40-020569 S40.05.17.	59	A18	ｶﾌﾞｼｷｶｲｼｬ ﾋﾀﾁｾｲｻｸｼｮ JP
JPP S43-013571 S43.06.08.	ｼﾝｸｳｼｬﾀﾞﾝｿｳﾁ	JP S40-043110 S40. 7.17.	59	A18	ﾄｳｷｮｳｼﾊﾞｳﾗｲﾃﾞﾝｷ ｶﾌﾞ ｼｷｶﾞｲｼｬ
JPP S43-006195 S43.03.07.	ﾋﾟｽﾄﾝｳｴﾙﾝ ﾆ ｵｹﾙ ｼﾝｸｳｶﾝｿﾌﾟﾘﾖｳ ｶｽﾞ ﾉｿﾒｼｮｳｷ ﾉ ﾋﾟ ﾂｳｶｲﾍｲﾌｾｯﾁ	JP S40-059891 S40.09.29.	123	E9	ﾐﾂﾐｼﾝｺｳ ｶﾌﾞｼｷｶﾞｲｼｬ
JPP S43-003329 S43.04.17.	ｼﾝｸｳｲｶｲﾛｾﾝｻｰｼｮｷ	JP S40-069808 S40.11.15. US	59	A18	ﾋﾞﾃｵ& ｴﾚｸﾄﾘｯｸ ｶﾝﾊﾟﾆｰ
JPP S43-003242 S43.02.06.	ｼﾝｸｳｼｬﾀﾞﾝｷﾖｳﾃﾝｷｾｲ	JP S40-076874 S40.12.15.	59	A18	ﾄｳｷｮｳｼﾊﾞｳﾗｲﾃﾞﾝｷ ｶﾌﾞ ｼｷｶﾞｲｼｬ
JPP S44-006084 S44.03.14.	ｼﾝｸｳｼｬﾀﾞﾝｿｳﾁ	JP S41-032709 S41.05.24.	59	A18	ﾄｳｷｮｳｼﾊﾞｳﾗｲﾃﾞﾝｷ ｶﾌﾞ ｼｷｶﾞｲｼｬ
JPP S43-021209 S43.09.11.	ｺｳｿｸｶｲﾛｼﾝｸｳｼｬﾀﾞﾝｷ	JP S41-034195 S41.05.28. US	59	A18	ﾋﾞﾃｵ& ｴﾚｸﾄﾘｯｸ ｶﾝﾊﾟﾆｰ
JPP S44-006085 S44.03.14.	ｼﾝｸｳｼｬﾀﾞﾝｿｳﾁ	JP S41-057003 S41.08.31.	59	A18	ｶﾌﾞｼｷｶｲｼｬ ﾋﾀﾁｾｲｻｸｼｮ ｼﾞｰﾋﾞｰ
JPP S44-006086 S44.03.14.	ｺｳｿｸｼﾝｸｳｼｬﾀﾞﾝｷ	JP S41-076878 S41.11.25.	59	A18	ｶﾌﾞｼｷｶﾞｲｼｬ ﾋﾀﾁｾｲｻｸｼｮ ｼﾞｰﾋﾞｰ

図Ⅱ－３－２９　検索結果解答リスト

れたもののみが回答となって出てくる。

(3)－4　かばんについての検索例

次に「かばんのさげて」についての検索では次のような結果を得た。指定した出願期日は，昭和39年4月から昭和44年6月までの5年3月である。

分類指定：　132B1051　OR132A24

論理式　　鞄AND(提手OR肥手具)

この質問の意味は，鞄の提手(肥手具ともいう)についての情報が欲しいという意味である。

この結果，12件の回答を得た。別の調査によると，「鞄のさげ手」を要旨とするものは，上記期間中に46件あることがわかっているので，この場合はもれが非常に多いことになる。もれた例には，たとえば実願昭40－3217(実公昭42－9319)「携帯用肥手装置」のように，「カバン」という用語がないものである。

上記の検索論理式で，分類を指定せずに質問すると，24件を回答として得た。この中で，上記分類以外のものを一例としてあげれば

108G11　かばんなどの計量器付さげ手

132B0　　かばん，袋物におけるさげ手装置

132B1052　かばんにおけるさげ手のとりつけ機構

これらを最初の質問と比較すると，分類を指定しなかったために，回答として得られなかった例である。このように，分類を指定せずに用語のみで質問するときは，かなり回答数が多くなるが，それでも対象とするものの全部の公報をうることができない。

(4)　問題点

イ　発明の名称は，発明の内容を正確に表現していない。そのために，精緻な検索要求には応じることができない。しかし，検索の要求が精緻なものでなく，大まかなものであれば，分類を指定しない質問により，多分野にまたがってサーチできる利点がある。ただし，この場合にはもれとノイズが非常に多くなるから，目的のものをさがし出すのに他の手段をも用いる必要が少なからずあると考えられる。

ロ　電々公社電気通信研究所が行なった自然語と分類による日本特許情報ある

いは英文特許情報の検索（昭和４４年電気四学会連合大会３３０５，３３０７）によると，日本特許に対する特許の名称部分に含まれる検索語数は，全体の１５～２０％程度であり，名称部分の検索寄与度は極めて低く，検索語の８０～８５％は明細書本文の㈠詳細な説明の冒頭部分，㈡目的記述部分，㈢特許請求の範囲などに分散潜在していると指摘している。また，英文特許（対称は米国特許のofficial Gazette)に対しては，表題に含まれている語のみで検索できるものは４７７％であると指摘している。

ハ　出願マスターテープは出願事務の処理のためのものであり，検索用としてはいくつかの欠点がある。すなわち，検索用語となりうる事項の不足と，分類および名称をインプットする際の正確度などが問題である。後者のことから，出願マスターテープをそのまま検索用に一部利用するにしても，データのチェックは欠くことができない。

ニ　ロ，に示される日本特許と米国特許の双方における表題部分の検索寄与率の差はかなりのものである。日本特許において，発明の名称を現在よりも発明の内容に忠実に書くようにすれば，検索効率の向上を図ることができる。

参　考　文　献

1. Automatic Information Organization and Ritrieval
   Gerard Salton Cornell Univ, 1969, McGrdw-Hill 54485-P
2. 志村秀雄他３名，昭和４４年電気四学会連合大会，No.３３０５，No.３３０７
3. 中村幸雄　通研にみる情報検索システム，エレクトロニクス，Ｐ１１７７～１１８０（昭４３１０）
4. 草間　基　自然語検索で効率化，事務と経営，Ｐ２３～２５（昭４３９）
5. 志村秀雄他２名　英文特許電子検索の一実施例，第５回情報科学技術研究集会論文集
6. 中村幸雄　ＲＥＷＤＡＣ方式の改良点の検討，第６回情報科学技術研究集会論文集

3.5 統　　計

この実験の目的の一つであるシステム設計に必要な各種データを得るために，2000件の蓄積データに基き，次のような各種統計を作成した。

3.5.1　一つの項目に対するデータ数が不定のもの

(1)　出願人数

表Ⅱ－3－23に示すように，単独出願が96.6％を占めている。

表Ⅱ－3－23　出願人数と特許件数表

出願人数	1	2	3	4	5	6	合計
件数	1933	41	14	1	11	0	2000
割合％	96.6	2.1	0.7	0.0	0.6	0	100

5人共同出願が多いのは，電気部門で比較的出願の多い特定の企業が，他の企業と共同出願する場合が多いので，その影響によるものと思われる。6人以上は皆無であった。

これをグラフで示すと次図Ⅱ－3－30の如くなる。

(2)　複数優先権の数

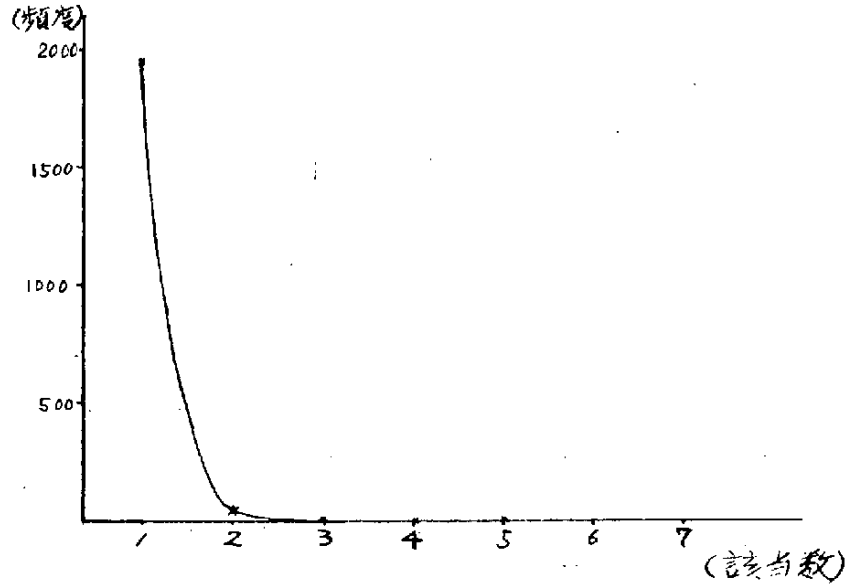
表Ⅱ－3－24　優先権の数と特許件数

優先権数	0	1	2	3	4	5	合計
件数	1517	454	25	2	2	0	2000
割合％	75.9	22.7	1.3	0.1	0.1	0	100

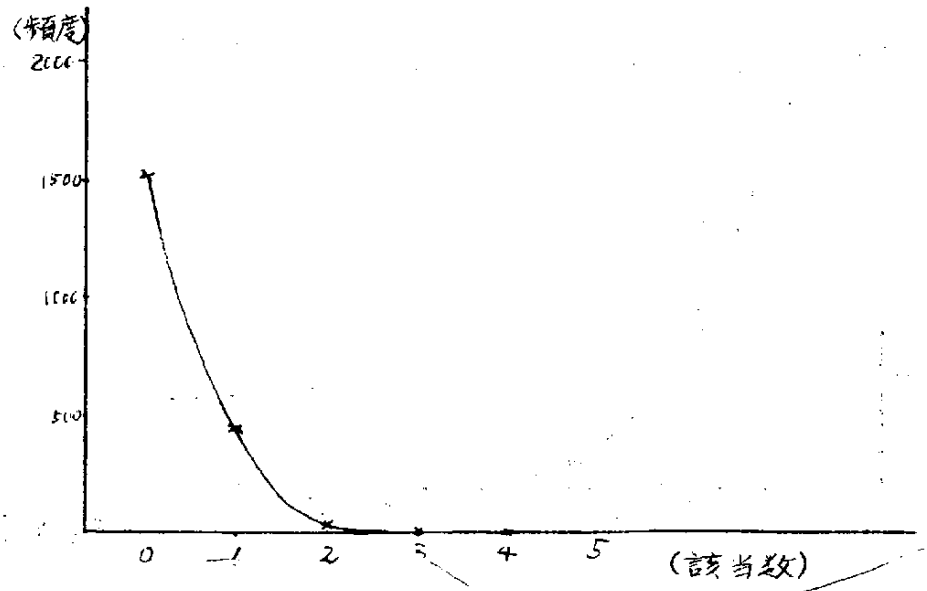
優先権主張は通常外国人が外国の出願に基いて日本に出願する場合よく用いられる出願形式で，優先権主張を伴なわない出願もある。

最高4個の優先権主張を伴う出願があることが示されている。5個以上の優先権主張を伴う出願は皆無であった。

これをグラフで示すと図Ⅱ－3－31の如くなる。



図Ⅱ－３－３０　出願人数頻度表



図Ⅱ－３－３１

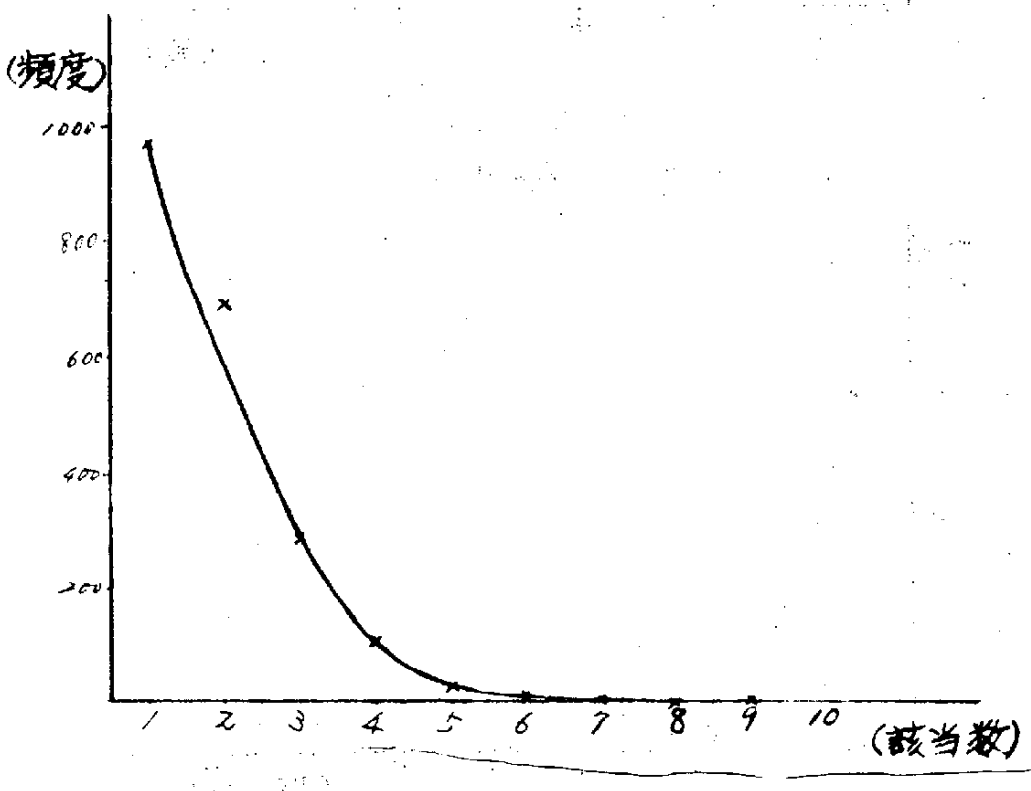
(3) 日本特許分類数

表Ⅱ－3－25 分類数と特許件数

分類数	1	2	3	4	5	6	7	8	9	10	合計
件数	911	648	268	125	32	11	4	0	1	0	2000
割合 %	45.6	32.4	13.4	6.3	1.6	0.6	0.2	0	0.1	0	100

