

50-S001

日本語情報処理システムの研究開発

昭和 51 年 3 月



財団法人 日本情報処理開発協会

この報告書は、日本自転車振興会から競輪収益の一部である機械工業振興資金の補助を受けて 昭和 50 年度に実施した「日本語情報処理システムの研究開発」の成果をとりまとめたものであります。

序

情報化社会の進展に伴って経営者が常に安定した企業経営を維持していくためには、経営のための関連情報を十分に活用して適格な意思決定を行なう必要がある。このため内外の企業・研究機関においてはいわゆる経営情報システムに関する研究開発が活発に行なわれている。当財団においてもコンピュータを利用した合理的な経営情報システムを確立するという観点から、あらゆるレベルの管理者に理解しやすい情報を提供するための情報処理システムについて調査研究を実施してきた。その一環として、48年度以来、我が国ではコンピュータからの出力情報がアルファベットやカタカナで表示されるため読みにくく、意味にあいまいさがあるなど理解度に難点があるとされてきた文書情報をとりあげ、カタカナ入力漢字かな混り文出力方式の情報処理システムを開発し、事例研究を通して実用性の検討等を行なってきた。

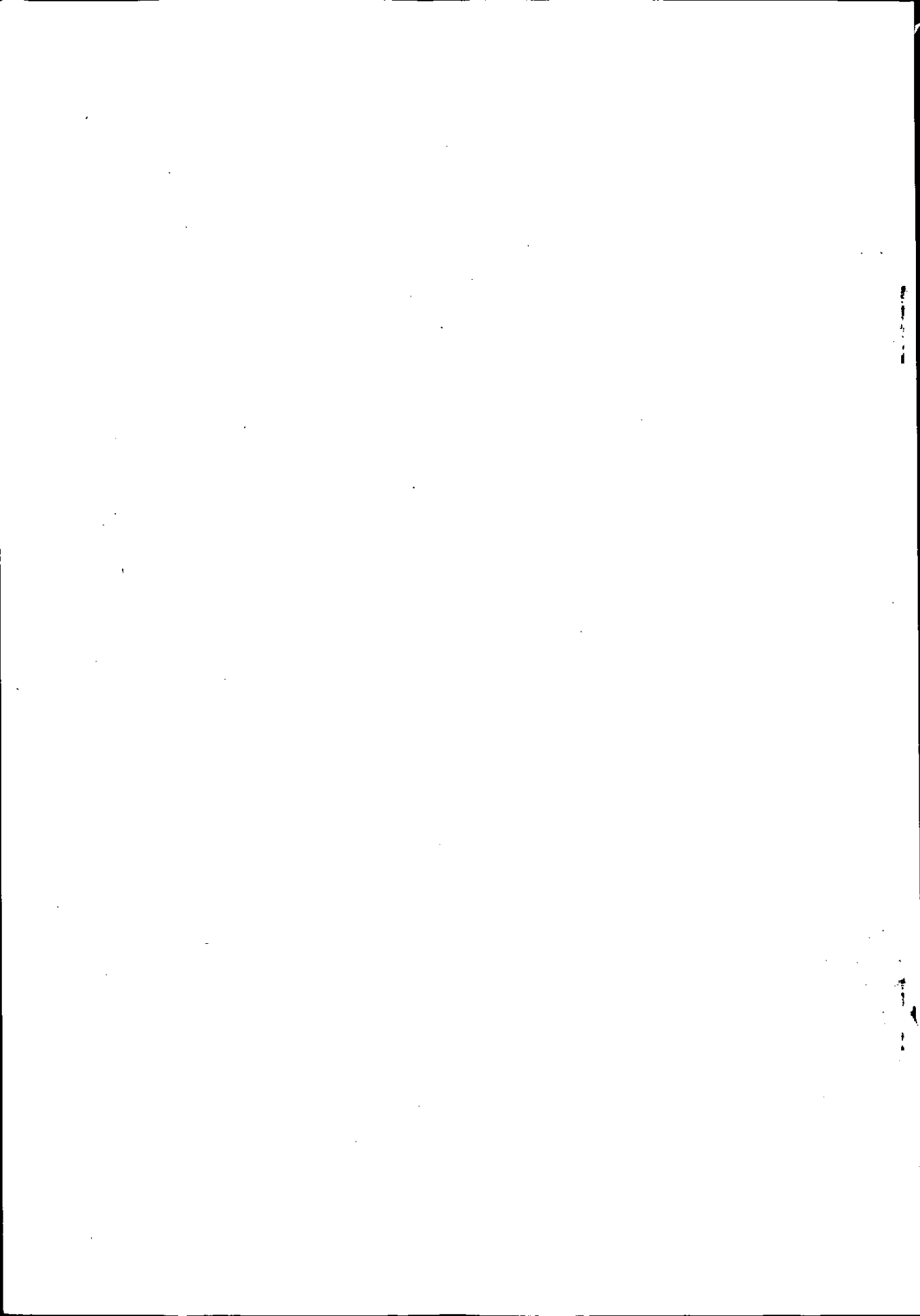
本年度はこれを更に進め、コンピュータを用いた日本語情報（和文情報）の取り扱い、特に情報の蓄積・検索等について検討すると共にカタカナ入力漢字かな混り文出力方式を応用した日本語情報検索システムのモデルを開発し、科学技術文献の検索に適用して2年計画で事例研究を行なおうとするものである。

本報告書は昭和50年度に実施した日本語情報の検索技術等に関する検討とカタカナ入力漢字かな混り文出力方式を応用した日本語情報検索システムの設計を中心にとりまとめたものである。

ここに、今年度の研究開発にご尽力ならびにご支援を賜った高橋達郎、植村俊亮、後藤榛男、上田陸奥夫の各氏および関係各位に心から感謝の意を表しますと共にこの報告書が各方面に活用され、我が国情報処理産業の発展の一助として寄与できれば幸いです。

昭和51年 3月

財団法人 日本情報処理開発協会
会長 植 村 甲 午 郎



日本語情報処理システムの研究開発

目 次

1. 総 論	1
1.1 研究の目的	1
1.2 報告書の概要	5
2. 日本語情報検索の技術動向	7
2.1 英数カナ・モードと漢字モード	9
2.1.1 SDI と RS	9
2.1.2 情報検索と索引誌	10
2.1.3 電算植字システムとの関連	12
2.1.4 情報検索システムにおける字種	12
2.1.5 両モードの取扱い	14
2.1.6 両モードのデータの比較	16
2.2 日本語情報検索のデータベース	21
2.2.1 データベースの設計	21
2.2.2 わかち書き, 漢字字種	24
2.2.3 漢字コード	26
2.2.4 漢語処理	28
2.3 日本語情報のキーワード/シソーラスの管理	33
2.3.1 情報検索とシソーラス	33
2.3.2 日本語キーワード	35
2.3.3 日本語シソーラスのコンピュータによる用語管理	38
2.3.4 キーワードの自動チェック	46
2.4 バッチ検索とオンライン検索	49

2.4.1	文献検索の本質と両方式	49
2.4.2	オンライン検索	52
2.5	日本語情報検索のニーズと運用	57
2.5.1	文献検索のニーズ	57
2.5.2	文献検索システムの運用	61
2.6	日本語情報検索システムの開発動向	63
2.6.1	国立国会図書館の索引編集システム	63
2.6.2	電子技術総合研究所の木かげシステム	72
2.6.3	通産省情報管理課の日本語情報検索システム	81
2.6.4	大蔵省主計局の国会議事録検索システム	88
2.6.5	JICSTにおける日本語情報検索	92
3.	日本語情報の編集と出力	101
3.1	編集システムの必要性	101
3.2	出力装置と編集機能	110
3.3	編集システムの設計	113
3.3.1	システムの概要	113
3.3.2	システムの機能	115
3.3.3	システムの制限	121
4.	事例研究—日本語情報文献検索システムの設計—	125
4.1	システムの概要	125
4.1.1	システムの目的	125
4.1.2	システムの特徴	126
4.1.3	システム構成	127
4.1.4	ハードウェア構成	136
4.1.5	ソフトウェア構成	137

4.2	検索サービス・サブシステム	140
4.2.1	処理概要	141
4.2.2	処理ブロックの構成	144
4.2.3	処理ブロックの概略	146
4.2.4	ファイル概要	173
4.2.5	検索対象項目	176
4.2.6	検索コマンドとパラメータ	177
4.2.7	検索情報の入力形式	179
4.2.8	表示・出力のレイアウト	182
4.2.9	エラー・メッセージ	185
4.2.10	EBCDIC コードの扱い	186
4.3	データベース管理サブシステム	188
4.3.1	処理概要	188
4.3.2	処理フロー	190
4.3.3	データ作成	194
4.3.4	データ格納	200
4.3.5	ポスティング・テーブルの作成	201
4.3.6	データベース・メンテナンス	202
4.3.7	ポスティング・テーブル・メンテナンス	205
4.3.8	ファイル概要	206
4.4	キーワード管理サブシステム	209
4.4.1	処理概要	209
4.4.2	処理フロー	211
4.4.3	キーワード・テーブルの作成	217
4.4.4	インバーテッド・ファイルの構成	219
4.4.5	キーワード・テーブル・メンテナンス	225
4.4.6	ファイル概要	226

4.5	索引誌作成管理サブシステム	231
4.5.1	処理概要	231
4.5.2	処理フロー	233
4.5.3	著者名索引	236
4.5.4	キーワード索引リスト	239
4.5.5	キーワード統計リスト	241
4.5.6	ファイル概要	243
4.6	システムの利用法	245
4.6.1	端末の構成	245
4.6.2	漢字ディスプレイ装置の盤面配列	246
4.6.3	各種キーの説明	247
4.6.4	表示・出力の様式	248
4.6.5	検索コマンド	250
4.6.6	データ入力方法	252
4.6.7	検索の方法	255
4.6.8	エラー・メッセージ一覧表	264
4.7	サービス・プログラムの概要	265
4.7.1	TSSP	265
4.7.2	ETAM	266
4.7.3	KDP制御プログラム	267
4.8	付 録	268
5.	今後の課題	297

1. 総論

1. 総 論

1.1 研究の目的

当財団では、昭和46年度の事業「データ・マネジメント・システムの研究」の成果としてとりまとめた報告書「記事情報検索のためのデータ・マネジメント」において、カナ漢字変換の事例研究をとりあげて以来、日本語情報処理技術に関する調査研究を推進してきた。従来わが国ではコンピュータからの出力情報が英字、数字あるいはカタカナで記述されていたために、情報を利用する者にとっては読みにくく、意味にもあいまいさがあるとされてきた。そのため経営情報などの意思決定には正確な情報伝達がしにくばかりか、場合によっては重大なミスにつながる誤解を与えないとも限らない。

そこで当財団では、コンピュータからの出力情報を漢字かな混り文で提供するための調査研究に着手し、カタカナ入力方式による日本語情報処理システムの研究開発を行なってきた。これまでに各メーカーや研究機関で研究開発されている各種漢字情報入力方式について調査検討したが、字種が多く、文字パターンも複雑で、類似した文字の多い漢字を入力するための合理的な入力機器はいまだに開発されていないといえる状態である。市場にもっとも多く実働している漢字けん盤の入力方式も、規模が大きく、操作も複雑でとても一般の人に適用することは困難である。しかも操作を行なうオペレータの訓練にも長期間を必要とするなど不十分な面が多く、その後も各機関において試行錯誤を繰り返して研究が進められている。このようにわが国における日本語とくに漢字情報をコンピュータで処理するには、日本語文の合理的な入力方式を確立することがもっとも重要な課題であることを認識するに至った。こうした技術動向を背景として、カタカナ入力方式に関する実験研究を行なったが、当財団で考案した方式は、日本語情報をカタカナけん盤を使って入力操作を行ない、カタカナ・コードで情報の蓄積などの処理をして、出力をカタカナ漢字コード変換処理に

よって漢字かな混り文で提供するものである。このカナ漢字変換システムの特徴としては、カタカナから漢字に変換するアルゴリズムを簡略化するために、変換テーブル（カナ漢字変換辞書）の構成を専門分野別に持つこと、および漢字単語の区切り、いわゆる分ち書き単位を原則として2字にしたことである。またカタカナ入力方式で問題となる同音異義語の判別は、カタカナ・データ作成時に識別情報を作成者が付加する方法をとって、内部での変換処理を容易にしている。しかしながら、登録単語の不足などで不変換がいくらか生じることには止むを得ず、その対処方法として漢字に変換できなかった用語を太字（アンチック体）のかな文字で出力するという工夫をしている。こうすることで、たとえ変換率が低くとも太字のパターンを認識することで理解度は十分効果をあげている。

いっぽう、日本語情報の処理に対する気運も増々高まりつつある現状であるが、これまでに見受けられる日本語情報（漢字情報）の適用分野といえば、ただ単に漢字データを作成して編集を加えた程度の特種な目的でしか利用されていない。その中で最も多く使われているのは、氏名、住所などのいわゆる固有名詞を扱う業務であり、これはほとんどが顧客サービスを目的としたものである。また出版・新聞業界等における電算植字という使われ方もあるが、いずれにしても情報処理という観点からは、あまり際立つ効果は現われていないように思われる。日本語情報の処理は、こうした一部の分野には適用できるが、それ以外の分野や業務には適用できないという性質のものではなく、コンピュータによる情報処理の出力結果として漢字かな混り文が要求されるものや、ローマ字やカタカナに比べて漢字かな混り文がより効果的であるという分野や業務に対してはすべて適用することが出来よう。いろいろな視点から見て利用価値が大きいと想像される分野として、

- ① 文献検索などのように大量のデータを迅速に読んだり、照合したり、検索したりする必要のあるもの
- ② 顧客サービスの向上につながるもの
- ③ 読み誤りをなくし、正確性を重視するもの

- ④ ローマ字やカタカナではまぎらわしいもの
- ⑤ 氏名・住所など固有名詞に漢字表記が望まれるもの
- ⑥ 同一データを効率よく何度も使用されるもの
- ⑦ 現在 EDP化されていながらも、出力結果を漢字かな混りにするために手書きによる再加工をしているもの

⑧ 漢字でなければならぬために人手によって処理しているもの

など、その処理の内容が省力化、正確性、迅速性、サービス向上につながる業務について効果的な適用が可能である。その代表的な適用分野としてあげられるものが日本語文の情報検索であろう。

今日のように企業を取りまく環境がつぎつぎに変化したり、技術革新の進歩がはげしい時代に、企業が外部から取り残されることなく環境の変化に順応し、時代とともに歩んでいくためには、企業をとりまく内外の情報をいかに効率的にコントロールしていくかが重要な問題であり、そのため企業活動としての情報管理をどうするかという事が重要視されるのは勿論のことである。その手段として、コンピュータを利用した情報検索が考えられ、内外の情報を収集・整理して貯えておき、必要に応じて問合せに対する情報を探索し、適格な情報を提供するところに情報検索システムとしての意義があるものと思われる。しかし、これまでコンピュータを利用した情報検索といえば、検索結果を英数字あるいはカタカナでコンピュータから出力し、提供するのが一般的であり、漢字かな混りの日本語文で提供されるのは索引情報を得てからの原文参照あるいはコピーにすぎなかった。これでは、せっかく大量のデータベースの中から利用者の要望する情報が提供されたとしても、英数字やカタカナばかりで印字されたものでは利用者の理解度は半減し、しかも価値ある情報を得たという満足度に関しても十分な効果はあげられないであろう。そのため、ここに漢字かな混りの日本語情報を適用することによって、

- ① 利用者が読みやすい。
- ② 文脈がはっきりする。

③ 学術名や技術用語などまぎらわしいものが明確になる。

④ 読み誤りをなくし、正確度を向上できる。

などのメリットが得られることになる。

以上のようなことから、当事業ではこれまであまり研究開発がなされていなかった日本語情報検索の動向を調査して、その技術的問題点を究明するとともに、漢字かな混りの文献情報を扱った検索システムのモデルを開発して、その効果や実用性等に対する評価を実験結果から得ようとするものである。

なお、当事業は2ヶ年計画で実施し、本年度は技術的問題点に関する調査検討と、モデル・システムの基本設計を行なっている。

1.2 報告書の概要

当財団では、昭和43年度以来経営情報システムに関する調査研究を推進してきたが、昭和46年度の「データ・マネジメント・システムの研究」においてカナ漢字変換の事例研究を開始して以来、46-S006「記事情報検索のためのデータ・マネジメント」、47-S001「階層構造の情報処理システム」、47-R001「日本語情報処理の技術動向調査報告書」、48-S004「文書情報処理に関する調査研究」、49-S006「文書情報処理に関する研究開発」、といった日本語情報処理に関する一連の報告書を取りまとめている。

今年度の報告書は5つの章からなっており、各章でそれぞれつぎのような内容について記述している。

第1章「総論」では、研究の目的を記述して、カナ漢字変換システムを応用した日本語情報処理システムに関する研究事業の背景を明らかにしている。

第2章「日本語情報検索の技術動向」では、日本語情報処理システムの事例として情報検索を設計するのに必要な情報処理技術を考察するとともに、日本語情報を扱った検索システムの開発動向についていくつかの事例を紹介している。

第3章「日本語情報の編集と出力」では、日本語情報を処理した結果を出力装置にアウトプットする際に必要な編集機能および処理方法などについて概説している。

第4章「事例研究—日本語情報文献検索システムの設計—」では、漢字かな混りの文献情報を検索サービスするシステムの基本設計について全容を解説している。

第5章「今後の課題」では、日本語情報処理全般についての展望と、事例にとりあげた文献検索システムの実用性などに関する考察について述べている。

2. 日本語情報検索の技術動向

2. 日本語情報検索の技術動向

情報検索（ここでは主として文献情報検索）はいよいよ実用期に入り、特に各種情報機関から情報磁気テープの提供が開始されるに及んで、急速に普及段階に入りつつある。またここ2,3年文部省から巨大学術情報システムの構想が発表されて以来、大学内および大学間における、情報検索システムの実用化研究も活発化しつつある。

過去十数年にわたる情報検索の理論的および実用化研究の結果、情報検索技術においても、「できること」と「できないこと」がはっきりしてきた。とくにできないことの代表は、意味論、構文論などの自然言語処理の分野で、実用システムにおいて、これらの処理を組入れているものは皆無に等しい。したがって現在の情報検索システムは、「できること」をいかに実用上便利のように、また費用効果の面から採算が合うように組むかにかかっている。

ところがここに「できること」にもかかわらず、実用化が非常に遅れているものに日本語の問題がある。わが国のコンピュータ・システムがほとんど非言語情報処理であったために、日本語に対する関心が薄く、その方面の技術開発が遅れていたことが、最大の原因と考えられるが、情報検索のように言語情報を中心とする処理システムの実用化となると、いやでも日本語の問題と対決しなければならなくなる。

もっとも今日、日本語の問題が取上げられるようになったということは、これまで情報検索について多くの議論が行なわれてはきたが、実用システムはほとんど無かったということもできよう。検索技術はアメリカを中心として発達し、わが国においてもほとんどそれらのうけ売りとして発展してきた。したがって可能性のみに興味が集集中、いよいよ実用化の時期となって、思わぬ伏兵が現われたというわけである。

今日でもなお、アメリカのCAS (Chemical Abstracts Service) の発行するCAC (CA (Chemical Abstracts) Condensates) 情報テープで検索を行なっ

ているとか、カタカナで行なっているとかの発表が行なわれているが、学内とか研究所内とかの局所的な場合はともかく、汎用システムとしては疑問といわねばならない。特に企業内の汎用システムとなると、検索システムそのものを脅かすほどの大きな問題といえることができる。

しかし一方漢字かなまじりの日本語に対し、英語またはカタカナにもそれなりの利点があり、それぞれの利点を生かしてシステムを組むことこそ、実用システムといえるのではなからうか。以下これらの点に留意しつつ論を進めることにする。

2.1 英数カナ・モードと漢字モード

この両モードを比較するに当っては、2つの面から論じなければならない。1つはシステムの便利性、すなわちコンピュータ出力と人間との間のインタフェースの良否、もう1つはコンピュータ処理の効率面である。まず前者の問題を文献検索の面から考察してみよう。

2.1.1 SDIとRS

情報検索システムはその運営面からみて2つの方式、すなわちSDI (Selective Dissemination of Information: 選択的情報提供) とRS (Retrospective Search: 遡及検索) があることはよく知られているところである。これを機能的にみると、SDIは前もって登録されている情報要求(これを質問プロファイルという)に従って、定期的に検索を行なって回答するもので、質問の情報内容はそれほど深くなく、むしろ網羅的に関連文献を回答し、最終的な選択は利用者の通覧(ブラウジング)による判断にまかされている。これは網羅的な提供によって、利用者に新たなる着想の機会を与えることが、必要であるとの判断に基づくものである。

この場合通覧のための道具立てとして、一般には文献の抄録を提供するという方式をとる。したがって通覧をより容易にするためには、読み易さが最も重要な要素となってくる。英文の場合でいえば、プリントを美しくするとか、大小文字を用いるとか、イタリック体やゴシック体を加えるとかの方法が要求され、日本語の場合では当然のことながら、漢字かなまじり文が要求されることになる。ある企業の例でいうと、カタカナによる文献の標題のみでSDIを行なったところ、ほとんど利用されなかったという例さえある。このように読むための情報としては、漢字かなまじり文が必須条件であるみてよいであろう。

RSにおいては事情は異なる。RSではある情報要求が発生した時点において、その要求に全く一致した情報の検索が要求される。したがって質問の

情報内容は深く、厳しい再現率と適合率が要求される。（これを満たすには高精度の索引作業が必要となる）

このような検索においては求まる文献の数も少なく、またその内容も信頼性の高いものである。したがって回答としては原文献を提供すればよい。原文献を記述している言語がたとえばロシア語であったとしても、それは情報検索システムの範囲外ということで許容される。

なお RS においては上述のように高精度のシステムが要求され、しかもそれが索引作業の精度に依存しているわけであるが、この精度を保持することが極めて困難であるため、検索側で何らかのカバーをしようという気運に向いつつある。この1つの具体的な現われがオンライン検索であるが、これについては2.4で詳しく論ずることとする。

以上を要約すると情報検索における漢字モードの問題は、システムの運営に大きく依存していることがわかる。なおキーワードの漢字モードの問題は、キーワードのソーラスによる統制と大きなかかわりあいがあるテーマであるので、2.3において論ずる。

2.1.2 情報検索と索引誌

最近では情報検索というと、すぐコンピュータ・システムと結びつけて論じられがちであるが、従来からの索引誌との関連も十分に考察しておく必要がある。一言でいえば情報検索システムでは、データベースがコンピュータの磁気ファイルにセットされているのに対し、索引誌ではデータベースが人間可読の形で、印刷されて大量に利用者に配布されているわけである。

情報検索が磁気データベースのシーケンシャル・ファイルで実行されていた時代には、この両者の親近感はほとんど無かったが、最近のように RAM (Random Access Memory) を用いたインデックス・ファイルによる検索が定常化されてくると、結局同じ結果に落ち着いたということになってきた。つまり両者ともファイル構成の面からみて、ほとんど差異がなくなってきた。したがってここで再度両者の利害得失を比較してみるのも意義があることで

あろう。両者の比較を表 2-1 に示す。

このように索引誌方式の利点もいろいろあるわけであるが、索引誌は直接人間が利用するものであるから、その便利性からみて漢字モードであることが必須条件となる。

表 2-1 情報検索と索引誌との比較

	情 報 検 索	索 引 誌
ファイルの配置	集 中 型	分 散 型
コミュニケーション	カスタム型	マスコミ型
ファイル構成	シーケンシャル型	ランダム型
ファイル更新	ランダム型	
検 索	容 易	困 難
対 話 性	コンピュータ	利 用 者
通 覧 性	オンラインで可	あ り
利用者の満足度	困 難	容 易
アクセス・ポイント	バッチ：やや小 オンライン：大	中
即 時 性	多	少
簡 便 性	オンラインであり	あ り
SDI（選択的情報提供）	小	大
RS（遡及検索）	実用的、ただし量的負担大	抄録誌で対処
	投資の負担大	アクセス・ポイント に依存

一方、索引誌の編集は一般の文書とは異なり、膨大な量の情報の配列、同一情報の副出など困難な問題が多い。ところがこれらの問題はコンピュータ処理の側からみると得意な面が多く、漢字モードによる索引誌の編集はコンピュータ処理のメリットが大きい。事実わが国のコンピュータを利用した漢字処理は、索引誌の編集処理から始まっているのである。

以上を総合すると、情報検索と索引誌とは、検索という行為を実行するための手段の表と裏との関係にあり、一心同体をなすものである。つまり情報検索システムは索引誌編集機能を具備することによって、その能力を倍にす

ることができるのであり、この面からの漢字モード処理の考察を怠ってはならない。

2.1.3 電算植字システムとの関連

印刷技術は活字による凸版印刷から、写植によるオフセット印刷に移り、最近ではキーパンチ→コンピュータ編集→電算植字→オフセット印刷の時代に入りつつある。これは鉛公害や植字の熟練工の不足に起因し、コンピュータ処理のメリットの認識が高まるとともに、促進されてきた。

文献情報の分野においても、情報量の増大や印刷コストの上昇に伴って、一次情報（文献）の出版の段階からのコンピュータ処理への関心が高まっている。特にアメリカにおいては、米国化学会や米国物理学会などの大学が、真剣にこの問題に取り組みつつあり、一部は実施に移されている。もちろんこの場合単に一次情報の出版のみではなく、最初からその対象として二次情報ひいては情報検索をも包含して計画が立てられている。

わが国においては漢字処理のコストの面から、未だ一次・二次情報を一丸としたシステムの開発にまでは至っていないが、一部の情報機関においては、二次情報の電算植字システムが稼動しており、その産物として漢字モードの情報テープの入手が可能となってきた。今後文献情報および特許情報の分野において、この傾向はますます強まるものと思われる。

これにより漢字情報の入力の問題が解消され、漢字モードによる情報検索の気運は急速に高まるものと考えられる。この場合1つの大きな問題は、漢字を含めた日本語の標準化の問題であり、字種・コード・入出力機器の標準化が急がれている。

2.1.4 情報検索システムにおける字種

文献情報検索システムにおけるデータ要素は非常に多岐にわたり、20項目以上にも及んでいる。これらの中には英数カナモードで取扱えるものも多いので、日本科学技術情報センター（JICST）の抄録を例にとって、それぞれの要素を字種モード別に考察してみよう。

抄録年月日 (数)

抄録者コード (数)

原稿番号 (数)

科学技術部門コード (英数)

資料番号 (英数)

資料名 国内雑誌 (漢, テーブル付加)

外国雑誌 (英, テーブル付加)

発行年 (数)

発行国名 (英, ISO 準拠, テーブル付加)

巻, 号 (英数)

記事番号 (英数, コンピュータ付加)

記事区分 (英数)

使用言語 (英, ISO 準拠)

写・図・表・参の数 (数)

ページ (英数)

著者名 外国人 (英)

日本人 (漢)

日本人ふりがな (カ)

著者所属機関 外国 (英)

日本 (漢)

日本ふりがな (カ)

文献標題 外国 (英)

日本 (漢)

和訳 (漢)

日本語標題ふりがな (カ)

抄録文 (漢)

分類コード (英数)

UDC 標数 (数, テーブル付加)

キーワード (漢)

キーワードふりがな (カ, テーブル付加)

以上がファイル中のデータ要素であるが、これからも明らかなように、本質的に漢字モードで処理しなければならないのは、国内雑誌名、日本人著者名、日本の著者所属機関、日本語標題、抄録文のみであり、さらにこの中で前3者はカナまたはコードで実用上処理可能であり、結局のところ漢字モードが必須なデータ要素は、日本文で表記される文献標題と抄録文のみということになる。

なお JICST では情報提供磁気テープとして、漢字モードと EBCDIC モードの2種類のものを発行しているが、漢字モード・テープでは上記のデータ要素すべてを含み、EBCDIC モード・テープでは (漢) と記したもの以外のすべてのデータ要素を含んでいる。

2.1.5 両モードの取扱い

英数カナ・モードと漢字モードの取扱いにおいては、漢字モードでは字種が3千種にも及ぶため、英数カナ・モードの要素すべてを漢字モードで処理可能であるわけである。しかし実際の処理運営上では、処理の効率面からいろいろ工夫が必要となってくる。この問題は処理を入力処理、コンピュータによる内部処理、出力処理に分けて考察する必要がある。

まず入力であるが、上の例のようにデータがレコードごとに独立しており、また各レコード内での文字のモードが定まっている場合には、それぞれに適した入力方式を取るべきであろう。両モードの入力方式の最も大きな相異点は検孔方式にある。英数カナ・モードではダブルパンチによる検孔が容易であるのに反して、漢字モードでは、グラによる校正方式によるため一般にそれが困難であり、入力作業が極端に複雑となる。

このような理由から漢字モードのレコードも、英数カナ・モードによって入力しようとの考えが生れてくる。これには2つの方法があり、1つはコー

ドによって入力し、テーブルによる変換で漢字モードに置換えるもので、繰返し現われる固定的なデータに適し、他はカナ文字文（またはローマ字文）によって入力し漢字モードに変換する、いわゆるカナ漢字変換方式である。

カナ漢字変換方式は、単語単位の変換の場合は前のコード変換方式とそれほど異なるところはないが、文単位の場合には語の分割や分かち書きなどの問題があり、精度を高めるのは非常に困難である。この方式の実用性は、漢字モードによる直接入力とのコストの比較や操作性によって論ずべき問題でもある。また、カナ漢字変換方式は、コンピュータ処理の入力としてばかりではなく、高精度を要求されないそれ自身で完結する場合、たとえば、モノタイプの入力テープの作成や電文の翻字などの場合には、即実用性が期待できよう。

さらに将来の問題としては、高精度なテーブルと変換アルゴリズムが完成した場合には、複雑な数値計算と同様に、TSS (Time Sharing System) によるリモート処理の大きな応用面としてとらえることはできよう。

次にコンピュータによる内部処理の面であるが、両モードがレコードによって区分されていれば、両者が混在しても不都合は生じない。混在方式によればメモリは節約でき、またプログラム上有利な点も多い。一般にデータ処理は英数カナ文字のレコードに偏在し、漢字のレコードには少ないのでこの利点は助長される。ただし編集植字処理の場合には、編集上の多くのファンクションが存在し、また英数カナ文字といっても、いろいろな書体や大小文字、特殊文字などが混在してくるので、すべて漢字モードのコード系に変換して処理が行なわれる。

漢字モードによるコンピュータ処理において、その方式が確立していないのは漢字情報のソーティングである。これはとりもなおさず漢字の配列に関する正書法が確立されていないことである。英数カナの場合には、その配列原理は明確ではないが、配列順は固定しており、慣用かつ容認されているので、コンピュータ処理側でその取扱いについて考慮しなければならないこと

は、ほとんどないといってよい。しかし漢字の配列順はコンピュータ処理のシステム設計上での、1つの大きな問題となっている。

漢字の配列順としては、次のような方式が考えられる。

- (1) 読みの五十音順配列：一般に国語辞典に採用されている方式で、慣用配列といわれるが、出版社により配列規則に小差がある。入力時にふりがなを入れなければならない。国語辞典では漢語、かな語、カナ語を混在させているが、分離も可。
- (2) 電話番号簿配列：文字単位の五十音順配列。特に人名の配列に有利。入力時に字単位に区切ったふりがなの入力が必要。
- (3) 漢字順配列：総画数順、部首別画数順、音訓順などがある。前2者は漢字が読めなくてもひけるが、当用漢字などでは後者が有利といわれている。漢字コードを上記のいずれかの順で定めておけば、コードによるソーティングで配列可能。現在進められている「情報交換のための漢字符号」JIS 原案の作成では、第1水準漢字(常用的漢字)は代表音訓順、第2水準漢字では部首別画数順による配列で、コード化の作業が進められている。

以上が代表的な配列方法であるが、それぞれソート・キーが異なるため、いずれの方式でも配列可能とするためには、すべてのキーを用意しておかねばならない。なお細部にわたれば、濁点、半濁点の処理、かなづかい、外来語の表記法など多くの難問をかかえている。

最後に出力処理であるが、この部分は全く出力装置に依存する。しかし英数カナ・モードの情報の出力のみの場合は、ラインプリンタを使用するのが効率的である。将来は多字種を扱える汎用プリンタの出現が期待されるので、出力処理の負担は大幅に軽減されるであろう。また漢字コードの標準化も効率化の大きな要素である。

2.1.6 両モードのデータの比較

英数カナ・モードと漢字モードにおける2,3のデータを比較しておこう。

- (1) 語の長さ(文字数)

キーワードを例にとると、大体キーワードは「情報検索」のように、漢字4字からなるのが平均的である。一方漢字の音訓をカナで表記すると次のような分布となっている。

カナ文字数	漢字の割合
1	7%
2	68
3	22
4	3
5	1%以下

これから漢字1字当りの平均音訓カナ文字数を求めると、約2.2文字となる。キーワード平均ではこの4倍で8.8字（約9字）となり、「情報検索」はぴったり一致している。

またキーワードをローマ字で表記すると、平均的にこの1.5倍となり、13文字を要する。「情報検索」の場合、ヘボン式で11字、訓令式で12字である。

一般の漢字かなまじり文と比較すると次のようになる。

漢字かなまじり文	1.0	
カナ文字文	1.8	1.0
ローマ字文	2.7	1.5

以上は文字数での比較であるが、バイト数で比較するとカナ文字は1バイトに対し漢字は2バイトを要するので、キーワードの場合はどちらにしても大体同じとみてよいであろう。またローマ字の場合はいずれと比較しても5割増ということになる。

(2) 同形語

漢字には同音・同訓のものが多くはよく知られている。したがってキーワードのように特に語単位の場合には、十分注意しなければならない。当用漢字の中で同音の多いものを次に示す。

音	漢字数
コウ	60
ショウ	59
シ	46
カン	44
トウ	36
キ	35
セイ	30
ソウ	30
ケン	29
カ	28

このように同音の多い漢字を組合せるのであるから、当然のことながら同音の漢語が多数出現する。漢字モードであれば異形異語であるが、カナ・モードとなると同形異語となり、何らかの区別をしなければならなくなる。この区別の仕方については2.3で述べるが、語数約3万語のシソーラスにおいて、カナの同形語が500語存在している例がある。

(3) 文中における字種の割合

漢字かなまじり文というと、漢字とひらがなから構成されているように思えるが、実用文となるとこれ以外の字種が多数含まれている。その極端な例として、文献検索の対象となる科学技術文献を分析したデータを次に示す。

字種	%
漢字	37.5
ひらがな	29.4
カタカナ	12.4
英字	7.6
記号	8.0
数字	5.1

このデータからも明らかになるように、特定の分野の文を対象とすると、漢字・ひらがな・その他の割合が、それぞれ約 $\frac{1}{3}$ ずつを占めている。これは漢字が少ないことによる入力処理の容易さと、特殊な字種の多さによる処理の複雑さを示している。