分類は1つの特許について，主分類と，場合によっては副分類が付けられる。したがって，限ず1つまたはそれ以上あるので，分類数0は理論的にはあり得ない。分類数1は主分類のみのものを示し，それ以上は副分類が1つ以上ついたもので，ここでは最高8（分類数9と表示されている）である。

これをグラフで示すと図Ⅱ－3－32の如くなる。



図Ⅱ－3－32

(4) 文献1件当りのキーワード数

特許文献1件当りの抽出キーワード数とその全体に対する割合を次表Ⅱ－3－26に示す。

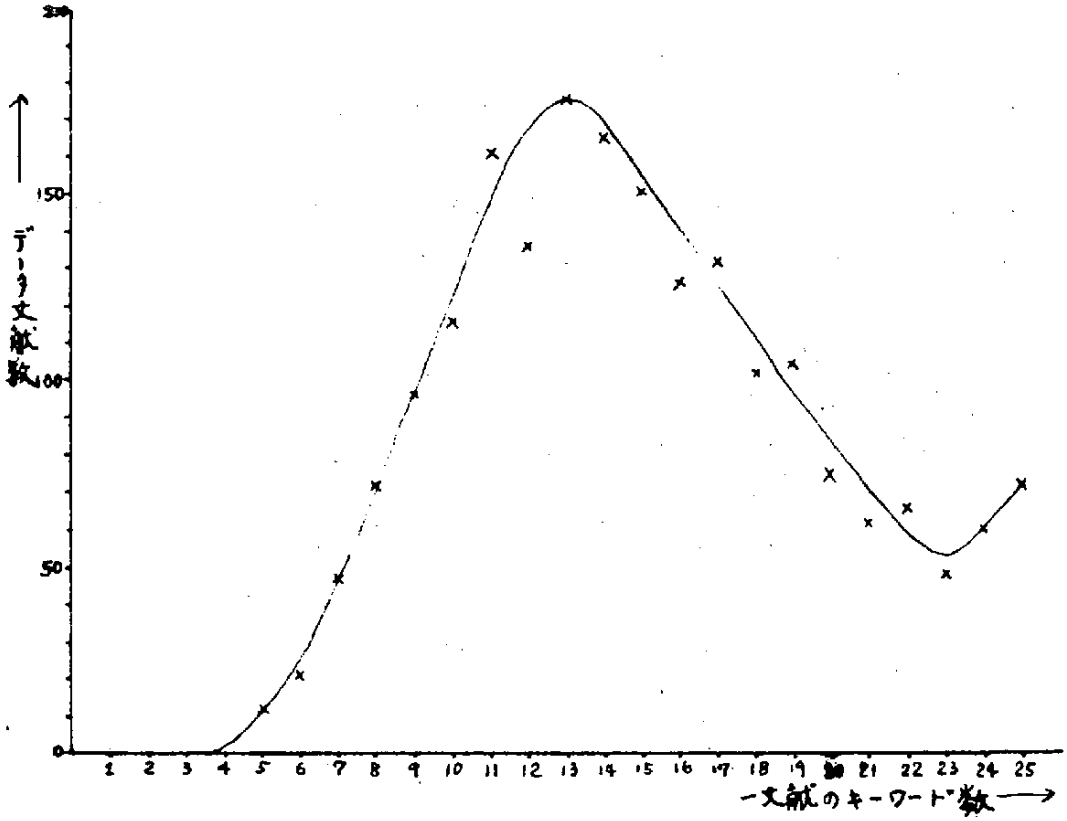
なお，3.2.6(1)「データの作成・解析」の項で述べた如く，このキーワードは16人の解析者によって抽出されたものである。

表Ⅱ－3－26　文献1件当りのキーワード数

語数	文献数	割合
1	0	0
2	0	0
3	0	0
4	0	0
5	12	0.6
6	21	1.1
7	47	2.4
8	72	3.6
9	96	4.8
10	116	5.8
11	161	8.0
12	136	6.8
13	179	9.0
14	162	8.1
15	151	7.6
16	126	6.3
17	132	6.6
18	102	5.0
19	104	5.2
20	75	3.7
21	62	3.1
22	66	3.3
23	48	2.4
24	60	3.0
25	72	3.6
合計	2000	100

これをグラフによって示すと図Ⅱ－3－33の如くなる。

この図でもわかるように，もっとも頻度の多いのは，キーワード数13，全体の平均は15である。キーワード数24，25また上昇しているのは，25語に制限したため，25語以上抽出できる文献のものが，25に集められたためと思われる。



図Ⅱ－３－３３　文献一件当りのキーワード数

３５２　データの長さが不定のもの

(1)　出願人名

インプットの段階で，最大５７字とし，それ以上のものは省略形式でデータを作成したので統計はとらなかった。

(2)　発明の名称

発明の名称の字数について統計をとった結果を次表Ⅱ－３－２７に示す。

表Ⅱ－3－27　発明の名称と字数に関する統計

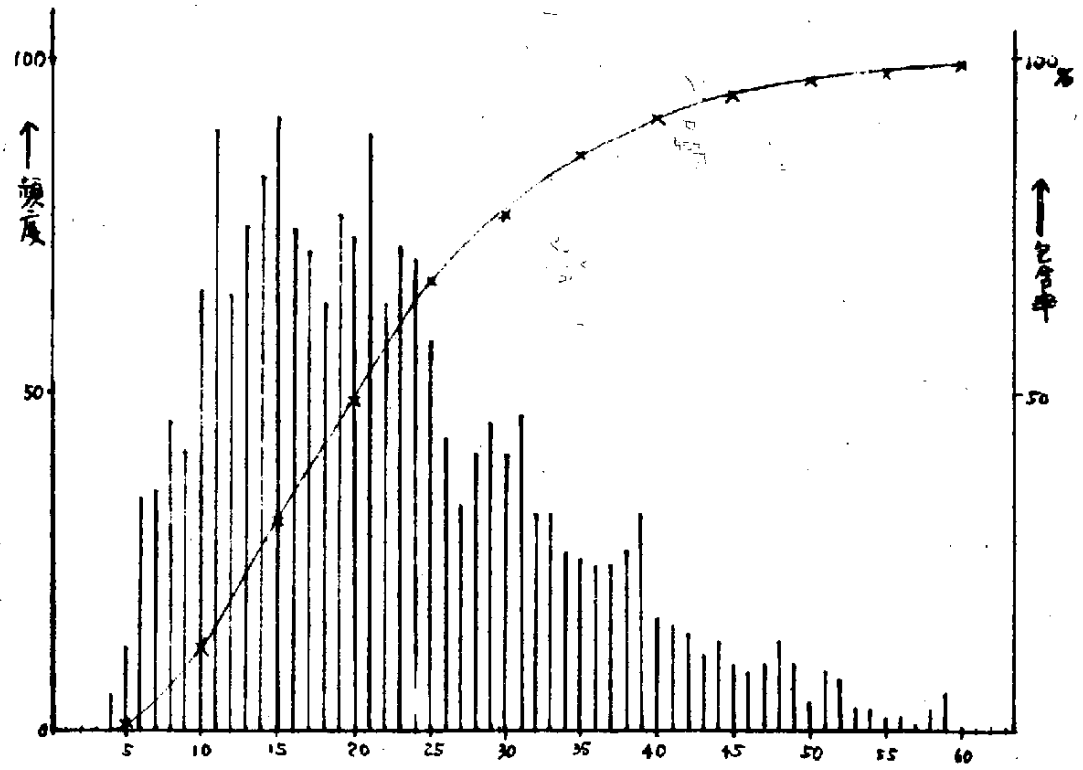
字数	件数	小計	割合	包含率
1	0			
2	0			
3	0	16	0.8	0.8
4	4			
5	12			
6	34			
7	36			
8	46	222	11.1	11.9
9	41			
10	65			
11	78			
12	63			
13	74	388	19.4	31.3
14	82			
15	91			
16	74			
17	71			
18	64	359	18.0	49.3
19	77			
20	73			
21	89			
22	63			
23	72	352	17.6	66.9
24	70			
25	58			
26	43			
27	33			
28	41	204	10.2	77.1
29	45			
30	42			
31	47			
32	32			
33	32	162	8.1	85.2
34	26			
35	25			
36	24			
37	24			
38	27	124	6.2	91.4
39	32			
40	17			
41	15			
42	14			
43	11	63	3.7	94.5
44	13			
45	10			
46	9			
47	10			
48	13	46	2.3	96.8
49	10			
50	4			
51	9			
52	8			
53	3	25	1.3	98.1
54	3			
55	2			
56	2			
57	1			
58	3	12	0.6	98.6
59	6			
60	0			
61以上	27	27	1.4	100%
合計	2000	2000	100%	

注１）　割合とは，５字単位で区切った時その単位に含まれる件数の全体件数２０００件に対する割合を示す。

２）　包含率とは，各区切り内に包含される件数の全体に対する割合を示す。

３）　６０字以上のものは，インプット段階で打切ったので，それ以上の字数のものについては，６０字以上として一括計上した。

名称の字数とその頻度，および包含率を表わすグラフを図Ⅱ－３－３４に示す。



図Ⅱ－３－３４　発明の名称字数頻度及び包含率

この図で非常に興味のあることは，ある間隔を置いて，頻度の高いものと低いものが周期的に出現し，全体として右裾のなだらかな山形をなしている点である。この特殊な形状は，常用言語の音節または，人間の言葉のリズムと何等かの関係があるのではないかと思われる。また，２７字目の非常に低い点が目立っている。発明の名称は，名詞または名詞句である。したがって比較的短い句は２７字目で，また長い句はその倍数の６０字目で終予する場合が多いことを示しているのではないかと思われる。

包含率は，４０字で９０％以上，５０字で９５％以上，６０字で，ほとんどカバーできることが示されている。この実験では，たまたま６０字に設定したが，６０字を越えるものは２０００件中わずかに２７件，１．４％にすぎなかった。

(3) キーワードの字数

キーワード３０１００語について，各語の長さ（字数）の頻度の統計をとった結果を表Ⅱ－３－２８および図Ⅱ－３－３５で示す。

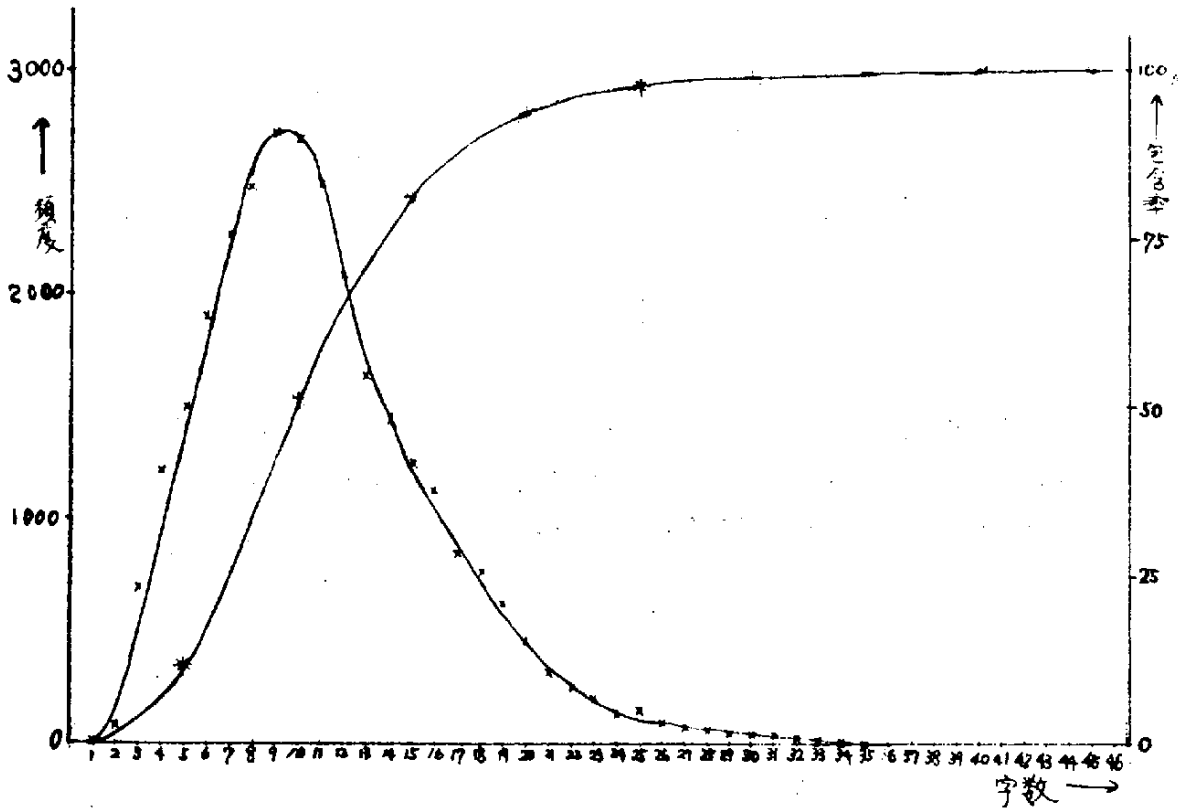
表Ⅱ－３－２８　キーワードの字数の頻度表

字数	頻度	累計	包含率	字数	頻度	累計	包含率
1	4			11	2491		
2	91			12	2087		
3	689	3489	11.6	13	1635	8891	81.0
4	1213	(11.6)%		14	1441	(29.5)%	
5	1492			15	1241		
6	1901			16	1128		
7	2251			17	850		
8	2446	12007	51.5	18	773	3831	93.7
9	2712	(39.9)%		19	626	(12.7)%	
10	2697			20	454		