さらにこれは日本語の中への英語（原つづり）の侵入と、カタカナ表記の外国語の侵入を意味しており、日本語の包容力の大きさを示すとともに、日本語の造語力の衰退を示すものである。

(4) 入力速度

各種タイプライタの入力速度の比較を次に示す。

英文タイプ	250 字/分
カナタイプ	200
和文タイプ	40
漢字テレタイプ	60
漢字入力機（ペンタッチ）	40

このように入力速度は、タイプライタの種類によって大きな差があり、これがいろいろな入力方式のアイディアの源となっている。

一方前述したように、同一情報を表記するのに、字種によって文字数が異なるから、これらの値を考慮に入れて補正した、情報入力速度を次に示す。これは漢字4字からなるキーワードを1分間に何語入力できるかを示している。

英文タイプ（ローマ字）	19 語/分
カナタイプ	22
和文タイプ	10
漢字テレタイプ	15
漢字入力機（ペンタッチ）	10

このように機種によって2倍以上の差が出ているが、これが情報の入力速度であると単純に考えてはならない。入力処理は入力のエラーを修正す

る段階まで含むものであるから、その効率を考慮しなければならず、また英文・カナタイプの場合は、漢字かなまじり文から翻字しつつ入力する場合には、相当割引いて考えねばならない。ともあれ漢字モード処理を対象とした入力システムは、非常に複雑な要素を含んでいるのである。

2.2 日本語情報検索のデータベース

2.2.1 データベースの設計

データベース設計上の留意点をあげる。

(1) 総合ファイルへの指向

情報検索という従来ともすると文献のサーチに重点が置かれているが、検索システムはあくまで情報流通システムの一環であるとの立場に立って、ファイル設計を進めねばならない。このためにはデータベースの原点に戻って、汎用性つまり処理と独立した立場から、情報の倉庫としての観点からの設計が肝要である。

情報検索のフェーズとしては 2.1 でも既に触れたように、SDI と RS、検索と索引誌、バッチ処理とオンライン処理、データ要素の異なる多種ファイルの処理などがあり、これらを総合的に考慮することはたいへんなことである。したがって設計に当っては、どのフェーズを対象とするかをまず明確にすることが出発点となる。

(2) サブファイルの分離

情報検索用のデータベースの内容は、文献情報のデータ要素からなるわけであるが、実際にこれを蓄積・運用していく上で、これをバックアップしていくサブファイルが必要である。その最たるものは、文献情報の書誌事項を管理する資料管理ファイルと、キーワードを管理する用語管理ファイルである。

これらのファイルは、それぞれ大量の情報を収容するとともに、情報検索とは別個にひんばんなメンテナンスを必要とする。特にこのメンテナンスの面からファイルを分離し、インタフェースを介して検索用データベースと接続するのが有利である。

(3) RAM への指向

RS またはオンライン検索を実行するためには、RAM として現状では

磁気ディスクが不可欠の要件となっている。RSの場合はほぼ10年間くらいの情報が必要とされ、その大量な情報を処理するには、磁気ディスクベースでなければ、その運営は不可能である。

ディスクを用いる場合、情報へのアクセスにはインバーテッド・ファイルを用いるが、アクセス・キーの決定が大問題となる。一般的にはキーとして、キーワード、分類コード、著者名などがあげられるが、この他にも著者所属機関、資料名（または資料番号）なども重要であるとの論も強い。特にオンライン検索において、利用者自身によるコンピュータとの対話が強調されると、キーの拘束による利用価値の減少は大きい。

日本語情報検索の場合には、さらに考慮しなければならない問題がある。データベースを作るまでの過程は、漢字モードによる日本語の処理が必要であるが、検索過程においては、アクセス・キーとして2,3のデータ要素のみが対象となるため、英数カナ・モードで処理しても大差がないと考えられるからである。特にキーワード、著者名にふりがながある場合には、そのカナを利用し同形語の処理を付加すれば、ほぼ同じ結果が得られると期待される。この問題に関してはオンライン検索の項で、さらに詳しく触れることにする。

文献検索システムの例を図2-1に示す。

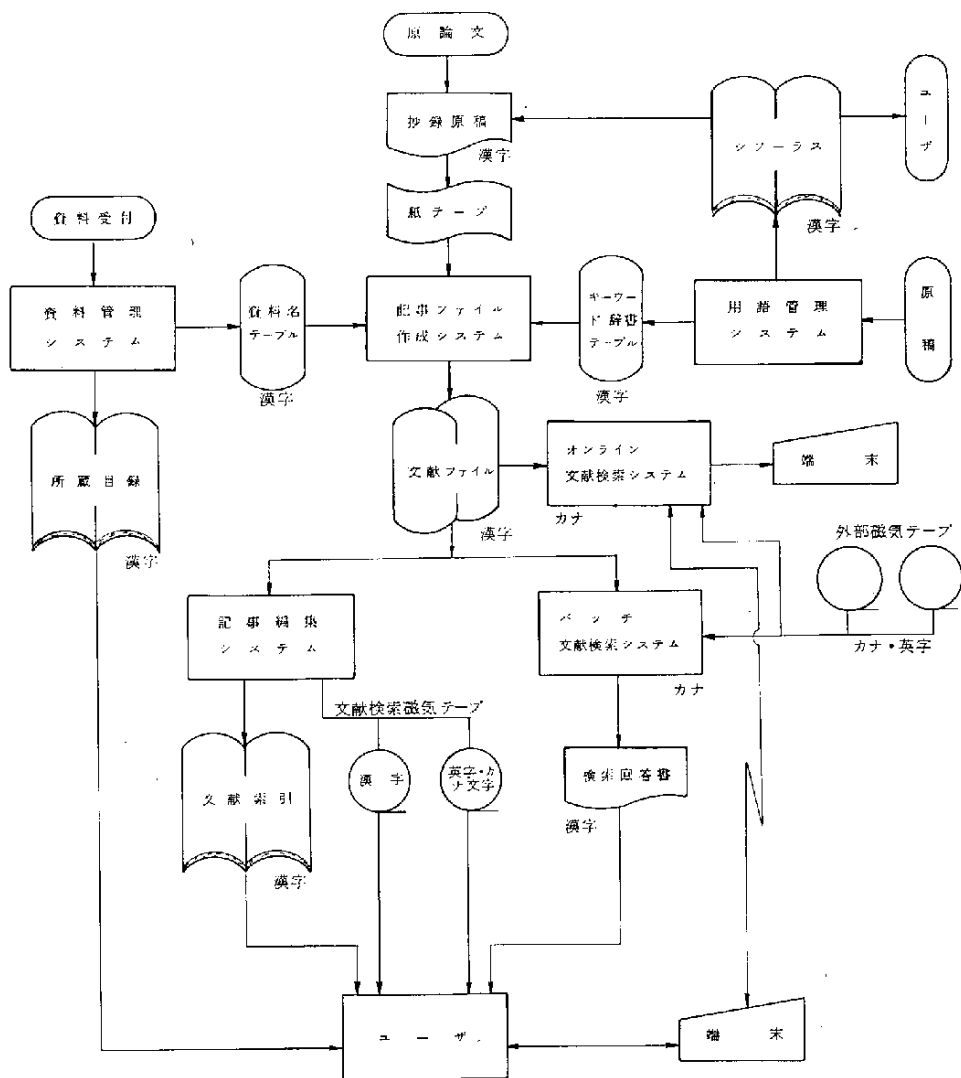


図 2 - 1 文献検索システムの例

2.2.2 わかち書き、漢字字種

わかち書きは日本語の正書法が確立していないため、その規則が非常に複雑である。わかち書きに関してはカナモジカイの規則があるが、複雑で相当の訓練を要する。一方言語というものはどの言語においても、非常に緩い規則で情報の伝達が可能なようにできているものであるから、情報検索のデータ要素という観点からは、わかち書き規則はできる限り簡潔なものとし、 unnecessary 努力は極力省くようにすべきであろう。

データ要素のうち文献標題に限っていえば次のような簡単な規則によれば実用的である。

「自立語（名詞、形容詞、動詞）は他の語と分けて書く」
いくつかの補足的な規則としては、語尾は自立語と分けない、助動詞は動詞と分けない、特に名詞において長いものは適宜分けて書く、とする。

（例） 燃焼器のモデル化 I 性能におよぼす周囲の状況の効果

ネンショウキ ノ モデルカ I セイノウ ニオヨボス シュ
ウイ ノ ジョウキョウ ノ コウカ

上の例は文献の標題として平均的なものを示したわけであるが、これを見ると漢字のカナ表記を含めて、非常に簡単な規則で行なわれていることがわかる。このことから当然のことながら、自動カナふり、つまり漢字－カナ変換の問題が出てくる。漢字－カナ変換で最も面倒な手順は、漢字の音訓が複数であるためのその選択の手順である。これを一般論として取上げ、汎用の変換アルゴリズムを開発するとなると大変なことであるが、文献検索に限れば比較的単純な手順が考えられる。

その原則は「訓読みの漢字はかなで書く」というものである。つまり原稿の作成に当ってこれを守るか、または翻訳に当ってこれを守れば、自動変換は容易である。自動変換とかかわりなく、この原則を主張する人々は意外と多い。これは日本古来の語には漢字を使用しないという主張であり、現代風にいえば、文中に Computer と書いてコンピュータと読むことはさしつか

えないが、ケイサンキとは読ませないということである。

前掲の例でいえば、「及ぼす」と書かないで「およぼす」と書いてあるの
で、標題中に訓読みの漢字は一字もなく、簡単なわかり書き規則とあいまっ
て、フリガナは至ってシンプルな形となっている。しかし実際問題としては、
主題をなす専門用語に訓読みの語があったりして、理想的にはいかないが、
自動変換においては、人間側にもある程度協力してもらうという発想が必要
な例である。

次に漢字の字種であるが、これも日本語の正書法との関連の問題である。
日本語では「この語は漢字で書かねばならない」という指定は1つもない。
漢字で書かないと読みにくい、威厳がない、恥かしい、意味が正確に伝わら
ないであろう、などという理由で漢字を使用しているに過ぎない。このよう
な状況から、コンピュータを利用した漢字処理システムを設計するに当って
は、漢字の取扱いについて何らかの明確な主張を打出す必要がある。漢字の
字種もその1つの要素である。

あるがままの日本文を処理するシステムは非常に高価である。コスト面か
らみれば漢字の字種を押えるのが最も効果が大きい。幸いなことに当用漢字
が定まっているので、これを基本として主張を打出すのがよい。1例を次に
示す。

- (1) 字種の基本は当用漢字とする。ただしその中で不要と思われるものは削
除する。
- (2) 人名の姓に多用される漢字は補足する。これは著者索引を想定するため
である。
- (3) 人名の名に多用される漢字は上位のものを少数補足する。
- (4) 地名は県名、市名に出現する漢字を全部補足するが他は無視する。
- (5) その他システム特有の漢字は補足する。

このように思い切った方針を打出すと、字種は大体1,800字程度となり、
コスト面からの効果が大きい。

なお字種とともに索引誌などの冊子体を想定すると、字体や文字のサイズも重要な要素となるが、これはフアンクションによって解決できる問題である。ただしこれに余りこると、出力面でのコストが大きくなるので注意を要する。

2.2.3 漢字コード

現在漢字コードの標準化が進められているが、現状では標準漢字コードはない。しいていえば新聞関係でCO-59コードがあるが、データ伝送面に重点があって漢字処理用としては不向である。漢字コード設定に当っては2つの側面があり、1つは情報交換を前提とした標準化であり、1つはコンピュータ処理の効率面である。

漢字処理の面からは、基本的に次の条件を満たすことが望ましい。

- (1) 字種（漢字、かな、カナ、英字、数字など）による大分類が必要である。

これはまた字種ブロックの配列順をも与える。

- (2) JIS, ISO, EBCDIC など既存のコード体系に与えられている字種については、それらのコードとの整合性のあることが望ましい。

- (3) 漢字についてはある利用面を想定した配列順に、コードが与えられていることが望ましい。

(1)のブロック順であるが、これには1つの原則がある。索引のように見出語によってアクセスする場合には、ブロックの大きさの小さいものを前に配列するのがよい。これにより小数文字へのアクセスが容易となり、また見落す危険が少なくなる。ある用語リストの例をあげると

数字見出し	1 / 6 ページ
英字見出し	8
日本字見出し	270

となっている。さらにこの論でいくと、日本字の中でもカタカナは前へ出したほうがよく、ひらがなは漢字と混在でよい。

(2)は英数カナ・モードとの使い分けを行なうことを想定すると当然のことであろう。

(3)は従来全く無視されていたものであるが、これは配列順についての正書法がなかったためである。しかし漢字処理が進み各種の処理やリストが必要となってくると無視はできない。単純にコードでソートして利用できることは、今後情報検索や語彙の分析・調査などで不可欠の条件となるであろう。

以上はコンピュータ処理の面からみた場合であるが、一方入力装置の面からはなかなか理想的にいかない多くの問題がある。今日の入力装置は、一般にハードウェアからの制約によって、文字盤上の位置とコードが固定している。したがって盤面上の文字配列を人間工学的に理想的に設計すると、前述のコードの条件を満たすことができない。これを解決するにはコード変換するしかない。よってこの両面の要求を満足させるには、始めからランダムなコード変換を前提としてシステム設計を行ない、悔を残さないことが賢明である。

これらの問題を入力装置側で解決する方法は、文字盤上の位置とコードとの固定を解放し、文字とコードを固定する方式にすればよい。この方針により設計・試作した入力装置も発表されているが、信頼性の面からなかなか実用には至らない現状である。

次にコードで注意しなければならないことは字体の取扱いである。たとえば漢字のゴシック体が必要であるとして、明朝体とゴシック体に全く別のコードを与えたりすると、いろいろの面で不都合が生じる。1つは配列であり、1つは照合である。情報検索では多くの場面で語の照合が必要となり、その最たるものはキーワードの照合による探索である。字体の違いによるアンマッチというナンセンスを防ぐためにも、字体の指定はファンクションコードかビットによるべきであろう。

2.2.4 漢語処理

ここでは漢語を扱う上で、いくつかの考慮すべき論点について述べることにする。

日本文の語彙は、和語、漢語、外来語、混種語および数字、記号からなる。国研（国立国語研究所）が行なった新聞の語彙調査の結果からのデータの一例を示すと、次のようになっている。

和語	1 7,0 9 6 語
漢語	1 7,0 7 8
外来語	4,5 6 4
混種語	1,8 5 1

このうち和語は形容詞、動詞が多く、名詞のみに限ってみると次のようになっている。

和語	4,7 4 9 語
漢語	1 5,8 8 4
外来語	4,3 9 6
混種語	1,4 6 2

これからもわかるように、文中の名詞は和語以外の語彙が圧倒的に多数を占めていることがわかる。さらに新聞でなく文献検索の対象となるような専門的文献となると、この傾向はますます著しい。

文献検索においては、文献標題や抄録のように句または文のものもあるが、コンピュータ処理の対象に限ってみると、キーワードを中心とする語がほとんどであるので、さらに限定してキーワードのみについてみてみよう。ある理工学用の日本語ソーラスにおいては、語種は次のようになっている。

和語	あらさ、鋤物、マダケ
漢語	石炭鉱業、集積回路要素、けい酸
外来語	ポリイソブチレン、BHC
混種語	和・漢 誤り制御

外・漢 システム工学, MHD 発電

和・外 色ガラス

このように語種については多種にわたっているが、異り語数の面からみると和語は極めて少ない。このことからいろいろと処理について考慮すべき問題が生ずる。

まずひらがなの問題であるが、和語が少ないということから、キーワード中にひらがなを混在させないという可能性がでてくる。上の例ではひらがなは2つの理由から現われている。1つは訓読みのための送りがなとして現われ(誤り)、1つはこのシソーラスを含む文献検索システムで、使用できない漢字があるために現われている(けい酸)。したがってひらがなを排除するには、前者では無理して送りがなをやめるか、漢語に置換えるか、後者ではその漢字を使用可能にするか、他の語たとえばSi(珪酸の原子記号)に置換えるかすれば可能である。幸いなことに日本語では外来語をカタカナ表記または原語表記で導入可能であるので、ひらがなの排除は工夫次第で比較的容易である。

なお上例の和語の中にマダケ(真竹)があるが、これは動植物名はカナ書きにするという原則に基づいており、これによりひらがなキーワードは大幅に減少している。このようにキーワードつまり専門用語からひらがなを排除することにより、文献標題または抄録文からの自動索引や、日本文KWIC索引の作成などが、極めて容易となるのである。

次に和語であるが、キーワードから和語を排除し漢語で置換えることにより、いろいろと有利となる面が多い。一般に和語は訓読みされるが、同一の漢字が音読みと訓読みをされることにより、同じカテゴリーの語がシソーラス上で離れ離れになり、またこのことが漢字の代表音による配列(一般に漢字コードによる配列)を困難なものにしている。その一例を次に示す。

チュウ	イ
鋳造	鋳はだ

鑄造欠陥	鑄ばり
鑄造性	鑄物
鑄造設備	鑄物仕上
鑄造組織	鑄物砂
鑄造用具	鑄物砂処理
鑄鉄	鑄わく

この場合には右側の語を鑄造品、鑄造用などの語を用いて置換えが可能である。一方和語の排除つまり一語一音への専門用語の整理は、今後のカナ→漢字変換または漢字の入力に大きな福音をもたらすものである。

次に上の例で示したシソーラスは約3万語からなるものであるが、そこに収録されている語は大部分複合語である。複合語は漢語+漢語が多く、ついで外来語+漢語が多い。この複合語が検索に当たっていろいろ問題となるのである。漢語の複合については国研から多くの研究が発表されているが、その中からいくつかの問題を検討することにする。

漢語は漢字2字からなるものが基本型であり、これが結合して複合語となるが、さらに接頭字、接尾字が結合して多様な複合語を形成する。一例を次に示す。

安定	基本型
不安定	接頭+基本
不安定性	接頭+基本+接尾
安定条件	基本+基本
不安定性解析	接頭+基本+接尾+基本
安定燃焼装置	基本+基本+基本

外来語+漢語の場合も全く同様な構造を示す。さて上の例に示した語がキーワードであるとして、これのマッチングにより検索を行なうと仮定すると、キーワードの完全一致のみでは不十分であることがわかる。たとえば、「適応制御系の安定性」についての文献を探す場合、安定性のみでなく不安

定性で探索する必要があるし、また燃焼装置の文献を探索する場合には、安定燃焼装置の文献も必要である。

一般にこのような場合には、キーワードと質問ワードの任意一致を可能にするアルゴリズムを用意して行なうわけであるが、日本語の場合には語が分かち書きされていないため、必要以上にむだなマッチングを行なわなければならない。またシソーラスにおいても安定から上例のすべてのキーワードを見出せると非常に便利である。

英語の場合には単語単位で分かち書きされているので、上の要求を満たすために単語の順を逆にして索引を作るか、KWIC索引の手法で索引を用意しているものがある。前者を特に循環索引(Permuted Index)とっている。

① Information Retrieval

Retrieval, Information

② Programming Language

Structured Programming

日本語の場合にもこれと同じ手法を用いればよいわけであるが、そのためには前もって人為的に分かち書きをしておくのと有利である。特にシソーラスで語の統制を行なう場合には、次のように語を区切って入力しておくのがよい。

不・安定・性・解析

日本語の場合キーワードのKWIC索引が有利な例を次に挙げておく。

関数	索引
応答関数	引用索引
解析関数	著者索引
三角関数	KWIC索引
指数関数	索引語
対数関数	索引作業
代数関数	抄録索引誌
相互相関関数	自動索引法
直交関数展開	件名索引語
関数近似	件名索引カード

複合語に関してはまだ多くの論点があるが、2.3においてさらに述べることにする。

最後に漢語のマッチングについて少し触れておく。英語またはカナのキーワードの場合には、語を構成する文字の字種も少なく、また文字そのものも特定の意味を持たないため、語単位のマッチングが常識である。さらにシリアル・サーチにおいて、質問ワードが論理積の関係にある場合には、サーチの効率化のために、出現度数の少ないキーワードを優先させてサーチするのがよい。

しかし漢語の場合には、漢字の字種が多くキーワードにおけるその使用度数にも大きな差があるので、文字単位の使用度数に従って、字単位のマッチングをするのが効率的と思われる。しかしこれに関しては未だ研究の発表もなく、また使用するコンピュータの機種にもよることであるので、決定的なことはいえないが、将来の研究課題であろう。さらにこれは転置ファイルのファイル構成においても、考慮しなければならない課題でもある。

2.3 日本語情報のキーワード/シソーラスの管理

2.3.1 情報検索とシソーラス

ここでは特に日本語情報とは限らず、シソーラスの機能の概要を説明する。まずシソーラスの定義としてユネスコ（UNESCO：国際連合教育・科学・文化機関）のものを挙げておこう。

「シソーラスとは意味的に関連ある語を、統制をとりながら、しかもフレキシブルに集めた用語集であって、ある特定の知識領域を、包括的にカバーするものである。」

上の機能を実現するためにとられているシソーラスの構成は、次の各要素を含んでいる。

(1) 用語の限定

キーワードとして許される語を限定する。用語をディスクリプタと非ディスクリプタに分け、USE, UF 参照で指定する。（USE: Use ～ を用いよ。UF: Used For ～ の代りに用いよ。ともに同義語の関係）

(2) 用語の定義

用語のシソーラス内での意味を限定する。スコープノートとして表示する。

(3) 用語間の関係

用語間の意味的な包摂関係および用語間の実用上の関連関係を、BT, NT, RT 参照によって指定する。（BT: Broader Term 上位語, NT: Narrower Term 下位語, RT: Related Term 関連語）

(4) 用語の分類

用語の使用の限定および意味の明確化のために、用語を分類しグループの名称を主題カテゴリーとして表示する。

以上が現在実用に供されているシソーラスの一般的構成であるが、この構成のもとに使用の便をはかって、次のような各種のリストが提供される。

(1) 主シソーラス

上述の各要素をすべて表示した基本リスト。

(2) 主題カテゴリー別索引

特定分野の用語の探索に便利。

(3) タームチャート

主題カテゴリー別に用語およびその関係をグラフ的に表示した図式索引。

(4) 循環索引

2.2.4 で述べた循環索引を応用した用語の索引。

(5) 階層索引

最上位の語(トップ・タームまたはルート語)からの階層表示のみを表示した索引。

シソーラスを最も多用するのは、文献の内容を分析し索引作業を行なうインデクサであるので、利用の便および索引作業の均質性を保つために、上記のような各種リストが用意される。また用語の語数が多くなると、当然のことながら、その編集作業はコンピュータ処理の1つの応用分野となってくる。さらに将来的には、シソーラスのオンライン表示が重要なテーマとなるであろう。

以上はシソーラスの各種機能であるが、次にシソーラスの具備すべき要件を考察しよう。この問題はシソーラスの実用化が進み、その使用経験が蓄積されるに従って、クローズ・アップされてきたものである。

(1) 便利性

シソーラスはインデクサにとっては日常頻繁に使用するものであるから、あらゆる意味において便利でなければならない。索引語として最も適切な用語を見出せるか否かは、インデクサの専門知識の深さに負うところも大きい。シソーラスの便利さも大きな要素である。

(2) 網羅性

シソーラスは与えられた検索システムに対して、内容的に網羅性がなけ

ればならない。これはシソーラス作成段階においての用語の選定手順に誤りがあると、応々にして生ずる問題であるが、一方分類コードにおける「一般、その他」に該当する機能を持たせることによってもある程度防ぐことはできる。

(3) 整合性

これは2つ以上のシソーラスの存在のもとでの問題であるが、当然のことながらその標準化には十分注意しなければならない。特に各専門領域ごとにシソーラスを作成する場合などでは、最重点項目である。

(4) 保守性

メンテナンスの容易性である。シソーラスの定義にもあるように、分類体系と異なりその柔軟性がシソーラスの身上であるから、メンテナンスの困難さによって、その特徴が失なわれるようなことがあってはならない。この問題の解決には、編集作業をコンピュータ処理で行なうのが1つの解決法であるが、追加、削除、訂正などの編集作業のみでなく、リスト化の電算植字までも含めて、システム設計を行なっておくべきである。さらにオンライン・メンテナンスに進めば理想的である。

2.3.2 日本語キーワード

シソーラスを構成する意味的要素である用語、とりわけ日本語による用語の問題をまとめて考察することにする。シソーラス作成の第一歩は用語の選定から始まる。一般に用語選定基準を設けて作業に入るが、その場合用語の網羅性と特定性について、しっかりした基準を設けておかないと、満足するシソーラスは作り得ない。ここではそのうち用語の特定性についてさらに考察を進めよう。

日本語のキーワードは漢語を主とし、それに外来語および混種語をもって構成されることは、前に述べたとおりである。漢語は漢字2字をもって基本とし、それに接頭・接尾字が付加され、さらに漢語が結合して複合語が構成

され、キーワードはこの複合語を主体として選定されている。

語は複合されることによって、その表わす概念の範囲は狭くなる。一例を次に示す。

最適化

運転最適化

原子炉運転最適化

FBR型原子炉運転最適化

このように語を複合することにより、いくらでも特定な用語を作り出せ、それに伴って用語の数は増大する。したがって語の特定性を定めるには、語の複合度を活用するのが1つの方法である。

基本漢語をキーワードとすると、その表わす概念があまりにも広すぎて、検索における適合率を著しく低下させる。これはアメリカにおいて十数年も前に、ユニターム (Unit Term, 最小単位の語) で失敗したのと同じことである。これが2語の複合語となると概念の範囲が非常に狭くなり、適合率の問題も解消される。しかし3語以上の複合語となると、語の中での主たる概念が語の中間に埋没して、検索上いろいろと別の問題を派生する。

以上のような理由によりシンソーラスにおけるキーワードは、漢語2語の複合語を基準とするのが最も効率がよいとされている。3語以上の概念は2つ以上に分割して索引化するが、その場合でもキーワードは2語の複合語を用いて表わす。

たとえば

コンピュータ・システム信頼性—コンピュータ・システム

システム信頼性

となるようにキーワードを設定しておく。

なお接頭・接尾字はそれ自身1つの語とみなさず、語の一部分として処理する。つまり上例のコンピュータ・信頼性は1語とみなす。

以上の基準により用語の選定を行えば、一応目的を達することはできる

が、個々の具体面においてはいろいろの問題がある。その1つの例を示そう。キーワードによって文献の主題を索引化するにあたって、主題のニュアンスを適切に表現しようとする、いくらでも多くのキーワードが必要となってくる。一例として燃焼というプロセスを取上げ、これが技術として具現化される場合の用語を列挙してみる。

燃焼システム	高温燃焼
燃焼方式	安定燃焼
燃焼構造	定温燃焼
燃焼制御	異常燃焼
燃焼機械	液中燃焼
燃焼装置	拡散燃焼
燃焼設備	完全燃焼
燃焼条件	空間燃焼
燃焼コスト	固定層燃焼
燃焼状態	重油燃焼
燃焼温度	蒸発燃焼
燃焼ガス	振動燃焼
燃焼過程	層化燃焼
燃焼効率	超音速燃焼
燃焼速度	不安定燃焼
燃焼特性	微粉炭燃焼

これをみると、燃焼という漢語はいわゆる用言性（サ変動詞の語幹）の語であり、これと同類の漢語の場合にも、同様のキーワードがそれぞれ必要となるものと考えられる。また上例の左列で後置されている基本漢語（これを後語基と名付ける）は、特定の主題カテゴリーに属せず、各カテゴリーに横断的に使用されうる語である。

このように考えてくると、上例左列の特に後語基については、別途の考慮が必要と思われる。たとえば、後語基として自由に付加可能語のリストを用

意しておき、必要に応じてインデクサの判断により付加させる方式をとれば、シソーラスの語数は少なく済み、またキーワードがなくて索引化に苦心することも軽減するものと考えられる。つまり特定の用語については統制を緩めるという方式である。この場合でもキーワードの中心概念の用語は前語基として前方にあるので、検索上それほどの困難もなしに処理可能と思われる。またオンライン会話型の検索となると、ある程度のキーワードの通覧が可能となるから、付加により特定性を高めておくことは、かえってプラスとも考えられる。

以上日本語キーワードについて問題点の一端を述べたが、まだまだ日本語そのものについての研究が不十分であり、かえって情報検索の面から、日本語研究へ問題提起をしなければならない時代であると思われる。

2.3.3 日本語シソーラスのコンピュータによる用語管理

(1) データ要素

表2-2に用語管理におけるデータ要素の例を一覧表として示す。これらの要素のうち若干のものに説明を加えておく。

シソーラス区分：2種以上のシソーラスを管理する場合の識別コード

語番号：見出し語に与えた識別番号。コンピュータ処理により付加

見出し語：シソーラス上で見出しとなる語で、ディスクリプタおよび非ディスクリプタよりなる

用途区分：ディスクリプタ・非ディスクリプタ・非見出し語の区分コード

関係子：語間の関係の区分コード。(関)は関係先のデータを示す

索引度数・検索度数：実際に使用された回数の記録

ソート番号：見出し語のソート順コード

以上のデータ要素から明らかなとおり、用語ファイルにおける漢字のデータ要素は、見出し語とスコープノート(Scope Note:注記)のみである。

表 2 - 2 データ要素の例

デ ー タ	字 種	F・V	レコード 長	レコード 数
シソーラス区分	数	F	2	1
語番号	数	F	6	1
見出し語	漢	V	max40	1
見出し語フリガナ	カ	V	・・40	1
用途区分	数	F	1	1
主題カテゴリ	英数	F	4	6
スコープノート(注記)	漢	V	max 1000	1
関係子	数	F	1	1
シソーラス区分(関)	数	F	2	1
語番号(関)	数	F	6	500
ソート番号	数	F	6	1
登録年月	数	F	4	1
索引度数	数	F	4	1
検索度数	数	F	4	1

(2) 用語管理の機能

基本的機能は次の7項目である。

- ① 見出し語とその付随情報の入力および校正
- ② 見出し語とその付随情報の登録またはメンテナンス
- ③ 見出し語間の基本的(直近的)関係の入力およびチェック
- ④ 見出し語間の関係のシソーラス構造上の矛盾チェック
- ⑤ ディスクリプタ間の階層関係の展開
- ⑥ 索引および検索時の付加およびチェック用の用語ファイルの作成
- ⑦ 冊子体シソーラスその他各種リストの編集および植字

これらの内④と⑤について説明する。

(3). 語間の関係の矛盾チェック

前項(2)の機能からもわかるように、用語管理システムの入力、見出し語とその付随情報の入力、見出し語間の基本的関係の入力は、別の処理となっている。この理由は前者は見出し語とかスコープノートのような、言語情報を含んでいるのに対し、後者は関係のデータのみを入力すればよいので、入力処理の効率化をはかるためである。特に日本語の入力を必要とするシステムでは、この効果は大きい。したがって矛盾チェックもこの入力方式に基づいて行なわれる。

矛盾チェックは次の各項について行なわれる。

- ① 見出し語ファイルに登録されていない語に対して関係付けがなされた場合
- ② 非ディスクリプタがUSE先のディスクリプタを持たない場合
- ③ 見出し語が自分自身に關係付けがなされている場合
- ④ ディスクリプタがUSE先のディスクリプタを持つ場合
- ⑤ NT・BT・RT関係について関係する語の中に非ディスクリプタがある場合
- ⑥ 2つの見出し語の間に矛盾する関係が存在する場合

たとえば、A、Bがいずれもディスクリプタである場合

A USE B A UF B

は共に矛盾した関係であり、また

A NT B A BT B A RT B

が同時に存在することも意味的にあり得ない。

上の6項目をみると、ごく当りまえのことで、入力時に注意すれば、このような意味的に矛盾する関係が入力されるのがおかしいのであるが、実際問題として数万語にも及ぶ見出し語が、複雑にからみ合っているシソーラスにおいては、入力時に人間側で完べきを期するのが非常に困難なのである。逆にいえばこのような矛盾チェックこそ、コンピュータ・システム

の面目であるといえよう。

ただここで注意しなければならないのは、コンピュータによるチェックは、形式的に矛盾を判定できる場合のみのチェックであり、実際にシソーラスを作成してみると、このチェックにかからない矛盾が往々にして現われてくる。たとえば、作成時に全く独立に、次の2つの階層関係を作って次のように入力したとすると、

原子炉構成要素	制御要素
制御要素	遅れ要素
原子炉安全系	流体素子

コンピュータによる編集過程で制御要素を仲介として、2つの構造は一体となり、すべて原子炉構成要素の下に包含されてしまう。これは左の階層において原子炉制御要素とすべきところを、ただ単に制御要素としたために生じた矛盾である。このような意味的な矛盾は、最終的には目検によるチェックをまたねばならない。

実際にシソーラス作成段階においては、このような矛盾が往々にして生ずるので、シソーラス編成の結果を少なくとも1回は校正するよう、始めからスケジュールに組込んでおくべきであろう。この場合矛盾の修正には、メンテナンスのルーチンを使用することになるので、特に日本語シソーラスの作成に当っては、メンテナンスおよび漢字出力が、これらの作業のネックにならないよう、十分検討を加えておくべきである。

なお、ここで逆関係の矛盾チェックについて触れていないのは、逆関係は一方の関係からシステム内において発生させることを前提としているためである。たとえば

A NT B

と入力すると、B BT Aはシステム内で発生させ、省力化をはかっている。

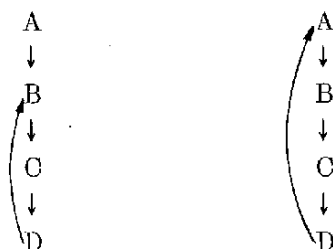
しかしこれは入力を誤ると二重に誤りを生み出す原因となり、また入力

におけるパンチ・ミスを見逃す原因ともなる。したがって、もし入力に余裕のある場合には、むだなようでも逆関係も入力し、その関係の矛盾をチェックした方がよい。特にメンテナンスにおいてはデータも少量であり、また既成のファイルを誤データによって汚染させないためにも、逆関係の二重入力を行なうべきであろう。

(4) ディスクリプタ間の階層関係の展開

ディスクリプタ間の関係は、直近的な基本関係のみで入力されているので、これをたどって木構造を作り出す。これによりルート (root, 根) 語を最上位とする階層関係が展開されるが、この過程においても種々の矛盾がチェックされる。

まず下位語から上位語に戻る回路 (サーキット) を削除する。この場合、次図のように木構造の中間の節に戻る場合には、その木における下位関係



を優先させて $D \rightarrow B$ の関係を削除する。右図のようにルート語に戻る場合には、木構造が成立しないので、その階層全体を矛盾するものとして削除する。