21	353		
22	312		
23	247	1249	979
24	203	(4.2)%	
25	134		
26	135		
27	97		
28	81	417	993
29	56	(1.4)%	
30	48		
31	39		
32	39		
33	31	149	998
34	21	(0.5)%	
35	19		

36	13		
37	12		
38	10	51	100
39	12	(0.2)%	
40	4		
41	3		
42	3		
43	2	10	100
44	2	(0.0)%	
45	0		
46	2	(0.0)%	
合計	30100		

この表でわかるように，もっとも多いのは９字で２７１２件，となっている。



図Ⅱ－３－３５　キーワード字数頻度

３６　シソーラス

８５実験システム評価と問題点の項で述べたように，蓄積のためのシソーラスおよび，質問作成に必要なるシソーラスの両方を作成する必要があるが，まず，シソーラスをどのようにして作成するかが問題である。

この項では，今回の実験で得たデータ中，シソーラス作成に関係のあるものについて述べる。

３６１　キーワードリスト

３０１００件のキーワードを次表Ⅱ－３－２９の順序に配列してリストを作成し，同時に，同一語についての頻度を積算した。

表Ⅱ－３－２９

空白	¬	⌙	C	・	＜	(	＋	｜	&	！	¥	*	)
；	∧	ー	／	．	％	＿	＞	？	：	＃	@	ヽ	＝
・	ア	イ	ウ	エ	オ	…	ワ	ン	〃	°	A	B	C
D	…	Z	0	1	2	…	9						

その結果得られたリストの一部を図Ⅱ－３－３６で示す。なお，同一語を整理した結果，３０１００語が２１２８１語となり，約⅔に圧縮された。

この程度のリストでも質問語を作成する場合には非常に有効である。

キーワード (PAT,67,9,1-11,15)	ヒント	キーワード (PAT,67,9,1-11,15)
トツメン カタ アミジョー チンキヨク	001	トランク シュージョー イタ シキベツ カイロ
トツメン*コーカク カラス	001	トランク ソータ
トヒ*コシ ソータ ナレヒ*シヨン シンコー	001	トランクインバーグ*ルーフ* シ*ヨーホー リート*
トヒ*コシ ソータ ハイレツ	001	トランクフ*ロクコネクタ
トヒ*コシ ソータ ヨー ヘンコー チンリュー ハツトイ カイロ	001	トランクリングフレーム
トフ カクサンホー	001	トランシ*エント
トフ サキヨー	089	トランシ*スタ
トフ フ*レート*	001	トランシ*スタ アンテイ チンケ*ン ソータ
トフ ヤキツケ	005	トランシ*スタ カイロ
トメカ*ネタ*	001	トランシ*スタ ケト*ー
トメハ*キ	001	トランシ*スタ ケイチンキ
トメフ*	001	トランシ*スタ ケト*ー パ*ラメータ フッテイ ソータ
トヒエシ* シ*ヨー	001	トランシ*スタ シキ シンコー チンリュー ソータ
トラツキング*	001	トランシ*スタ シキ ソーフクキ
トラツキング* スレ	001	トランシ*スタ シキ テヨクリュー ヘンカン カイロ
トラツキング*エラー	001	トランシ*スタ シキ テヨツハ* カイロ
トリサンタル チン	001	トランシ*スタ シキ テレヒ*シ*ヨン
トラック	001	トランシ*スタ シキ*テレヒ*シ*ヨン シ*ユソーキ
トラック ケイシキ カクト*	001	トランシ*スタ シュツリヨク ソーフクキ
トラック ヒンタク キコー	001	トランシ*スタ シュツリヨクカ*ワ
トラック ハンキ	001	トランシ*スタ ソーコ*キタイアツ
トラハ*ーキ	001	トランシ*スタ ソータ
トラハ*ーキ*ヘント*	001	トランシ*スタ ソーフクキ
トラヒックコントロール ヨー リレー	001	トランシ*スタ タル
トランク	001	トランシ*スタ ツルベン

4. 英語による検索システム

4.1 システムの概要と特徴

英語による検索システムは，日本特許を対象とした場合と，英語によって記載されている外国特許を対象とした場合とがある。

日本特許を対象とした場合は，前記，日本語による検索システムの項において述べたデータの一部を使用して，試験的に行なった経過を述べる。

外国特許については，米国特許について行なった結果を述べる。

4.1.1 日本特許の英語ファイル

日本語のキーワードと英語のキーワードとが，機械によって自由に変換できたならば，前項で述べたカナ文字によって蓄積した日本特許の蓄積データは，たゞちに英語のファイルに変換でき，国際的にデータを交換する場合に便利である。もちろん，その逆に外国のデータも必要に応じて日本語のファイルに変換することもできる。

このような目的に使用する英語－日本語の辞書を作成するために次のような作業を行なった。

日本特許の検索実験に使用した２０００件の特許のうち，図Ⅱ－４－１に例示した英文抄録（ジャパン パテント センター発行，Japanese Patent Abstracts）の発行されているもの1319件について，日本語キーワードに対応する英語キーワードのみを拾い出し，リストを作成した。（図Ⅱ－４－２参照），このリストには，８８５１の日本語－英語の対になるキーワードが収録されている。

このリストによってわかるように，日本語キーワードの特定のものに対応する英語キーワードが全く同じものばかりであれば，その日本語と英語とは１：１の関係にあり，そのまま翻訳してもよいこととなる。もしもそれが2以上に分散するときは，二つ以上の訳がありいずれかを選択しなければならない。選択の方法は，関連語によって行なう方法と，分類によって行なう方法とがある。この作業を行なうために，キーワードには1件ずつその抽出された文献の特許分類を付けてある。しかし，この作業は蓄積データ数が不足のため，いまだ行なっておらず，今後の問題として残されている。

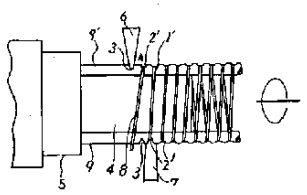
また，英語を中心として配列したリストも日本語－英語の辞書作成に役立つものと思われる。

なお，今回は，日本特許の抄録と，日本人の作成した日本特許の英文抄録を材料として使用したが，米国人の作成した米国特許資料と，その内容と同じで日本人の作成した日本語の米国特許抄録を材料として使用すれば，さらによい結果が得られるであろう。

CLASS: 99 A 13
Pub. No. 20321/67; Non-Conv.; April 23, 1965; TOSHIBA DENKI CO. ($42)

A METHOD FOR MANUFACTURE OF ELECTRON TUBE GRID comprising the steps coiling a metallic wire by engaging it in depressed parts provided around a metallic pole, said depressed parts being made perpendicular to the axis of the side wall thereof, and fixing said metallic wire being engaged with said depressed part on the metallic pole by causing plastic deformation with the aid of the pressure on the wall part between said depressed parts adjacent to each other after said metallic wire is wound.

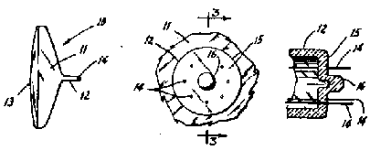
20321



CLASS: 99 A 9
Pub. No. 20322/67; Conv.; September 5, 1963; PHILCO FORD CORP. (USA) ($94)

In a combination of neck of a cathode rays tube having an exhaust pipe tip-off device which is extended from a tube neck and formed in line in axial direction with respect to said tube neck, and a plurality of terminal pins having specific interval therbetween and being adjacent to said tip off device and having stretch similar to said device, A CATHODE RAYS TUBE DEVICE comprising a body part made of an insulation base material, said body is of hollow and approximately cylindrical shape and being coaxial with said exhaust tip-off device, and a multiplicity of grooves in an axis direction formed outwardly with the provision of interval on the circumference so as to give electric connection at the side surface and to receive said terminal pins, whereby said exhaust pipe tip-off device is substantially surrounded by said hollow insulation base material.

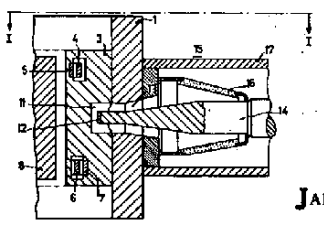
20322



CLASS: 99 B 21
Pub. No. 20323/67; Conv.; December 23, 1964; PHILIPS (GERMANY) ($42)

In a magnetron provided with a multiplicity of resonance hollow cylinders surrounding symmetrically a centre cathoda, at least one of said resonance hollow cylinder having a hole on the rear wall or side wall, A RESONANCE HOLLOW CYLINDER TYPE MAGNETRON characterized in that the wall adjacent to said hole is caused to deform for other resonance hollow cylinders, and electric characteristic and thermal characteristic of said whole resonance hollow cylinder are equalized each other.

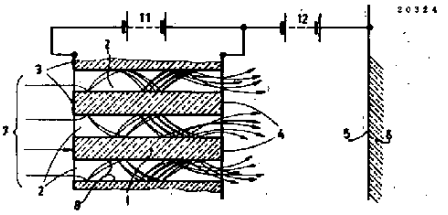
20323



CLASS: 99 B 3
Pub. No. 20324/67; Conv.; April 3, 1963; PHILIPS (GERMANY) ($53)

In A CATHODE RAYS TUBE comprising an electron radiation source, fluorescent screen, and a matrix being disposed in parallel with said fluorescent screen and being made of nonconductor material, said matrix forming narrow passages adjacent to each other extending in an advancing direction of an electron, the surface of walls of inlet and outlet of said passages being coated with a conductive layer, and the surface of inner walls of said passages are coated with the secondary electron radiation surface, the improvement characterized in that a conductive layer is coated integrally with the conductive layer on said end surface over a short length from the end surface of electron outlet side of said passage to the inner wall surface.

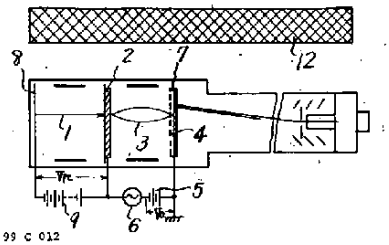
20324



CLASS: 99 B 3
Pub. No. 20325/67; Non-Conv.; April 30, 1964; NHK ($71)

A METHOD FOR STABILIZATION OF ELECTRON IMAGE MULTIPLE TUBE characterized by the steps comprising suppressing the generation of self-sustaining electron radiation with theperiodical suppression of the ascension of electric potential on the surface of secondary electron discharging by making the secondary collective voltage in the secondary electron multiplier part at a proper period below an ordinary operation voltage or by making it a negative voltage, whereby large secondary electron discharging ratio just prior to the radiation thereof is maintained.

20325



CLASS: 99 C 012
Pub. No. 20326/67; Conv.; August 16, 1963; SIEMENS ARTIENGESELLSHAFT (GERMANY) ($147)

A MAGNETIC FIELD TYPE LENS for use in particle rays device characterized in that an ultra conductive part forms a ring-shape magnetic covering member surrounding lens windings entirely, said member having a ring-shape opening at a position of the flat surface perpendicular with the particle rays, said ring-shape opening is provided with a ultra conductive orifice plate leaving sufficient leeway for the passage of flux, said orifice plate forms a wall which prevents the passage of flux at the inner part of said magnetic covering member, the orifice disposed at the outside of said magnetic covering member forms the passage for the flux coming out to pass through the particle rays and ring-shape opening, the ultra conductive part is thermally connected to low temperature cooling substance installed, iron does not exist in the flux passage of said substance, said part is connected to a pump like an object lens device of electron microscope having function to focus the flux in the zone of the particle rays, said flux being made by using a suitable number of lens windings in which current is conducted.