次に階層関係を展開した結果、階層関係と関連関係が重複している場合には、階層関係を優先させて関連関係を削除する。

以上が階層関係の展開における主たる矛盾のチェックであるが、これ以外にも矛盾と考えられる関係 (たとえばバイパス的な下位関係) が存在するが、いずれを矛盾とするかは、そのシソーラスを適用する情報検索システムに依存するので、システム設計の段階で十分検討することが必要であろう。

なお前項をも含めて、これら発見された矛盾についてはエラー・リストを出力し、人間側に矛盾の修正を促がさねばならない。

(5) システム運用上の問題点

前項までで用語管理システムの構成と機能の要点を述べたが、これをシステムとして運用するとなると、いろいろの問題が発生する。そのいくつかのものに触れておこう。

まずデータの精度であるが、シソーラスは情報検索システムにおけるキーワードを統制する典拠リストであるから、そのデータは絶対の精度が要求される。特に日本語シソーラスにおいては、見出し語およびその付属情報の入力には漢字入力を必要とする。もともと漢字入力の精度には問題が多いので、シソーラスの場合これをいかに効率的に乗切るかを、十分検討しておかねばならない。新たにシソーラスを作成する場合には、大量のデータを入力しなければならないので、既成のシステムを活用することになるが、メンテナンスの場合には少量であるから、小回りのきく精度の高い方式を考えるべきである。1つの解決法としては、漢字ディスプレイ付のペンタッチ方式の装置を用い、シソーラスの内容に詳しい専門家が自身で入力することも考えられる。

一方、関係情報の入力については、見出し語を直接用いずに、語番号またはフリガナを利用して、英数カナ・モードで入力すべきであろう。

次にシソーラスのメンテナンスであるが、もともとシソーラスは分類コードとは異なり、自然語を用い、構造の規則をある程度自由にして、日進月歩の領域に対応できるように、その柔軟性を特長としている。したがって、始めからそのメンテナンスの柔軟性にも重きを置いて設計しなければならない。これは処理が容易であることと共に、処理の自由度も大きくして置かねばならないことを意味する。これはシステムの効率および経済性と相反する要素であるので、設計上苦心の多いところであるが、最新のファイル構成技術とオンラインの即時処理とを駆使して、利用側に立った設

計が望まれるところである。

しかし、メンテナンスに対し、情報を蓄積する側から見ると、キーワードの追加・修正・削除は、情報ファイルの維持または情報の検索に大きな負担をかけることになる。シソーラス側としては見出し語のヒストリーを完備しなければならず、文献ファイル側では、これによって全ファイルを常にメンテナンスしていかなければならない。

この問題に対しては世界的にも運用経験が浅く、未だ定説はない。アメリカNLM(National Library of Medicine, アメリカ国立医学図書館)のMEDLARS システムがこれを徹底して実行しているので、その運用待ちというところである。情報検索の権威であるF W Lancaster氏も、「用語の統制はほどほどにすべきである」といっているように、シソーラス指向の文献検索システムは、コンピュータがいかに面倒な処理を経済的にしかも手ぎわよくこなしてくれるかにかかっているといっているようにやろう。

次に矛盾チェックの問題であるが、数万語にも及ぶシソーラスともなると、無矛盾のデータの入力を期待する方が無理であって、矛盾を発見することで、人間側をバックアップするという態度が必要である。このためには矛盾を発見した場合エラーの表示をして、新規まき直しを要求するのはいかにも非能率である。その時点で修正可能なオンライン・メンテナンスとすべきであろう。

もう1つの問題はコンピュータ処理のチェックにかかわらず、最終出力で矛盾が発見される場合である。これは結局メンテナンスの再処理ということになるので、前に述べた形式的な矛盾チェックのみでなく、ある程度意味的チェックの可能なアルゴリズムを開発する必要がある。

最後に用語管理システムのシステム・フローの例を図2-2に示す。

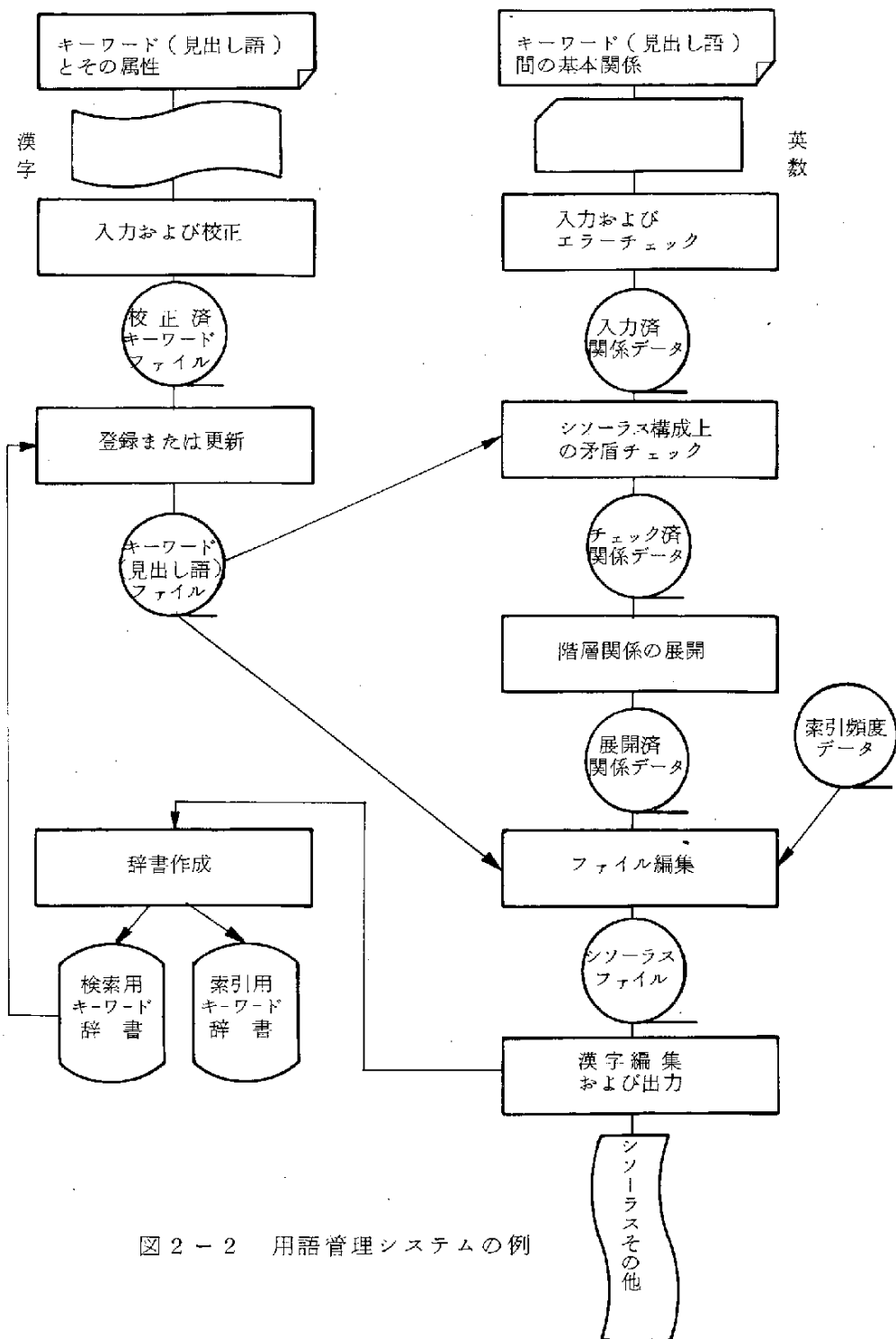


図 2 - 2 用語管理システムの例

2.3.4 キーワードの自動チェック

シソーラスによって統制される文献検索システムでは、キーワードとしてはディスクリプタ以外は認められない。したがって索引作業によって文献に付与されたキーワードのチェックが必要となる。しかるにシソーラスはこの統制語彙の典拠リストであるから、キーワードとシソーラスとの照合によって、自動チェックを行なうことができる。さらにこの照合の時点で、その他の処理も同時に行なうことができる。それらの機能を次に示す。

- ① キーワードがディスクリプタとして登録されているかどうかのチェック
- ② キーワードに表記上の誤りがないかどうかのチェックおよびその修正
- ③ コードによるキーワード入力 of 自然語ディスクリプタへの変換
- ④ キーワードへの付属情報の付加
- ⑤ 上位ディスクリプタの付与

これらの機能を順を追って説明しよう。

インデクサは索引作業にあたっては、文献の主題を表わす概念にあてはまるディスクリプタを、シソーラスから探して付与するわけであるが、経験によって常用するディスクリプタは記憶の中に入っている。したがってこれらの語を付与する際には記憶にたよることになるが、往々にして勘違いを起して登録されていない語を記入することがある。主として上記の理由によって生ずる非登録語または非ディスクリプタの記入は、自動チェックによって排除される。

なおシステムとしては、非ディスクリプタの記入をディスクリプタに自動変換する方がベターである。この場合シソーラスに非ディスクリプタを大量に登録しておくと、キーワードの記入の自由度は大きくなる。また次の例のように、略記名を非ディスクリプタとして登録しておくと、記入の能率は向上する。

LSI	大規模集積回路
LP	線形計画法
PVC	ポリ塩化ビニール

インデクサは往々にしてディスクリプタの記入時に誤りを犯すことがある。またインデクサは正しく記入したのに、パンチャがミス・パンチをすることがある。これらのミスは自動チェックによって排除される。

しかしこの場合ただ単に排除してしまうのでは、あまりに非能率的であり、これら誤りを自動的に修正するアルゴリズムを付加すべきであろう。この手順はいろいろ考えられるが、たとえば、ディスクリプタの1文字のみがアンマッチの場合には、その字のみを修正してパスさせるのも1つの方法であろう。かかる手法については、英単語については発表されているが、日本語については皆無に等しい。今後の1つの研究課題であろう。

この問題について1つ注目すべきことは、カタカナに誤りが多いことである。文字に冗長性が少なく他の文字と混同しやすいこと、ひらがなと混同しやすいものがあること、ディスクリプタの1文字1文字に意味がなく、パンチャが意味的に文字を判断できないこと、拗音・促音の表記法が不安定であることなどによって、誤りが多くなる。極端な例としては、「ウオータ」と「ウォータ」など全く表記がまちまちである。したがって、よほどよくシンソーラスを見ないと書き間違い。

これを防ぐにはカタカナについては、拗音・促音の区別なしにパスさせるとか、また文字数が多いから1文字違いの場合はパスさせるとか、カナの「へ」とかなの「へ」とは同一視するとかのアルゴリズムを設定すべきであり、またこれによる誤チェックも皆無とみてよいであろう。

日本語の入力は特にそのハード面でいろいろと面倒な問題があるので、できる限り英数カナ・モードによる入力望ましい。シンソーラスが完備すると、キーワードに関してはこれが可能となる。コードをシンソーラス・ファイルに入れておくことにより、自動変換が可能となる。索引作業においては大量の

キーワードを取扱う関係上、直観性のない暗号的なコード（数字コードなど）では、かえって記入の労力が増し、また誤記入も多くなるから、フリガナ程度のコードが無難である。

一方、検索においては扱うキーワードが少量であるから、どんなコードでもさしつかえないが、オンラインの場合には会話性を高める意味から、直観性のあるコードが望ましい。この問題は広い観点からみれば、カナ-漢字変換の一種であり、完備された辞書と文法による変換の一例である。

キーワードは検索あるいは語の管理などの面から、多くの属性情報を必要とする。前に示したデータ要素からも明らかなように、語番号、フリガナ、主題カテゴリー、ソート・コードなどが必要であるが、これらはもちろん自動付加の対象である。

最後に上位ディスクリプタの付与であるが、これは検索上の要求から行なわれるものである。一般に分類コードではコードの各桁が概念の階層のレベルを示している。したがって検索においては前方からの桁を指示することにより、いかなるレベルでの検索も自在である。しかし自然語にはこの階層性がないから、文献ファイル中に上位のキーワードも入れておき、その欠点を補っている。

2.4 バッチ検索とオンライン検索

2.4.1 文献検索の本質と両方式

ここではバッチ検索とオンライン検索との比較を論ずるが、それに先立って両方式の定義をはっきりさせておかないと混乱が生ずる。まずバッチ検索であるが、これは2通りに解釈されている。1つは単純に多数の検索要求（つまり質問）を、一括して同時に処理するという解釈である。他は探索というプロセスをコンピュータにまかせて一括処理するという解釈である。

ここで注意しなければならないのは、文献検索という一連の処理の中で、コンピュータが介入するのは、ファイルをサーチして照合条件に合ったデータを抽出する処理のみであり、これを一般に探索といっている。バッチ検索では照合条件つまり探索条件を組立てて入力し、探索過程はすべてコンピュータにまかせて出力を得る方式をとるので、探索をバッチ処理すると解釈される。ここではこの後者の解釈をとる。

この定義からもわかるように、バッチ検索といった場合、その処理方式がオンライン処理か否かは本質的な問題ではない。オンライン処理なら即時処理が可能であり、そうでなければ遅延処理となるだけである。

そうすると上の定義からオンライン検索という用語がおかしくなる。正確には会話型検索というべきであって、必然的にオンライン会話型となる。つまり探索過程の中へ人間が介入して、人間が考えてコンピュータが探すという過程を繰り返して、より適合する文献を探そうという主旨である。したがってここではバッチ検索とは探索の一括型処理、オンライン検索とは探索の会話型処理ということで論を進める。

さてこの両方式を比較するに当って、両者の情報処理技術のみを比較しても、検索の本質に触れることはできず、また将来への見通しを誤まる恐れがあるので、文献検索における人間の思考過程を考察し、それに基づいて両方式の機能を論ずることとする。まず図2-3に探索の思考過程の流れを示す。

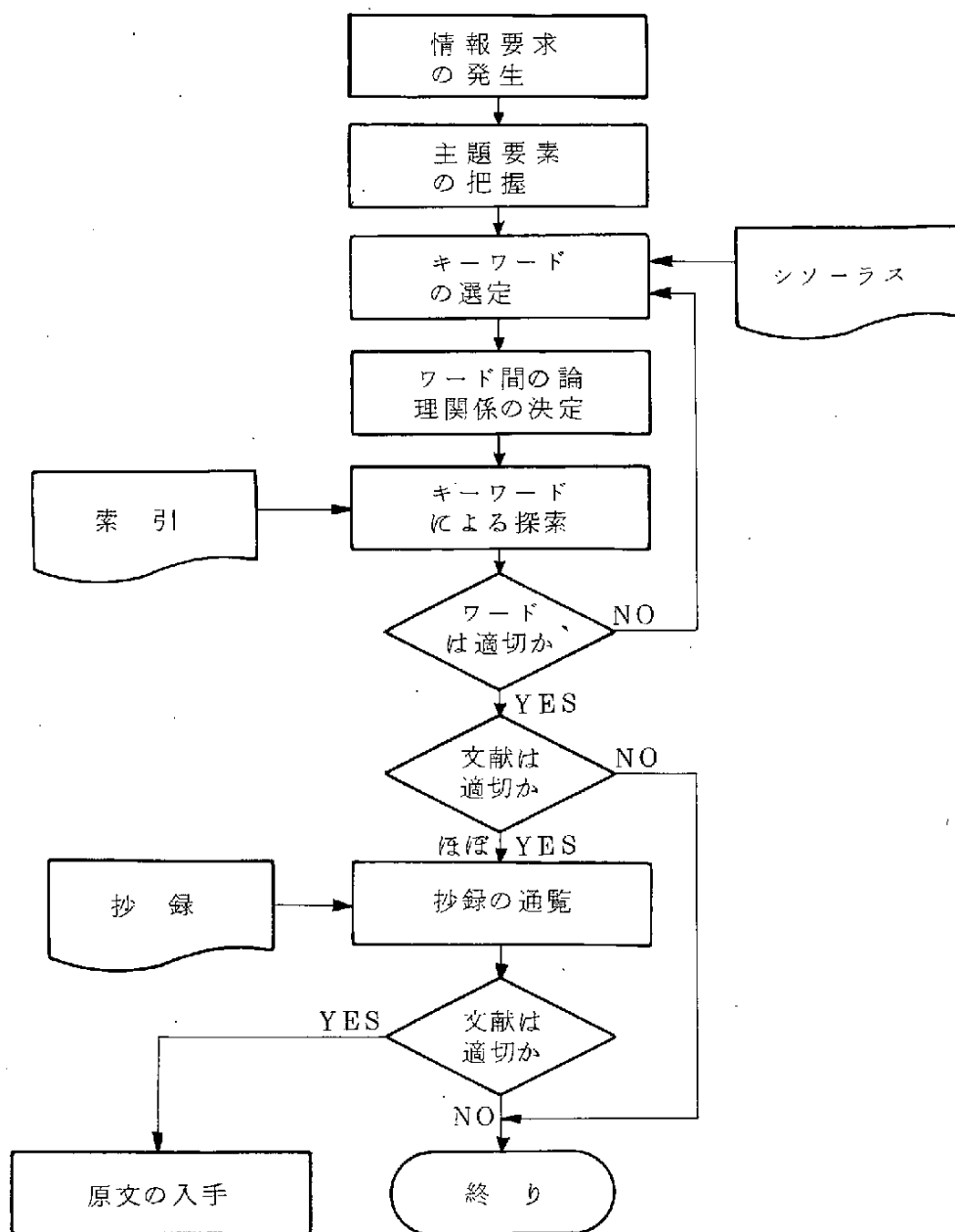


図 2 - 3 文献探索における思考過程

この図からも解るように、探索過程の中には多くの試行の繰返しが含まれている。これは文献検索に特有の性質であって、事項検索と非常に異なる点である。その原因をいくつかあげておく。

- (1) 文献の主題をキーワードにより適確に表現するのが困難である。
- (2) (1)の過程が個人差が大きい。索引作業者と探索者で判断が異なる。
- (3) (1)の理由で蓄積されている文献の品質が不安定である。
- (4) キーワードのみでは適合文献かどうかの判断が困難である。

要するに文献情報という大量の情報を含む対象を、キーワードという極く少量の情報で取扱うために生ずる矛盾なのであるが、人間にとってもコンピュータにとっても、現状ではこれ以外の方法が考えられないのである。一方この方式(タグ方式とでもいうべき方式)でも、探索における試行をいとわなければ、ほぼ満足すべき結果に到達することはできる。

以上のことから情報検索システムにおける、人間とコンピュータとの役割りの分担がはっきりしてくる。つまりシステム設計における要点はここに存在することになる。

さてバッチ検索における探索過程を分析してみると、キーワードの選定および質問の設定、また探索出力から適合文献の選定のプロセスは、まったく人間側の負担において行なわれる。また前述したように、文献検索においては結果に対する信頼性が低いために、検索結果からの判断により検索の繰返しが行なわれることが多いが、バッチ検索は特にこの点において不備である。

バッチ検索が実用的に成り立つのは、2.1.1で述べたようにSDIにおいてである。大量の質問があること、および同一質問いわゆる質問として完成の域にあり試行を伴わない質問を、繰返し使用する場面においてである。つまりこのような意味におけるバッチ処理においてである。

このようにバッチ検索における欠点が目立つのは、オンライン検索と比較されるためである。かつてバッチ検索全盛の時代があったが、情報処理技術の進歩により、その席をオンライン検索に譲りつつあるのが現状である。こ

れは人間側からみれば、より多く人間の手助けとなる方式の方へ、移行しつつあるということになる。

前述のようにバッチ検索においては、探索の前後の過程にノータッチであるが、それなりにこれらの過程に貢献する方式を追及する必要がある。前処理つまりキーワードの選定と質問の設定については、次のようなことが考えられる。まず第一はキーワードのシソーラスによる自動チェック、および非ディスクリプタのキーワードへの自動変換である。キーワードとして認められていない語で誤って質問を設定し、無意味な探索を行なう無駄を防止するためである。

第二はオンライン・バッチ方式である。この方式でターンアラウンド・タイムが極めて短くなれば、バッチ検索でもある程度の会話ができることになる。従来ともすればこの方式への関心が薄いが、企業内システムではこの方式で十分である場面も多いと思われる。

後処理つまり探索出力から適合文献の選定のプロセスについては、人間が出力を通覧して選定する作業がやり易いように、バックアップしなければならない。1つには出力の情報量を増すことがあり、文献番号のみよりは標題と書誌項目がある方がよく、さらに抄録があればなおよい。他は出力情報を読み易くすることで、このためには日本語による出力が望ましい。すでにSDIの出力として日本語抄録を提供して、好評を得ている例がある。さらにマイクロフィッシュによる原文献提供という論もあるが、検索の精度の面からみて行過ぎであろう。

2.4.2 オンライン検索

オンライン会話型検索は結局のところ、人間による探索プロセスの思考過程をシミュレートしたものである。その中で思考部分は人間が、機械的な探索部分はコンピュータが分担する。したがって会話といってもそのつど必要に応じて、作業をコンピュータへ指示するという形をとっている。会話型に

もシミュレートのレベルによって、機能はいろいろ差がある。

ここである会話型システムのコマンドを例にとって、それと人間の探索過程とを対比してみよう。

EXPAND：オペランドとして語を入力すると、その語が関連するシソーラスの部分が表示され、これによりキーワード選定の援助を行なう。会話型システムでこのコマンドの部分が最も複雑で遅れている。その理由は完備したシソーラスが少ないこと、表示の要求が多種多様にわたっていること、これに対応するには高度なファイル処理技術が要求されることなどによる。たとえば「汚染」という語に対して、この語を部分として含むすべてのキーワードを表示するのはたいへんなことである。このへんは人間がマニュアルで行なうとすると、ほぼ絶望的なところであるので、コンピュータへの要求も強いところである。

TRY：選定したキーワードで試みの探索を行なう。該当文献数程度の簡単な情報を表示し、キーワード修正の援助をする。またときにはこの試行探索で探索自身が終了することもある。いくつかの選定キーワードについて試行の後、それらの論理式による試行も可能である。

DEFINE：TRYにより質問式が決定したら、それを登録しておく。探索・表示の結果により再度質問を修正することもある。

LIMIT：キーワード以外の探索キーを指定する。たとえば発行年、使用言語などを指定して探索を有効にする。ここで指定するキーは第二次キーであって、それ自身でTRYを行なうことはできない。したがってシステム設計上、第一、第二キーの決定をせまられ、またこれがその会話型システムの能力に直接影響を与える。

OUTPUT：探索された文献の書誌項目および標題程度を表示し、最終的な文献の選定を援助する。抄録の表示があればベターであるが、これはファイル容量と経済性の問題である。

以上が会話型システムの探索の部分であるが、これで大体思考過程の手順

をふむことができよう。これ以上にレベルアップするには、キーワード間または文献間の意味的関連、統計的関連を加味しなければならないが、これは会話型システム以前の問題である。

なお会話型システムの全機能は、大体次のような機能からなっている。

会話の開始／終了

検索準備：EXPAND・TRY

検索：DEFINE・LIMIT

出力：OUTPUT・ディスプレイ・プリンタ・ラインプリンタ・磁気テープなど

補助機能：検索結果の保存、プライベート・ファイルの設定・使用、出力への演算処理、利用者へのサービス情報の提供など

さて次に本題であるオンライン検索と日本語の関係について考察を進めよう。まず第一にすべてが漢字モードで行なわれれば、それに越したことはないが、リモート処理の場合、現状ではコスト的に問題が多すぎる。一方漢字モードを全く無視するのも会話という面から問題が多い。となると現状で許される範囲内で、日本語を巧みに活用するということになる。ただし漢字処理で最大の問題は、漢字パターンの発生部分にあるわけであるから、ローカルに文字発生装置を多重利用できるシステムの場合には、問題はある程度解消される。

さて会話型システムで日本語が要求される理由は、表示・プリントした出力を読み易くするためである。したがって、まず表示情報の日本語化が優先する。次にコスト面からみると、文字発生部分を制限することである。これは字体の簡略化と字種の制限があるが、当面後者の方が実現性が高い。

以上の観点に立って会話型システムをみてみると、次のようになる。

- (1) 質問情報：キーワードは思考を重ねていく上で日本語がよい。字数は少ないが字種はシステムの全字種に及ぶ。
- (2) 文献情報：通覧する関係上もちろん日本語が望ましいが、字数が多く字

種も多いから実現はむづかしい。前もって印刷物にするとか、マイクロ化して用意しておく方法もある。またセンター側で出力する方法もある。

- (3) システム側のメッセージ：会話の主旨から日本語が望ましい。字数も字種も設計上から制限可能であるから、最も実現の可能性は高い。

なおキーワードに関してであるが、従来のシソーラスはオンライン検索に対する考慮が少なく、自由に漢字を用いているが、その字種をある程度制限することは可能である。たとえば字種を10ビットで表現可能な範囲(1024字)に制限することなども考えられる。

以上の考察から、まずシステム側のメッセージを日本語化することが手始めとなる。その場合の表示の例を次に示す。

*利：◎◎検索システム

計：◎◎検索システムです。

昭和◎年◎月◎日 ◎時◎分◎秒

番号と名前を入れてください。

利：×××××× ヤマダ タロウ

計：登録終わりました。コマンドは？

利：試行 オンラインケンサク

計：該当文献は10件です。コマンドは？

利：検索 オンラインケンサク

計：該当文献は10件です。コマンドは？

*利：表示 標題

(*入力 は英数カナで行ない変換して表示)

この例でもわかるように、ほんの少数の漢字字種を用意することにより、格段に会話機能が向上する。あとはコマンド「表示」による、文献の書誌事

項・標題・抄録などの出力であるが、前述のように前もって配布しておくか、センター側で出力するか、ファクシミリで送信するか、端末側に漢字プリンタを設置しておくかすれば、漢字出力が可能である。または当面英数カナモードでがまんすることになる。

2.5 日本語情報検索のニーズと運用

2.5.1 文献検索のニーズ

まず文献検索のニーズについて、段階的に考察を加えよう。

(1) 文献検索のニーズ

文献検索のニーズについては、これまでに研究開発の面や企業活動の面から、多くの論評が行なわれている。これらの面において先人の残した業績や研究結果を文献により吸収し、活動の効率化をはかろうという行動は、情報の本質的な役割りを示し、また人間活動の当然の帰結であるといえよう。ここにおいて文献をより速かに、より適確にキャッチしたいという要求も当然の帰結であるといつてよい。

ここでニーズの問題を適確に把握するためには、人間行動の本質に由来する情報入手の欲求と、検索システムに対する需要とを、はっきり区別して考察することが重要である。つまり潜在的なニーズと顕在的なニーズとは別問題である。

人間活動は衣食住と同等に情報に対しても常に欲求を持っている。この欲求を分析してみると、

- ① 良質の情報を求める
- ② 即時性への要求が強い
- ③ 必要な情報のみでよい

となる。すなわち「適確な情報のみをすぐ欲しい」という要求であり、専門的にいえば情報検索への欲求ということになる。

これを満たすために、現状では百科事典を備えるとか、電話をひくとか、新聞・雑誌を常に読むとか、ラジオ・テレビを視聴するとかの、いろいろの手段を講じているが、まだまだ不備の点が多い。

たとえば、ある種の情報については、これらの要求が非常に強いというものがある。不動産売買とか求人求職に関する情報などがそれである。し

たがってこれらの分野では、すでに情報検索システムが実働を始めている。今後このような形で必要度の高いものから、順次システム化が行なわれていくことであろう。このような意味からいって、情報検索の潜在的なニーズは人間活動の本質的欲求であり、ますます高まっていくことであろう。

次に検索システムに対する需要となると話はまた別である。需要を支える要素はその便利性と価格である。つまり費用効果と費用便益の多寡が需要を左右する。文献検索については現状では企業活動との関連が主であるから、その面からみると価格よりも便利性の面でより問題が多い。

これまで文献検索技術の理論面について多くの研究開発が行なわれ、実用システムとしては大体固定したレベルに落付いてきた。しかしそのレベルは人間の思考過程と比較すると、非常に低いレベルであるといわねばならない。このようなレベルで固定するとなると、ますますその便利性が追求されることになる。便利性によって効果や便益を生み出すという方向へ向いつつある。

(2) 日本語文献検索のニーズ

日本語の問題は全く便利性の観点からのものである。これまでたびたび述べたように、わが国の文献検索技術は、アメリカの全くの模倣といつてよい。したがって開発の段階においては情報そのものが議論の対象であつて、情報を表記した言語に対する関心は極めて低かつた。しかし実用段階となると別である。検索の結果求めた文献がロシア語であつた場合、人によっては求めなかったのと同じである。

日本語についてはコンピュータのハードウェアの問題もあるが、今後の技術開発によって徐々に解決されると思われるので、日本語化は必然的な方向であり、これによる需要喚起も大いに期待されよう。これは文献検索のみに限らず、情報検索の普遍的なテーマである。要するに言語は文献検索とか情報処理とかと無関係に不変に存在し、したがってそれに無関心であつては、システム自身が無視されることにもなりかねないのである。

結論として日本語化は、ある価格の限界点を割ることにより、文献検索の需要を喚起する点において、オンライン化と双壁をなす要素である。

(3) オンライン文献検索のニーズ

オンライン検索についてはすでに2.4において、その必然性について述べたが、ここでは観点をかえて需要の面から考察してみよう。

まずオンライン検索がいかに需要を喚起するものであるかを、アメリカにおける例によって試みる。NLMでは、医学文献の情報検索システムとして、バッチ方式によるMEDLARSをいち早く開発し、1964年からそのサービスを開始した。その後システムの改良、サービス・ネットの

整備と拡張に努力を重ね、着実にその需要を開拓していった。その間の需要の伸びを図2-5に示す。図からも明らかなように、毎年着実に需要は伸び、ピークの1971年においては年間約2万件にも達した。

一方この間オンライン・システムMEDLINEの開発を進め、1970年からサービスを開始したところ、爆発的な需要を喚起し、1975

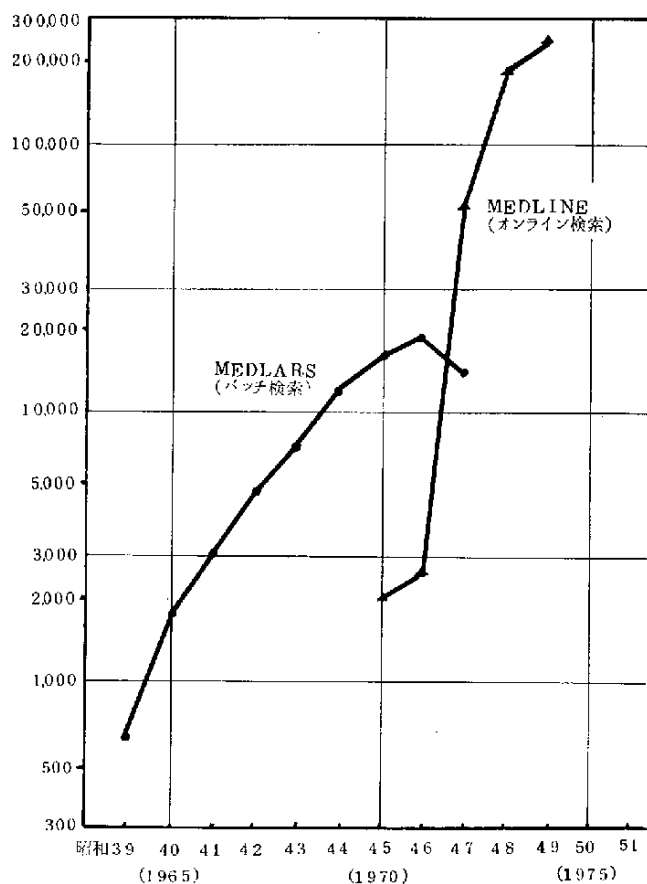


図2-5 オンライン検索の需要の伸び

年には約30万件(推定)にも達している。もっともNLMはアメリカの国立機関であるため、他の商業ベースのサービス(SDC社など)に比して、料金は安くなっている。

この例のようにオンライン化によって需要が全く変化したことは、サービスが改善されたというようなことでなく、両方式の間に本質的な違いがあるとみるべきであろう。その差については、すでに2.4で述べたとおりである。

さてわが国においては、未だ本格的なオンラインによる文献検索サービスは開始されていないが、開始に対する要望とか需要の予測はいろいろと行なわれている。それによるとわが国においても、年間30万件の需要があるとしているものもある。しかし要望と需要とは直接結び付かない危険もあり、割引いてみなければならないが、一種のあこがれムードも含めて需要は極めて強いといえよう。

さてここで問題となるのがそのサービスの内容である。サービスのポイントは、会話機能、端末の利用の便、日本語である。まず会話機能についてはアメリカで現用されている機能で大体満足が得られているので、その流用でよいであろう。次に端末であるがアメリカでは電話を利用し、設置場所は任意となっているが、わが国ではこの状態になるまでは相当制限されるとみなければならない、需要を阻害する大きな要因である。

最後に日本語であるが、前述のMEDLINEによるサービスは、すべて英語に翻訳された情報で行なわれている。わが国においても英語で通用する人々は多いが、当面カタカナによる表記が中心となろう。その場合でも2.4.2で述べたように、最少限の漢字表記は可能なように努力すべきである。

このようにみると、わが国ではまだオンライン検索を阻害する要素がいろいろあるので、当面の需要については相当割引いて考えておくのが無難であろう。アメリカにおける研究歴のある人々の意見を、全面的にう

のみにするのは危険である。このことはすでに抄録誌において実証済みである。

2.5.2 文献検索システムの運用

運用については検索システムの構成によって大きな違いがある。まず文献のソースであるが、すでに機械可読情報となっているソースを導入する場合と、自家製のファイルを使用する場合との違いがある。まず前者については近い将来オンライン・サービスを受けられるようになる可能性が強い。後者についてはシステムの大部分の運用努力は、ソースの作成に費されることになる。これまでにシステムのみ先行して、実用化が伴わなかった例が非常に多い。

次に企業内などのローカルなシステムと、サービス機関とでは大きな違いがある。前者ではある程度の不便をしのいでも、コスト面を重視しなければならず、一般に利用者オリエンティッドな理想に近い形のシステムは実現がむずかしい。一方サービス機関では需要の面からサービスの向上に重点が置かれ、これをスケール・メリットで補っている。しかしコンピュータのコスト・プッシュによって、他のサービスに比して割高となっている。

ここで登場するのがオンライン化であり、スケール・メリットを生かしたコスト・ダウンによって、文献検索の問題を解決の方向へ持っていくことを試みる。したがって公開された文献情報に関しては、このような方向は必然的な成りゆきとみななければならないであろう。かつて文献の抄録についてもこれと同じような歩みがあり、現在ではローカルな抄録作業はほとんどみられなくなってしまった。

結論的にいうと、ローカルなシステムは、そこに特有な情報のみにとどめ、公開情報は情報機関に頼るという傾向はますます強まるであろう。またローカルなシステムも、文献検索向きのシステムを用意したTSSセンターの利用によって、運用されるようになるものと考えられる。

このような状況となると、情報の機密保持、プライバシーの保護、情報独占に対する監視などという問題が、運用上の最重点になるとも考えられる。

最後に日本語情報に関する運用上の問題点をまとめておこう。日本語の処理についてはまだ多くの開発テーマが残っている。これらが実用的に解決されて始めて、日本語文献検索システムの開花期を迎えることとなる。このためにも多くの要求を出して、開発を促進させるという主導性が必要である。

次に日本語情報処理については、その標準化が非常に遅れている。現在進められている漢字コードのJIS化の次に、情報処理上の多くの標準化が必要である。さし当っては、入出力装置・端末装置の標準化を進めることが、文献検索システムにとって、その普及を促進させる要点である。

またシステム運用上の問題点として、その入力作業の改善も見逃すことはできない。日本語処理の進歩に対して、入力作業は活字時代、写植時代のままだに取残されている。この作業の革命はコンピュータ処理の問題のみに止まらず、印刷産業にも革命をもたらすものである。

2.6 日本語情報検索システムの開発動向

2.6.1 国立国会図書館の索引編集システム

(1) 図書館業務の索引誌作成

国立国会図書館における図書館業務の機械化は、1969年の準備室設置、翌年コンピュータおよび漢字プリンタ・システムの導入以来、日本語情報処理に長い経験を有している。

国会図書館で機械化された業務のうち、とくに日本語情報処理に関係するのは、

- ・国会会議録総索引の編集
- ・和雑誌所蔵目録の編集

である。前者については、簡単な情報検索システムも一時実験されたが、漢字データが適量に満たないことなどから中断され、将来計画として残されている。現在までは再び編集業務が中心に処理されている。

索引誌や所蔵目録の編集は、いわば情報検索用データ・ファイルの作成に相当する作業でもあり、実際の検索・索引はこれらの索引誌を利用する読者によって行なわれることになる。

(2) 国会会議録総索引の編集

国会会議録とは、衆・参両院本会議、常任委員会、特別委員会における会議の速記録を整理・収録したものである。国会図書館では、これらの速記録をもとにして探索に便利な各種索引を自動編集して、国会会議録総索引を作成している。

会議録総索引の構成は次のようなものからなっている。

- ・案件索引……………本会議および委員会の会議に付された案件を、両院の議事の進行に即して会期ごとに配列し、会議録の掲載号、頁を指示したもの。
- ・発言者索引……………すべての発言者の発言を収録し、会期を通して、五十音順あるいは官公庁順に配列し、会議録の発言箇所の指示、発言内容を示

す事項名を挙げたもの。

- ・ 事項索引……………発言内容を表現する事項名をすべて見出し語とし、その見出し語の下に、会議録掲載号、月日、頁、発言議員氏名を記載し、会期を通して、五十音順に配列したもの。検索を便利にするために「事項名一覧表」を作成してある。

- ・ 法律案等審議経過表……………法律案、条約、予算、決算等の案件について、提出番号順に付託年月日、審議経過、結果および公布年月日、公布番号等を記載し、さらに、会議録の掲載号、頁を示して一覧表としたもの。

これらの索引のうち、コンピュータを利用して自動編集しているのは、発言者索引、事項索引および事項名一覧表である。これ以外の案件索引と法律案等審議経過表等については従来どおり手作業で作成している。

国会会議録の入手から索引誌の完成までの流れ図を図2-5に示す。会議

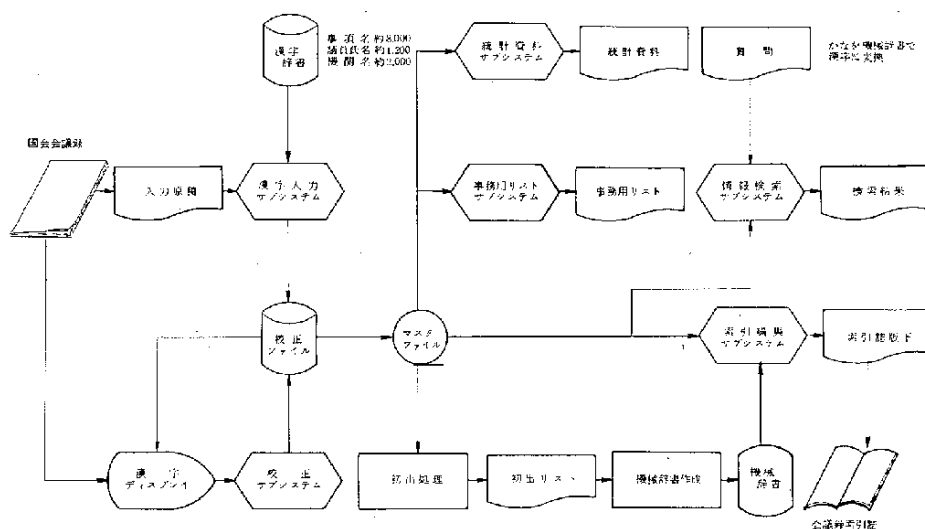


図2-5 国会会議録編集システムの概要

録に対する入力原稿の作成は、館内の専門の職員が行なっている。また、索引編集システムのソフトウェアは、全体で約60本のプログラム（合計約3万ステップ）からなっている。

コンピュータ処理によって自動編集された国会会議録総索引の第1号は、1972年6月に刊行されており、これは1968年度に開かれた、第58回から第60回の国会分の会議録索引であった。自動編集によって作成された各索引の例を図2-6、図2-7、図2-8に示す。

この索引誌は、第39回国会分から刊行がはじめられているが、刊行当初、わが国初の国会議事録の索引として脚光をあびたものである。索引誌はその後も順次刊行されており、最終的には第1回国会（1947年）までさかのぼって作成することが計画されている。

なお、この索引誌編集にあたって、ふりがな、分類コードの復号化、キーワードの統一などのために、つぎの3種類の変換用辞書を作成して利用している。

- ・事項名辞書（約8,000項目）
- ・人名辞書（約1,200項目、国会議員名）
- ・機関名辞書（約2,000項目、行政機関名）

(3) 和雑誌所蔵目録の編集

国立国会図書館では、国内で刊行された全雑誌（同図書館所蔵分）の目録作成、各種データ・リストの編集の自動化も行なわれている。雑誌目録の作成は、一見単純な作業のように思われるが、実際には、雑誌の刊行ペース（月刊、季刊、……etc.）が異なるばかりか、雑誌の休刊、廃刊、増刊、改題など変化にとんだ複雑な作業である。

和雑誌所蔵目録編集システムは、これだけの単独システムではなく、図2-9に示される大規模な和雑誌処理システムの一環として計画され、開発されたものである。

衆議院議員・参議院議員

(あ)

相沢重明君(参・無)

- 58参建 4(3.7)12 砂利資源
5(3.12)5 し尿処理場(厚木)
7(3.28)7 有料道路(鎌倉)、有料
道路料金
8 道路管理
18(5.14)11 関法152〔55〕、関法
56〔58〕、都市計画
12 日本住宅公団
13 団地建設、電波障害指定区域
14 米軍本牧住宅
20(5.16)8 広域行政
9 政令指定都市、米軍基地
10 東京湾開発

愛知揆一君(衆・自)

国務大臣の項を見よ

青木一男君(参・自)

- 58参予 4(3.22)22 関予1〔58〕、関予2
〔58〕、関予3〔58〕、三派全学連
28 阿波丸事件

青木正久君(衆・自)

- 58衆本 26の1(4.19)10 小笠原、沖縄、関法
104〔58〕、関条18〔58〕、北方領
土、防衛
38(5.22)25 関条9〔58〕委員長報
告、関条10〔58〕委員長報告、関条
16〔58〕委員長報告、関条17〔58〕
委員長報告、関条8〔58〕委員長報
告

衆地 32(5.23)7 関法84〔58〕、消防
8 消防(高層建築)

衆物特 1(1.30)1 委員長推薦動議、理事
指名動議

59衆物特 1(8.2)1 委員長推薦動議、理事
指名動議

60衆物特 1(12.11)1 委員長推薦動議、理事
指名動議

青島幸男君(参・第二)

- 59参予 2(8.10)30 佐藤内閣、選挙制度

青田源太郎君(参・自)

58参法 20(5.23)6 関法93〔58〕附帯決議
案説明・討論

参災特 1(12.27)1 委員長指名動議

8(5.23)2 派遣報告

19 委員長推薦動議

政府委員、説明員(農林省)の項をも見よ

青柳秀夫君(参・自)

- 58参本 5(3.1)6 関法1〔58〕委員長報
告
7(3.13)14 関法15〔58〕委員長報
告
9(3.30)30 関法22〔58〕委員長報
告、関法31〔58〕委員長報告、関法
32〔58〕委員長報告
11(4.10)8 関法5〔58〕委員長報
告
14(4.19)92 関法6〔58〕委員長報
告、関法7〔58〕委員長報告、関法
34〔58〕委員長報告
15(4.24)27 関法3〔58〕委員長報
告、関法4〔58〕委員長報告
17(4.27)12 関法14〔58〕委員長報
告
18(5.8)16 関法47〔58〕委員長報
告
21(5.15)13 関法20〔58〕委員長報
告
14 関法21〔58〕委員長報告
22の1(5.17)20 関法16〔58〕委員
長報告
23(5.22)13 関法36〔55〕委員長報
告

参大

- 2(2.29)1 理事の補欠互選
5(3.12)8 関法15〔58〕討論
7(3.22)1 参考人出席要求
8(3.26)1 理事の補欠互選
9(3.28)11 関法22〔58〕討論
12 関法47〔58〕討論
14(4.12)1 理事の補欠互選
19(4.25)1 理事の補欠互選、連合
審査会、参考人出席要求
25(5.16)1 理事の補欠互選、参考
人出席要求
26(5.17)1 理事の補欠互選
27(5.21)5 請願
28(5.24)6 継続調査要求
29(6.3)1 請願

赤沢正道君(衆・自)

国務大臣の項を見よ

赤路友蔵君(衆・社)

- 58衆農 15(4.23)17 漁港整備
16(4.24)10 原子力発電所

図2-6 国会会議録総索引から発言者索引の例

事項索引

(あ)

IMF

58衆外 14(4.23) 6 田中武夫君(衆・社)

IMF蔵相代理会議(昭43.3)

58衆大 8(3.12) 6 武藤山治君(衆・社)

IMCO

58衆運 7(3.19) 5 米田東吾君(衆・社)

ILO公務員部会

58衆地 8(3.14) 11 山口鶴男君(衆・社)

ILO事務局長モース氏書簡(1967.2.6 外務大臣宛)

58衆予 10(3. 1) 36 久保三郎君(衆・社)

ILO条約

60衆社 3(12.20) 3 山田睦目君(衆・社)

ILO条約第26号

58衆本 15の1(3.26) 9 田畑金光君(衆・民)
参本 11(4.10) 19 鬼木勝利君(参・公)
11(4.11) 15 梅岡秋夫君(参・社)
衆社 18(4.23) 4 海部俊樹君(衆・自)
19(4.24) 2 枝村要作君(衆・社)
22(5. 7) 12 加藤万吉君(衆・社)
23(5. 9) 10 加藤万吉君(衆・社)

ILO条約第30号

58参社 14(5.16) 6 藤田藤太郎君(参・社)

ILO条約第32号

58衆運 5(3. 8) 10 米田東吾君(衆・社)

ILO条約第114号

58衆運 6(3.15) 9 久保三郎君(衆・社)
衆予 10(3. 1) 36 久保三郎君(衆・社)

ILO総会(第52回)決議

59参社 1開(9.17) 28 小野明君(参・社)
2開(10. 8) 14 小野明君(参・社)
4開(11.21) 11 小野明君(参・社)
60衆社 3(12.20) 3 山田睦目君(衆・社)

相川簡易裁判所(移転)

58参法 19(5.21) 1 亀田得治君(参・社)
秋山長造君(参・社)

愛国心

58衆文 11(4.19) 4 有島重武君(衆・公)
15(5. 8) 8 有島重武君(衆・公)
参文 4(3. 7) 11 楠正俊君(参・自)
参決 8(4. 3) 2

アイシン・ワナー合併会社

59衆商 2(8. 9) 2 板川正吾君(衆・社)

愛知県労働金庫一宮支店

58衆大 25(4.17) 12 石田幸四郎君(衆・公)

愛知少年院

58衆内 4(3. 8) 3

愛知用水

60衆農 3(12.19) 16 佐々栄三郎君(衆・社)

愛知用水公団

58衆大 5(3. 5) 11 村山喜一君(衆・社)
衆建 18(4.24) 12 佐野憲治君(衆・社)
19(4.26) 3 内海清君(衆・民)
阿部昭吾君(衆・社)、石川次夫君(衆・社)、佐野憲治君(衆・社)
20(5. 8) 1 井上善方君(衆・社)
衆予第4分科 2(3.13) 28 加藤清二君(衆・社)
参建 20(5.16) 23 沢田政治君(参・社)
21(5.21) 15 田中一君(参・社)
沢田政治君(参・社)、鈴木一弘君(参・公)
59衆社 2開(8.27) 7 大橋敏雄君(衆・公)
3開(9. 9) 8 大橋敏雄君(衆・公)

会津線(雪害)

58衆運 8(3.22) 1 唐橋東君(衆・社)
衆災特 4(3.13) 1 唐橋東君(衆・社)
6(4. 3) 5 川村経義君(衆・社)

青色申告

58衆予第3分科 1(3.12) 9 只松祐治君(衆・社)
参決 2(2.28) 22 柴谷要君(参・社)

青空職業紹介(横浜)

58衆予第3分科 3(3.14) 3 大出俊君(衆・社)

赤川

59衆災特 3開(9. 6) 17 森田重次郎君(衆・自)

赤字バス会社

図2-7 国会会議録総索引から事項索引の例

憲 法

憲法第9条
憲法第21条
憲法第22条
憲法第25条
憲法第27条
憲法第29条
憲法第73条
憲法第85条
憲法第87条
憲法第92条
憲法第99条
自衛権
財産権
国家賠償法
国家賠償
国家賠償（松川事件）
憲法改正
公式制度連絡調査会議
元号制度

皇 室

天皇の海外旅行
皇室典範
皇室経済
皇室費
皇室費（外国旅費）
皇室私有財産
皇居造営
新宮殿
皇居移転
下田御用邸
下田御用邸用地
高輪旧御用邸払い下げ
沼津御用邸
下総御料牧場
高根沢御料牧場
恩赦法
恩赦（選挙違反者）
恩赦（明治100年）

国際法・外交

戦時国際法（追跡権）
追跡権
安全保障
安全保障（極東）

安全保障（東南アジア）
安全保障体制
集団安全保障体制
公海
領海
領海（無害通航権）
領海3海里
領海12海里
領海12海里（ソ連）
専管水域（ニュージージーランド）
接続水域
領空権
領空侵犯
米軍機（領空通過）
大陸棚
大陸棚（日本海）
大陸棚開発
国際海洋法会議

条 約

世界人権宣言
世界人権宣言20周年記念式典
国際人権規約
児童憲章
難民の地位に関する国際条約
外交関係に関するウィーン条約
国際商航海水線条約
国際商航海水線条約（1930）
国際商航海水線条約（1966）
アジア＝オセアニア郵便条約
米比条約
日本国との平和条約第3条
通商航海条約
日米安全保障条約
日米安全保障条約（延長）
日米安全保障条約（極東の範囲）
日米安全保障条約第4条
日米安全保障条約第6条
日米安全保障条約第10条
事前協議（安保条約）
事前協議（核兵器持込み）
事前協議（在日国連軍の行動）
随時協議（安保条約）
日米安全保障協議委員会
地位協定
地位協定（国連軍）
地位協定（米軍人）
地位協定第5条
地位協定第7条
日ソ平和条約
アンザス条約

国際連合

国連
国連憲章前文
国連憲章第55条

国連憲章第102条
国連総会決議（朝鮮問題）
中国代表権
重要事項指定方式
難民救済
在日国連軍
軍縮
18カ国軍務委員会
国連貿易開発会議

外 交

国際情勢
外交姿勢
非武装中立
非武装中立宣言
外交政策
佐藤総理外遊（昭42）
ド・ゴール大統領
請求権（オーストラリア）
請求権（オランダ）
米中会談（ワルシャワ）
プエブロ号事件
チェコ動乱
オーストリア賠償

日米外交

日米安保体制
日米関係
ニクソン政権
佐藤・ジョンソン会談
佐藤・ジョンソン共同声明（昭42.11）
吉田・アチソン交換公文
日米協議委員会
日米協議会（昭43.9.12 基地問題）
阿波丸事件

日ソ外交

北方領土
北方領土（根室市財政援助）

アジア外交

アジア外交
アジア外交（ベトナム戦争後）
アジア政策（アメリカ）
アジア太平洋関係会議
アジア情勢
極東の平和と安全
極東の範囲
A S P A C
日韓外交
日韓関係
尖閣列島
日韓閣僚会議（昭43.8）
韓国人（日本人妻）
日台関係
朝鮮問題

図2-8. 国会会議録総索引から事項名一覧の例

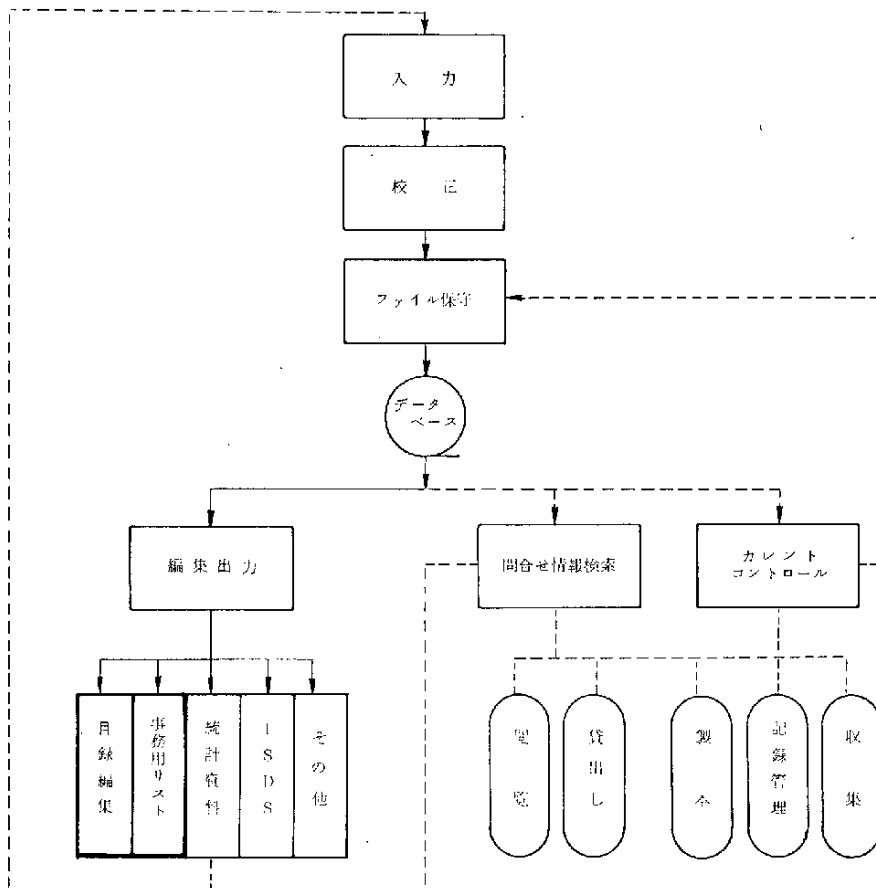


図 2 - 9 和雑誌処理システムの全容

コンピュータ処理による最初の和雑誌所蔵目録は、1974年に刊行された。これは1973年（昭和48年）版であり、昭和48年12月末までのデータが収録され、以後、常時マスタ・ファイルの更新を行なって、3ヶ月ごとに補遺版（累積形式）が刊行されている。昭和50年10月現在27,613件の雑誌が登録されている。

和雑誌所蔵目録の編集途中のゲラ刷りを図2-10に示す。

愛知図書館協会会報

2~28号（昭26.3~33.12）

Z21-528

欠多し

(ID: 26781)

葵 静岡県立中央図書館葵文庫

2~21号（昭35.12~44.2）

Z21-548

別誌名：静岡県立中央図書館報

以後「文化」と改題

(ID: 27121)

葵 静岡県立中央図書館

「文化」より復題

1号（昭49.9）~

Z21-550

別誌名：静岡県立中央図書館報

(ID: 27122)

蒼い共和国

→ 季刊蒼い共和国

青森県議会情報と資料 青森県議会議事録局

1号（昭49.7）~

Z2-456

(ID: 26782)

青森県教育こうほう

→ 教育こうほう

青山語文 青山学院大学日本文学会

1号（昭45.12）~

Z13-1526

(ID: 27556)

図2-10 和雑誌所蔵目録のゲラ刷り

(4) 日本語MARCの開発

国会図書館は、アメリカの議会図書館（LC；Library of Congress）を
ならって設立された図書館であるが、アメリカ議会図書館のMARC
（MAchine Readable Cataloging：機械可読目録）テープ形式の、日本語
MARCを計画中といわれる。MARCは、アメリカ議会図書館が1969
年に開発した図書目録であり、その後改良が加えられてMARCⅡになっ
ている。このフォーマットは、書誌データ伝達用のアメリカ規格となり、さら
に国際規格案となっている。

日本語MARCは、日本語の図書目録の書誌記述が標準化されることによ
り、完成のあかつきには日本の図書館業務のみならず、日本語情報処理全般
に大きな影響を及ぼすものと期待されている。

(5) ハードウェア環境

国会図書館のコンピュータ・システムは、1970年に、HITAC-8400
（65KB）を中心にJEM-3800 漢字処理システムとともに設置された。そ
の後1973年および1975年に機器の増設、機種変更が行なわれ、現在
では、以下のような大規模なシステム構成に拡張されている。

- | | |
|------------|---|
| ・中央処理装置 | HITAC-8450（262KB） |
| ・周辺機器 | 磁気テープ装置……………8デッキ |
| | 磁気ディスク装置……………9モジュール |
| | ラインプリンタ……………2台 |
| | カード読取装置、紙テープ読取せん孔装置、コンソ
ール・ディスプレイ（プリンタ付）……………各1台 |
| ・日本語情報処理機器 | JEM-3100Ⅱ/D（20KW） |
| | ディスプレイ・ターミナル……………6台 |
| | ソフトプリンタ……………1台 |
| | 漢字ラインプリンタ JEM-7600……………1式 |
| | 漢字テレタイプライタ……………11台 |

(磁気ドラム, 磁気テープ, データライタ, 紙テープ読取装置……各1台)

2.6.2 電子技術総合研究所の木かけシステム

(1) 日本語索引の検索システム

電子技術総合研究所では, 自然言語処理の研究が多方面かつ長期間にわたって行なわれている。当研究所ソフトウェア部言語処理研究室では, 1972年から1974年にかけて「国研データ自動索引計画」を遂行した。

「国研データ自動索引計画」は, 1972年3月に国立国語研究所が公開した新聞語彙調査データをもとにして, 大規模な日本語用例KWIC索引を, コンピュータを利用して自動作成するものである。しかしながら, データ量がぼう大なため, 索引リストをとるには全体で約750,000行にも達することになる。そのため, 印刷物として出力するかわりに, 情報検索システム化とマイクロフィッシュ化の計画が実行された。これを木かけシステムと呼ぶ。

木かけシステムは, 全体が日本語中心であること, 索引の検索を対象にしたこと, ミニコン・ベースのシステムであること, などに特徴をもっている。

(2) システムの設計方針

索引データは, 原則として磁気テープに記録され, その内容について検索を行なうシステムが木かけシステムである。

索引全体を印刷物として出力する代りに, このシステムを使うのであることから, あたかも本のページをめくって読んでいくような感触の, 使い易いシステムの開発がめざされた。しかしながら, 2次記憶装置として磁気テープを, 漢字入力装置として旧来の漢字テレタイプライタを使用しており, そのためハードウェア面の制限が, システム設計に大きな影響を与えている。

システム開発の基本方針は, 以下の4項目であった。

- ① 日本語を活用したシステムにすること。また言語学者など, 情報処理を専門としない人にも使い易いシステムにすること。

いわゆる、検索システムと利用者との間のインタフェースは、漢字ディスプレイ装置を活用して日本語で行ない、しかも利用者にとって操作のめんどろな漢字入力キーボードを、できるだけ使わなくてすむように配慮する。すなわち、検索用コマンドの入力はメニュー方式（あらかじめ画面に表示されているものの中から、必要なものを選択指示する方式）とし、検索用キーワードの入力は、ひらがなで行なうようにすることである。

② 磁気テープに記録されている索引の内容を、本を読むような感覚で検索できること。

利用者が選択・指定する動作は、ライトペンを使って、検索、ページ送り、出力、使用法表示、終了の5つとする。これらの動作指示は、任意の順序で指定できるが、検索動作に対しては検索用キーワードを与え、出力動作に対しては出力装置の選択をする、などの入力操作が必要である。

③ 2次記憶装置（磁気テープ）には、索引結果をファイルとして記録する。

検索対象のデータとしては、漢字かな混り文の索引を想定し、できる限り多様な検索が行なえるように配慮する。

④ 検索システムの管理者と利用者との間の責任分担をはっきりさせる。利用者はあくまでも索引サービスを受ける者として、複雑なハードウェアの操作やソフトウェアの細部についてはシステム管理者にまかせる。

(3) システムの使用手順

木かげシステムが動作する漢字入出力装置の全容を図2-11に、システム動作の流れを図2-12に示す。

木かげシステムのソフトウェア、ハードウェア両面を管理・担当するシステム管理者は、システム・プログラム全体の管理を行っており、利用者の依頼にもとづいてシステムの生成を行なう。

利用者が、オペレーションを行なう漢字ディスプレイ装置の前に着席すると、システム管理者の準備によって、あらかじめ図2-13の画面が表示さ

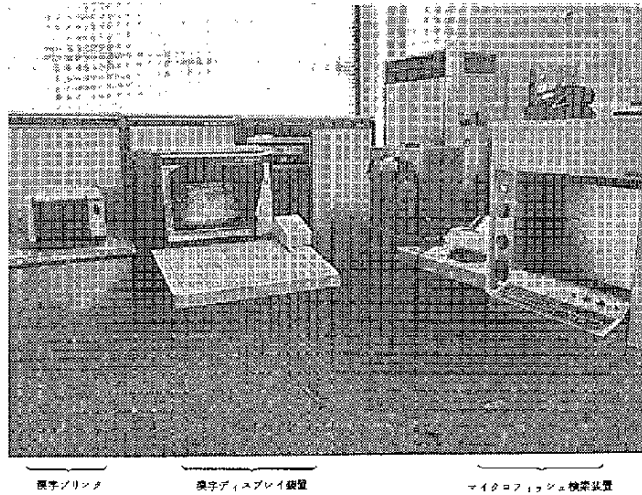


図 2-11 木かげシステムのハードウェア環境

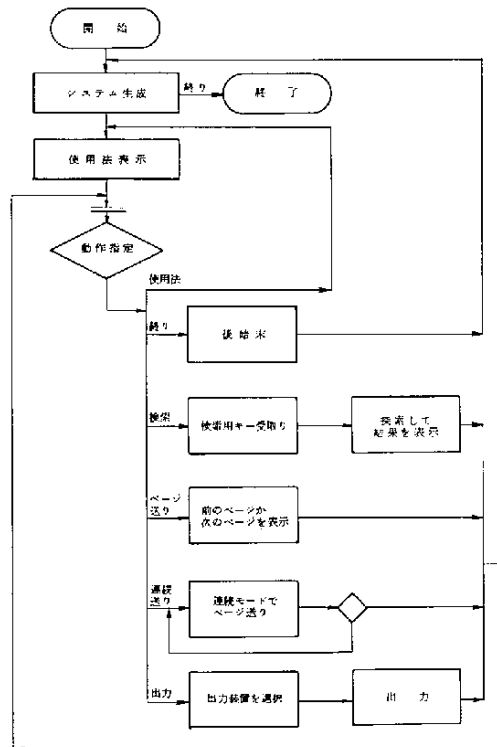


図 2-12 木かげシステムの動作の流れ

れている。画面構成は16行の情報からなり、そのうち上部15行の領域を内容領域、最下行1行分の領域を動作指定領域として使用する。図2-13の画面では、内容領域に使用法が、動作指定領域には基本動作指定行が表示されている。利用者は、この基本動作指定行に示されている中から、任意の動作を指定できる。指定方法は、ライトペンでカーソルを移動させて、制御ボックス上のLP(ライトペン)ボタンを押下するという手順によって行なわれる。動作指定は、間接的に日本語を介して行なわれるが、原則として漢字そのものを入力しなくともよい選択方式である。

最初に利用者が与える動作指定は、通常は「検索」である。これを指定す

画面①

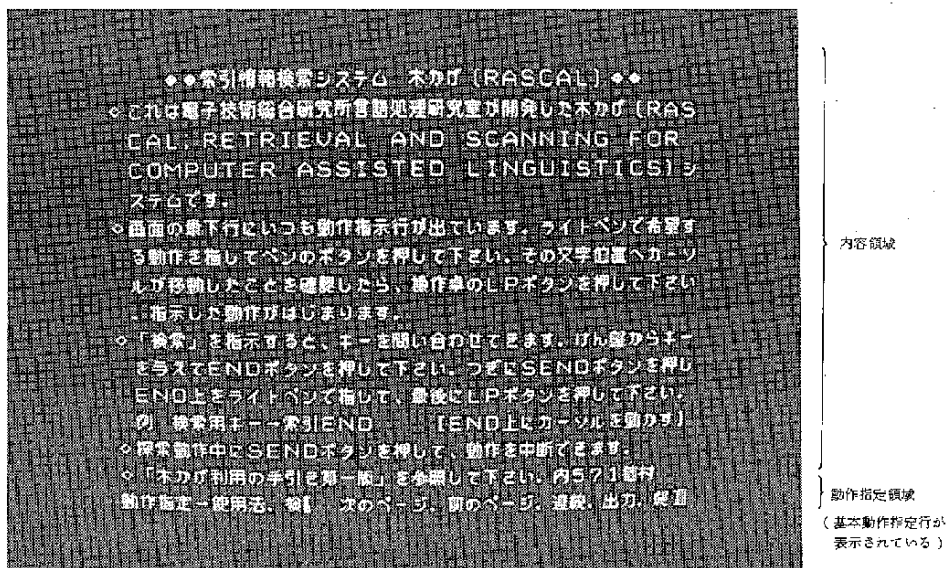


図2-13 木かげシステム起動後の第1画面

ると、動作指定領域に補助動作指定行①が表示される。この画面に対する応答については、利用者は漢字入力けん盤から検索用キーワードをひらがなで入力しなければならない。キーワードを入力したあと、再度LPボタンを押すと磁気テープの探索動作が開始される。なお、検索用キーワードと漢字デ

ディスプレイ画面のどの部分とを比較するかは、システム生成時に前もって指定しておく。また、磁気テープの探索であるために、検索性キーワードとして与える単語によっては、かなりの探索時間がかかることがある。このとき、利用者に不安感を抱かせないように、磁気テープの各ブロックの先頭レコードの内容を順次漢字ディスプレイ画面に表示して、実行中であることを知らせるようにしている。

利用者が指定した検索性キーワードと一致するブロックが見つかるか、あるいは見つからなくても検索性キーワードの内部符号に最も近いブロックでシステムは停止し、画面の内容領域にはそのブロックの内容が表示され、動作指定領域には基本動作指定行が表示される。図2-14には、「けいさんき」による検索性手順および結果の表示例を示した。

一方、検索性した情報をハード・コピーとしてプリント・アウトする機能も必要であり、「出力」の動作指定はこのために使用される。動作指定領域に表示されている基本動作指定行の中から「出力」をライトペンで指示すると、補助動作指定行②が表示される。利用者は、図2-14の画面④に示される補助動作指定行②の中から、出力装置を選択する。いずれを指定しても、内容領域に表示されている内容だけが出力され、このうち「ハード・コピー」を指示すると漢字プリンタに画面の内容が印刷される。

しかしながら、検索性結果が1画面の内容領域におさまることは少なく、そのため利用者は「次のページ」あるいは「前のページ」を指定して、前後いずれかの15行分の情報を内容領域に表示することができる。図2-14の例では、画面③に対して「前のページ」を指定した場合の結果が画面④である。このようにして、次々と前のページへもどってもよいし、次のページへ進んでもよい。また、「連続」ページ送りを指定することもできる。

検索性操作の工夫としては、ふりがなによる検索性の他に、補助的な機能として磁気テープ探索を中断させる機能がある。

圖面(3)

5代巻*音楽事務乃	計算係	茶舞3級程度18歳-3	C64008
*お好みカクテルを	計算機	が課合(P4S1T1)	A08707
1T1)*加藤機・	計算機	・海合計算・olymp	A53283
し星/北川敏男通風	計算機数学夜話	ノ森口繁一数学	A35593
【一八六八年】から	計算し	た百周年日は昭和四十三	A30682
・商品別に書き上げ	計算し	たものどろ、それによる	C49803
校は人學院でもまだ	計算し	ていないが、非普通職校	C02933
じめた。*地合いを	計算し	てみて、やはり足りないと	C30515
年の一戸当り単位で	計算し	ても約三兆円(うち政府	A06078
・配達十二名。女子	計算事務	・成徳五名高大学35	A09818
数学*一口入門」	計算尺	による乗算)・1口*打	J06109
ような反応はすべて	計算ずみ	のことであって、いつ	A02461
・ホームラン記録も	計算ずみ	*「そうなること	C26616
P4)・えらい。	計算書	・すねか、それから	K30860
以上の商品目を含む	計三巧点	近くの商品が出品され	A52522
製品を第一使用天、	地金、次のページ、	前の頁、リ、建設、	出力、軍事

「前のページ」を
指定している

画面④

する(P3)★新井	警察庁長官	は二十二日、全国の	C84961
、国家公安委員会の	警察長官規則	で、1口警察功分	J57655
ていたそうだが、	警察長官	について説明して下	J57655
の集金日に、9月の	計算	ご使用量と、9月に集金に	J13146
子事務員若千名職種	計算	事務員格中平20歳★定年	C35324
ル、これを計量制で	計算	すると九十八円となって、	C23959
、居った通りの演技	計算	でいけると思うた」と早く	J02753
らぞりぞり聞に合う	計算	ではあるが、実は、この手	A36470
である。★この誤さ	計算	にいたた嶋方氏は、無記名	C62466
った意図がどうい	計算	に立っていたかは別として	C08820
、鉄鋼輸出の動化を	計算	に入れたも、年間二〇パー	K87006
もめてはまる、この	計算	もあるようだ。★	A85101
◇一〇入門「分枝の	計算	」同業券多25たのしいフ	K54567
★等号と括弧のない	計算	ノ大矢直一★怒るべき受難	A35533
★一〇入門整数の	計算	★00フランス語入門丸山	K48290
動作指定一使用法、検索、次のページ、前のページ、連続、F7、終了			