JAPAN PATENT CENTER, INC.
MAIL ADDRESS:
SHITAYA POST OFFICE, BOX 72.
TOKYO, JAPAN

図Ⅱ－4－1 英 文 抄 録

ﾄﾗﾝｼﾞｽﾀ ｿｰﾀ	TRANSISTOR DEVICE
ﾄﾗﾝｼﾞｽﾀ ﾁｮｸﾘｭｰ ｺｰﾘｭｰ ﾍﾝｶﾝｷ	TRANSISTOR DC AC CONVERTER
ﾄﾗﾝｼﾞｽﾀ ﾃﾞﾝﾘｮｸ ｿﾞｰﾌｸｷ	TRANSISTOR ELECTRIC POWER AMPLIFIER
ﾄﾗﾝｼﾞｽﾀ ﾃﾞﾝﾘｮｸ ｿﾞｰﾌｸｷ	TRANSISTOR POWER AMPLIFIER
ﾄﾗﾝｼﾞｽﾀ ﾆｭｰﾘｮｸ ｶｲﾛ	TRANSISTOR INPUT CIRCUIT
ﾄﾗﾝｼﾞｽﾀ ﾆｭｰﾘｮｸｶﾞﾜ	TRANSISTOR INPUT SIDE
ﾄﾗﾝｼﾞｽﾀ ﾊｯｼﾝｷ	TRANSISTOR OSCILLATOR
ﾄﾗﾝｼﾞｽﾀ ﾊｯｼﾝｷ	TRANSISTOR OSCILLATOR
ﾄﾗﾝｼﾞｽﾀ ﾌﾞﾌﾞﾝ	TRANSISTOR SECTION
ﾄﾗﾝｼﾞｽﾀ*ｶｲﾛ	TRANSISTOR*CIRCUIT
ﾄﾗﾝｼﾞｽﾀ*ｾﾝﾀｸ ﾏﾄﾘｯｸｽ	TRANSISTOR*SELECTION MATRIX
ﾄﾗﾝｼﾞｽﾀ*ﾅｲﾌﾞ ﾃｲｺｰ	? *INTERNAL RESISTANCE
ﾄﾗﾝｼﾞｽﾀ*ﾋﾃｲ ｶｲﾛ	TRANSISTORS* ?
ﾄﾗﾝｼﾞｽﾀ*ﾋﾖｰｭｰ ﾘｱｸﾀﾝｽ	TRANSISTOR*FLOATING REACTANCE
ﾄﾗﾝｼﾞｽﾀ*ﾍﾞｰｽ ﾃﾞﾝｲ	TRANSISTOR*BASE POTENTIAL
ﾄﾗﾝｼﾞｽﾀ*ﾗｯﾁ ｶｲﾛ	TRANSISTOR*LATCH CIRCUIT
ﾄﾗﾝｼﾞｽﾀ*ﾛｼﾞﾘ ｶｲﾛ	TRANSISTOR*THEORY CIRCUIT
ﾄﾗﾝｼﾞｽﾀｰ*ｹﾞｲﾝｺﾝﾄﾛｰﾙ ｿｰﾀ	TRANSISTOR*GAIN CONTROL DEVICE
ﾄﾗﾝｼﾞｽﾀｲﾝﾊﾞｰﾀ	TRANSISTOR INVERTER
ﾄﾗﾝｼﾞｽﾀﾗｯﾁ ｶｲﾛ*ﾄﾘｶﾞ ｶｲﾛ	TRANSISTOR LATCH CIRCUIT*THIGGER CIRCUIT
ﾄﾗﾝｽ	TRANSFORMER
ﾄﾗﾝｽﾃﾞｭｰｻ	TRANDUCERS
ﾄﾗﾝｽﾃﾞｭｰｻ	TRANSDUCER
ﾄﾗﾝｽﾃﾞｭｰｻ	TRANSDUCER
ﾄﾗﾝｽﾃﾞｭｰｻｰ	TRANSDUCER

図Ⅱ－4－2　日本語 → 英語キーワードリスト

4.1.2　米国特許の英語ファイル

蓄積工程は，図Ⅲ－4－3に示すように，日本語による場合とは若干そのシステムが異なる。

英語によって記載されているU.S. Official Gazette（特許請求範囲または，それに代る簡単な抄録が掲載されている。）から抽出された英語キーワードと，米国特許明細書全文を見て作成した邦文抄録から抽出した日本語キーワードを英訳したものとを合せて，磁気テープに蓄積データとして読込む。

この際，日本特許について行なったように，さらに分類表のキーワードを追加することも可能であるが，今回の実験では一応除外した。

分類は，米国分類と，邦文抄録に記入されている日本分類の両方をインプットする。

検索方法は3.1の日本特許において述べた方法とほぼ同じである。

なお，図Ⅲ－4－3に点線で示されているように，前記4.1.1における日本語－英語の辞書が完成すれば，現在人手によって行なっているキーワードの英訳も，機械によって自動的に行なうことが可能

このシステムの特徴は，

① データ抽出材料として2つ以上の資料，すなわち，米国特許公報（U.S. Official Gazette）と日本語の米国特許抄録誌を使用して，偏った原データによるキーワードの抽出もれを防いだ。

② 通常，英語による検索の場合は単語を用いているが，複数の単語よりなる複合語はそのままキーワードとして採用している。

③ ノイズ防止の手段は，日本特許において述べた方法と同様に，質問語のウエイト付けと，部分マッチングによる評価とを併用して関連性の薄い文献を除くようにした。

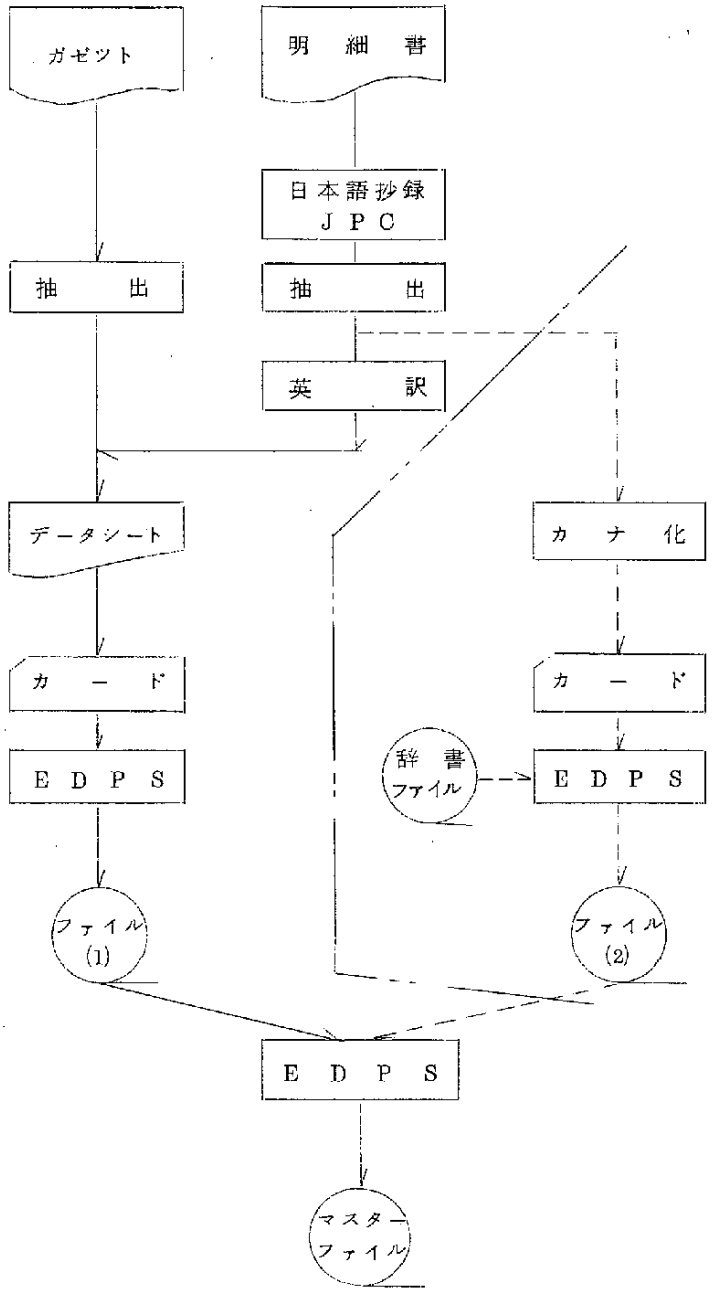
④ ネガティブ条件の設定

試験的に，ネガティブ条件についてもインプットデータを作成した。ネガティブ条件とは，発明がその条件を使用しないこと，または，そのような条件が起らないことを特徴としている場合の条件のことで，例えば「感温制御器を使用しなくても，従来以上の効果をあげることができた。」という場合の，「感温制御器」はネガティブ条件と呼ぶことにした。

インプットする場合には，その内容を表わすキーワードの前に，特殊記号を入れ，ネガティブ条件であることが判別できるようにし，さらに，そのキーワード自体でも検索できるようにした。

⑤ 分類には，米国特許分類と日本特許分類の両方をインプットした。したがって，分類としては，米国，日本の両方から検索できるばかりでなく，日本特許の検索方式で述べたように，分類表のキーワードを追加する場合には，両方の分類からのキーワードを使用することができる。

米　　国



図Ⅱ－４－３　蓄積工程図

4.2 データ

4.2.1 原資料

米国特許のデータ蓄積のために使用した抽出材料は，外国特許抄録誌と米国特許公報である。以下，上記の文献からキーワードを抽出し，インプットカードを作成した工程迄の経過を述べる。

① 外国特許抄録誌

米国特許の主要部門について抄録を作成し，部門別に編集し抄録誌として市販されている。この抄録の作成は㈱特許データセンターが行ない，発行は㈱発明協会で行なっている。内容は図Ⅱ－4－4に示すように，発明の内容の要約と図面の外，米国分類と日本分類とが併記されている。

② 米国特許公報（U.S.Patent Official Gazette）

米国の特許が登録されると同時に発行される特許公報で，書誌的事項と主要請求範囲および図面が記載されている。近年は請求範囲の代りに，抄録が掲載されている。

その内容は，図Ⅱ－4－5に示す通りである。

3,296,459
SUPERCONDUCTOR CIRCUIT WITH PROTUBERANCES
Stanley Phillips Frankel, Los Angeles, Calif., assignor to General Electric Company, a corporation of New York
Filed Jan. 13, 1964, Ser. No. 337,322
3 Claims. (Cl. 307—88.5)

3. A superconductor circuit comprising a first superconductor strip forming a gate, a second superconductor strip forming a control, and means for providing electric current in said second strip, said second strip being disposed in proximity to said first strip for the magnetic field produced by said current to couple to the first strip, said first strip being formed with a plurality of ridged deformations, each of said ridged deformations being relatively small compared with the width of said second strip, said deformations being oriented such that a substantial portion of said magnetic field is applied to said deformations in a direction substantially transverse to said deformations.

図Ⅱ－4－5　米国特許公報

米国特許　3,296,459　　Cl. 307－88.5
特許日　1967.1.3　　☆100 D 6　98(5) G 291
出願日　1964.1.13
発明者　Stanley Phillips Frankel
権利者　G.E. Co. (米)

超伝導回路

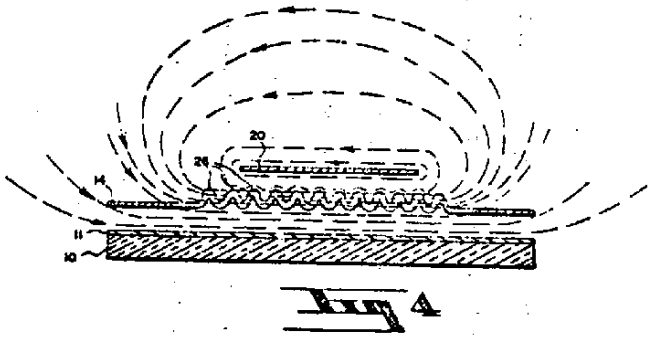
この発明は利得が大きく、かつ時定数が小さい超伝導ゲートおよびスイッチング回路に関し、ゲート導体に膜または溝をつくったことを特徴とする。

従来の平らなゲートで構成した超伝導スイッチ装置では、ゲート導体は第1および第2の変化点をもっている。第1の変化点はゲート導体に直交する制御導体の電流がある値に増加したとき、制御導体の線の部分に近接したゲート導体の極く狭い領域が誘導磁束によって抵抗性になるために生じるもので、全体としてゲート抵抗の変化は小さくしたがって時定数は大きいが、利得は大きい。第2の変化点は制御電流をさらに増加したときに生じるもので、制御導体の影響をうけるゲート全体が抵抗性になるので、ゲート抵抗の変化は大きく、したがって時定数は小さいが、利得も小さい。この発明は利得が大きい第1の変化点で大きなゲート抵抗の変化を実現するものであり、利得が大きく、時定数の小さい超伝導装置をうるものである。

図において、10は基板、11は薄膜超伝導物質のシールド、14はゲート導体で直交する制御導体20の下部に幅方向に溝26を設けている。図では省略しているがシールド11とゲート14の間およびゲート14と制御導体20の間には絶縁層がある。

ゲート14に溝または穴を設けると制御電流によって生じた磁束が穴の面にほぼ直角に何回もゲート14を貫通するので、小さい制御電流によって抵抗性に変化し、かつ抵抗性に変化するゲートの領域が広くなり、したがって時定数が小さくなるから、高利得、高速動作の超伝導ゲートあるいはスイッチング装置に適する。

（全4頁　全8図　3クレーム）



第4図

参考文献
米　2,966,647　3,115,612　英　908,704

6－151. P 32. (1968.8.5)

③　対象分野とデータ数

日本特許分類１００Ｄ「電気的諸装置」に該当するもののうち，１９６７年１月～４月１１日に特許になったもの１００件を対象とした。

4.2.2　データ作成

①　解　析

解析は，米国特許公報，および外国特許抄録上において，直接，キーワードにアンダーラインを引く（図Ⅱ－４－４および図Ⅱ－４－５参照）ことによって行なった。

米国特許公報の解析は，６人の電気または物理の専門技術者によって行ない，１人１０～２０件を数日で完了した。

外国特許抄録の解析は，米国特許公報に記載されていないと思われるキーワードのみを抽出した。またこの作業には１人で数日を要した。外国特許抄録誌から抽出された和文キーワードは，３人の文科系の作業員が英語に翻訳した。翻訳段階で，英文キーワードと付き合せ，重複していないキーワードのみを翻訳したが，それに要した日数は２～３日であった。

②　コーデング

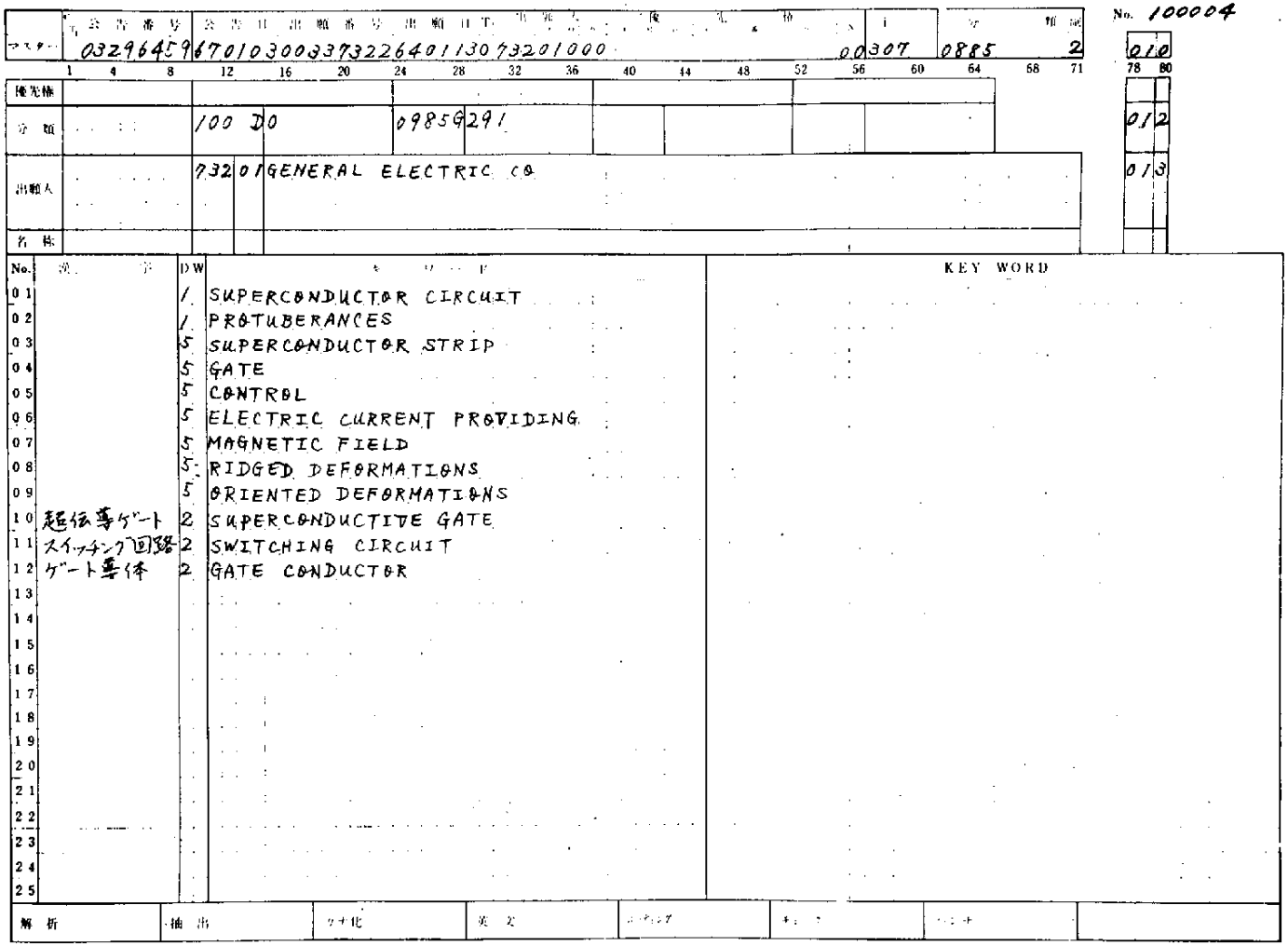
コーデングは図Ⅱ－４－６で示すコーデングシートを使って行なった。

書誌的事項のコーデングは，日本特許の場合とほぼ同様であるが，主分類欄に米国分類を，副分類欄に日本分類を記載した。なお，米国分類は，主分類のみであるが，日本分類には副分類も含まれている。

書誌的事項のコーデングに要した日数は２人で１日で，日本特許の場合に較べて，カナ化の問題がないので，遙かに楽である。

チェックは，１人１日で完了した。

チェックの結果発見されたミスの種類と件数は次表Ⅱ－４－１の通りである。



	公告番号	公告日 出願番号 出願日	分類	頁数	No. 100004
マスター	03296459	67010300 33732264 01130 73201000	00307	0885 2	010

1 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 71 78 80

優先権								
分類		100 D0		0985G291				012
出願人		732 01 GENERAL ELECTRIC CO						013
名称								

No.	文節	DW	キーワード	KEY WORD
01		1	SUPERCONDUCTOR CIRCUIT	
02		1	PROTUBERANCES	
03		5	SUPERCONDUCTOR STRIP	
04		5	GATE	
05		5	CONTROL	
06		5	ELECTRIC CURRENT PROVIDING	
07		5	MAGNETIC FIELD	
08		5	RIDGED DEFORMATIONS	
09		5	ORIENTED DEFORMATIONS	
10	超伝導ゲート	2	SUPERCONDUCTIVE GATE	
11	スイッチング回路	2	SWITCHING CIRCUIT	
12	ゲート導体	2	GATE CONDUCTOR	
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				

解析	抽出	ワク化	英文				

図Ⅱ—4—6　コーディングシート

表Ⅱ－4－1　書誌的事項のミスの種類と件数

欄	ミスの内容	件数
マスターカード	出願人コード脱落	3
	優先権コード脱落	47
出願人カード	ミススペル	1
	コード脱落	2
	表記法誤り	1
コントロールコード	脱　落	1
合　　計		55

註　1）ミスの件数とは，ミスの発生件数である。したがって，1枚のコーデングシート上に数個所ミスのある場合は，その数だけ計上した。

この表で分るように，ミスの大部分は，特殊コードの記入もれで，書き違いはほとんどなかった。

キーワードのコーデングについては，米国特許公報の場合，解析者が直接コーデングシート上に記載し，米国特許抄録の場合は翻訳者が翻訳と同時にコーデングシート上に記載した。

抽出されたキーワードの語数は次のとおりである。

米国特許公報から	1117語
外国特許抄録から	346語
合　　計	1463語

このキーワードのチェックは1日で完了したが，チェックの結果発見されたミスは次表Ⅱ－4－2の通りである。

表Ⅱ－4－2　キーワードのミスの種類と件数

ミスの内容		ミスの件数		
		抽出者	翻訳者	合計
抽出ミス	不足	4	0	4
	不適当	1	0	1
	和文キーワード転記ミス	0	1	1
コーデングミス	書き忘れ	4	2	6
	ミススペリング	46	5	51
	語順ミス	6	0	6
	語間ミス	6	0	6
	単語分解	1	0	1
	翻訳ミス	0	2	2
	特殊コードミス	172	5	177
a．合計		240	15	255
b．総キーワード数		1117	346	1463
ミスの発生率 a/b %		21.5	4.3	17.4

註　1）of，in などの不要語は削除しているが，不要語を除いた時，語順を変えれば名詞句を形成する場合，語順を置換えている。この規約に反したものを語順ミスという。

2）語間ミスとは，原文が単語を分けて記載しているにもかかわらず，それを続けて記載した場合をいう。

3）単語分解とは，当然一単語として採用すべきキーワードを，二単語として分けて記載したものをいう。

この表からわかるように，抽出者がコーデングシート上にキーワードを転記する場合，意外とミススペリングが多いことである。

なお，特に多かった特殊コードのミスとは，抽出者の1人がキーワードの出所（D欄）の記載方法をほとんど全部間違えたためのもので，このミスを除けば，抽出者によるミスの発生率は，約6％，全体で，約5％となる。

ⓐ カードパンチ

米国特許１００件について作成したカード枚数は次表Ⅱ－４－３の通りである。

表Ⅱ－４－３　カードの種類と作成枚数

カードの種類		枚数
書誌的事項	マスターカード	１００
	副分類カード	１０３
	優先権カード	１２
	出願人カード	１０２
	発明の名称カード	１００
	小計	４１７
キーワードカード		１４６３
合計		１８８０

このカードパンチは，書誌的事項については社内で，キーワードカードについて社外に外注した。パンチ作業は，英数字のみであるので，特に問題になる点はなかった。

4.3　プログラム

3.3 の日本語ファイルによるものとほゞ同じであるので省略する。

4.4　検索実験

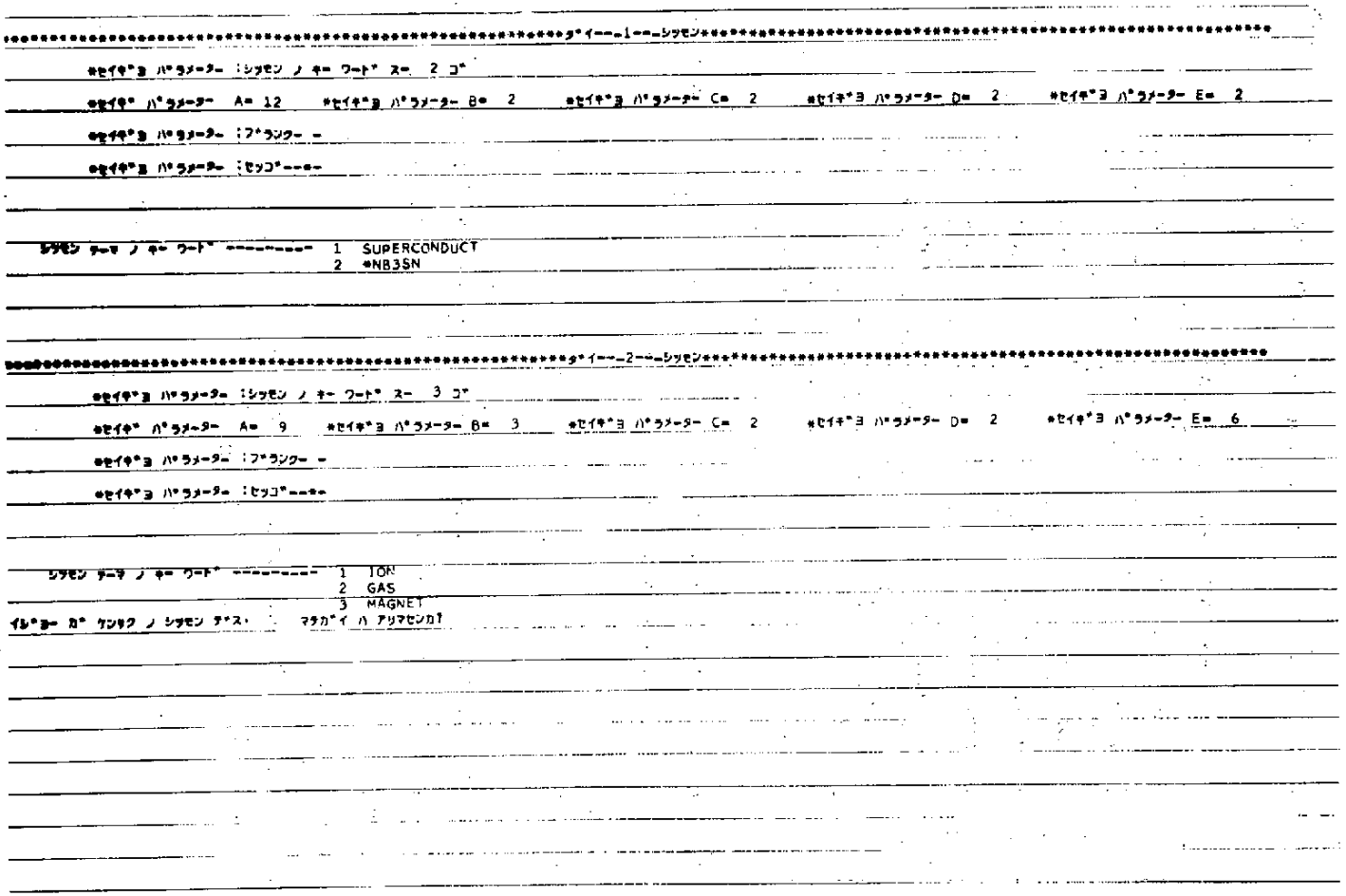
米国特許の検索の場合には，①英語キーワードがカナキーワードと検索上同等に取扱われるか否か，②日本語の外国特許抄録誌からの追加が有効であったか否か，の２点を主として調べることとした。

なお，この実験では分類表からのキーワードを付加せずに行なった。

その代表的例については次に説明する。

この実例では，「超電導材料」と「ＭＨＤ発電」の２質問を同時に質問した。

その質問内容および回答のアウトプットの一部を次図Ⅱ－４－７で示す。



図Ⅱ－4－7　質問および回答のアウトプット

ﾃﾞﾊﾞ ｺﾚｶﾗ ｲﾝｻｸ ｦ ﾊｼﾞﾒﾏｽ.