画面の15行分を表示

出力を指定している

画面⑤

もめてはまる、この	計算	もあるようだ。★	A85101
◇一〇入門「分枝の	計算	」同業券多25たのしいフ	K54567
★等号と括弧のない	計算	ノ大矢直一★怒るべき受難	A35533
★一〇入門整数の	計算	★00フランス語入門丸山	K48290
出力装置一磁気テープ、紙テープ、ハードコピー、その他の装置			

補助動作指定行②

ハードコピーを指定している

図2-14 「計算機」の検索例

さらに検索手法としては、つねに表示画面を把握・管理する機能を有し、利用者が検索用キーワードを与えると、前回の検索結果を表示した画面の情報とまず比較を行なって、探索動作の前に磁気テープを巻戻すかどうかを判断して、検索の効率をはかっている。

利用者が基本動作指定行の中から「終了」動作を指定すると、検索システムの処理作業が1区切り終了する。次の処理は、システム管理者と利用者とが相談して決めなければならないが、もし同じ形式の別ファイルを使って処

理するならば、磁気テープを取り替えるだけで簡単に再開できる。また異なる形式のファイル进行处理するためには、システム生成の段階にもどって最初から再実行する。

(4) データ・ファイル

1976年2月現在、木かけシステムに蓄積されているデータは、以下のとおりである。

- ・新聞データの用例K W I C索引……………全体で744,248行(レコード)
- ・新聞データの語尾配列用例K W I C索引……………ニュース部分だけ、
176,810行(レコード)
- ・新聞データのE X C O R D索引……………ニュース部分だけ、176,810行
(レコード)
- ・漢字ふりがな長単位語表……………96,720語
- ・語尾配列長単位語表……………96,720語

ここに新聞データというのは、1966年度の朝日、毎日、読売3紙の朝夕刊全紙面から、1/180に標本抽出したデータをいう。その後の新しいデータの追加は行なっていない。

(5) システム構成

木かけシステムを運用するハードウェアの構成概要は次のようである。なお、漢字パターンはドット式で表現され、U I 変成器型固定メモリに記憶されている。

- ・漢字ディスプレイ装置……………16インチのブラウン管に512字(32字×16行)の文字を表示できる。ライトペンによる簡単な編集機能付き。
- ・漢字入力装置……………漢字テレタイプライタ型の入力けん盤。2,600字種。
- ・漢字プリンタ装置……………感熱方式の漢字プリンタ(サーマル・プリンタ)で、漢字ディスプレイの1画面をハード・コピーする。
- ・磁気テープ装置……………1200フィート・テープまで装着可能。

制御装置……… OKITAC-4300 C。サイクル・タイム-0.6マイクロ秒
／語。記憶容量-16 K語(16ビット／語)。

木かげシステムのソフトウェアは、制御装置(OKITAC-4300 C)の
アセンブリ言語 OKISAP-43 で作成された。プログラム・ステップ数約
4,800(命令), サブルーチンの個数にして約38個である。

(6) 日本語COMによるマイクロフィッシュとその検索装置

この研究計画では、用例索引を漢字COMによってマイクロフィッシュ化
する実験も行なっている。漢字COMで作成したマイクロフィッシュは、図
2-15に示す自動検索装置に収容されており、必要な部分をただちに読む
ことができる。

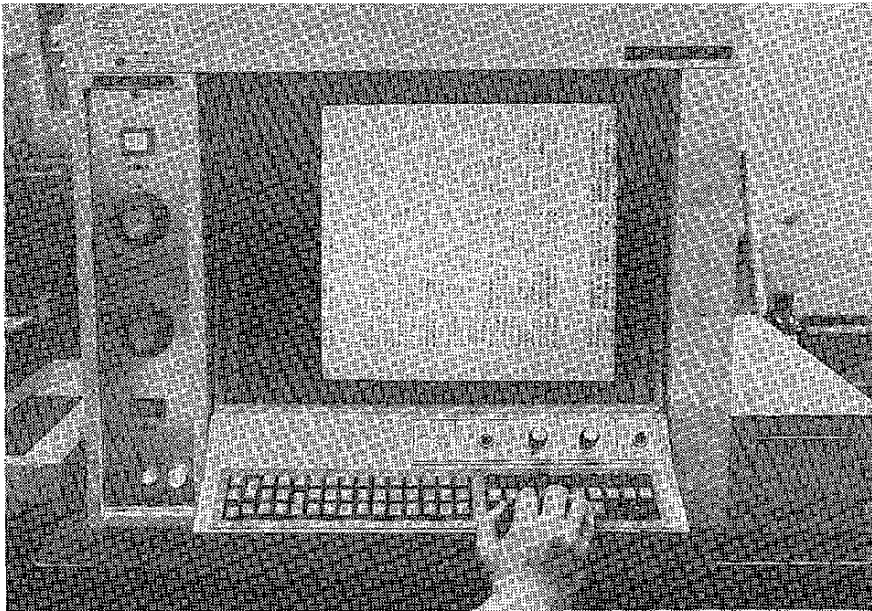


図2-15 日本語マイクロフィッシュと検索装置

2.6.3 通産省情報管理課の日本語情報検索システム

(1) 通産省における情報検索

通産省では、昭和37年ころからコンピュータを利用した情報検索の必要性が検討されており、当初開発された情報検索システムは英数字モードで処理されていた。その後、昭和47年に漢字入出力装置が導入されて以来、漢字モードの情報検索システムが検討されて、ここに日本語情報検索システムが開発された。

このシステムの運用は、通産省情報管理課で行なっており、検索サービスはオンラインに結ばれた漢字ディスプレイを操作して行なわれる。システムを利用する場合、まず検索質問式の作成から行なわれるが、これは利用者が直接作成する場合と、システム担当者に依頼して作成する場合とがある。検索質問式は、システムとの会話により漢字ディスプレイのキーボード上から打けん入力する場合と、漢字入力装置によって漢字モード紙テープをせん孔して入力する場合と、2とおりの方法で入力できる。

当システムで検索対象としているデータは、長年の実績を反映して多岐にわたっているが、日本語文の情報に限ってみると表2-3のようになる。これらのうち、一部は外部関係機関で作成したデータを入手している。

情報検索手法としてのシステムの特徴は、漢字・かな、英数字などにかかわらず、自然語を1字ずついいねいに探索していることである。そのため、法律文の検索のようなやっかいな情報検索をも、時間はかかるが比較的容易に実現している。

表 2 - 3 蓄積データの種類

蓄積情報の種類	蓄積状況	蓄積件数	発 行	内 容
通商産業六法	最新	270 法令	通商産業調査会	全文
情報要約	64年～最新	約10,000件	官房企画室	要約, キーワード
海外情報インデックス	74年～最新	約 100件	日本情報処理開発センター	全文
鉄鋼業景気指標日誌	53年～最新	約 4,000件	日本鉄鋼連盟	全文
鉄鋼一般情報 (鉄鋼需要関連資料 記事索引)	46年～71年	約24,000件	日本鉄鋼輸出組合	全文
海外鉄鋼情報	68年～71年	約 300件	日本鉄鋼連盟	全文
情報処理ニュース	64年～74年	約 1,500件	日本情報処理開発センター	要約
数理科学	66年～最新	約 500件	ダイヤモンド社	コンピュータ・ニ ュース全文

(2) 検索サービスの実際

検索サービスの窓口である情報管理課では、情報検索の依頼を受けると、まず依頼者と相談して質問内容を分析し、質問文をキーワードに分解して、システム用の検索質問式を作成する。この質問式は、紙テープにせん孔して入力するか、漢字ディスプレイのキーボードから直接入力する。システムの操作および会話は、システム担当者が原則として行なうが、利用者自身にも開放している。

漢字ディスプレイによる対話型の日本語情報検索の画面の流れを図 2-16 に示す。このシステムでは、キーボードの機能キーや数字入力を利用して検索指示を行なっている。

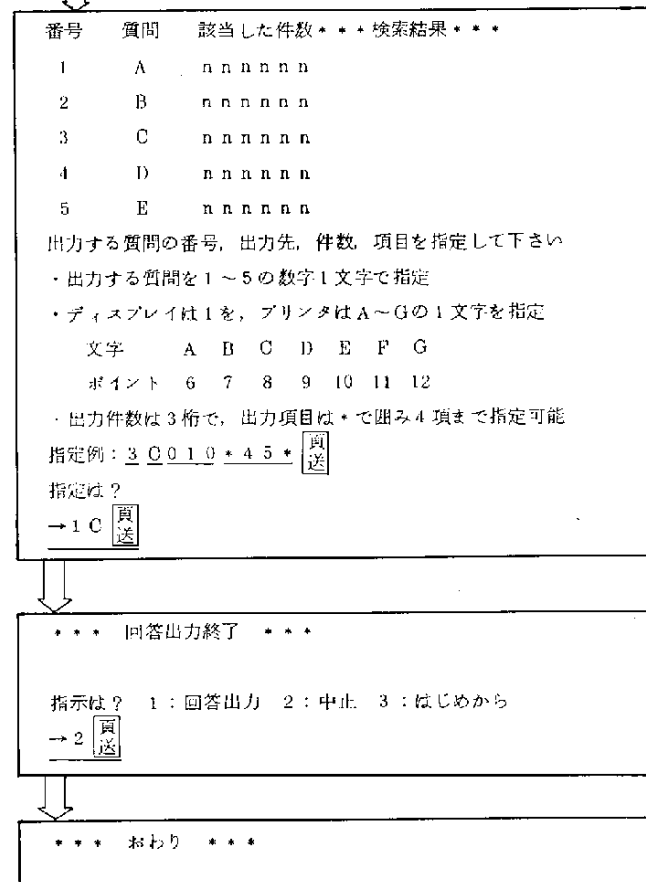
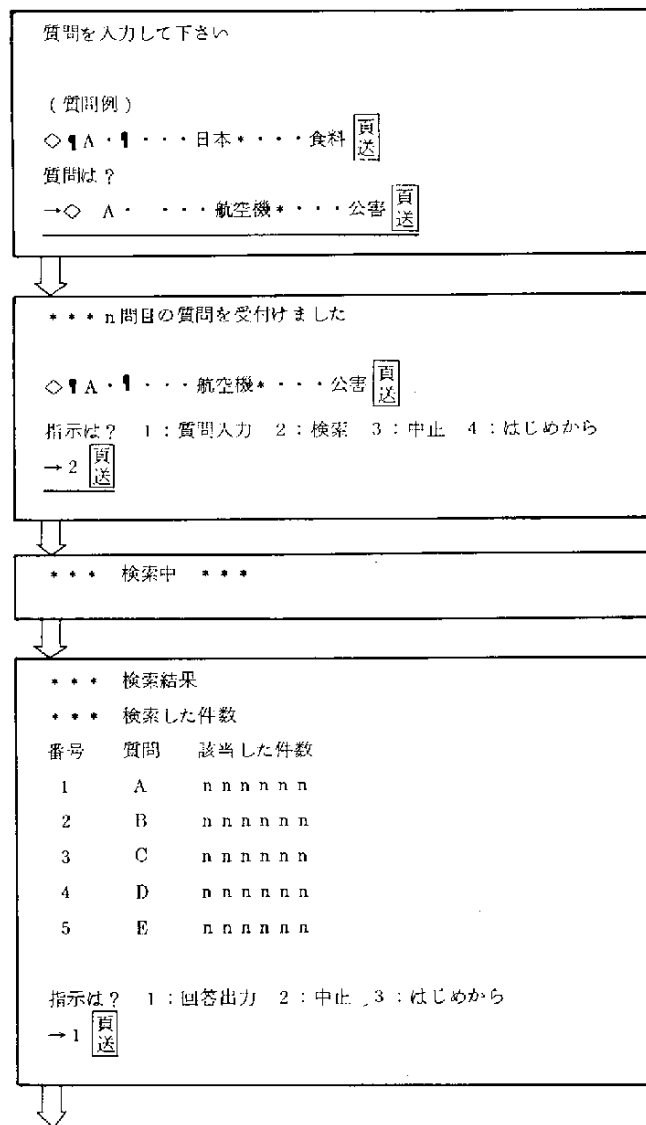
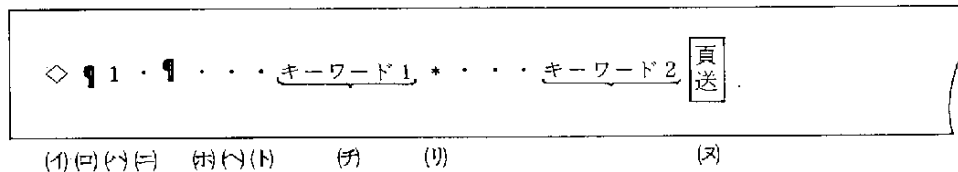


図 2 - 1 6 情報検索の画面の流れ

検索質問式の形式を図2-17に示す。また、法律文データに対する具体的な検索質問式の例と検索結果の一部を図2-18に示す。

情報検索の方法は、データの種類にかかわらず、いずれも逐語的にファイルを探索して行なわれ、そのためデータの件数に比例して最終結果を得る時間は多くかかる。



- (イ) 質問の開始記号
- (ロ) 項目区切り記号
- (ハ) 質問番号
- (ニ) 緩和条件
- (ホ) 項の指定
- (ヘ) 桁の指定
- (ト) 判断条件の指定
- (チ) キーワード
- (リ) 論理記号
- (ヌ) 質問の終了記号

図2-17 検索質問式の一般形

◇オ. 5. 計画 届け出

… 検索結果

… 検索した件数 5228

… 該当した件数 20

… 解答 1

◇ 昭二六法二〇四 高圧ガス取締法 昭四二法一四九

§ 第三章 保安

§ 第二十七条 保安教育

§ 第二項

§ 第一種製造者は、高圧ガスの製造を開始したときは、遅滞なく、その従業者に対する保安教育計画を定め、都道府県知事に届け出なければならない。これを変更したときも、同様とする。

… 解答 2

◇ 昭二六法二〇四 高圧ガス取締法 昭四二法一四九

§ 第三章 保安

§ 第二十七条 保安教育

§ 第二項

§ 第一種製造者は、前項の規定により届け出た保安教育計画を忠実に実行しなければならない。

… 解答 3

◇ 昭三六法一五九 割賦販売法

§ 第三章の二 指定受託機関

§ 第三十五条の八 事業計画書等の提出

§ 第二項

§ 指定受託機関は、事業計画書に記載した事項を変更したときは、遅滞なく、その旨を通商産業大臣に届け出なければならない。

… 解答 4

◇ 昭四三法七四 砂利採取法 改正四七法五二

§ 第三章 採取計画の認可等

§ 第二十条 変更の認可等

§ 第二項

§ 第十六条の認可を受けた砂利採取業者は、当該認可に係る採取計画について前項ただし書の通商産業省令、建設省令で定める軽微な変更をしようとするときは、その旨をその認可をした都道府県知事または河川管理者に届け出なければならない。

… 解答 5

◇ 昭三七法一七八 石油業法

§ 第二章 石油精製業等

§ 第十条 石油製品生産計画

§ 第一項

§ 石油精製業者は、通商産業省令で定めるところにより、毎年度、石油製品生産計画を作成し、通商産業大臣に届け出なければならない。これを変更したときも、同様とする。

図 2 - 18 法律文データの検索例

(3) 日本語情報の蓄積

蓄積すべき情報は、すべて次の形式に整理して磁気テープに入力している。

◇ ① (第1項)

② (第2項)

⋮

① (第n項)

頁送

項の数は蓄積する情報ごとに固定し(9項以内)、各項のデータの長さは自由に、全体で900字以内になるように要約される。

磁気テープ上には、2バイトで1文字を表わす漢字モードですべて構成されるが、この漢字の種類やコードを決定するために、昭和36年～37年の1年間に行なわれた、省内の業務、会議資料、起案文献を中心に、出現漢字の頻度調査を行なっている。その結果、省内で取扱われている文書、資料、刊行物等では当用漢字使用という規定が比較的守られているためか、当用漢字外漢字の使用はわずかであった。そのため、当用漢字1,850字、人名漢字92字、それに加えて当用漢字に準じて多く使われる文字を追加して、文字種を次のように決定した。

漢字 2,181

カタカナ 82

ひらがな 77

英欧字 80

記号 196

数字 { 丸いり 10

ローマン 10

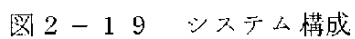
ボールド 10

ゴシック 10

アラビア 13

合計 2,669字

(4) システム構成



2.6.4 大蔵省主計局の国会議事録検索システム

(1) 国会議事録の検索

大蔵省では、主計局、関税局、理材局などにそれぞれコンピュータが導入されており、各種の業務についてコンピュータ処理が行なわれている。このうち主計局では、総務課予算分析管理室が中心となって、国会議事録検索システムを開発した。

このシステムは、大蔵省において予算の作成および国会審議の各段階で必要な情報を、コンピュータを利用してすばやく取得することを目的として計画された情報検索業務の一部である。これまでに情報提供を行なっているのは、国会議事録のうち、衆参予算委員会およびその分科会における議事内容についてである。将来計画には拡張が予定されている。

(2) 議事録の検索サービス

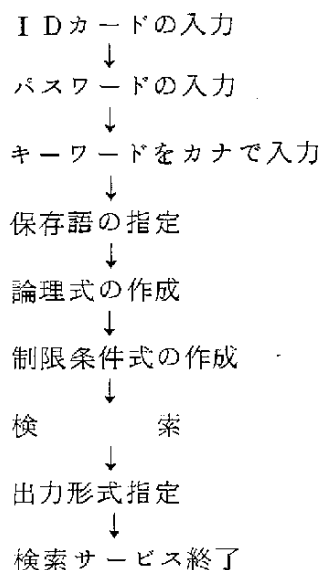
国会議事録検索システムは、オンラインで国会議事録情報を検索する事を目的として作成されたシステムであり、利用者は端末装置を通じ、センター・ファイルに蓄積されている情報を簡単な操作によって検索できる。

システムの特徴として次のようなことがあげられる。

- ・漢字の使用。端末装置としては、漢字ディスプレイ装置を、またセンター出力機器として漢字ラインプリンタを使用している。とくに漢字入力の難点を克服するため、漢字ディスプレイ装置のキーボードは、従来のタイプライタ・キーボードを使用して、入力カナで、出力は漢字で行ない、利用者の負担を軽減している。また、端末画面に表示されたデータのハード・コピーをとることができる。
- ・自由会話方式の採用。利用者がシステムと自由に会話を行ない、いろいろな角度から情報の検索ができるように、検査指令手段(コマンド言語)を用意している。この指令はすべてファンクション・キーとなっているので簡単なボタン操作で行なうことができる。
- ・キーワードの自動変換。利用者の思いついたキーワードを、辞書ファ

イルによって標準キーワードへ自動的に変換している。

検索操作手順は、おおよそ次のような流れで行なわれる。



「ヒカクサンゲンソク」というキーワード入力に対する国会議事録の検索例を図 2-20 に示す。

質問番号-回答番号	01-0001
事 項	非核三原則のうちの持ち込みを許さないというのは有効なものなのか
キーワード	非核三原則 * 核廃絶 * 核抑止力 * 核脅威
元 名	S 源田実 (自民) / K 宮沢喜一 (外務大臣) * 三木武夫 (総理大臣) * 坂田道太 (防衛庁長官)
回次/委員会	15 参予委
事 項 分 類	核兵器 日米安保条約
開催・年月日	50.6.3.53
掲載・議事録	65 号 29-21 ページ
	前 (B) 後 (E) のページを見ますか?

図 2-20 国会議事録の検索例

(3) 関連する索引サービス

国会議事録検索サービスに関連して、データ・ファイルからいくつかの索引を編集して提供するサービスも行なわれている。

その1つは、事項名のKWOC (Keyword out of Context) 索引で、キーワードを手がかりにデータを読み込むと、そのキーワードに関係する質問事項と質問者名、解答者名が得られる。図2-21にその例を示す。

また同様の趣旨で、答弁者索引も用意されている。これは国会で政府側の誰がどのような事項について答弁したかを、答弁者ごとに整理した索引である。図2-22に例を示す。

この他、シソーラス編集用のデータを得る目的で、同義語リストも作成されている。これは比較的単純な手法による同義語リストであるが、シソーラス編集の手だすけには十分役立つリストである。図2-23に例を示す。

S50・10・01		事項名索引(KWOC)	PAGE 000003
アメリカ	事項	日本の領海通航に関して、アメリカ以外には領海航行を認めないと言明するの か、又、その旨外交手続をとったことはあるのか	1113-75-0269 500206 75衆予委 08/27
衆引語	氏名	アメリカ・シン連・非核3原則・領海航行 答 岡田喜夫(社会)／答 宮沢喜一(外務大臣)	外交方針 日米問題 日ソ問題(北方領土)
アラブ	事項	PLOの代表と交渉する用意はあるか	1113-75-0174 500203 75衆予委 05/22-23
衆引語	氏名	PLO・アラブ・アルバニア・民族自決 答 安宅常彦(社会)／答 宮沢喜一(外務大臣)・三木武夫(総理大臣)	外交方針
アラブ外交	事項	アラブ諸国諸日本に対する態度に変化があるか、また従来の外交は間違ってい たのではないか	1162-75-0972 500331 75衆予委 02/19-20・23-25
衆引語	氏名	サウジアラビア・アラブ・イスラエル王 答 平賀和郎(自民)／答 宮沢喜一(外務大臣)	外交方針
アラブ外交	事項	わが国のアラブ外交は非常に所産が多いと思うが、政府の今後の外交方針を述 べよ	1113-75-0219 500205 75衆予委 07/10
衆引語	氏名	アラブ外交・PLO・石油問題 答 渡部一郎(公明)／答 三木武夫(総理大臣)	外交方針
アラブ諸国	事項	わが国はIEAに参加しているが、これはアラブ諸国の強い反発をうける可 能性があるのではないか、また石油問題に関し米国の政策と大筋では同じか	1113-75-0223 500205 75衆予委 07/13-14
衆引語	氏名	IEA・IEP・OECD・石油問題・アラブ諸国・石油国 答 正井成二(共産)／答 宮沢喜一(外務大臣)・三木武夫(総理大臣)	外交方針 エネルギー
アラブ外交	事項	サウジシヤー氏は石油問題解決のためには武力行使もありうる発言したが、 わが国はこのような発言に反対すべきではないか	1113-75-0228 500205 75衆予委 07/17-20
衆引語	氏名	サウジシヤー氏・石油問題・アラブ諸国・アメリカ・OPEC・石油 国連常任第7機関・第3相長研 答 正井成二(共産)／答 三木武夫(総理大臣)・宮沢喜一(外務大臣)	外交方針
アラブ外交	事項	総理、外務大臣の演説で、パレスチナ問題に関する見解の違いを解決せよ 中東問題・アラブ外交問題・国連安全保障理事会決議242号・自決権・パレ スチナ・エルド・レム問題	1113-75-0560 500221 75衆予委 18/5-7
衆引語	氏名	答 岡田喜夫(社会)／答 宮沢喜一(外務大臣)・三木武夫(総理大臣)	外交方針

図2-21 事項名KWOC索引の例

岩間英太郎（文部省初等教育局長）

事項	通信教育の施設費に文部省は努力せよ	1113-68-0624-470321
		68条予1 02/7
事項	小中学校の環境整備に学校の緑化を組み入れよ	1113-68-0625-470321
		68条予1 02/7-8
事項	臨時教員を定め定員拡大により教員の身分平等をせよ	1113-68-0627-470321
		68条予1 02/8-10
事項	沖縄の教職員の給与制度	1113-68-0702-470321
		68条予1 02/20
事項	同和教員の基本方針と予算	1113-68-0712-470321
		68条予1 02/23-26
事項	日本図書教材協会から小中学校への副教材の販売	1113-68-0715-470321
		68条予1 02/30-31
事項	小中学生向けの学習雑誌の学内販売の禁止	1113-68-0718-470321
		68条予1 02/28-29
事項	幼稚園の養護費、法人化にその影響等について	1113-68-0818-470323
		68条予1 04/21-23
事項	公立と私立の幼稚園の格差是正の問題について	1113-68-0812-470323
		68条予1 04/23
事項	幼稚園の教員の給与と退職金について	1113-68-0827-470323
		68条予1 04/36-37
事項	個人立宗教法人立の私立幼稚園に対する助成と振舞	1113-68-0830-470323
		68条予1 04/35-36
事項	学校教員に対する国からの助成について	1113-68-0832-470323
		68条予1 04/46-47
事項	教員の不当労働行為思想調査と労務委員会について	1113-68-0840-470323
		68条予1 04/45-46
事項	教員への思想調査及び不当労働行為に関する問題	1113-68-0841-470323
		68条予1 04/44-45
事項	教科書の地図が国境線が日本の領土と明示している現状	1113-68-0843-470323
		68条予1 04/48-49

図 2 - 2 2 答 弁 者 索 引 の 例

R77777

話 番 号 同 義 語 リ ス ト (日)

DATE 75.12.22

PAGE 00005

区分	漢 字 用 語 情 報	頻度 番号	区分 番号	同 義 語
アブダビ		5	093726	055221 アブダビ石油
天下り		22	000502	192419 天下り官道 055242 天下り重複 192635 天下り人事 192643 天下りポスト 060119 官修天下り 197891 公社公団への天下り
奄美群島		19	000591	055255 奄美大島
奄美群島振興開発特別措置法		3	055271	000612 奄美群島振興特別措置法 093709 奄美振興特別措置法 093793 奄美復興特別措置法
奄美群島振興計画		1	000604	055263 奄美群島振興開発
アムチトカ核実験		1	093807	093823 アムチトカ島地下核実験
アメリカ経済		6	000621	093802 アメリカ経済成長率
アラブ		12	094021	094056 アラブ諸国 094066 アラブ石油戦略
有明海沿岸		1	094001	156901 有明海 157007 有明海
アリナミン		2	157015	157023 アリナミンA 157031 アリナミンF
並核動ガス		6	050402	094111 並核動ガス公害

図 2 - 2 3 キーワードの同義語リスト例

(4) ハードウェア構成

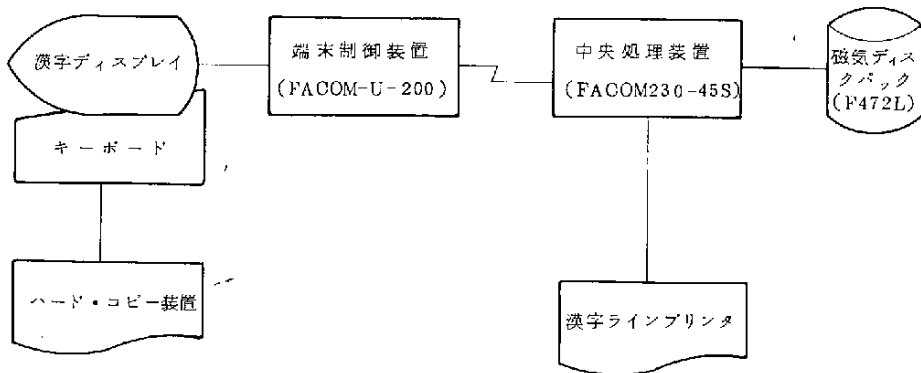


図 2-24 ハードウェア構成

2.6.5 JICSTにおける日本語情報検索

(1) JICSTにおける情報検索の経緯

JICSTにおける科学技術情報の機械的処理の歴史は、昭和36年にさかのぼる。すなわちこの年の10月に、情報検索のうちの遡及検索の専用機として、JEIPACを導入したときから始まる。ついで昭和37年3月には、IBMのPCSを導入して、索引の機械編集を開始した。

そして、昭和42年12月には、汎用大型コンピュータFACOM230-50を導入し、科学技術文献速報の編集処理の機械化が開始され、機械処理は本格化した。機械化した主な業務は、年代順に次のとおりである。

44年 4月：文献速報の編集

46年 4月：タイトル速報（物理編）

- 46年 4月：環境公害文献集の編集
- 46年 4月：逐次刊行物所蔵目録の編集
- 47年 4月：文献検索磁気テープの提供
- 47年 6月：MEDLARS医学文献検索
- 47年10月：理工学文献SDI
- 48年 6月：CAC化学文献SDI
- 49年 4月：文献速報国内化学編の編集

さらにJICSTの部内用として、所内オンラインシステムDOORを47年3月に完成し、また用語管理システムにより、シソーラス暫定版を48年4月に作成した。

この間JICSTのサービスを一層拡充強化するために、各種の問題点を解決した新システムを建設することになり、47年3月より検討を開始した。

新システムにおける主な改善点は次のとおりである。

① 各種サブシステムの有機的統合

理工学文献ファイルを作成するシステムを中心として、各サブシステムを有機的に結合し、データの重複入力 of 排除、プログラムの共用化、例外処理の排除を達成する。主なサブシステムは次の5個からなる。

理工学文献ファイル作成システム

文献速報編集システム

用語管理システム

資料管理システム

情報検索システム

② 遡及検索サービスへの対応

JICSTファイルのみでなく、CACファイル、MEDLARSファイルまで含めて、大量の文献を対象とする遡及検索に対処する。このた

め大容量磁気ディスクを積極的に活用する。

③ 国際標準化への対応

最近磁気テープ・ベースの二次情報ファイルが、国際的に流通するようになったため、その標準化の機運が急速に高まってきた。これにともなってISOを中心として各種の規格や標準化の勧告が出されている。新システムでは極力これらの標準化を取入れることとし、データ要素に関しては次の諸点を改善した。

資料名、発行国コード、使用言語コード：国際標準に準拠

著者所属機関名、ISSN（国際標準逐次刊行物番号）、CODEN

（科学技術逐次刊行物コード米国規格）：新たに追加

提供用磁気テープのフォーマットについては、ISO/DIS 2709にできる限り準拠し、ディレクトリ方式に変更した。

④ 検索精度の向上

キーワードについては、JICST科学技術用語シンソーラスによって、完全に用語の管理を行ない、また付与されたキーワードの上位語を自動付加する。分類については、新たにJICST理工学統一分類表を作成し、その機能をキーワードと明確に分離した。この分類は文献のみでなく、資料管理、クリアリング・サービスにも適用する。

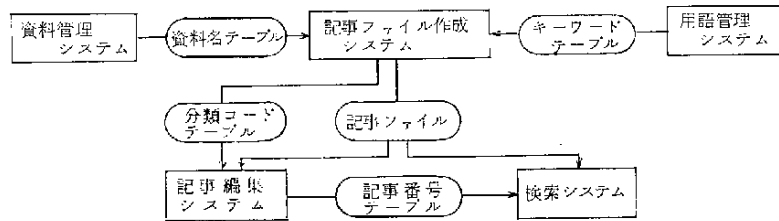
オンライン会話型検索システムについては、現在実験的な小規模システムが完成し、そのフィールドテストを行なっている程度で、本格的な開発は今後に残されている。

以上がJICSTにおけるこれまでの経緯の概要であるが、各サブシステムおよびそれらの関連フローを図2-25に示す。

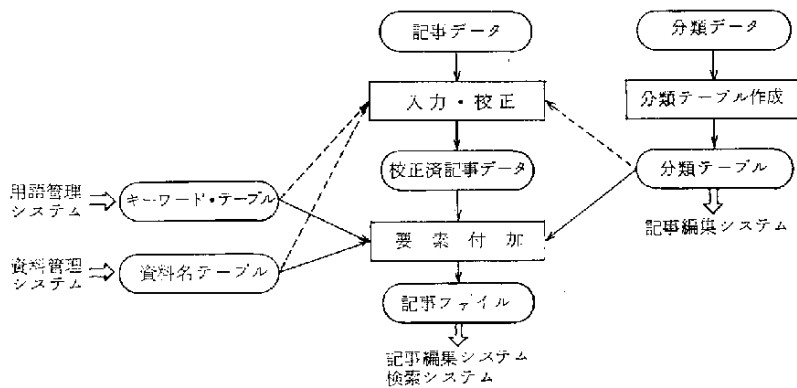
(2) 日本語情報検索サービス

JICSTにおける検索サービスの対象は、独自に作成している理工学文献ファイルと、CACおよびMEDLARSである。このうち日本語を用いているのは前者のみであるので、理工学文献ファイルに限り紹介する。

① システム関連図



② 記事ファイル作成システム



③ 記事編集システム

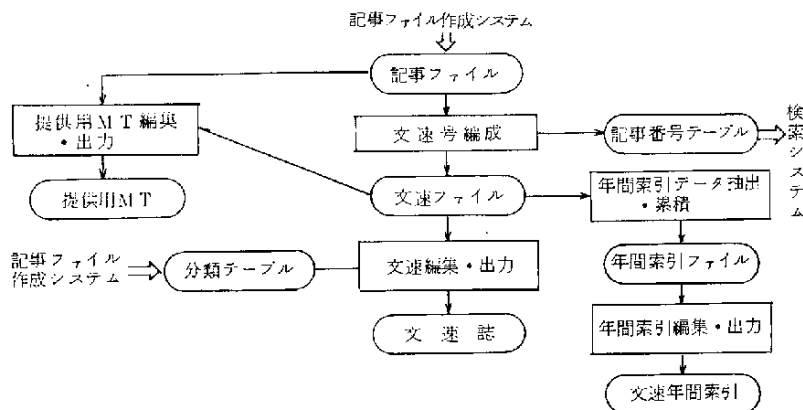
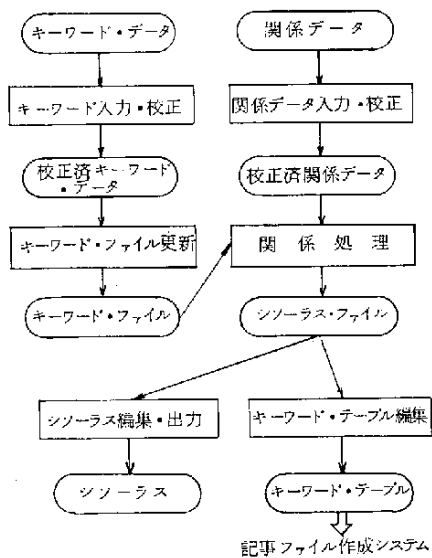
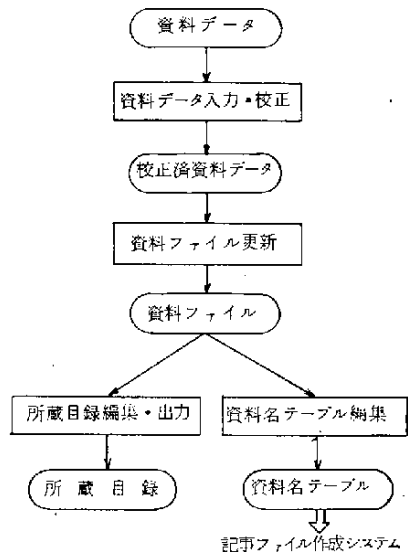