********** ｺﾉ ﾌﾞﾝｹﾝ ﾊ ﾀﾞｲｰｰ1ｰｰｼｭﾓﾝ ﾉ ｶｲﾄｰ ﾀﾞｽ **********

*ｺｰｺｸ ﾊﾝｺﾞｰ *ｺｰｺｸﾋﾞ *ｼｭﾂｶﾞﾝ ﾊﾝｺﾞｰ *ｼｭﾂｶﾞﾝﾋﾞ *ｼ ｭ ﾌﾞ ﾝ ﾙ ｲ *ﾌ ｸ ﾌﾞ ﾝ ﾙ ｲ *
3296684 29
*ﾊﾂﾒｲ ﾉ ﾒｲｼｮｰ :
*ｼｭﾂｶﾞﾝﾆﾝ

*ﾕｰｾﾝｹﾝ ｼｭﾁｮｰ :

*ｼｭﾓﾝ ﾀｰﾏ ﾉ ｷｰ ﾜｰﾄﾞ 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

*ﾏﾂﾀﾝｸﾞ ｼﾞｮｰﾀｲ A 1 16

*ｺﾉ ﾌﾞﾝｹﾝ ﾉ ｷｰ ﾜｰﾄﾞ: 1 1 INTERMETALLIC SUPERCONDUCTORS FORMING METHOD
3 5 DUCTILE BASE
5 5 TIN LAYER
7 5 INTERDIFFUSION LAYER
9 5 ELEVATED TEMPERATURE
11 5 X-RAY DIFFUSION
13 5 METALLURGICAL FURNACE
15 5 *NB3SN BRIDGES
17 2 SOLENOID COIL
19 2 SUPERCONDUCTION PARTS
2 5 *NB3SN
4 5 NIOBIUM LAYER
6 5 NIOBIUM SURFACE
8 5 HEAT TREATMENT
10 5 INTERMETALLIC LAYER
12 5 SPOOL
14 5 REFRACTORY WASH
16 2 MAGNETIC INSULATION
18 2 ARMATURE

********** ｺﾉ ﾌﾞﾝｹﾝ ﾊ ﾀﾞｲｰｰ2ｰｰｼｭﾓﾝ ﾉ ｶｲﾄｰ ﾀﾞｽ **********

*ｺｰｺｸ ﾊﾝｺﾞｰ *ｺｰｺｸﾋﾞ *ｼｭﾂｶﾞﾝ ﾊﾝｺﾞｰ *ｼｭﾂｶﾞﾝﾋﾞ *ｼ ｭ ﾌﾞ ﾝ ﾙ ｲ *ﾌ ｸ ﾌﾞ ﾝ ﾙ ｲ *
3297890 310
*ﾊﾂﾒｲ ﾉ ﾒｲｼｮｰ :
*ｼｭﾂｶﾞﾝﾆﾝ

*ﾕｰｾﾝｹﾝ ｼｭﾁｮｰ :

*ｼｭﾓﾝ ﾀｰﾏ ﾉ ｷｰ ﾜｰﾄﾞ 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

*ﾏﾂﾀﾝｸﾞ ｼﾞｮｰﾀｲ B 0 1 4

*ｺﾉ ﾌﾞﾝｹﾝ ﾉ ｷｰ ﾜｰﾄﾞ: 1 1 HIGH ENERGY ELECTRICAL IMPULSES
3 5 MHD GENERATING MEANS
5 5 UNIDIRECTIONAL ELECTROMOTIVE FORCE
7 5 MAGNETIC FIELD
9 5 INDUCTOR
11 5 HIGH IMPEDANCE
13 2 ELECTRIC PULSE
2 5 ELECTRICAL IMPULSES GENERATING APPARATUS
4 5 ELECTRICAL CURRENT
6 5 OPPOSED ELECTRODES
8 5 HOT ELECTRICALLY CONDUCTIVE GAS
10 5 UNIDIRECTIONAL CONDUCTING MEANS
12 5 ELECTROMOTIVE FORCE REDUCING MEANS

図Ⅱ－4－7　質問および回答のアウトプット

実例1.

① 質問の主旨

Nb3Sn を合金成分とする超電導材料

② 質問に使用したキーワード

1. SUPERCQNDUCT

2. ＊NB3SN

③ 論理式

1×2

④ 検索結果

検索の結果次の3件が抽出されたが，3件とも正解であった。（表Ⅱ－4－4参照）

表Ⅱ－4－4 検索結果

No.	特許番号	英語KW	英語KW 抄録KW	適否
1	3296684	○	○	○
2	3309179	○	○	○
3	3310862	×	○	○

なお，蓄積データ100件について，マニュアルで全数チェックした結果は，超電導材料に関するものは，7件あり，その中Nb3Snについて記載のあるものは，上記の3件のみであり，したがって検索もれはなかった。

なお，外国特許抄録誌からのキーワードを除いた場合を考えると，No.3の特許は脱落することになり，抄録誌からのキーワードが有効に働いていたことがわかる。

実例2.

① 質問の主旨

電磁流体発電

② 質問のキーワード

1. ION

2. GAS

3. MAGNET

③ 論理式

(1 + 2) × 3

④ 検索結果

次の6件の特許が検索された。これを全数チェックしたところ，正解は8件あることが分った。これを内容的に検討するために，質問に使用したキーワードの出所と検索結果の関係を次表Ⅱ－4－5に示す。

表Ⅱ－4－5　キーワードの出所と検索結果

	特許番号	(ION+GAS)	×MAGNET		MHD	適否	備考
1	3297890	－	○	○	△	○	
2	3309545	△	○	○	－注2	○	
3	3309546	△	○	○	－注3	○	
4	3310689	△	△	○	－	○	
5	3310807	△	△	○	－	○	MHDに使用するとの記載はないが関係がある
6	3311762	○	○	○	－注2	○	
7	3303683	－	－注4	○	○	○	
8	3303684	△	○	－	△	○	

註　1）表中△印は，抄録からのキーワード，○印は，ガゼットからのキーワード，－印は，該当キーワードが蓄積されていなかったことを示す。

2）抄録の名称に該当キーワードがあっても抽出しなかった。

3）抄録中に該当キーワードがあったが，抽出しなかった。

4）抄録中で抽出したが，コーデング中のミスにより脱落した。

この表からわかるように，抄録からのキーワードを追加しなかった場合は，*No.*4，*No.*5および*No.*7が脱落することになり，抄録のキーワードを追加した効果がはっきり表われている。

次に脱落した2件について，原因を調査したところ，*No.*7は，キーワード「GAS…」について，原データ上では抽出されているにも拘らず，コーデングシート上には転記されておらず，明らかなコーデングミスによる検索もれである。*No.*8は，MHD発電に使用される材料に関する特許で，ガゼット上には，その記載は全くないが，抄録上では，MHD発電との記載があり，キーワードとしても抽出されている。

したがって，この例では，次の2点が指摘される。

① 質問作成上の問題

結果的に考えれば，MHDを含めて次の論理式

（ION＋GAS）×MAGNET＋MHD

で検索すれば全数8件抽出され，またMHDという特定的なキーワードの場合は，ノイズはおそらく増えないと思われる。

② 蓄積データ作成上の問題点

No.7の検索もれは，コーデングシート作成上のミスによるものであり，質問作成上このようなエラーの発生を常に念頭においておかねばならないことが，実際上起っている。この場合，若しも上記のように念のためMHDで検索しておけば，このエラーをカバーすることができた。

しかし，検索されたもののうち，No.2，3，6は，いずれも，抄録上の発明の名称または冒頭部分に記載されているにも拘らず，抽出されていない。これは，抄録からのキーワード選定方針を，発明の名称は当然，ガゼットの発明の名称からキーワードを抽出するので，省略した結果によるもので，（抄録作成の方針として，発明の名称は原文の名称をそのまま翻訳することになっているが，例外として略称を記載してもよいことになっている。）抄録作成上の例外規定に気が付かず，抽出しなかったためである。

4.5 統　　計

この実験の目的の一つであるシステム設計に必要な各種のデータを得るために１００件の蓄積データに基き，次のような各種統計を作成した。

4.5.1 一つの項目に対するデータ数が不定のもの

(1) 出願人数

表Ⅱ－４－６に示すように単独出願が９８％を占めている。蓄積データが100件と少ないので正確には解明出来ないが２人共同で出願しているものは全体の２％にすぎない。

表Ⅱ－４－６　出願人数分布

出願人数	1	2	3	合　計
頻　　度	９８	2	0	１００

(2) 優先権の数

優先権主張については表Ⅱ－４－７に示すように最高２個の優先権主張を併う出願が１件あり，これについで１個が１１件，１１％を占めている。

表Ⅱ－４－７　優先権数分布

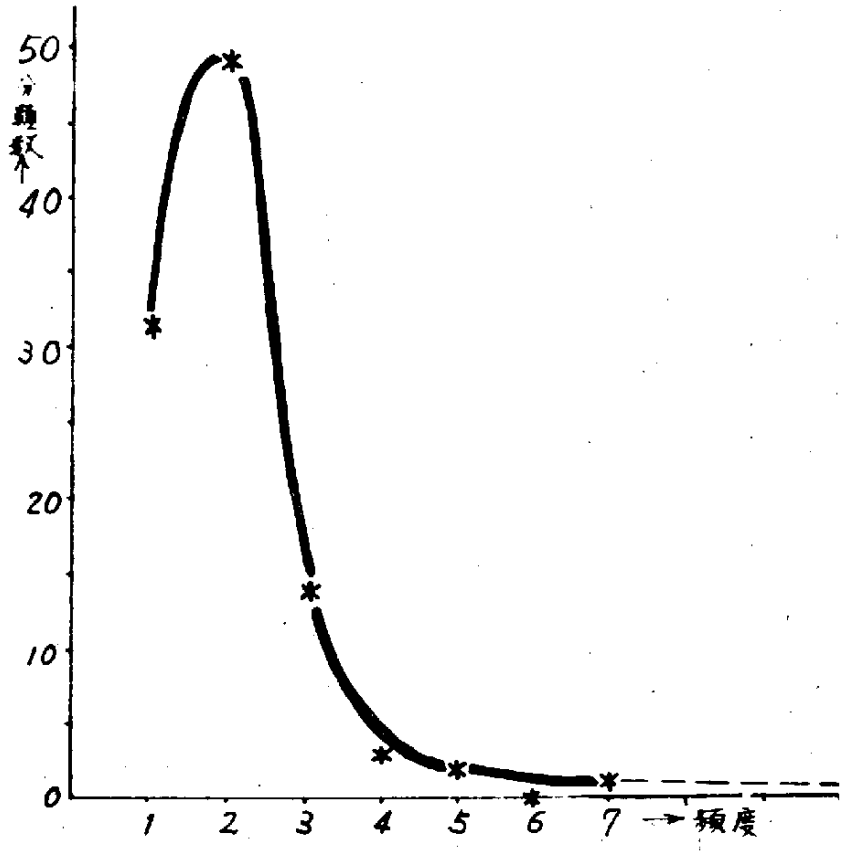
優先権数	0	1	2	3	合　計
頻　　度	８８	１１	１	０	１００

(3)　分類の数の分布

米国特許の検索システムには，米国分類と日本分類とをインプットした。米国分類は主分類のみであるので，特許１件につき１分類である。日本分類は副分類もつけたので，その分類数の分布状態について統計をとり，その結果を表Ⅱ－４－８および図Ⅱ－４－８で示す。

表Ⅱ－４－８　日本特許分類の数分布

分　類　数	1	2	3	4	5	6	7	8	合　計
頻　　度	３１	４９	１４	３	２	０	１	０	１００



図Ⅱ－４－８　日本特許分類の数の分布

この図でわかるように，分類数２のものがもっとも多く，この状態は通常の分類数の分布状態（例えば，日本特許検索システムにおける統計表，表Ⅱ－３－２５参照）のものとを較べると変則的である。これは，このシステムで採用した対象が日本特許分類１００Ｄに属するものだけであり，しかも１００Ｄは電気部門の雑に相当する分野であるためであろう。

(4) 文献１件当りのキーワード数

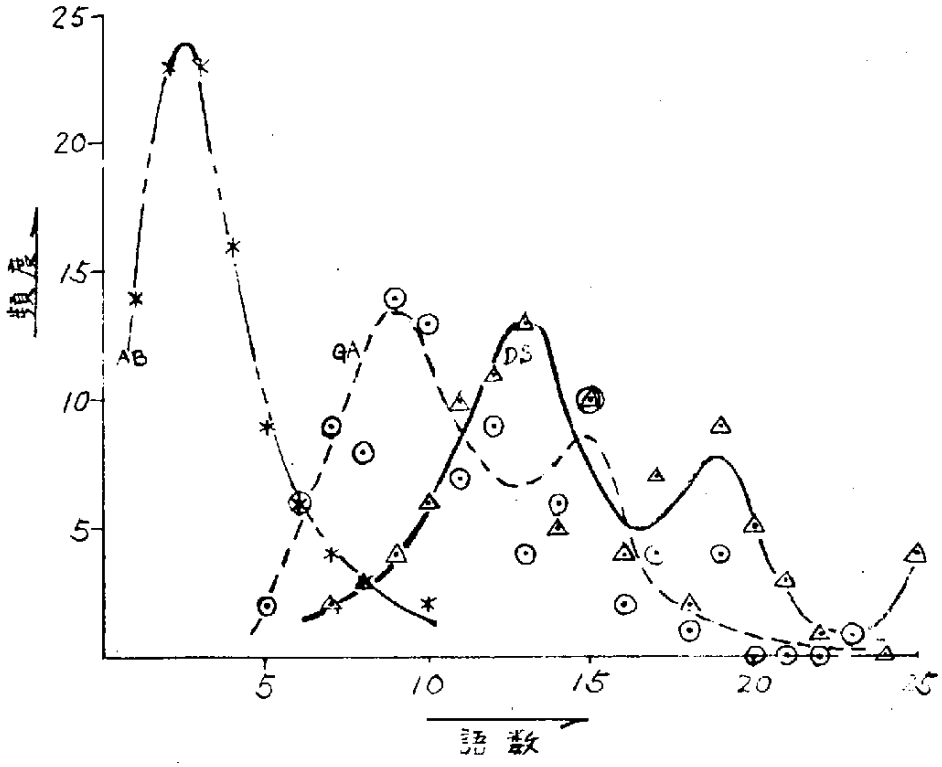
蓄積データ１００件から抽出したキーワードの数の分布を調べた結果は次表Ⅱ－４－９に示す如くである。

表Ⅱ－４－９　キーワード数分布

語数	頻度		
	GA	AB	DS
1	0	14	0
2	0	23	0
3	0	23	0
4	0	16	0
5	2	9	0
6	6	6	0
7	9	4	2
8	8	3	3
8	14	0	4
10	13	2	6
11	7	0	10
12	9	0	11
13	4	0	13
14	6	0	5
15	10	0	10
16	2	0	4
17	4	0	7
18	1	0	2
19	4	0	9
20	0	0	5
21	0	0	3
22	0	0	1
23	1	0	1
24	0	0	0
25	0	0	4
	100	100	100

注　GA：ガゼットからのキーワード
　　AB：抄録からのキーワード
　　DS：データシート上のキーワード

表Ⅱ－４－９を図Ⅱ－４－９に示すと下記のようになる。



図Ⅱ－４－９　キーワード数の分布状態

註　（GA）　ガゼットからのキーワード

（AB）　抄録からのキーワード

（DS）　データシート上のキーワード

図Ⅱ－４－９についての説明

GA：ガゼットから抽出したものの内容のバラツキと抽出時点における作業者７人の語数抽出のバラツキが考えられる。（表Ⅱ－４－１０参照）

AB：外国抄録より抽出し，作業者は１人である，抽出基準は抄録に特徴的に表われる言葉を中心

としてガゼットとの重複が予想されるものは抽出しなかった。ガゼットからの抽出と重複した場合には外国抄録のキーワードを捨てゝある。図でも分るようにデルタ関数的な形をとっている。

ＤＳ：ＧＡ，ＡＢ，とのキーワードをデータシート上に書き込み，入力カードを作成したものである。ここで曲線上からはＧＡのカーブとほぼ一致する。ただしここでＧＡを加えた結果横軸方向に主軸が約4語のずれを生じている。ＤＳで語数２５附近での立上りはＧＡを加えたためと２５語以上を打切ったために２５語附近に集約されたことに起因する。

表Ⅱ－４－１０　作業者別キーワード語数抽出状況

語数＼作業者	A	B	C	D	E	F	G	合計
5				2				2
6		2		3	1			6
7		3		4			2	9
8	3		1	2	2			8
9	2	1	1	2	4	3	1	14
10		2	1	2	2	3	3	13
11	1	1			1	2	2	7
12	3	2			1		3	9
13	1				2		1	4
14		2			2		2	6
15	2					3	5	10
16	1					1		2
17	1	1					2	4
18						1		1
19	1					2	1	4
20								
21								
22								
23		1						1
合計	15	15	3	15	15	15	22	100
平均	12.3	11.0	9.0	7.8	10.3	14.1	12.9	

4.5.2　データの長さが不定のもの

(1)出願人名

インプットの段階で，最大４６字とし，それ以上のものは省略形式でデータを作成したので統計はとらなかった。

(2)発明の名称

発明の名称の字数については件数が１００件と少いので統計はとらなかった。

(3)キーワードの字数

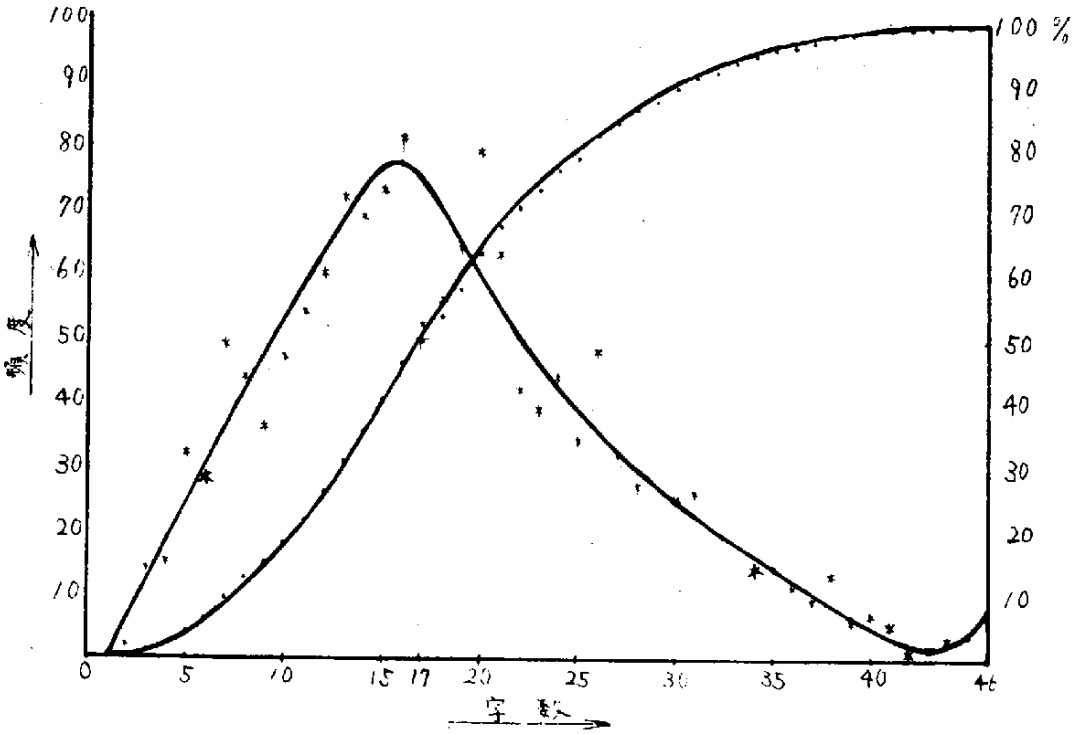
キーワード１４６３語について，各語の長さ（字数）の頻度の統計をとった結果を表Ⅱ－４－１１図Ⅱ－４－１０，１１で示す。

日本特許3.5.1で述べたごとく，日本特許の場合は９字が最高となっているのに対し英語の場合は16字が最高となっている。また日本語の場合は３６字以上は減衰しているのに対し英語では４６字で立上りをみせている。これは，英語の場合は４６字以上を切り捨てたためであろう。したがって英語の場合は日本語の場合よりも字数を増やす（推定５４字）必要が有ると思われる。包含率については英語，日本語共ほゞ一致している。

表Ⅱ－４－１１　米国キーワード字数統計

字数	頻度	累計	包含率	字数	頻度	累計	包含率
1	0			26	48		
2	2			27	32		
3	14	63	4.3	28	27	155	79.1
4	15	(4.3%)		29	28	(10.6%)	
5	32			30	25		
6	28			31	26		
7	49			32	15		
8	44	204	18.2	33	21	90	95.3
9	36	(13.9%)		34	14	(6.2%)	
10	47			35	14		
11	54			36	11		
12	60			37	9		
13	72	328	40.6	38	13	46	98.4
14	69	(22.4%)		39	6	(3.1%)	
15	73			40	7		
16	81			41	5		
17	52			42	1		
18	56	332	63.3	43	2	15	99.4
19	64	(22.7%)		44	3	(1.0%)	
20	79			45	4		
21	63			46	8	8 (0.6%)	100.0
22	42			計	1463	1463 (100%)	
23	39	222	78.5				
24	44	(15.2%)					
25	34						

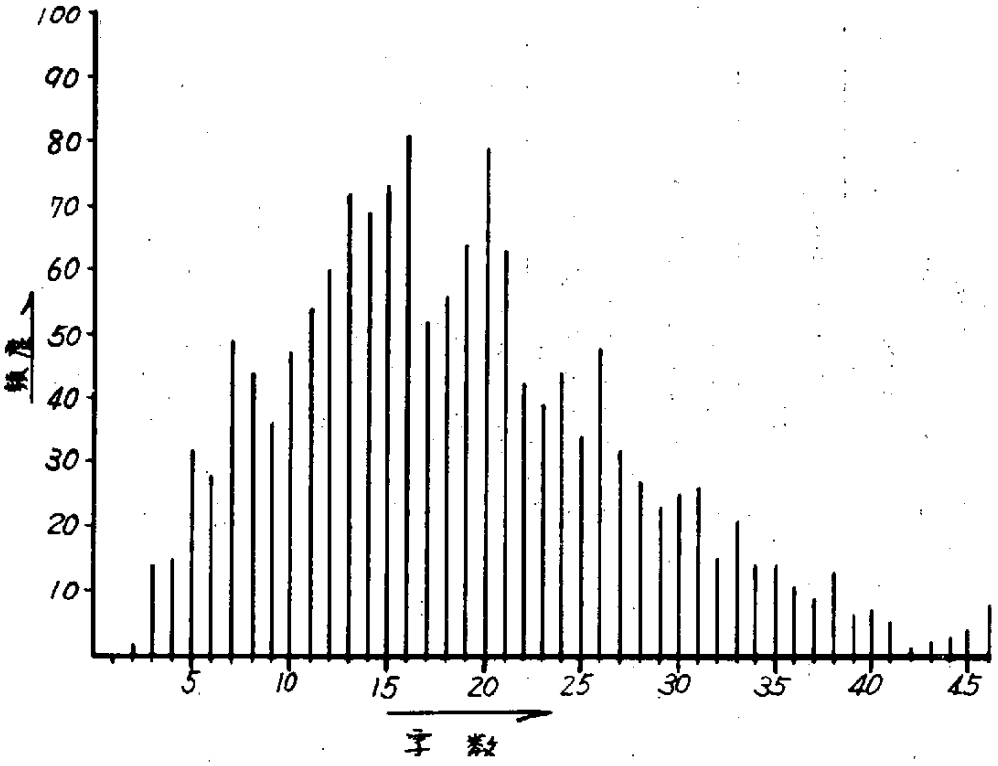
※．{完結－1
　　未結－7



図Ⅱ－４－１０　英語の字数頻度

注　(A)　字数頻度

(B)　包含率



図Ⅱ－4－11　英語の字数頻度

4.5.3　シソーラス

4.4 の検索実験の項で述べたように，蓄積のためのシソーラスおよび，質問作成に必要なるシソーラスの双方を作成する必要があるが，まずシソーラスをどのようにして作成するかが問題である。この項では，今回の実験で得たデータ中，シソーラス作成に関係のあるものについて述べる。

(1)キーワードリスト

1463件のキーワードを日本特許と同様の順序（表Ⅱ－3－29参照）に配列してリストを作成し，同時に，同一語についての頻度を積算した。

ヒント	キーワード (FROM USA PAT 67)	ヒント	キーワード (FROM USA PAT 67)
001	INERT ELECTRODES	001	ION PRODUCING METHOD
001	INFORMATION COUPLING ARRANGEMENT	001	ION SOLUTION
002	INFRARED RADIATION	001	ION SOURCE
001	INFRARED RAYS	001	IONIZABLE LIQUID STREAM
001	INHOMOGENEOUS MIXTURE	001	IONIZABLE SUBSTANCE
001	INITIATOR TRAIN	001	IONIZATION
001	INJECTOR ELECTRODE	006	IONIZED GAS
001	INLET MEANS	001	IONIZED*HIGH-TEMPERATURE GAS
001	INPUT MEANS	001	IONS
001	INPUT SOLID STATE MEANS	001	ISOLATOR CRYSTALS
001	INSULATING COATING	001	ISOTHERMAL EXPANSION STAGE
002	INSULATING MATERIAL	001	JOINING ENDS
001	INSULATING RING	001	KINETIC ENERGY
001	INTERCONNECTED FILAMENTARY NETWORK	008	LASER
001	INTERMETALLIC BOND	001	LASER BEAM
001	INTERMETALLIC SUPERCONDUCTORS FORMING METHOD	001	LASER CAVITY
001	INTERNAL PERMANENT SUPPORT	001	LASER CHAMBER
001	INTERNALLY SUPPORTED THERMOPILE ELEMENT	001	LASER DEVICES
001	INTERROGATION	002	LASER ELEMENT
001	INTRINSIC SILICON	001	LASER EMISSIVE LIGHT ENERGY
001	INVESTIGATION	001	LASER LIGHT
001	ION CHAMBER	001	LASER LIGHT OUTPUT
001	ION DISPENSING APPARATUS	001	LASER MATERIAL QUALITY
001	ION GENERATION APPARATUS	002	LASER OSCILLATOR
001	ION JET	001	LASER PUMPING

図Ⅱ－4－12　キーワードリスト

その結果得られたリストの一部を図Ⅱ－4－12に示す。なお，同一語を整理した結果，　1463語が約700語となり約½に圧縮された。この程度のリストでも質問語を作成する場合は非常に有効である。

5.　日本語，英語併用による検索システム

3.および4.において，日本語による検索システム，および英語による検索システムについて述べたが，国際的に情報の交換が盛んに行なわれている今日，日本語だけ，英語だけのシステムでは色々と障害があるので，いずれの言語でも自由に使用できるシステムを開発するのが好ましい。

特に，原資料が，英語と日本語の二種のものがあり，その内容が若干相違しており，両方からキーワードを抽出したい場合や，でき上ったものを，別の言語のファイルとして焼き直したい場合の外，両国語を混合してデータファイルを作成したい場合（技術用語は英語のままの方が使い易い場合もある。）など，両国語が，コンピュータの中で，自由に変換できれば非常に便利である。この具体的例をあげて次に説明する。

5.1　日本特許への応用

4.1.1「日本特許の英語ファイル」の項で述べたように，日本特許のうち，その英文抄録が発行されている分については，和文キーワードと対になる英文キーワード数は，8851で，そのリストは図Ⅱ－4－2に示す通りである。

このデータ数では，まだ不足であるが，これを累積し，単語における日本語と英語との関係を究明し，辞書ならびに変換の規則を作れば，英語と日本語との相互の翻訳が可能となる。キーワードは，単語であるので，通常の文を対象とした自動翻訳に較べてこの作業は遙かに容易である。

この辞書が完成すれば，日本特許のファイルを機械で自動的に英語ファイルに変換できるので，国際的なデータ交換の場合に非常に有益であろう。

5.2　米国特許への応用

4.2.1「原資料」の項で述べたように，米国特許のキーワード抽出材料として使用した資料は，英語で記載されている米国特許公報の外に，日本語で記載されている外国特許抄録誌をも使用した。外国特許抄録誌は，日本語で記載されているので，キーワードの抽出は容易であるが，そのキーワードを英訳しなければならない。現在は，人が辞書によって翻訳しているが，この作業は大変である。

しかし，上記5.1で述べた英語－日本語の辞書が完成すれば，この作業は機械が自動的に行なってくれる。

この翻訳工程を図で示すと，4.1.2「米国特許の英語ファイル」の項に掲載されている図Ⅱ－4－

3のうち，点線で示されたようになる。

5.3　シソーラス作成への応用

対になる英語と日本語のキーワードを，日本語を中心として配列すれば，同一の日本語に対応する英語が集まり，またその反対に，英語を中心として配列すれば，同一の英語に対応する日本語が集まる。この表の一例を図Ⅱ－5－1(1)（和→英キーワードリスト），図Ⅱ－5－1(2)（英→和キーワードリスト）で，それぞれ示す。