図 2-25 JICST情報処理システム(1/2)

① 用語管理システム



② 資料管理システム



③ バッチ検索システム

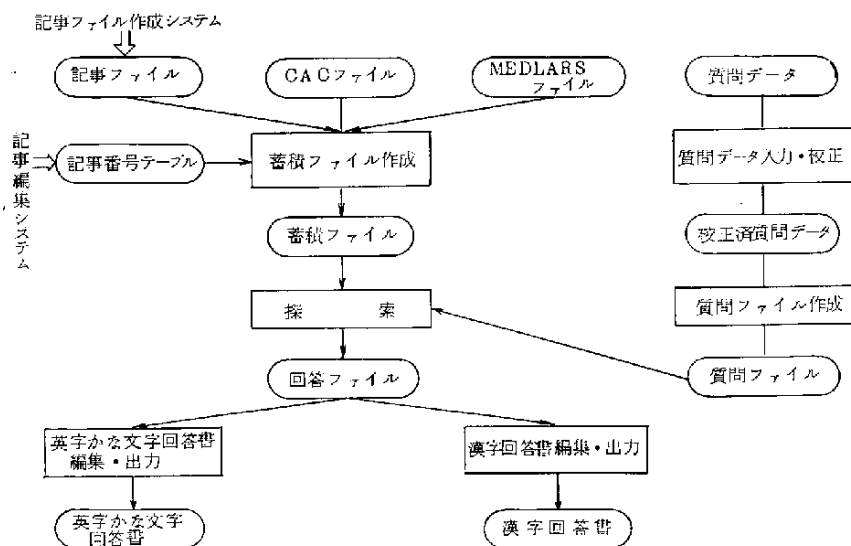


図 2-25 JICST情報処理システム(2/2)

検索サービスは

SDI (選択的情報提供)

RS (遡及検索)

磁気テープ提供 (漢字モード・EBCDIC)

いずれも対象とするファイルの内容は同一である。

① SDI サービス

特定のテーマについて、前もってキーワードや分類を用いて作成した質問式を用意しておき、毎月1回ファイルを検索して回答するサービスで、次の2種類のものがある。

● スタンダードSDIサービス

JICSTで前もって作成したテーマの中から選択して申込むもので、回答は漢字モードの抄録付きである。現在136テーマが用意されており、その一部を図2-26に示す。

共 通 部 門	87 熱電素子
11 超音波およびその応用	88 ノンインパクトプリンタ
12 真空技術の応用	89 調理回路
77 研究開発の計画・管理・評価	90 演算増幅器
79 レーザの計測への応用	91 医用エレクトロニクス
93 海洋開発	104 磁気材料
	116 セミコンダクタの電子工業への利用
物理・電気部門	133 安定化電源
1 ホログラフィー	134 フリント配線技術
2 ファイバー・オプティクス	135 磁気ヘッド
3 色・色彩論・色彩管理	136 電子写真・静電記録
4 結晶成長	
5 液晶	化 学 部 門
6 薄膜作成・薄膜装置	4 結晶成長
7 金属間化合物半導体	5 液晶
8 圧電素子・強誘電体	6 薄膜作成・薄膜装置
9 非晶質の特性	7 金属間化合物半導体
10 レオロジー	8 圧電素子・強誘電体
13 直接発光	9 非晶質の特性
14 燃料電池	10 レオロジー
15 超電導材料および装置	14 燃料電池
16 リニア集積回路(ICおよびLSI)	15 超電導材料および装置
17 デジタル集積回路(ICおよびLSI)	33 内蔵抵抗の掛りカス制御
18 レーザの通信への応用	37 アルミニウムの表面処理
19 フラクセンリ	39 金属材料の塗装
20 アレックリンク	40 プラスチックを利用した建設材料
21 記憶素子	49 エロゾル
22 フォトソレイ技術	50 融解温度の電気化学

図 2-26 スタンダードSDIのテーマ

回答は，利用者情報，検索情報，質問情報，回答書からなるが，質問情報と回答書の例を図 2-27 に示す。

ユアツ ト ソノオウヨウギジュツ

質問文
回答配列指定1
出力形式

42（統一分類順に配列されることを示す）
9（書誌事項、抄録、キーワード、統一分類が出力されることを示す）

[検索タグ]

統一分類を示す	G 0 0 1 → 11 E 1 MF 0 3 ← 統一分類コード
	G 0 0 2 10 E 0 ※ユアツキキ
	G 0 0 3 10 E 0 ユアツクラッチ
	G 0 0 4 10 E 0 ユアツケイ ← キーワード
⋮	完全マッチを示す
キーワードを示す	G 0 4 0 → 10 E 0 エキアツセイギョ
	G 0 4 1 70 E 0 01~40 ← 検索タグ01~40までの論理和集合を示す

(第1次検索タグ論理式)

H 0 0 1 41 ← —— 検索タグ41による検索
 (1 + 2 + 3 + 4 + + 40) を示す。

	【 0 0 0 1 】	← 回答一連番号
UDC標数	→ 624.15	
和文標題	→ 油圧式コンクリート破碎機器の開発研究〔a①〕 J A	
書誌事項	→ 田付茂男、かき崎正義、原田実、や野憲一、西田充行(鹿島建設・技研) : F 1 2 7 A 鹿島建設技術研究所年報 (J P N) 22 3 4 3 - 3 5 0 ('74)	
抄 録	→ 鉄筋コンクリート構造物を静的に破碎し、解体するために開発した 油圧式コンクリート破碎機器(回転加振せん孔機、拡張破砕器)の 性能、コンクリートに対する基本的な破碎特性を究明し、実物構造 物の解体工事へ適用する場合の諸項目について検討した; 写真14表1参7	
キーワード	→ 鉄筋コンクリート構造 解体 破壊 破壊用機械 コンクリート 油圧機器	
統一分類コード	→ M G 0 4 0 4 0 2	

図 2-27 S D I サービス回答例

● リクエスト S D I サービス

スタンダードにない特定のテーマについての S D I で、申込みにより質問式を作成して登録する。なおスタンダード、リクエストとも検索キーはキーワードと分類コードのみで、検索結果を著者名、資料番号、使用言語、発行国でしぼることは可能である。

② R S サービス

現在MEDLARSについてのみしか行っていない。遡及年数は1年または2年である。

③ 磁気テープ提供

この磁気テープは科学技術文献速報に収録された内容を、検索用に編集したもので、漢字モードのものと、EBCDICコードに変換したものの2種類がある。後者は漢字以外の情報をEBCDICコードに変換したもので、変換不能な抄録文などの情報は欠落している。なお現在漢字モード磁気テープは正規の提供は行っていない。

(3) オンライン情報検索サービス計画

オンライン情報検索システムは、将来における情報検索の主流をなすものと期待されているが、JICSTでは現在小規模な実験システムを開発し、テストを重ねている段階である。これは内部の専門家および外部の特定利用者の直接利用にとどめ、検索対象ファイルは、理工学文献ファイルのEBCDICモード、CAC, MEDLARSとし、ファイル別に時間帯を定めて運用する。原則として外部の利用者はオンライン処理を指定して検索の申込みをし、JICSTの専門家がオンライン検索を行なって回答するが、所内に設置した端末を一般の利用者の利用に供したり、特別の契約に基づいて所外に設置した端末を利用に供することも行なう。端末装置はディスプレイ型、タイプライタ型を合せて10セットまでである。

検索機能は次のとおりである。

検索キー：キーワード、分類コード、著者名

限定キー：著者所属機関、資料番号、ISSN、使用言語、発行国、記事区分

論理演算：AND, OR, NOT

マッチング：完全一致

前方一致（キーワード、分類コードのみ）

JICSTではこの小規模システムを土台として、複数ファイルの取扱いの問題、RSに耐えうる膨大な数の文献の処理の問題などを逐次解決し、利用者に十分満足されるサービスを、効率よく提供するシステムを建設していく、将来構想を発表している。その構想をまとめて表2-4に示す。

表2-4 オンライン・サービスの将来構想

	第1期 (小規模オンラインシステム) 1976年~1977年(S51~52)	第2期 (中規模オンラインシステム) 1978年~1980年(S53~55)	第3期 (大規模オンラインシステム) 1981年以降(S56~)
1. 使用業務	a. JICST所内における検索式の設定および修正 b. JICSTの緊急RSサービス業務への使用 c. JICST所内設置の端末による外部ユーザの使用 d. 少数の外部端末(首都圏、筑波支等)の試験的使用	第1期のa, b, cのほか、外部端末の商業的使用 (原則として首都圏および大阪・名古屋支所)	変更なし (外部端末の全般的配置)
2. 使用ファイル	a. JICST英字カナ文字ファイル, CACファイル 1~3年分 b. MEDLARSファイル 2~4年分	a. JICST英字カナ文字ファイル, CACファイル, MEDLARSファイル, 各4~5年分 b. JICSTクリアリングファイル c. その他の外部機関作成ファイル d. JICST漢字ファイルの試験的使用	a. 各ファイルとも 6~10年分 b. JICST英字カナ文字ファイル, 各種英字ファイルの統合 c. JICST漢字ファイルの商業的使用
3. オペレーションの方法	a. 各種データベースの日単位巡回使用(バック処理は随時) b. 端末は、プリンタ型を主とし、一部ディスプレイ型(プリンタ機併付)を使用 c. 端末は、所内巡回および、専用データ通信回線(特定通信回線)で接続 d. 最終回答はオンライン出力も可能とする	a. 各種データベースの同時使用、検索時間の延長 b. 公衆電話回線と一部で専用データ通信回線の使用 c. 端末はプリンタ型とディスプレイ型(ライトペン付)の両者とする	a. 昼夜間操業とする
4. 会話機能	a. 会話プログラムのコマンド数 9 b. 一次検索タグは分類とキーワードの2種 c. シソーラスリンクアップ(関係語の表示)は行なわない d. 1コマンド当たり平均応答時間10秒以内	a. 会話プログラムのコマンド数 約30 b. 一次検索タグは分類、キーワードのほか、著者名(クリアリングファイルでは機関組織名)とする c. JICSTファイルのシソーラス用箱を10万箱まで追加 d. シソーラスリンクアップを可能とする e. 1コマンド当たり平均応答時間 5秒以内	a. c. 一般科学技術を対象とするようコマンド、シソーラス用箱、その他会話機能を更に向上する
5. 使用装置	a. CPU HITAC8450 主記憶容量524KB~1,000KB b. 外部メモリ(ディスクパック型)容量 80.0MB/セット(100MB×8 スピンドル)×2~4 c. 端末台数 10台以内 d. 故障時のバックアップ装置なし	a. CPU高性能機と取換え b. 外部メモリ容量 5000MB~10,000MB c. 端末台数 50台程度 d. 故障時のバックアップ装置は主要部のみを設置	a. CPU外部メモリについては現在のところ予測不可能 b. 端末台数 数100台 c. 故障時のバックアップを完全にする
6. その他の事項	a. データファイルのメンテナンス不可能 b. 端末使用料金の即時表示不可能	a. データファイルのメンテナンスを行う b. 端末使用料金の即時表示 c. マイクロイメージによるJICSTファイル抄録の試験的使用 d. 検索以外のJICST諸業務のオンライン化 e. 計算機室(端末室を含む)の拡張	a. 外部の計算機との連動化

3. 日本語情報の編集と出力

3 日本語情報の編集と出力

3.1 編集システムの必要性

日本語情報を取り扱う業務を大別すると

(1) 一般事務用

(帳票類, 統計資料など)

(2) 文章情報用

(印刷, 版下を中心にしたもの)

に分類できる。

これらの業務について、日本語情報の編集を行なう場合は、それぞれのアプリケーションによって多少の違いはあるが、ほとんどはその文字を扱う機能の組合わせといえる。

文字を扱う機能の組合わせとは、次のようなものである。

- ① 入力処理
- ② 訂正処理
- ③ 編集処理
- ④ 出力処理

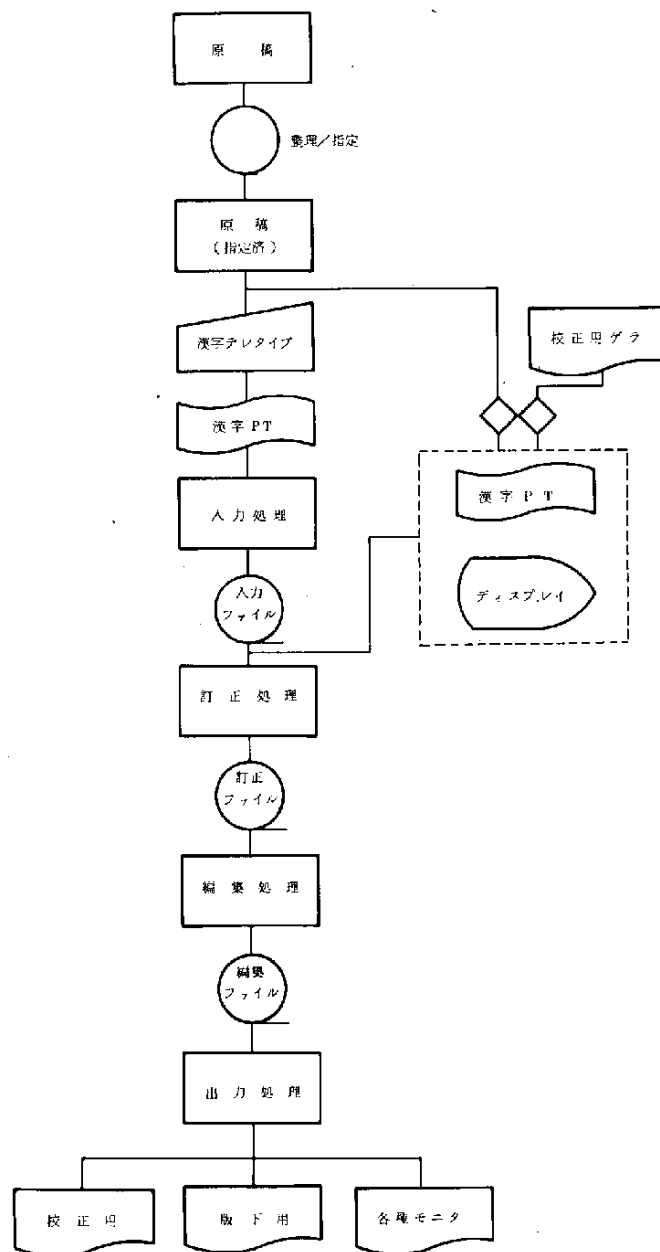


図 3 - 1 一般の編集処理フロー

従来から行なわれている編集について手順内容を考えると次のようになる。

(1) 一般の編集手順

- ① 編集プラン
- ② 原稿の依頼
- ③ 原稿の整理
- ④ 図版（さしえ，カット）
- ⑤ 目次，表紙
- ⑥ 割付とレイアウト
- ⑦ 装丁と造本
- ⑧ 印刷

(2) 一般の編集内容

- ① 本の大きさ，ページ
- ② 見出しレイアウト
- ③ 頁段行
- ④ 図版（さしえ，写真）のレイアウト
- ⑤ 指定の仕方
- ⑥ 割付，図版の指定
- ⑦ 写真植字の指定
- ⑧ 校正 —

{	原稿
{	初校
{	再校
{	校了
- ⑨ 製本

以上の編集手順および内容のうち

- ・原稿の整理，指定
- ・校正

などは従来の方法と編集システムであまり相違は感じられないが、実際には原稿整理指定においては編集システムのソフトウェアに処理させるパラメータの設定が必要になり校正においても赤字起し等に多少のちがいが生じている。

印刷物を作成する過程において文書情報の編集ほど忍耐の必要な作業はないと言われる。誤字、脱字の校正をする作業は単調であるが避けることのできない重要な仕事である。

現在に至るまで、こうした編集分野の作業方法は旧来から進歩が見られない。しかし、コンピュータや、漢字プリンタの出現によって合理化がはかれ、いかに編集者のエネルギーを単調な毎日の手作業から開放して、より内容のある作業に向けられるようになってきた。そのためにコンピュータのソフトウェアによる編集合理化が要求されてきた。

このようなことから図3-1にみられるような編集システムが設計され、次のような効果が得られるのである。

- ① 編集合理化（編集の容易性）
- ② 文書情報の伝達の迅速性
- ③ 情報の磁気貯蔵化
- ④ 情報の規則化、体系化
- ⑤ 分類の容易性
- ⑥ 追加、削除の容易性
- ⑦ 検索の容易性
- ⑧ 組版における人手不足の解除
- ⑨ レイアウトの修正が容易

以下に編集システムの入力処理、訂正処理、編集処理、出力処理における必要機能・項目について概述する。

入力処理

日本語データや各種の編集用ファンクション・データの入力情報の登録処理をいい、その際に入力コードのチェック、コード変換、データ・レコード・フォーマット変換、テーブル検索などを行なう。

(1) 入力データの種類

- ① 組版割付ファンクション
- ② 図版ファンクション
- ③ 本文データ

(2) 組版割付ファンクションの種類

- ① ページの大きさ
- ② 組方向
- ③ 行送り
- ④ 基本文字大きさ
- ⑤ ページ組
- ⑥ 基本文字書体

(3) 図版ファンクションの種類

- ① 図版の種類
 - 図（表、カットを含む）
 - 写真
 - 広告
 - 箱組み
 - 見出し

- ② タブ・ファンクション

(4) その他の入力ファンクションの種類

- ① 入力チェック

入力コードのチェック処理を行なう。

② コード・コンバート

特殊外部コードから内部コードにコード・コンバートする。

③ ピッチ付加

④ ルビ，割ルビ

⑤ キーワード

訂正処理

訂正処理には編集本来の校正処理と，コンピュータ処理による訂正処理（漢字テレタイプによる訂正データ作成方式とディスプレイによるダイレクト訂正方式）に分けることができる。後者はいわゆるファイルの訂正を行なって完全なデータを作成するファイル・メンテナンス処理である。

主な処理内容

- ① 原稿に対する校正，すなわち，文字情報を実際にプリントアウトしたモニタ上に，赤字で訂正を起す。
- ② 赤字の入った訂正原稿に対し，漢字テレタイプで訂正データを作成（あるいは漢字ディスプレイによってダイレクトに訂正）し，ファイルを容易に修正する。
- ③ この訂正処理は，ファイルのメンテナンス用として多大の効果をあげることができ，完璧なデータベース作成には欠かすことのできないソフトウェアである。

編集処理

データ・ファイルに入っている各種の編集ファンクション・データと編集パラメータにもとづいて，ページ割付，棒組調整，図版処理などを行なって編集ファイルを作成する。

(1) 棒組，頁組調整処理

この処理は入力ファイルの割付ファンクションおよび編集パラメータにもとづいた編集をおこなう。

① スペースの種類

固定スペース……入力指定されたもの

可変スペース……調整されうるスペース

図版スペース……図版のしめるスペース

② 欧文の始めと終わりには可変スペースを挿入する。(欧文の終りのスペースが行末の場合はそのままよい)

③ 字送り指定によるスペースは字送り固定スペースを挿入。

④ ？，！の記号の次には全角可変スペースを付加する。

⑤ 受けカッコの後，および起しカッコの前には，それぞれ半角可変スペースを付加。ただし，同じ種類のカッコが続く場合，および受けカッコの後に句読点がくる場合，その間はベタ詰めとする。

(2) 行頭行末処理

① n字下げの指定。

② 文章中の行頭に起しカッコがくる場合，ベタまたは半角あき。

③ 句読点，？，！，ハイフン，中黒，受けカッコ等，行頭禁則文字テーブルにある文字記号約物は行頭におかず前行と追込み処理を行ない，不可の場合は前行より1字移して追落とし，前文を字詰め処理する。

④ 行末処理は本来の行末およびタブ・エンドにおいても行なう。

⑤ 句読点がくるとき，ぶらさげる，または半角スペースでそろえる，あるいは付加される半角スペースは除いてそろえる。

⑥ カッコ，パーレン等の行末禁則文字テーブルにある文字記号約物は行末におかずその行で追込み処理を行ない，不可の場合次行に移し，その行を字詰め処理する。

(3) タブ処理

- ① タブ・パラメータがきたとき、前にタブ・セットがなされていないと、そのタブ・パラメータはその位置のタブ・セットとなる。
- ② タブ・パラメータ後のテキストが行末、あるいはタブ・エンドで行替えされる次行の頭は、そのタブ位置となる。
- ③ ジャスティフィケーションの中央揃え、行末揃え、均等分割を行なう。

(4) 欧文処理

- ① 「欧文」と「欧文終」のファンクション・コードの間のテキストは欧文扱いとなり欧文処理される。
- ② 欧文の前後には可変スペース処理がなされる。
- ③ 欧文語が行末処理される場合、下記によりハイフォネーション処理される。その行でその語全部が追込み、および、追落としができない場合、ハイフォネーション処理が行なわれる。

(5) 図版処理

- ① 図版処理は図、表、写真、見出し、カット、広告、箱組み等、一般文章の植字と別に処理されるスペーシング処理である。
- ② 図版位置割付けの優先度は入力 of 順による。
- ③ 同一ページで2ヶ以上の指示があり、図版がページ内に入らない場合は、位置指示から割付けて重なるものとする。

出力処理

完全なデータ・ファイルに対し編集処理の終わったファイルを出力するために必要なプリント・ファンクション変換、レイアウト変換を行なって出力する。

主な内容

- ① 編集されたデータ・ファイルを校正用ゲラ、および版下用プリントある

いは各種モニタや台帳などの出力処理が実施される。

- ② 実際にはこれらの出力処理は各種汎用出力装置や版下用出力装置の仕様に合わせるための処理に近いが、それぞれの出力装置を効率よく稼働させるための特殊コード変換等にはかならない。
- ③ 校正用ゲラの出力は文字や意味校正用でありレイアウトは考慮していない。
- ④ 仮レイアウト・ゲラの出力は、ページ組をする場合にページ内の全体の配列を調べるために、すべてのファンクション等が入ったページ・レイアウト用出力であり、編集指示に相違がないかの校正をするために出力される。
- ⑤ 版下用、各種モニタ用の出力は校了になった時点において各種編集ファンクション機能によって出力するものであり、最終アウトプットである。すなわち、ページ組の場合はページごとの出力がなされる。

3.2 出力装置と編集機能

前述したように、日本文情報を扱う場合は、そのアプリケーションの出力方法（最終出力形式）によって各種の編集システムを開発する必要がある。

この編集システムを開発するにあたっては、当然、使用する出力装置に編集機能がなければならない。

この節では、日本文情報の出力を満足させるための一般的な編集機能を取り上げて概説する。

(1) 基本機能

コンピュータを使って編集処理がなされたデータは、漢字出力装置に最終出力形式で作成される。この最終出力は、アプリケーションによって各種の形態がとられるが、ここでは、日本文情報という点に留意して基本的な機能を中心に述べる。

① ポイント（文字の大小）

日本文情報では、もっとも必要とされる情報である。この機能は文字の大きさを変えるもので、出力装置によって各種の能力を持つが、一般的には4ポイント～32ポイント位の単位が要求されよう。

② 縦、横

この機能は、文字を「縦書き」で出力するか「横書き」で出力するかの機能である。

日本文情報の場合は、縦書き横書きの文章が半々くらいあるので、日本文情報の出力をすべてカバーするためには、必要な機能である。

③ 字 体

日本文情報には、明朝、ゴシック、イタリック、ボールド、教科書体など同じ字に対して各種の字体がある。これらの字体の選択を行なう機能が必要である。

④ 合 成

この機能は、漢字の字種の多さや特殊記号などの不足をカバーする目的で用意されるものである。

⑤ 上つき

この機能は、□□ のような機能である。

たとえば、 $X \overset{\text{上つき}}{2} \rightarrow X^2$ というような場合に、使用される。

⑥ 下つき

この機能は、□□ のような機能である。

たとえば、 $H \underset{\text{下つき}}{2} \overset{\text{正常}}{O} \rightarrow H_2 O$ など、数式、化学式などに利用される。

⑦ 平 体

この機能は、全角文字を水平方向につめるための機能であり、50 % ~ 90 % の範囲で平体率が決められる。

⑧ 長 体

この機能は、全角文字を垂直方向につめるための機能であり、50 % ~ 90 % の範囲で長体率が決められる。

⑨ ビッチ

英数字などは、それぞれが決められた文字幅でデザインされている。この文字幅をビッチといい、これを処理するために全角の文字幅を削る機能が必要である。

⑩ 解 除

この機能は、①～⑨までの機能を解除するためのものである。

その他、出力装置によって、上記以外の機能が追加されたり、あるいは削除されたりして構成される。

(2) 応用機能

基本機能は、文字の1字1字をどうするかという機能が主であったが、文章情報では段組み、ページ組みなどの編集を行なうための機能が必要である。

① 段 間

段間とは，段組みと段組みの間の大きさを示すものである。

② 行 間

行間とは，行と行の間の大きさのことである。

③ 字間

字間とは，字と字の間の大きさのことである。

(3) データ処理機能

前述の他，機能的には，次のようなデータ処理の機能がある。

① E N D

データの終了を意味する機能である。

② L F, E O B

データを移動させるための機能である。

③ E O M

オペレーションをストップさせる機能である。

3.3 編集システムの設計

3.3.1 システムの概要

ここに設計された編集システムは、当財団に設置されている漢字入出力装置を使用して実行するシステムである。よって、一般的な編集機能を網羅することができず、アプリケーションとして機能するのは、入出力装置の制限下にごく限られた編集機能にすぎない。

この編集システムの処理概念としては、各種アプリケーションの処理によって作成された漢字データ・ファイルを、漢字プリンタおよび漢字ディスプレイに出力・表示するための編集処理を行なうことにある。

図3-2にハードウェアの概要を、図3-3にシステム・フローを示す。

なお、編集システムを構成する編集プログラムは、センター側のコンピュータ FACOM 230-75 の処理装置にもつものとする。

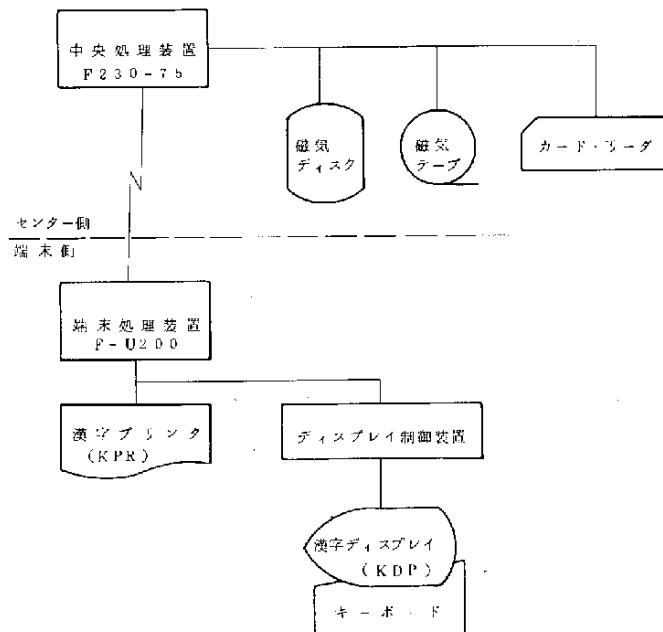


図3-2 ハードウェアの概要（編集システム関連機器）

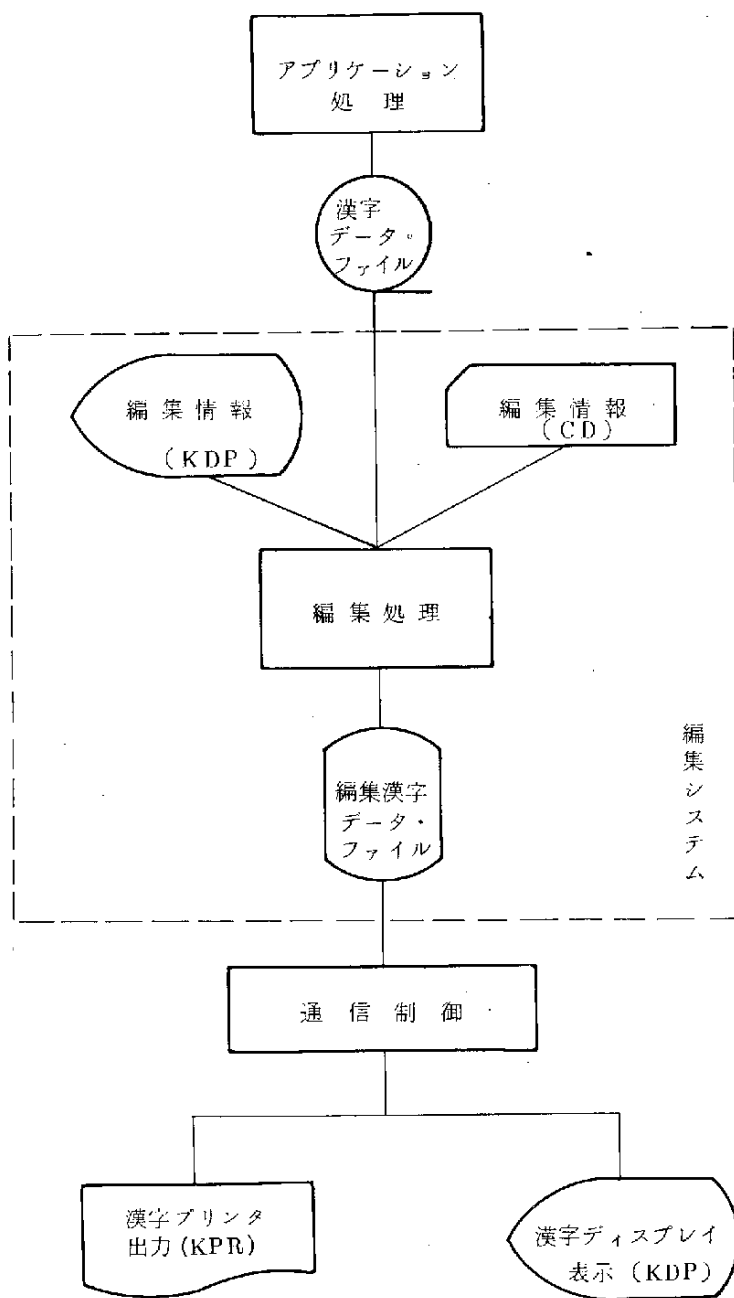


図 3 - 3 システム・フロー

3.3.2 システムの機能

この編集システムの機能は、大きく分けて、次の3項目の機能をもつことが必要である。

- ① KDPまたは、カードより入力される編集情報（出力機器、縦書・横書、ページ寸法、行数などの指定：以下、パラメータという）に従って、編集処理を行なう。
- ② 入力データである漢字データ・ファイル中に含まれているファンクション・コードの機能が果せるようにする。
- ③ 各禁則処理、文頭、文末処理などを行なう。

なお、この編集機能においては、出力機器の機能に影響されるため、KDP、KPRの仕様を表3-1にまとめた。この中で、もっとも機能に障害を生じるのは、文字ピッチが固定ということである。

表 3 - 1 性 能 表

項 目	K P R	K D P
印 字 速 度	65字/秒	約40フィールド/秒
1 行 印 字 数	max 55字/行	32字（半ピッチ：64字）
頁 寸 法	A5～A4（ソフト指定）35～70行 （1/6インチ間隔）	固定 16行
文 字 種 類	5376字種+117字	5376+5 （ケイ線表示など）
印 字 書 体	中ゴシック体	中ゴシック体
文 字 寸 法	約3.0×3.0mm	約ヨコ5.0×タテ7.0mm （半ピッチ 3.0×4.0）
文 字 ピ ッ チ	3.2mm（固定）	—
行 ピ ッ チ	1/6インチ、1/3インチ	—
縦 書 ・ 横 書	手動により切換可能	ソフト指定で切換可能
コ ー ド	富士通漢字コード	富士通漢字コード、EBCDIC

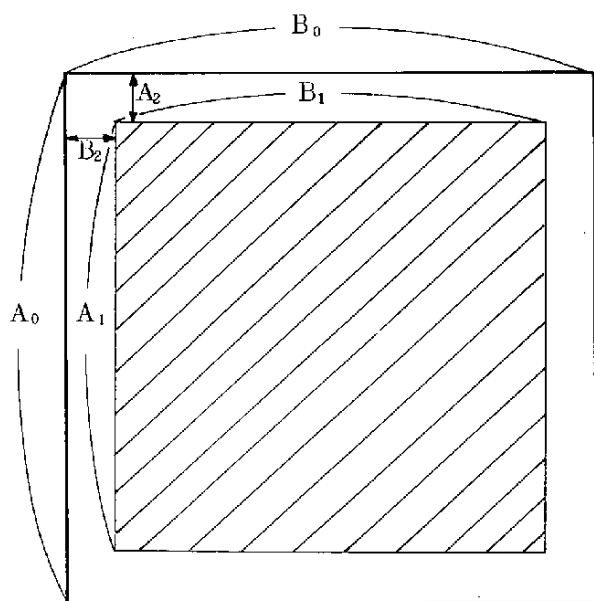
表 3 - 2 K P R 機能一覧表

項 目	コ ー ド						機 能
	1	2	3	4	5	6	
1 スペース・シフト	B	0					印字位置を1桁ずらす 印字動作なし
2 紙 送 り	B	8					改行を行なう 印字位置は先頭に戻る
3 紙 切 断	B	C					紙の切断を行なう
	2	0					

表 3 - 3 K D P 機能一覧表

項 目	コ ー ド						機 能
	1	2	3	4	5	6	
1 全画面消去	F	0					プロテクト・フィールドを含めた全画面を消去し、アドレス・ポインタを原点に位置する
	C	7					
2 行 消 去	F	E					アドレス・ポインタの位置から、その行の終りまでを消去する。
	C	7					
3 復 改	F	1					表示中のデータを復帰改行する。
	C	5					
4 座標ポイント指定	F	5		値		値	ケイ線表示に際し、スタート・ポイントの位置情報を与える。
	C	E					
5 文字アドレス・ポイント指定	F	7	値	値			文字表示の際のアドレス・ポインタの値を指定する。
	C	A					
6 カーソル位置指定	F	D					カーソルが、アドレス・ポインタに示されているロケーションに移動する。
	C	9					
7 文字像反転指定開始	F	7	7	0			この指定の開始から終了までの文字をネガ像で表示する。
	C	E	C	0			
8 終 了	F	7	C	0			
	C	E	5	0			
9 HDM指定	F	2					画面データ送信時のスタート・マークを表示する。
	C	1					
10 オペコール	F	2					端末のブザーを鳴らす。
	C	F					
11 プロテクト・フィールド開始	F	4	値	値			プロテクト・フィールドの開始アドレスを定義する。
	C	C					
12 プロテクト・フィールド終了	F	6	値	値			プロテクト・フィールドの終了アドレスを定義する。
	C	E					
13 全プロテクト・オフ・フィールド消去	F	E					プロテクト・フィールド以外の画面を消去する。
	C	4					
14 Kシフト指定	F	0					以後、漢字モードで表示する。
	C	F					
15 Aシフト指定	F	0					以後、半ピッチ文字で表示する。
	C	E					
16 ケ イ 線	F	F					座標アドレス・ポインタの示す位置からケイ線を引く。
	C	2					
17 NOP	F	0					NO operation
	C	0					