このリストでわかるように，「コンデンサ」に対応する英語は，

CAPACITOR	4
CAPACITORS	2
CONDENSER	48
CONDENSERS	4
CONDENSOR	1
CONDENSORS	1

の6つである。この表において，頻度の多い用語，すなわち，「CAPACITOR(S)」，「CONDENSER(S)」は，「コンデンサ」に対応する通常使用されている用語であり，非常に頻度の少い用語「CONDENSOR(S)」は，極めて例外的に使用されている用語，ないしは，データまたはデータ作成上のミスであろうということが推定される。

したがって，このリストを作成することによって，

① 日本語と英語の対訳関係を知ることができる。

② 同義語，同類語が集められる。

③ 同音異議の用語の識別ができる。

④ 表記法の差（例えば，単数，複数）が発見される。

⑤ データのミスが発見される。

などの効果がある。

今までは，日→英キーワードリストによる英語キーワードの処理について述べたが，英→日キーワードリストを使用すれば，上記と同様の項目について，日本語キーワードの処理を行なうことができる。

このように，上記のリストを上手に利用すれば，シソーラス作成の基礎データを機械的に作成することも可能であろう。

ｺﾛｶﾞﾘ ｾﾂｼｮｸﾗ	ROLLER CONTACTOR
ｺﾛﾅ ﾚｹﾝﾎｰ	CORONA TESTING METHOD
ｺﾝｺﾟｰ ｶｸﾊﾝ	MIXING STIRRING
ｺﾝｺﾟｰ ｷﾀｲ	MIXTURE GAS
ｺﾝｺﾟｰ ｹﾝﾊﾟ	MIXING DETECTING
ｺﾝｺﾟｰ ｹﾝﾊﾟ	MIXING DETECTING
ｺﾝｺﾟｰ ﾌﾝｾｲ ﾊﾝﾉｰ	MIXED PULVERRIZED ?
ｺﾝｺﾟｰｻ	MIXING DEVICE
ｺﾝｺﾟｰｻ	MIXING DEVICE
ｺﾝｻﾍﾞｰﾀ	CONSERVATOR
ｺﾝｾｲ ｶﾞﾀ ﾎﾞｺｰﾀﾞ	MIXTURE TYPE VOCORDER
ｺﾝﾀｸﾄ	CONTACT
ｺﾝﾀｸﾄｽｲｯﾁ	CONTACT SWITCH
ｺﾝﾃﾞﾝｻ	CAPACITOR
ｺﾝﾃﾞﾝｻ	CAPACITOR
ｺﾝﾃﾞﾝｻ	CAPACITOR
ｺﾝﾃﾞﾝｻ	CAPACITOR
ｺﾝﾃﾞﾝｻ	CAPACITORS
ｺﾝﾃﾞﾝｻ	CAPACITORS
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER

図Ⅲ－5－1(1)A　和→英　キーワードリスト

ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER

図Ⅱ－5－1(1)B　和→英　キーワードリスト

ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSER
ｺﾝﾃﾞﾝｻ	CONDENSERS
ｺﾝﾃﾞﾝｻ	CONDENSERS
ｺﾝﾃﾞﾝｻ	CONDENSERS
ｺﾝﾃﾞﾝｻ	CONDENSERS
ｺﾝﾃﾞﾝｻ	CONDENSOR
ｺﾝﾃﾞﾝｻ	CONDENSORS
ｺﾝﾃﾞﾝｻ ｷｺｰ	CONDENSER MECHANISM
ｺﾝﾃﾞﾝｻ ｺｰｾｲ ｶﾞﾗｽ	CONDENSER FORMING GLASS

図Ⅱ－5－1(1) C　和→英　キーワードリスト

MAGNETIC DISTORTION PHENOMENON	ｼﾞｶｲ ｹﾞﾝｼｮｳ
MAGNETIC ENERGY	ｼﾞｷ ｴﾈﾙｷﾞｰ
MAGNETIC FIELD	ｼﾞｶｲ
MAGNETIC FIELD	ｼﾞｶｲ
MAGNETIC FIELD	ｼﾞｶｲ
MAGNETIC FIELD	ｼﾞｶｲ
MAGNETIC FIELD DETECTING SYSTEM	ｼﾞｶｲ ｹﾝｼｭﾂ ｼｽﾃﾑ
MAGNETIC FIELD FACTOR	ｼﾞｶｲ ｾｲﾌﾞﾝ
MAGNETIC FIELD TYPE LENS	ｼﾞｶｲ ｶﾞﾀ ﾚﾝｽﾞ
MAGNETIC FIELD* ?	ｼﾞﾊﾞｷﾌﾖｳ
MAGNETIC FILM	ｼﾞｾｲﾏｸ
MAGNETIC FILM	ｼﾞｾｲ ﾊｸﾏｸ
MAGNETIC FINE PARTICLE	ｼﾞｾｲ ﾋﾞﾌﾝ
MAGNETIC FLUX LEVEL	ｼﾞｿｸ ﾚﾍﾞﾙ
MAGNETIC FLUX VARIATION	ｼﾞｿｸ ﾍﾝｶ
MAGNETIC FLUX VARIATION DELAY	ｼﾞｿｸ ﾍﾝｶ ﾁｴﾝ
MAGNETIC FLUXES	ｼﾞｿｸ
MAGNETIC HEAD	ｼﾞﾘｮｸ ﾍｯﾄﾞ
MAGNETIC HEAD	ｼﾞｷ ﾍｯﾄﾞ
MAGNETIC HEAD	ｼﾞｷ ﾍｯﾄﾞ
MAGNETIC HEAD	ｼﾞｷ ﾍｯﾄﾞ
MAGNETIC HEAD	ｼﾞｷ ﾍｯﾄﾞ
MAGNETIC HEAD	ｼﾞｷ ﾍｯﾄﾞ
MAGNETIC HEAD	ｼﾞｷ ﾍｯﾄﾞ
MAGNETIC HEAD*ABRASION	ｼﾞｷ ﾍｯﾄﾞ*ﾏﾓｰ

図Ⅱ－5－1(2)　英→和　キーワードリスト

6. 実験用システムの評価と問題点

自然語による広域検索システムの可能性を追及するために開発したこの実験用システムは，一応その目的を達したものといえよう。

すなわち，自然語による検索の精度，またその精度を高めるために用いた手段等いずれも，初め予想した効果を実験的に裏付けることができ，さらに，実験用システムとして設けたシステムコントロール上の各種のパラメータは，今後，このシステムを使用して，さらに，上記の問題を追及する上においてきわめて有効であると思われる。

また，試験的に蓄積したデータからえられた統計資料は，今後，この種のシステムを設計する上において十分参考になるデータであり，特に，日本語－英語のキーワードリストは，シソーラス作成の用語処理に糸口を付けたものとして，高く評価されるべきであろう。

検索システムを評価する上において重要な因子である検索時間は，相当複雑なマッチング方法，ならびに評価の計算をしている割合には，かなりスピードが早いとはいえ、数万件，数十万件のデータを対象とする場合にはかならずしも早いとはいえない，このために，この実験用システムでは，１０質問まで同時処理を行うことができるようにしたが，シーケンシアルサーチを行なっている限りこの問題は解決されないものと思われる。したがって，今後の問題としては，ランダムサーチの方式を確立すべきであろう。

次に，今回の実験において，特に留意すべき点は，システム分析に使用したデータ数の絶対量ならびに，実験回数の少なかったことである。

蓄積件数３０００件とは，用語の特性を検討するためには，ある程度有効であったが，検索精度を追及するためには，技術領域が広い割には少ないので，同一技術に関する文献数が少なく，これらの問題を統計的に処理するには不十分であった。したがって，今後，さらに蓄積データ数を増やして実験する必要を痛感している。

検索質問の作成の仕方は，他の検索システムにおいても同様であるが，そのシステムの良否を左右する重大な問題である。このシステムにおいても，実例をもって説明したように，一見同じような質問の仕方でも，その結果は非常に異ってくる場合がある。特に，このシステムでは，評価のためのパラメータの設定が，重要で，実験回数が少なかったため，質問語のウエイトと論理式との関係を十分実験的に解明できなかったことは残念で，今後さらに検討すべき事項である。

その後，データ作成，カナ化，シソーラス作成の問題点等，数多く残されているが，これらについては，他の項において述べたので，ここでは割愛する。

Ⅲ　汎用情報検索システムへの展開とその方針

Ⅲ. 汎用情報検索システムへの展開とその方針

1. シソーラスとデータの作成

この実験で行なった情報検索システムによる結果を分析し，本来のシステム全体である汎用情報処理サービスシステムへ応用発展させる基礎にしたい。

今回の実験で特に痛感したのは，データの自動処理および質問時点におけるシソーラスの必要性である。本来のシソーラスという言葉の情報管理に使われている意味は類語集であるが，ここではさらにその意味を拡張させて同類語ファイル，同義ないしは同意語ファイルおよび関連語ファイルとしたい。今回の実験ではあらかじめシソーラスなどが完備されていなかったので，抄録から手作業によってソースデータを作成し，検索時には比較的好結果を得たが，ソースデータ作成時の時間，労力，経費を十分少なくするという点，および作成されたソースデータが複数の人の手によって行なわれたためのバラツキなど，今後の研究に待つところが多い。データ作成作業の機械化としては，日本語の表現形態をカナなどの中間的状態に変換せずに，直接コンピュータに導入して処理する手法を開発しなければならない。

この足掛りとなる手法としては，当面は漢字タイプによるインプット方式，第2にはＯＣＲによる方式，さらに進んでは音声による方式などがあげられる。

次にシソーラスについては，ある程度の少数のデータを基本として他のデータを自動的に処理してゆく方式をとる方がよい。このとき，先に述べたように，シソーラスとしては単なる同類語的ファイルのみでなく，さらに機能的な構造を有する関連語ファイルなどを用意しなければならない。この際今回の実験で着眼し，実施した日本語データの英訳ファイル，および対照米国特許ファイルのデータを相互に関連使用して，自動的にシソーラスの基本を作成することも新しい試みとして有効な手法と思われる。さらに，今回の実験ではシステムのアルゴリズムを試行錯誤的に解明する方式をとったため，経済性，検索速度などはある程度無視して，システムの信頼性，使用上の簡便性などに重点を置いた。したがって，ファイル構成も，シーケンシアルファイルを主体とした。しかし，汎用性のあるシステムとして発展させるためには，検索の速度の向上，システムの経済性などを考慮して，先づ現在のシステムアルゴリズムをそのまま生かしたランダムアクセ

スファイルを作成し，大容量のデータに対処する必要がある。

2. 現在のシステムのアルゴリズムに加える数学的手法

これまでのシステムは一般に言語の物理的な対比にのみ主眼があり，それに，質問テーマの概念を現わす意味論的構成要素としての論理式を加えた手法だけでは，ソースデータが原データに比して非線型に次元が歪められていることもあって，文章集団によって表現される意味集合をとらえることは非常に難しい。

そこで本実験システムでは，キーワードの表わす意味論的機能を対比アルゴリズムの中に取り入れ，分ち書きの部分対比の評価判定を行なう新しい手法の導入を試みた。その検索結果は従来より比較的良かったが，このシステムも本質的には，従来の文献の意味を表わす文章構成要素としての言葉の物理的取り扱いが基本になった上での意味論的取り扱いと変りなく，完全にデータ作成時から意味論的な要素を主体としてのアルゴリズムの基礎の上には立っていない。そこで，シソーラスの作成およびその作業の自動化にともないデータ処理も自動的に数学的手法と意味論的考えを合わせた手法による検索アルゴリズムを確立する必要性がある。例えば，自動分類によるデータ分布表とかデータ集合の情報空間処理など，さらには現在研究されている自動適応系の理論を取入れて生かした方法に移行するようになるべきであると思われる。

3. 情報処理実行時におけるコンピュータと人間との関連

システムを汎用性のあるものにするために必要なもうひとつの大事なことは，実行事において機械と人間つまりハードとソフトとの間で随時，調整制御できるようにすることである。

これは一般にマンマシン応答コントロール方式と呼ばれているが，あるテーマにしたがって情報処理している過程でパラメータの交換ないしは変更が行なえるようにしておけば，質問の内容を中間段階の検索の結果と対比しながら質問の主旨にそって変えて行くことも可能であり，ちょうど人間が文献調査を行なう状態とほぼ同じ状態で検索できるので，信頼感のおける回答を得ることができるものと考えられる。

したがって，このためのシステムはその構成されているシステムが外部から容易に調整制御可能なものでなければならない。その試　の第一歩として今回の実験システムでも，その機能を附加してみた。

むすび

情報の氾濫は，適確な処理技術が確立されていなければ，情報の価値を損ない返って利用者と情報の関係を阻害する。

特許情報についても同様，過大なる情報量は，その物理的処理を困難にするばかりでなく，権利に直接つながっているので，特許権の所在，実態が把握できないために混乱をきたし，ひいては，特許制度そのものの危機にもなりかねない。

したがって，いかにして，大量の情報を適確，迅速に処理し，必要とする情報のみを，必要とする時点に，利用者に提供するかを，情報関係者のみならず，一般利用者も均しく切望するところである。

今回の研究では，大量の特許情報を処理するための一つの試みとして，自然語による広域検索システムの開発を行ない，併せて，日本語と英語との関係を究明した。

実際に，研究に着手して見ると，予期しない問題点が数多く発生し，その一つ一つについては，未だに完全に究明されていないものも多々あるが，問題点を提起したこと自体に意義があるものと考えられる。

昭和45年3月　発行

発行所　財団法人　日本情報処理開発センター
東京都港区芝公園21号地1番5
機械振興会館内
TEL（434）8211（代表）

印刷所　ジャパンプランニング株式会社
東京都中央区京橋2丁目13番地
貴仙ビル2号館
TEL（567）0801～3

44 S002