(項目 4, 7, 8, 16の機能は、機器不備のため動作しない)



(1) 漢字プリンタ編集機能

a. サイズ処理 (パラメータ指定)

$A_0 \times B_0 = A4, A5$

A_1 = ライン指定

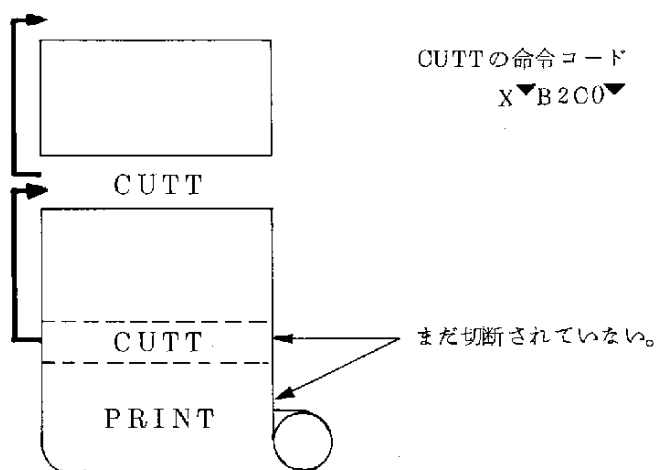
B_1 = 文字指定

A_2 = スタート・ライン位置

B_2 = スタート・文字位置

① 頁寸法処理；一頁の大きさは、A4、A5の大きさを指定できるので、パラメータにより、寸法を指定し、それに応じてCUTT命令を出す。

なお、最後の切断では、下図のことを考慮する。



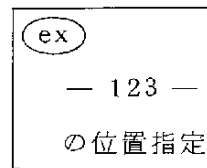
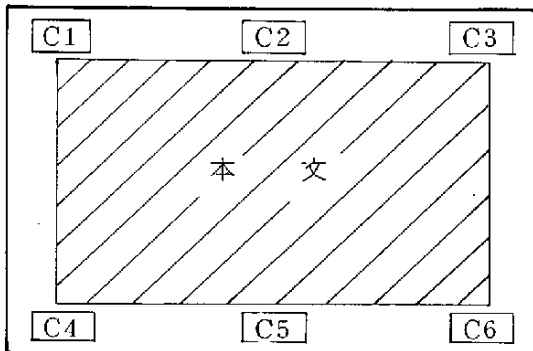
② 開始文字位置；印字をどの位置から始めるかをパラメータによって行ない、この指示によって、先頭文字を印字する場合

所まで、紙送り、スペース・シフトを行なう。

- ③ 文字・行カウント処理；パラメータにより指定された文字数、行数によって、復改、改頁をする。

ただし、行頭・行末の禁則処理や分割禁止文字がある時は、この限りではない。

b. ノンブル処理（パラメータ指定：ページ数挿入の位置指定）



◎桁数は3桁で表示

◎C1～C6のいずれかを指定する

パラメータにより指定された場所にページを入れる。

ページ数は3桁で表現する。ページ数はゼロ・サブレスを行なう。



c. 特殊コード設定

以下にあげるコードが入力データ中に入ってきたら、各々の処理を行なう。

- ① くりかえしコード（表作成のケイ線など）…… (FUNC)

ex. (FUNC) — (21) …→ —を21個つづける。

- ② タブ（帳票出力などの桁位置指定）…… (TAB始), (TAB終)

ex. (TAB始) ①×× ②×× ③×× (TAB終) ①×××××× ②×××××× ③×××× (復改) ①
.....
桁位置の指定 (先頭文字位置) データ (注)データが2行以上につづく場合もある)

- ③ 分割禁示（数式など2行にまたがって印字したくないもの）……

(分禁始), (分禁終)

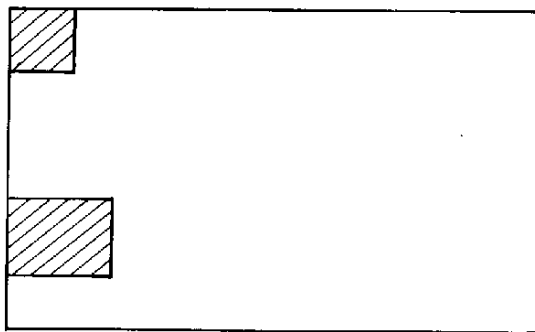
ex. (分禁始) ~ (分禁終) ...→~のデータは行をまたがない。

ただし、1 行に入りきらない場合は次行へ。

④ 特殊コード変換テーブル

横書から縦書になった時に使用

⑤ 特殊領域 (図版行) 処理 (後で図などを挿入するための領域を確保する) …… (空白) ex.



ex. (空白) (3 × 5)



3 行 5 列の領域確保

条件として、このデータの前は原則として (復改) とする。

d. 行間の指定 (パラメータ指定)

1/6 インチ間隔, 1/3 インチ間隔のどちらかを選ぶ。

e. 文頭処理 (パラメータ指定)

復改の後、先頭 1 文字あける。

f. 行頭・行末禁則処理

各行において、行の先頭文字 (行頭)、行の最後尾の文字 (行末) に使ってはいけない文字 (行頭・行末禁則文字) がこないように処理を行なう。

① 行頭禁則の文字は、前行の最後尾に移す。

(ex) xxx xxx } xxx xxx)
)xxxx。 } xxx。

② 行末禁則の文字は、次行の先頭に移す。

(ex) xxx xxx ■ } xxx xxx
 xxxx。 } ■xxxx。

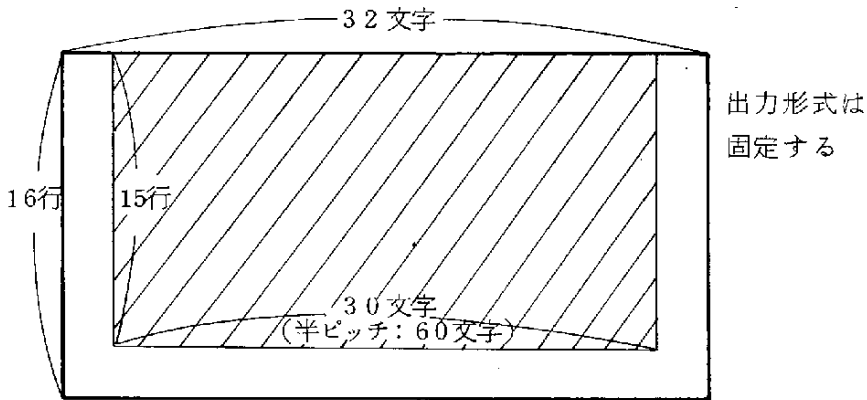
禁則の時は、前行の □ , ・ の次にスペースを送って行替し、行末をそろ

える。

g. タテ・ヨコ書き（パラメータ指定）

(2) 漢字ディスプレイ編集機能

a. サイズ（固定）



b. ノンブル（固定）

最下行，中央に固定

c. 特殊コードの設定

K P R と同様

d. 行頭・行末の禁則処理

K P R と同様，ただし，縦書は行なわない。

e. 半ピッチ文字処理

EBCDIC で表わせる英数字，カナなどは半ピッチとしてディスプレイする。

f. チェック・コード，ファンクション・コード（クワタを除く）の削除

g. 文頭処理

復改の後は，一桁あける。

3.3.3 システムの制限

キーボード・ファンクション・キーのうち、使用可能なファンクション・キーを表3-4に示す。

表3-4 ファンクション・キー

キー	機 能	キー	機 能
復改	復改(改行)を行なう。	終	レコードの区切り
クワタ	スペースとなる。	空白	図版行処理
FUNC	同一文字の繰返し	改 段	改 頁
TAB始	タブ機能の始めと終り	末 消	
TAB終		後 退	
分禁始	分割禁止文字の始めと終り		
分禁終			

その他次のようなことに制限がある。

- ① 欧米語の問題(ハイフオネーションなど)は取扱っていない。(将来は取扱い予定)
- ② 印刷の途中で頁寸法の変更はできない。
- ③ 現在、各漢字処理装置はU-200の下で稼動しており、編集プログラムは、このU-200内にあった方がよいのであるが U-200には、補助記憶がないため、F-230-75で編集処理をしている。従って、将来は、U-200にディスクなどの補助記憶装置をとりつけ、編集プログラムをU-200内にもつことになるう。
- ④ ディスプレイ上に表示された漢字データのハード・コピーは、サービス・プログラムで処理可能である。

また、ディスプレイ上のデータ修正はキーボードより行なり。

表 3 - 5 縦・横・変換テーブル

横 書		縦 書		横 書		縦 書	
文 字	コ ー ド	文 字	コ ー ド	文 字	コ ー ド	文 字	コ ー ド
】	0D77	┌	0D75	。	2475	。	2C73
、	2474	、	254E	」	2473	┐	2C73
。	2475	。	254F	【	0D76	┐	0D74
)	2471	、	2649	(	2470	(	2648
」	2473	┐	264B	「	2472	┐	264A
”	2863	”	2861	『	2966	『	264C
』	2967	』	264D	”	2862	”	2860
>	2A67	∨	2A65	<	2A66	∧	2A64
}	2B63	┐	2B61	[	2B62	┐	2B60

表 3 - 6 行頭・行末禁則文字表

	横 書		縦 書		横 書		縦 書	
	文 字	コ ー ド	文 字	コ ー ド	文 字	コ ー ド	文 字	コ ー ド
行 頭 禁 則 文 字	:	0E4A	..	0E62	?	2960	?	—
	;	0E4B	..	0E63	!	2961	!	—
	?	0E4E	..	0E66	/	2962	/	—
	!	0E4F	..	0E67	//	2963	//	—
	:	0E52	..	0E6A	』	2967	』	264D
	;	0E53	..	0E6B	>	2A67	∨	2A65
	?	0E56	..	0E6E	)	2B63	┐	2B61
	/	0E57	..	0E6F	,	2C76	,	—
	•	0F4E	..	0F66	、	2C77	、	—
	,	0F4F	..	0F67	拗 促 音 も 禁 則 処 理			
	.	0F56	..	0F6E	行 末 禁 則 文 字			
	,	0F57	..	0F6F	横 書		縦 書	
	”	0A5E	..	0D5D	文 字	コ ー ド	文 字	コ ー ド
	】	0D77	┐	0D75	“	0A5A	”	0D59
	、	254C	、	—	【	0D76	┐	0D74
	、	254D	、	—	(	2470	(	2648
	、	2474	、	254E	「	2472	┐	264A
	。	2475	。	254F	『	2966	『	264C
	)	2471	)	2649	”	2862	”	2860
	」	2473	┐	264B	<	2A66	∧	2A64
	”	2863	”	2861	[	2B62	┐	2B60

表 3-7 EBCDIC 扱い可能文字とコード

文字	漢字コード	EBCDIC	文字	漢字コード	EBCDIC	文字	漢字コード	EBCDIC	文字	漢字コード	EBCDIC
	D001	4 0	ア	2848	8 1	へ	2B4C	A 2	J	2B59	D 1
\$	764B	4 9	イ	2849	8 2	ホ	2B4D	A 3	K	2B5A	D 2
¢	7648	4 A	ウ	284A	8 3	マ	2B4E	A 4	L	2B5B	D 3
•	2C53	4 B	エ	284B	8 4	ミ	2B4F	A 5	M	2B5C	D 4
<	2A66	4 C	オ	284C	8 5	ム	2C48	A 6	N	2B5D	D 5
(	2470	4 D	カ	284D	8 6	メ	2C49	A 7	O	2B5E	D 6
+	0E70	4 E	キ	284E	8 7	モ	2C4A	A 8	P	2B5F	D 7
	2C54	4 F	ク	284F	8 8	ヤ	2C4B	A 9	Q	2C58	D 8
&	0F48	5 0	ケ	2948	8 9	ユ	2C4C	A A	R	2C59	D 9
/	2961	5 A	コ	2949	8 A	ヨ	2C4D	A C	S	2C5A	E 2
¥	0A73	5 B	サ	294A	8 C	ラ	2C4E	A D	T	2C5B	E 3
*	2E73	5 C	シ	294B	8 D	リ	2C4F	A E	U	2C5C	E 4
)	2471	5 D	ス	294C	8 E	ル	2850	A F	V	2C5D	E 5
;	0875	5 E	セ	294D	8 F	レ	2851	B A	W	2C5E	E 6
「	264A	5 F	ソ	294E	9 0	ロ	2852	B B	X	2C5F	E 7
—	0E71	6 0	タ	294F	9 1	ワ	2853	B C	Y	2D58	E 8
/	0A72	6 1	チ	2A48	9 2	ン	2855	B D	Z	2D59	E 9
,	2476	6 B	ツ	2A49	9 3	。	2862	B E	0	2E51	F 0
%	0E75	6 C	テ	2A4A	9 4	°	254F	B F	1	2D50	F 1
—	2C55	6 D	ト	2A4B	9 5	A	2A58	C 1	2	2D51	F 2
>	2A67	6 E	ナ	2A4C	9 6	B	2A59	C 2	3	2D52	F 3
?	2960	6 F	ニ	2A4D	9 7	C	2A5A	C 3	4	2D53	F 4
:	0874	7 A	ヌ	2A4E	9 8	D	2A5B	C 4	5	2D54	F 5
#	704A	7 B	ネ	2A4F	9 9	E	2A5C	C 5	6	2D55	F 6
@	7049	7 C	ノ	2B48	9 A	F	2A5D	C 6	7	2D56	F 7
/	0A5B	7 D	ハ	2B49	9 D	G	2A5E	C 7	8	2D57	F 8
=	0E74	7 E	ヒ	2B4A	9 E	H	2A5F	C 8	9	2E50	F 9
”	0D5D	7 F	フ	2B4B	9 F	I	2B58	C 9			

4. 事例研究 —日本語情報文献検索システムの設計—

4. 事例研究 — 日本語情報文献検索システムの設計 —

4.1 システムの概要

当システムは、漢字コード表現によって蓄積された文献情報を、漢字入出力機器を用いて、オンラインかつ会話型で情報サービスする情報検索（IR）システムである。

当システムは、通称「^{カインズ}KINDS」もしくは、「^{カインド}KIND システム」と言い、これは、Kanji Information Document Retrieval Systemの略称である。以降、KINDシステム設計の成果を概略的に記述する。

4.1.1 システムの目的

文献検索システムが現在直面している課題の1つは、漢字かな混り文で表現される日本語の情報を如何にして処理するかにある。こうした技術の開発は、「データ」処理から「情報」処理へと高度化するコンピュータ・システムにとって不可欠の条件であろう。

KINDシステムは、こうした日本語情報処理特有の課題に対し解決の糸口を求めるべく3つの基本方針をたてて設計された。すなわち、

- ① 漢字キーボード、漢字ディスプレイ（KDP）、漢字プリンタ（KPR）等の漢字入出力機器を端末装置として使用する。
- ② 検索語のタイプインはカナ文字で行ない「カナ漢字変換手法」を用いて漢字の検索語とする。
- ③ シソーラスを導入し、検索は会話型式でサービスする。

である。

これらのねらいは、ハード的には各漢字処理装置の使い方や、操作性など効果的な利用法へのアプローチであり、ソフト的には検索語（キーワード）

入力操作の一般化への試みであり、またシソーラスの導入によるデータベースの効率化および検索過程の効果測定にある。

4.1.2 システムの特徴

KINDシステムの特徴として、次の点があげられる。

- ① 漢字処理装置の利用により検索過程の日本語による表示が可能となり、検索結果の閲覧およびプリント出力もできる。
- ② オンライン会話型にて当初9種類のコマンドのファンクション・キーを利用することにより、納得いくまで誤操作の少ない形で繰り返しトライできる。
- ③ キーワードによる主題検索と発行年、発行国等書誌的事項による検索の分離によりシステムを簡略化してある。
- ④ 主題検索は、入力したキーワードに該当する文献集合作成の試行検索と、集合と集合および集合と制限条件間の演算を対象とする集合演算検索の2レベルを持ち質問の積み重ねによる複雑な検索が可能である。
- ⑤ キーワード入力はカナ文字で行ない、漢字へ自動変換された漢字キーワードの表示、保存語の指定といった過程を経る。
- ⑥ キーワード・テーブルは、シソーラス関係の情報を持たせた上でヨミカナ、漢字コードでソーティングされているので、キーワード・テーブルの通覧は、シソーラス形式でも入力したカナ文字による前方一致語の表示形式でも任意に選択できる。
- ⑦ 文献ファイル、キーワード・テーブル等のファイル創成時より格納されるデータベースに至るまで、漢字コードによる記録を主体としている。
- ⑧ FACOM U-200に制御されたターミナル・ステーションとFACOM 230-75に制御されたホスト・コンピュータとの間を通信回線で結ぶシステム構成で、前者には、検索サービスに伴う操作、表示、出力処理を、後者にはアプリケーション機能の処理を分担させている。

4.1.3 システム構成

KINDシステムは、検索サービスを主要な機能とし、他にサービスの準備機能にあたるデータベース管理、キーワード管理、さらにサービスの補助的機能を有す索引誌作成管理の各機能があり、おのこのサブシステムを構成している。処理はオンラインないしバッチでなされるが、いずれもホスト・コンピュータF230-75のオペレーション・システムであるM-VIIの制御下で、一般のアプリケーション・ジョブと併行に処理される。

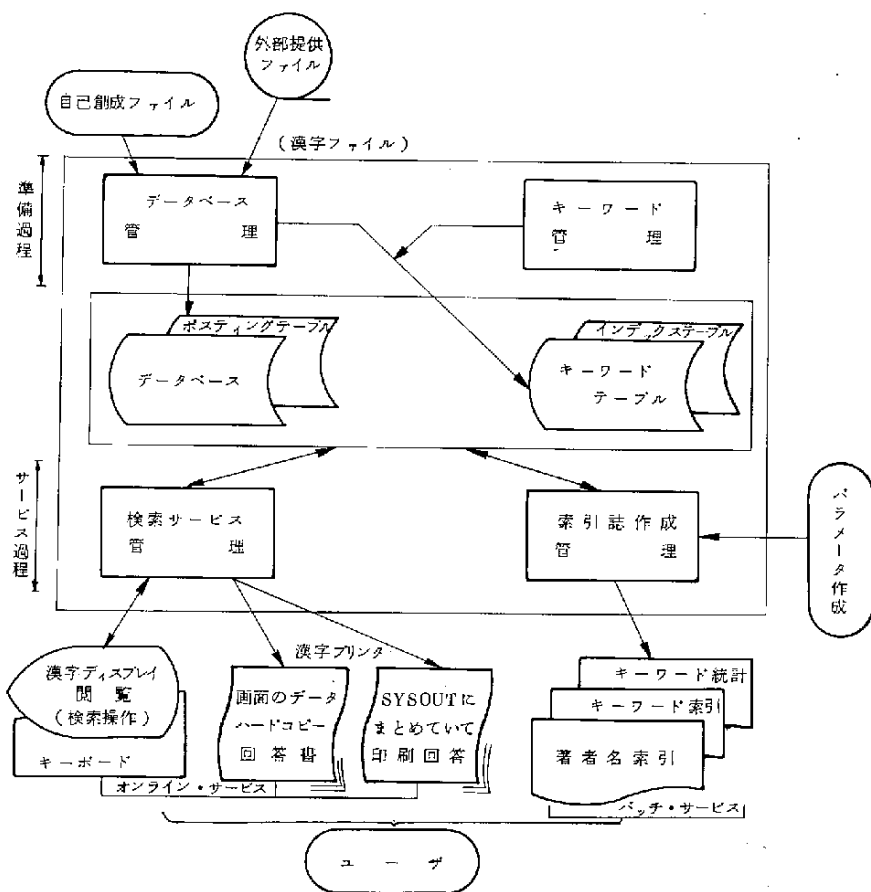


図4-1 KINDシステムの機能

次に各サブシステムの概略を示す。

(1) 検索サービス・サブシステム

オンラインで動作する処理プログラムで、ターミナル・ステーションからの呼びだしに応じオンライン状態となる。

利用者とシステム間との会話は、ファンクション・キーに定義された9種類の検索コマンドの操作と、キーワード、検索式、条件式などタイプインによるデータの直接入力とによってすすめられる。これらのコマンドに共通する動作パターンとしては図4-2のように整理される。

また検索過程における会話の流れはある程度想定可能であり、従ってコマンドとコマンドとの関係において一部にシステム・リードのプロセスもある。

(2) データベース管理サブシステム

バッチで動作する処理プログラムで検索サービスの準備段階に位置し、文献データの入力、ファイル化、ポスティング・テーブルの作成ならびに磁気ディスク・パックへのデータベース創成までを行なう。

データベースには、文献情報として情報処理分野の技術文献に制限し、当財団内で収集可能な文献は、自己入力し、他に外部サービス機関で作成提供される文献ファイルをも導入して統一したファイルを作成する。データベースを構成するデータ項目は表4-2に示すように文献データとして一般的なもので、このような文献データを100MB規模のディスク2台(約8万件)に創成し1ヶ月単位で更新する。

(3) キーワード管理サブシステム

バッチ処理にて運用するサブシステムであり、検索サービスの準備段階に位置し、シソーラス・ファイルの作成と、キーワード・テーブル、インデックス・テーブルなどインバーテッド・ファイルの作成ならびにディスクへの格納処理を行なう。

インバーテッド・ファイルは、データベースの作成、ポスティング・テ

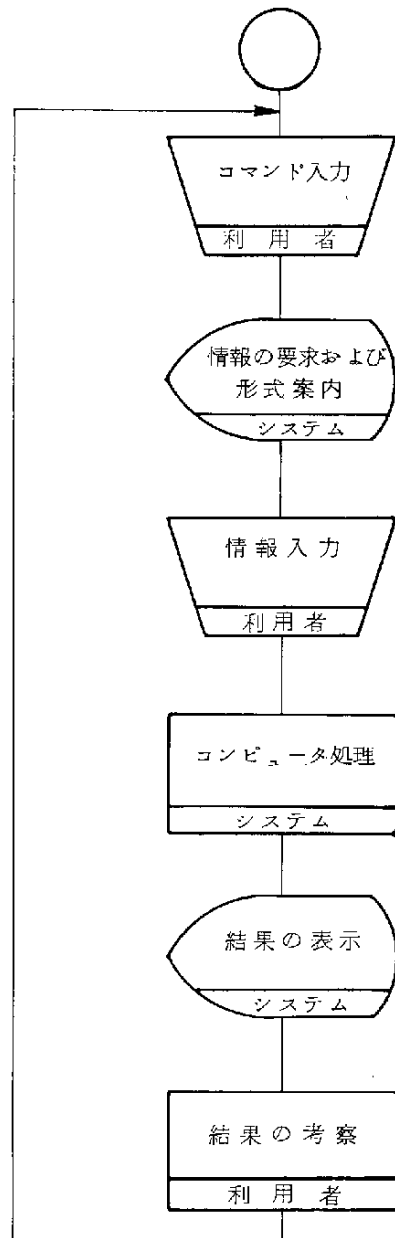


図 4 - 2 検索コマンドの基本パターン

表 4 - 1 検索コマンドの種類

検索コマンド	説 明
会 話 開 始	このコマンドは、検索サービスを受けようとするとき最初に入力するものである。これにより各種ファイル、データの初期設定を行なう。
キーワード入力	このコマンドは、目的とする文献情報を得るのに、検索の手がかりとなる語（キーワード）を入力するためのものである。
保 存 語 指 定	このコマンドは、キーワード入力によって得られたキーワード群のうち不要なものをとり除き、残ったキーワードを保存させるためのものである。
検 索 式	このコマンドは、正確な文献情報を得るために、キーワードを使って論理式を作成し入力するためのものである。
条 件 式	このコマンドは、検索式で得た文献情報に対し、さらに論理式および条件項目によって検索を限定するためのものである。
出 力	このコマンドは、検索した文献情報を出力させるためのものである。
ハード・コピー	このコマンドは、漢字ディスプレイ装置に表示されている文献情報と同じものを漢字プリンタ装置に出力させるためのものである。
継 続	このコマンドは、漢字ディスプレイ装置の 1 画面に表示しきれない情報を、継続表示させるためのものである。
会 話 終 了	このコマンドは、検索サービスを終了したいときに使用する。

表 4 - 2 データベースの構成表

項目	構成内容	蓄積モード	内 容
記事番号	EBCDIC		文献ごとに付られる唯一の番号
和文標題	漢 字		文献の日本語タイトル(和訳タイトル)
原文標題	漢 字		文献の原タイトル(外国語タイトル)
記事区分	EBCDIC		文献の記述内容の区分
資料略名	漢 字		掲載資料の略名
発行年	BINARY		文献の発行された年
巻	EBCDIC		資料の通巻
号	EBCDIC		資料の通号
言語コード	BINARY		記述されている言語のコード
発行国コード	BINARY		発行された国のコード
著者名	漢 字		文献の著者(個人, 団体)
所属機関名	漢 字		著者の所属する機関名
抄録文	漢 字		文献の要約

表 4 - 3 ファイル構成概要

項目	ファイル名	キーワード・テーブル	ポスティング・テーブル	データベース
レコード・モード		固 定 長	可 変 長	可 変 長
レコード・ファクタ		23	1	1
下位ポインタ数		1	複 数	—
上位ポインタ数		— (ブロック 1)	1	複 数
レコード・サイズ(バイト)		136	最大 13,024	最大 2,800
記 憶 媒 体		ディスク	ディスク	ディスク

ープルの作成，キーワード・テーブルのリンク情報の作成という手順で作られる。ポスティング・テーブルとは，文献データの格納アドレスをキーワードごとに集合したファイルであり，このポスティング・レコードの格納アドレスは，さらにキーワード・テーブルのリンク情報へとつながっている。またキーワード・テーブルのファイル構成に対してブロックごとにインデックスをとり，インデックス・テーブルを作成する。これは検索実行時にはインコアへ常駐して，キーワードによる探索のスピードを向上させるのに使用する。

インバーテッド・ファイルの構造を図4-4に示す。

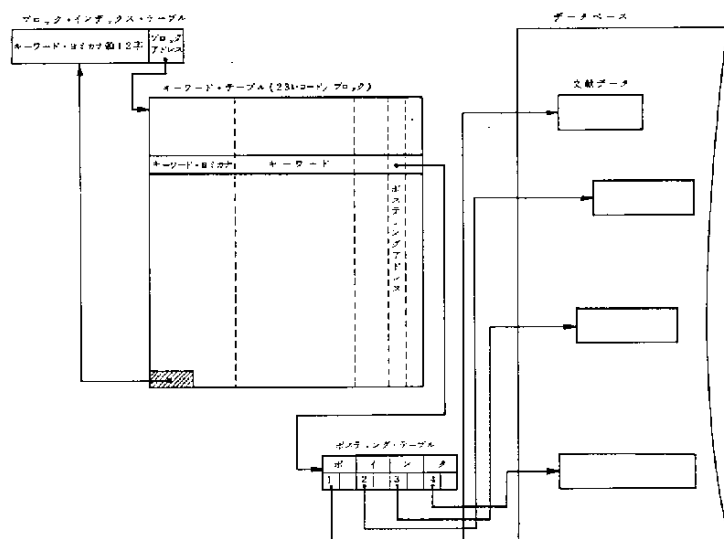


図4-4 インバーテッド・ファイルの構成図

(4) 索引誌作成管理サブシステム

検索サービスを補完する機能と，システムのレベルアップのための統計リスト作成機能を持ち，遂次拡充の予定である。

パラメータの入力によって得られる著者名索引など目的別に索引誌作成をすれば多くの人を対象とした情報サービスも可能である。

4.1.4 ハードウェア構成

(1) 機器構成

KINDシステムのハードウェアは、FACOM230-75に制御されるホスト・コンピュータと、FACOM U-200に制御されるターミナル・ステーションの2つの機器構成からなり、双方は通信回線を介して連絡する。

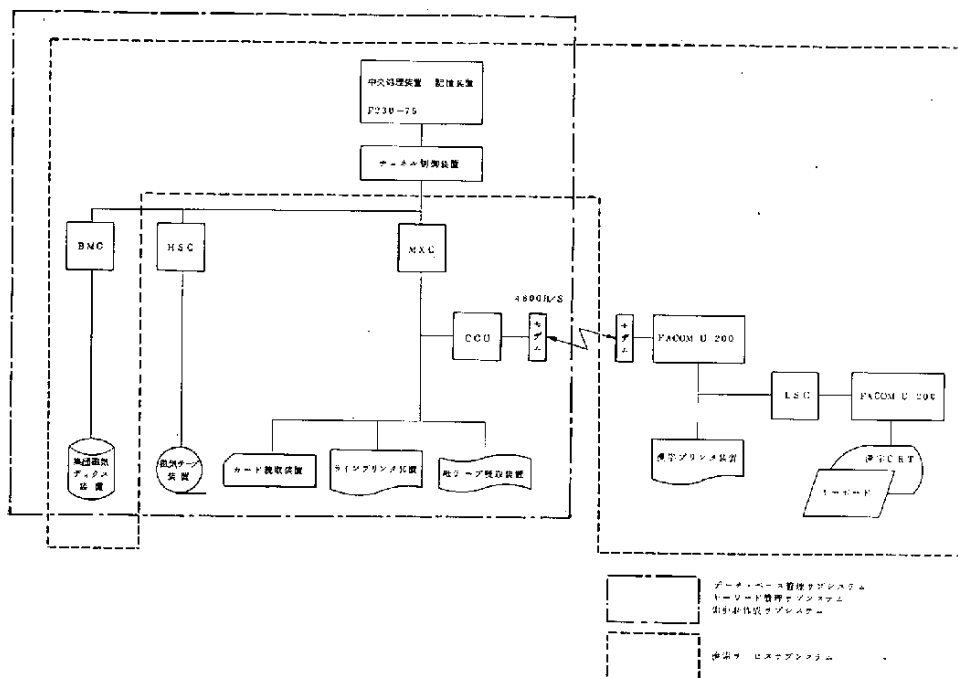


図4-5 KINDシステムの機器構成

(2) 装置概要

○ホスト・コンピュータ (FACOM230-75)

中央処理装置 1台 サイクル・タイム: 1ns/2w

記憶装置 記憶容量: 320KW

集積磁気ディスク装置 (F477K) 4パック アクセス・タイム: 26ms

情報転送速度: 806KB/S

磁気テープ装置 (F603N) 5台

記憶容量: 100MB/バック
 情報転送速度: 192/96KB/S
 トラック: 9
 記録密度: 1600/800BPI

○ターミナル・ステーション (FACOM U-200)

中央処理装置 2台 サイクル・タイム: 650ns
 記憶装置 記憶容量: { 32KW (ターミナル制御)
 16KW (ディスプレイ制御)

漢字ディスプレイ装置 (F6570) 1台 表示文字種: 5,376種+特殊5種
 表示文字数: 漢字512字,
 英/数1,024字
 プログラム・ファンクション: 16個×4ページ
 画面再生速度: 40フィールド/S
 コード: 富士通漢字コード,
 EBCDICコード

漢字プリンタ装置 (F6521A) 1台 収容文字数: 5,376字+ファン
 クション記号
 文字の大きさ: 9ポイント
 行ピッチ/文字ピッチ: 1/6インチ/3.2ミリ
 1行印字数: 最大55字
 印字速度: 65字/S
 コード: 富士通漢字コード

4.1.5 ソフトウェア構成

ホストおよびターミナル両構成とも、各々独自の機能を担当する。ターミナル側では、漢字キーボード、漢字ディスプレイ、漢字プリンタを用いて利用者のシステム利用窓口となり、検索指示コマンドおよび情報の入力、メッ

ページの表示、結果の出力など、メーカー提供の FACOM U-200 モニタ (COMOS-CRS) や KDP 制御、TSSP といったサービス・プログラムを用いて処理する。

ホスト側では、FACOM 230-75 のオペレーティング・システム (M-VII) の他、ユーザズ処理プログラムのすべてを適宜起動させ、KIND システムの各機能を実行する。

ユーザズ処理プログラムには、ETAM (Extended Telecommunication Access Method) を通して受渡しされるデータを対象に、オンライン処理される検索サービス関係プログラムと、データベース管理等バッチ処理形式のプログラムとの 2 通りに分類され、いずれも M-VII の下で動作する。

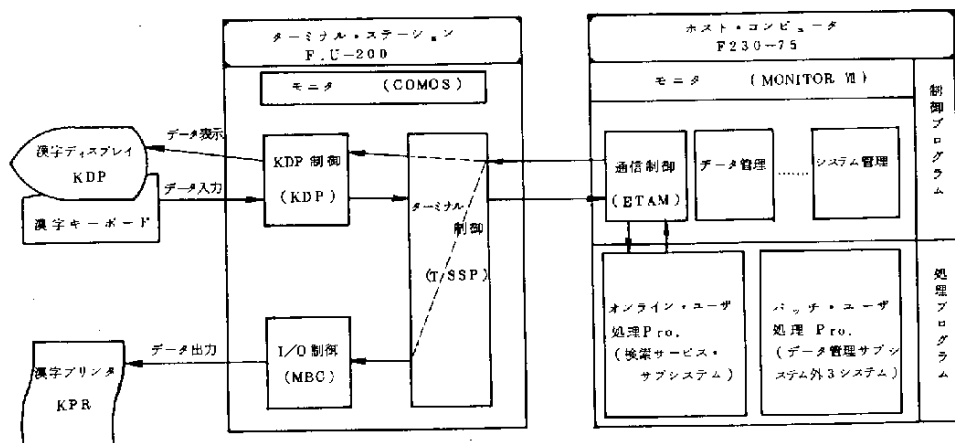


図 4-6 KIND システム・ソフトウェア構造

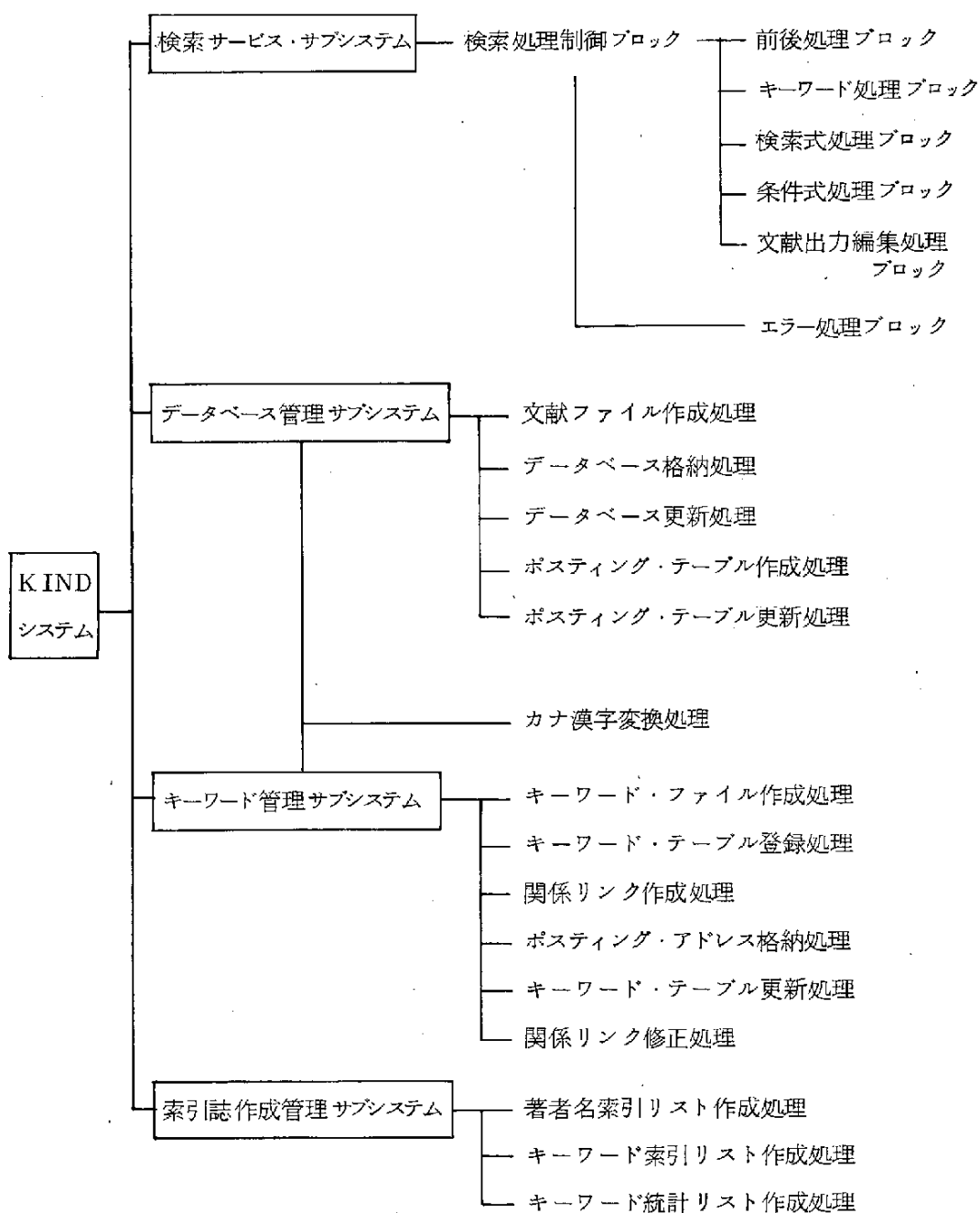


図 4-7 KINDシステムのソフトウェア

4.2 検索サービス・サブシステム

当サブシステムはKINDシステムの1サブシステムとして、データベース管理サブシステムおよびキーワード管理サブシステムで作成されたデータベース、キーワード・テーブル等を用い、利用者の要求する文献情報を検索し、端末に表示・出力するサービス・システムである。

利用者は、検索コマンドと対応する情報の入力を積み重ねて行くことにより合目的情報の検索をすすめることができる。

対応する情報とは求めたい文献を得るために

- ・キーワード : フリータームでよい
- ・検索式 : 論理式
- ・条件式 : 発行国, 発行年, 記述言語, 記事区分
があり、検索結果の出力媒体ないし出力形式指定のために
- ・出力様式
がある。

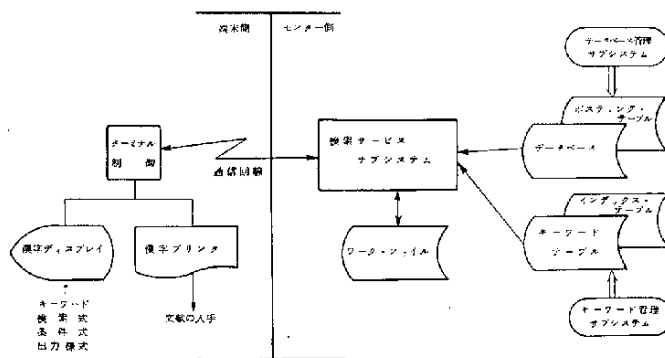


図4-8 検索サービス・サブシステムの概要図

4.2.1 処理概要

検索サービス・サブシステムの処理概要を図4-9に示し、それぞれについて概説する。

(1) 処理の制御

当サブシステムでは、処理についての管理は「メイン処理」によって集中的に行なわれる。

「メイン処理」では、端末から入力された検索コマンドを識別して何の処理を行なうか決定するとともに、制御に必要な情報を把握したりえて、該当する処理に制御が移される。それぞれに対応する一定の処理が行なわれたあと、再び制御は「メイン処理」に戻され、エラーの有無、制御情報のリセット等を行なって、次の検索コマンドの入力を持つ。

検索コマンドと処理の関連は表4-4のとおりである。

表4-4 検索コマンドと処理の関連

処 理	検索コマンド
キーワード入力	キーワード入力 保存語指定
検 索	検 索 式 条 件 式
文 献 出 力	出 力 ハード・コピー

(2) 処理の流れ

当サブシステムの処理概念は図4-9に示すように、次の3つの処理に大別される。

① キーワード入力処理

キーワード検索

キーワード保存

② 検索処理

検索式による文献検索

条件式による文献検索

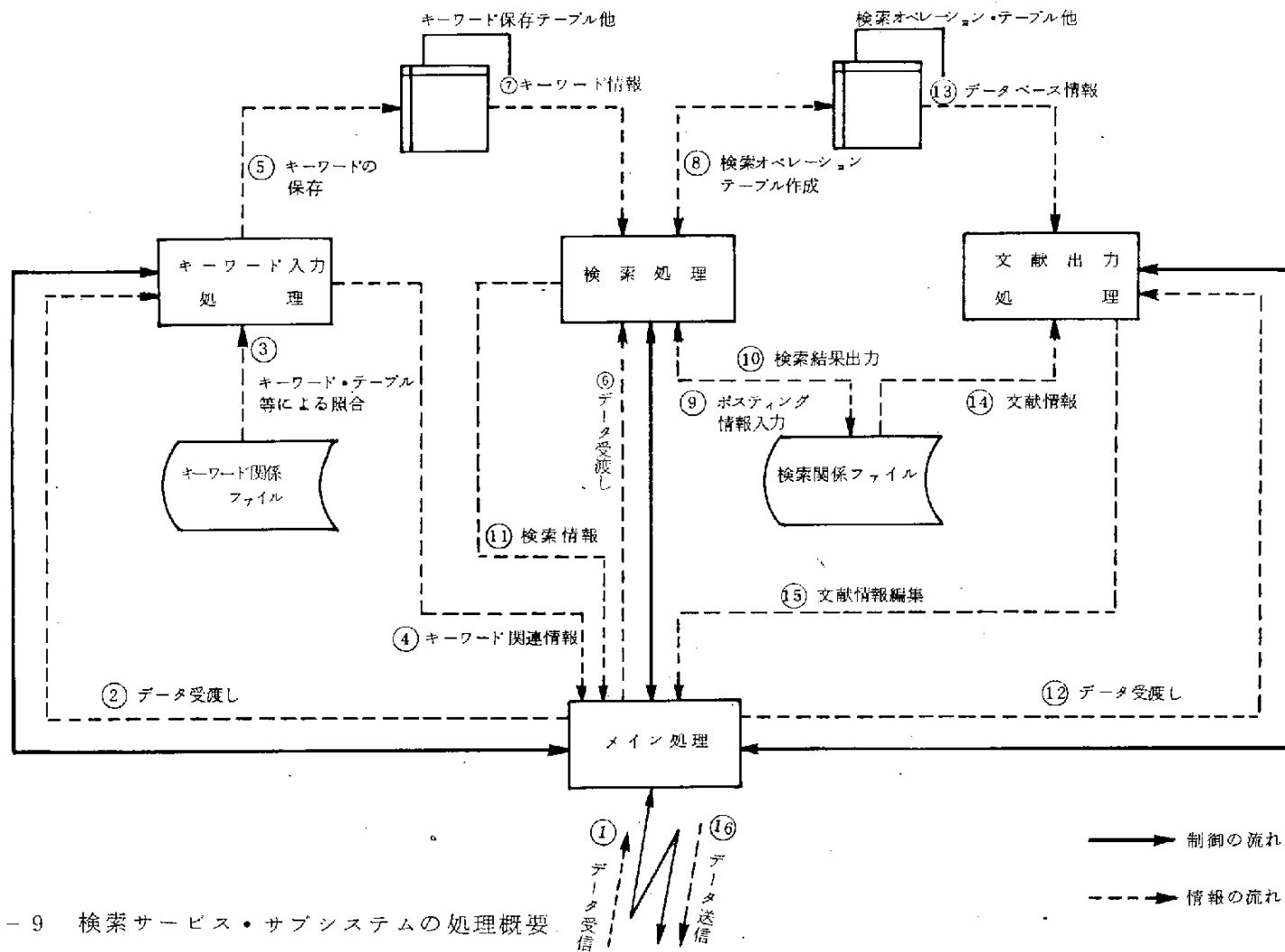


図 4 - 9 検索サービス・サブシステムの処理概要

③ 出力処理

文献情報の表示出力

端末から入力されるデータは、「メイン処理」によって検索コマンドが識別され、入力情報に適した処理へデータが渡される。検索コマンドの入力手順に従った処理の流れは以下のようになる。

最初に入力されるキーワード(カナ文字)は「メイン処理」によって、キーワード入力処理に制御が渡され、そこでキーワード関係ファイル(ブロック・インデックス・テーブル, キーワード・テーブル)を探索して漢字に置き換えられる。

さらに漢字キーワードから、下位概念語、あるいは前方一致によるキーワード検索を行なって、そこで得られるキーワード関連情報を端末の漢字ディスプレイに表示する。

つづいて、キーワード関連情報の中から利用者によってキーワードの保存語が指定されたら、それを主記憶内の保存語テーブルに登録する。

キーワード入力と保存語指定が繰返されたあと、保存したキーワードを使って、論理検索の指定がなされ、処理は検索に移る。

検索処理では、検索式と条件式による論理検索があるが、処理の流れとしては大体同じである。

論理検索を行なうには、まず入力された論理式を解説して、検索オペレーション用テーブルに検索順序などをスタックする。

検索に必要な情報はポスティング・テーブルなどから得られ、検索オペレーション・テーブルにしたがって検索が行なわれる。

その結果は中間ファイルに格納し、端末には検索情報を表示する。なお、検索式には保存キーワードに付けられた保存語番号を使って論理式が作られる。

条件式は保存語番号による論理式の他に、条件項目として発行年、発行国、記述言語、記事区分の指定が任意に行なえる。

検索された文献の表示出力は文献出力処理で行なわれ、出力コマンドに対する出力形式の指示を受けて、データベースから文献データを検索して画面編集あるいはプリント編集を行なって端末に送る。漢字プリンタへの出力には、SYSOUT出力と直接出力があり、前者のときはSYSOUT用ファイルに出力されて、会話終了後に一括して出力される。また直接出力の場合には、会話を中断して出力が行なわれる。

4.2.2 処理ブロックの構成

通常1つのシステムは、目的とする結果を得るためにいろいろな処理を行ない、各処理における結果（中間結果）にもとづいて有機的に解を導くものである。このことはシステムの動作を理解するうえでも、作成するうえでも重要なことであり、いまこの各処理の単位をブロックすると、当サブシステムは図4-10に示すように、いくつかの処理ブロックから構成される。

この図で主ルーチンとは常駐部分であり、副ルーチンとは非常駐部分と考える。これは、実行時において主記憶の使用効率の向上を計り、かつ目的の処理ブロックに対してデータの授受がしやすいように配慮したものである。各処理ブロックの関連を説明すると以下のようになる。

端末から送られてきた情報はまず検索処理制御ブロックに入り、そこでどの処理ブロックが適当であるか判断される。処理ブロックが決定されたならばその処理ブロックを呼出し（すでに呼出し済であるならその必要はない）て制御を移行する。前後処理ブロックは検索サービスに先だって各種テーブルのイニシャライズやファイルの開設などを行ない、サービスの終了に際してはシステムの後始末を行なうものであり、このブロックはそれ以外呼出されることはない。

キーワード処理、検索処理、条件式処理、文献出力編集処理の各ブロックは、前述したように入力情報により適宜呼出される。エラー処理ブロックは主ルーチンに含める。

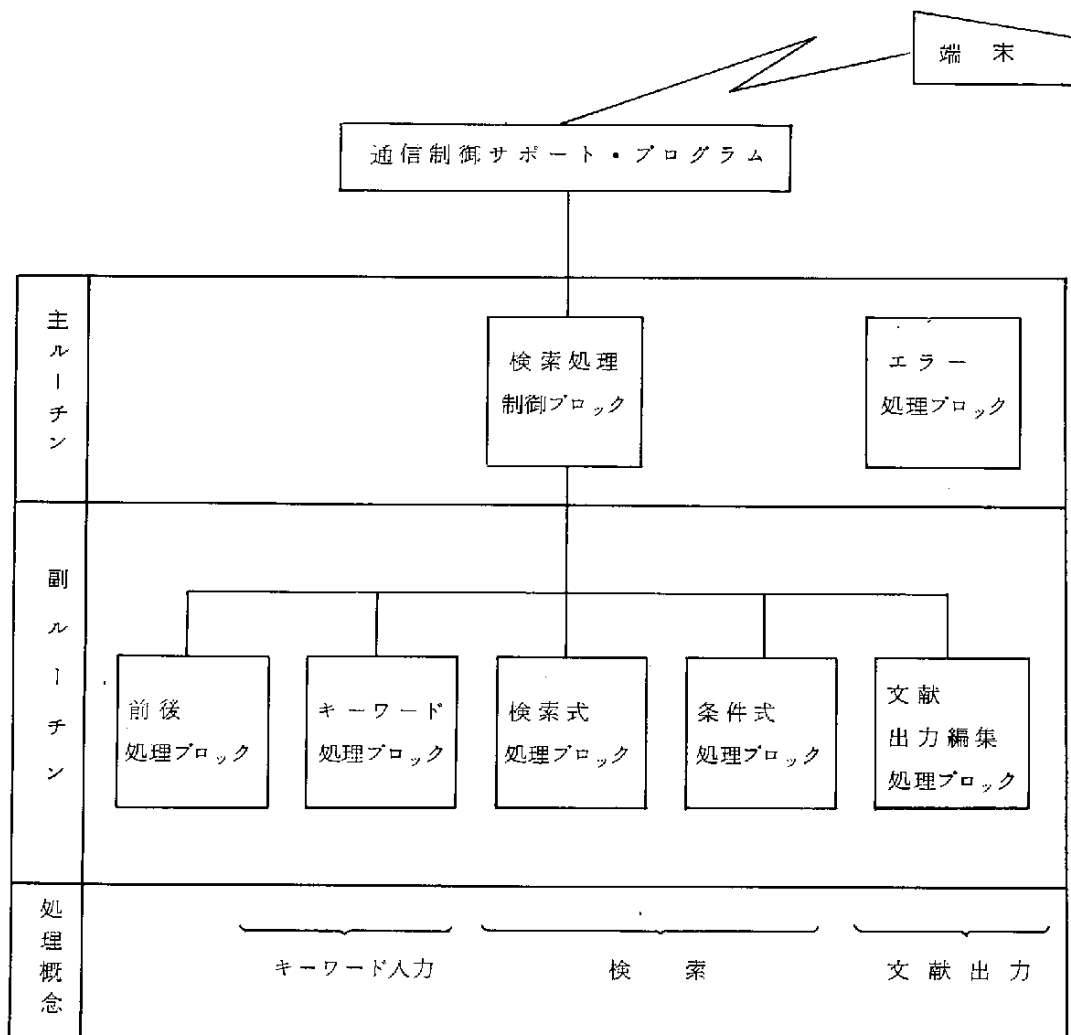


図 4-10 処理ブロック構成図

これはエラーが発生した場合、当サブシステムは端末側の利用者に対してなんらかの通知を行なう必要があるが、この通知（画面編集する）を各処理ブロックで別々に行なわせるのでは、プログラム作成上無駄が生じる。

エラー処理の扱いは各ブロックとも共通した部分が多いのでエラー情報を検索処理制御ブロックに通知することにより、エラー処理ブロックで一括して行なうようにし無駄を排除した。

また多くの場合エラーは端末からの応答を待つことがあるが、その場合、エラーの発生した状態にしておくことが望ましいと考え、エラー処理ブロックを主ルーチンにしている。

4.2.3 処理ブロックの概略

当サブシステムを構成する処理ブロックに関し、データの処理方法を中心に概要を説明する。

検索処理制御ブロック

(1) 機 能

- 各処理ブロックの制御を行なう。
- 検索の流れを監視する。
- 端末からの電文を受信する。
- 各処理ブロックで編集された電文の送信を行なう。

(2) 概 要

検索処理制御ブロックの概要を図4-11に示す。

この処理ブロックは、端末からシステムの起動がかけられたとき、最初に動作するブロックであり、受信電文（検索コマンド、データ）は処理ブロック管理表によって管理される。受信した電文は、内容を解析されチェックが行なわれたのち当該処理ブロックへ受継がれて、制御はその処理ブロックへ移る。

各処理ブロックは、オーバレイ構造となっており、該当する処理ブロックがインコアにローディングされていないときは、新たにロードしなければならない。

各処理ブロックでの処理が終了したならば、制御は当処理ブロックへ戻される。また処理ブロック管理表のエラー通知情報を検査して、エラー通知がなされていれば、該当処理ブロック中でエラーが生じたことになり、

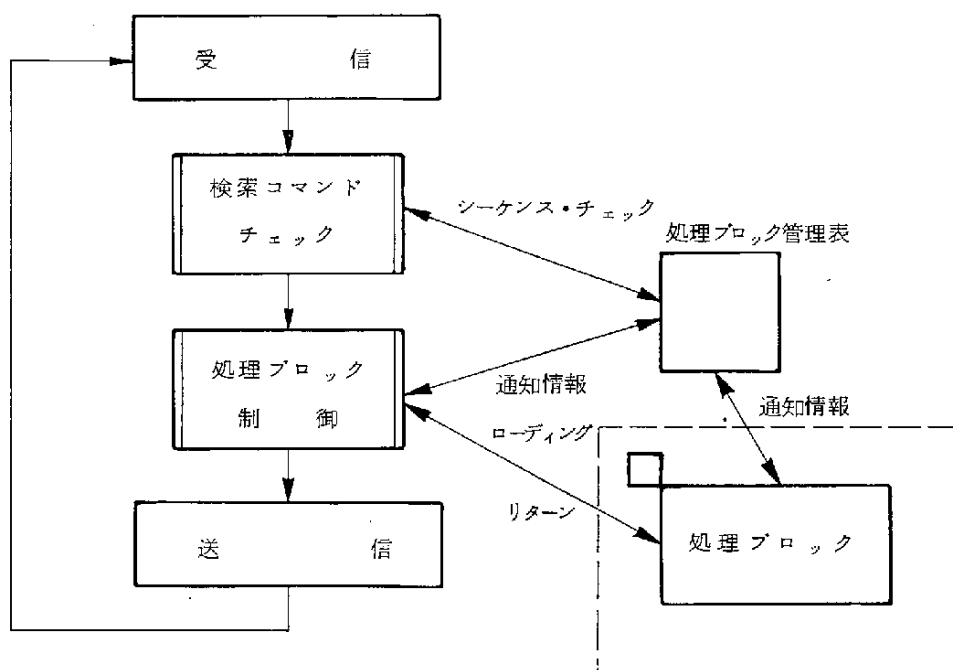


図 4 - 1 1 検索処理制御の処理概要

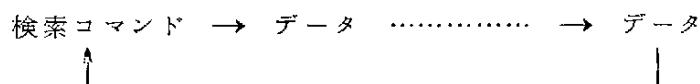
エラーに関する処理をエラー処理ブロックに依頼する。

各種の処理によって画面編集された電文は、当処理ブロックにおいて端末側に送信し、そのあと受信態勢に入る。

(3) 主な処理の説明

● 検索コマンドのシーケンス

端末から送信されてくる電文は、次のようなサイクルをもっている。



検索コマンドは、前の検索コマンドとの処理関係を見捨てず、常にどの処理ブロックを実行しているのか、把握しておかなければならない

が、当サブシステムでは、処理ブロックを管理するのに処理ブロック管理表を用いている。

検索コマンドは、一定の流れにそって端末から入力されるものであるが、それに反したコマンド入力も当然考えられる。それによる検索処理上のミス、無駄を、受信した時点で発見することが必要である。

当サブシステムの検索コマンドの関係は、表 4-5 のとおりである。

表 4-5 検索コマンドの処理関係

検索コマンド	入力が許される検索コマンド
会 話 開 始	キーワード入力
キーワード入力	キーワード入力, 保存語指定, 検索式, 継続
保 存 語 指 定	保存語指定, キーワード入力, 検索式, 継続
検 索 式	検索式, キーワード入力, 条件式, 出力, 継続
条 件 式	条件式, 出力, 継続
出 力	出力, ハード・コピー, 会話開始, 会話終了, 継続
ハード・コピー	
継 続	
会 話 終 了	

また、処理ブロック管理表における情報の受渡しを示したのが図 4-12 である。

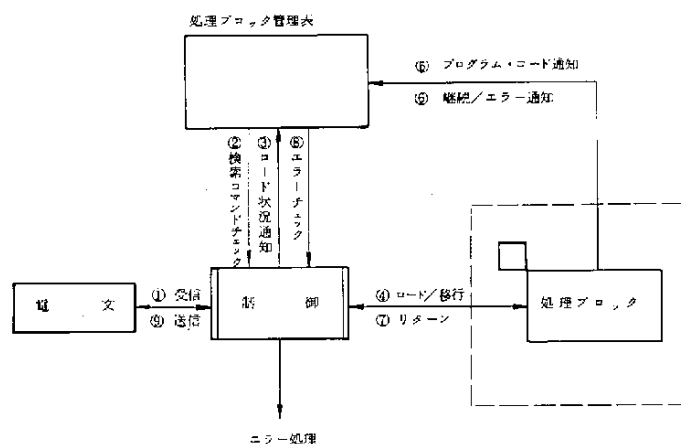


図 4 - 1 2 ブロック管理情報の受渡し

●画面編集された電文の送信

当サブシステムでは、画面の編集は各処理ブロックで行なうため、検索処理制御ブロックでは送信のみを行なう。

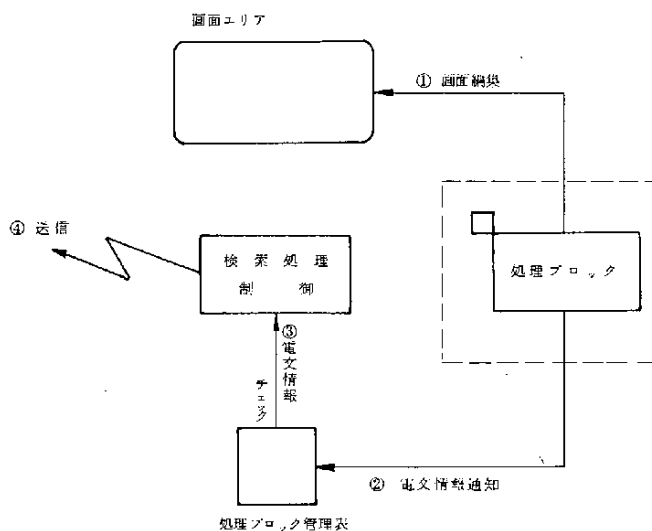


図 4 - 1 3 画面送信手順

エラー処理ブロック

(1) 機能

○各処理ブロックで検出されたエラーのメッセージ編集を行なう。

(2) 概要

エラー処理ブロックは、検索処理制御ブロックと同じく主ルーチンにあって、エラー・メッセージの処理を検索処理制御ブロックからの依頼を受けて行なわれる。

他の処理ブロックにおいてエラーが検出されると、処理ブロック管理表の該当エリアにエラー情報（エラー通知ビット、エラー・ステータス）がセットされる。これを受けて、エラー処理ブロックでは、エラー内容をエラー・ステータスから識別を行ない、エラー・メッセージ・ファイルの中からエラー内容に適したメッセージを読込んで、そのメッセージを画面編集したりして、検索処理制御ブロックに制御を戻す。

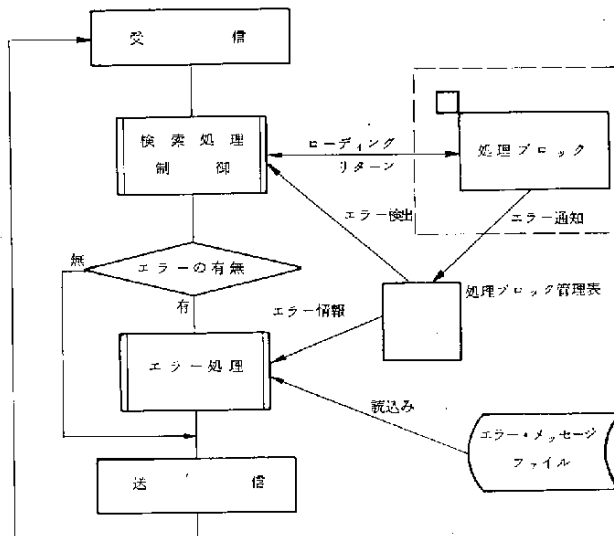


図 4-14 エラー処理概要

(3) 主な処理の説明

● エラーメッセージの編集

エラー・ステータスにセットされた情報（エラー番号）から，メッセージ・テーブルを使ってエラー・メッセージ・ファイルのアドレスに変換する処理を行なう。その手順は次のとおりである（図 4-15）。

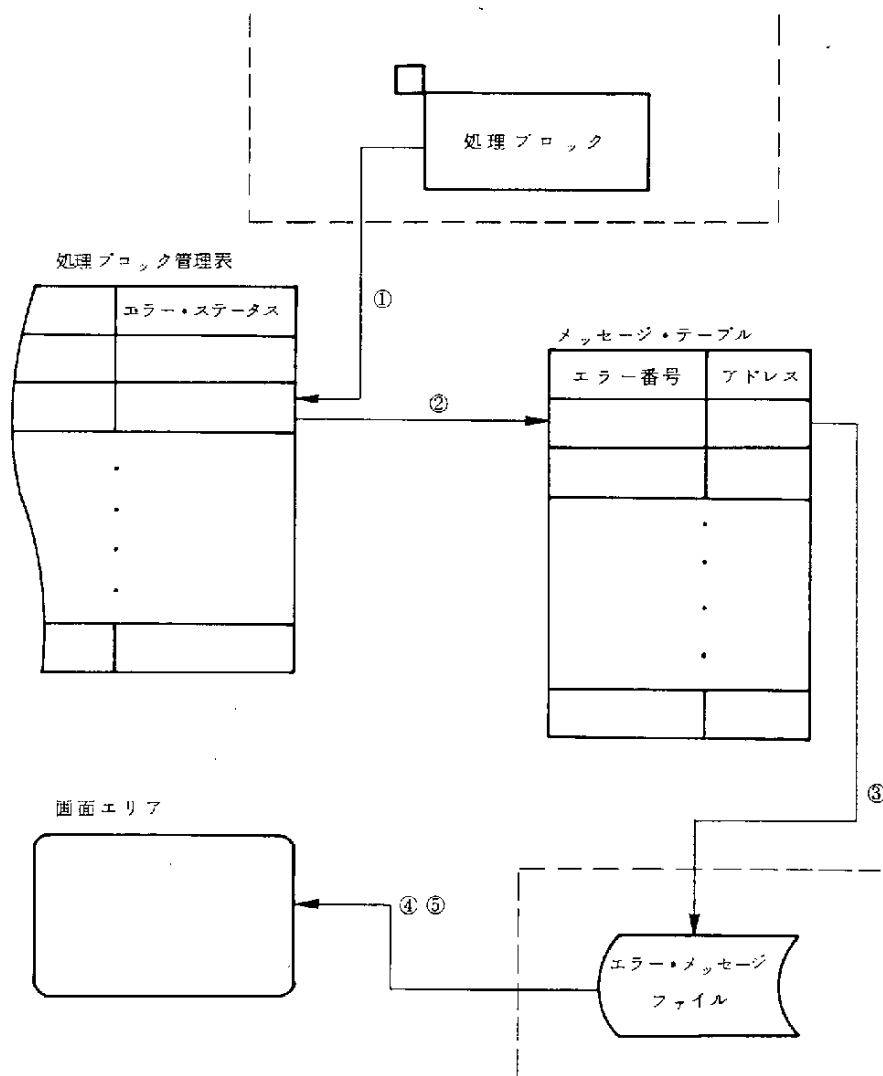


図 4-15 エラー・メッセージ編集の流れ

- ① 処理ブロック管理表を調べ、エラー発生処理ブロックを知る。
- ② 該当処理ブロックのエラー・ステータスによりエラー番号を知る。
- ③ ②で得られたエラー番号からメッセージ・テーブルをサーチしてエラー・メッセージ・ファイルのアドレス (T T R S) に変換する。
- ④ エラー・メッセージ・ファイルから該当するエラー・メッセージを読み込む。
- ⑤ メッセージを画面編集する。

前後処理ブロック

(1) 機能

- o 各種ファイルのオープンおよびクローズ
- o 各種テーブル、データの初期化
- o 必要な情報の読み込み

(2) 概要

前後処理ブロックは、“会話開始”コマンド、および“会話終了”コマンドによって起動され、検索サービスの開始と終了に関する処理を行なう。

各種ファイルのオープンは、その処理ブロックからいつでもファイルを参照できるように、前もってオープンしておく。

各種テーブルおよびデータの初期化の目的は、新規に当サブシステムがロードされたとき、あるいは再検索が行なわれるときに、前回までの不要な情報が残っていたのでは正しい処理ができない。

そのため検索に先だって必要な各種のテーブルやデータを、初期状態に設定する。

“会話終了”コマンドによる検索の後処理としては、検索サービスに要した時間や出力件数などを検索処理情報として端末に表示するとともに、使用した（オープンしておいた）各種ファイルのクローズを行なう。

キーワード処理ブロック

(1) 機能

- カタカナで入力されたキーワードを漢字に置換える。
- シソーラス、または前方一致によりキーワード関連語を検索する。
- キーワードを保存する。

(2) 概要

キーワード処理ブロックは、端末から“キーワード入力”コマンドが入力されたときに動作する処理ブロックであり、その処理概要を図4-16に示す。

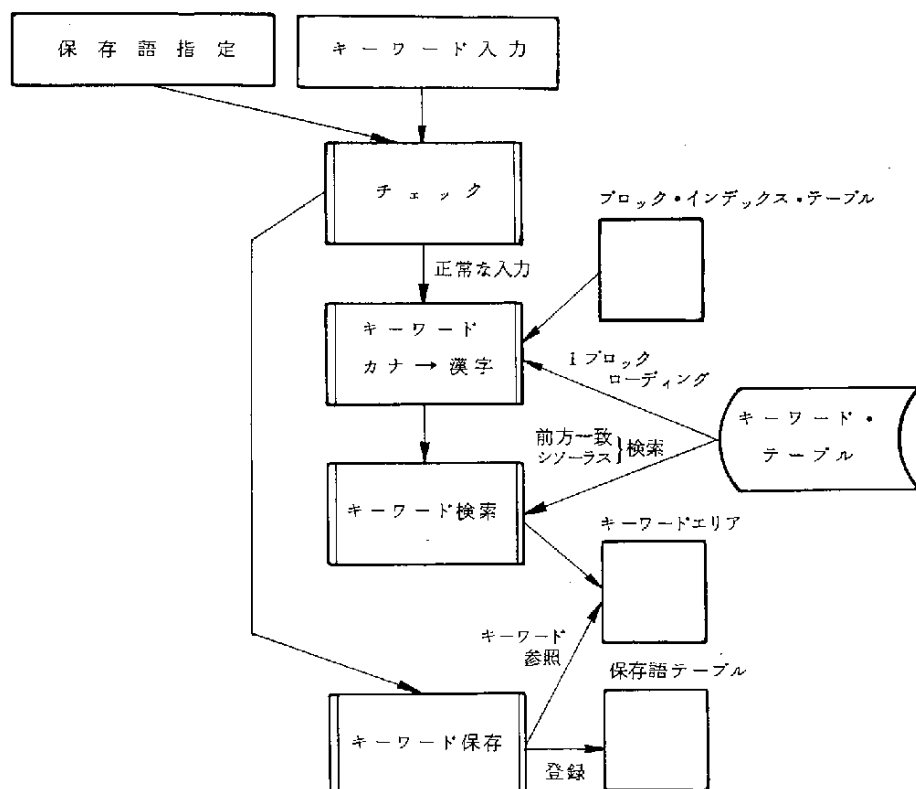


図4-16 キーワード処理概要

カタカナ（ヨミカナ）によるキーワードが入力されると、それが正しく入力されたかチェックが行なわれ、エラーが検出されると処理ブロック管理表のエラー・ステータスに、エラー内容を通知して主ルーチンに制御が戻される。

正しく入力されたキーワードについては、ブロック・インデックス・テーブルのヨミカナ（頭12文字）と大小の比較を行なって、キーワード・テーブルの該当するブロックをインコアにロードする。このブロックを検索することにより、キーワードのヨミカナから対応する漢字に置かえられる。検索されなかった入力キーワードに対しては、未登録キーワードとして端末に通知する。

また、入力されるキーワードについて関連語サービスを行なうが、これは利用者のサービス指定にしたがってシソーラス、あるいは前方一致で検索を行なって関連語を端末のディスプレイに表示する。シソーラス・サービスについては入力キーワードに対する下位概念のキーワードをサービスし、前方一致ではキーワードの漢字について前方2字が共通しているキーワードをサービスする。いずれかの方法によって検索されたキーワード関連情報は、検索語番号を付加して画面編集を行ない端末に送信する。

キーワードの保存語処理は、“保存語指定”コマンドによって行なわれ、検索語番号を使って保存語の指定がなされたキーワードについて、新たに保存語番号を付加して保存語テーブルに登録する。

(3) 主な処理の説明

● キーワード（ヨミカナ）の漢字変換

キーワードは漢字とヨミカナが対応した形で、キーワード・テーブルに格納されており、キーワード管理サブシステムによって約1万語近くのキーワードが登録されている。この中から目的の漢字に変換する流れは図4-17に示すとおりであり、次のような内容の処理を行なう。

① 端末の利用者からカタカナのヨミカナで入力されたキーワードをブ

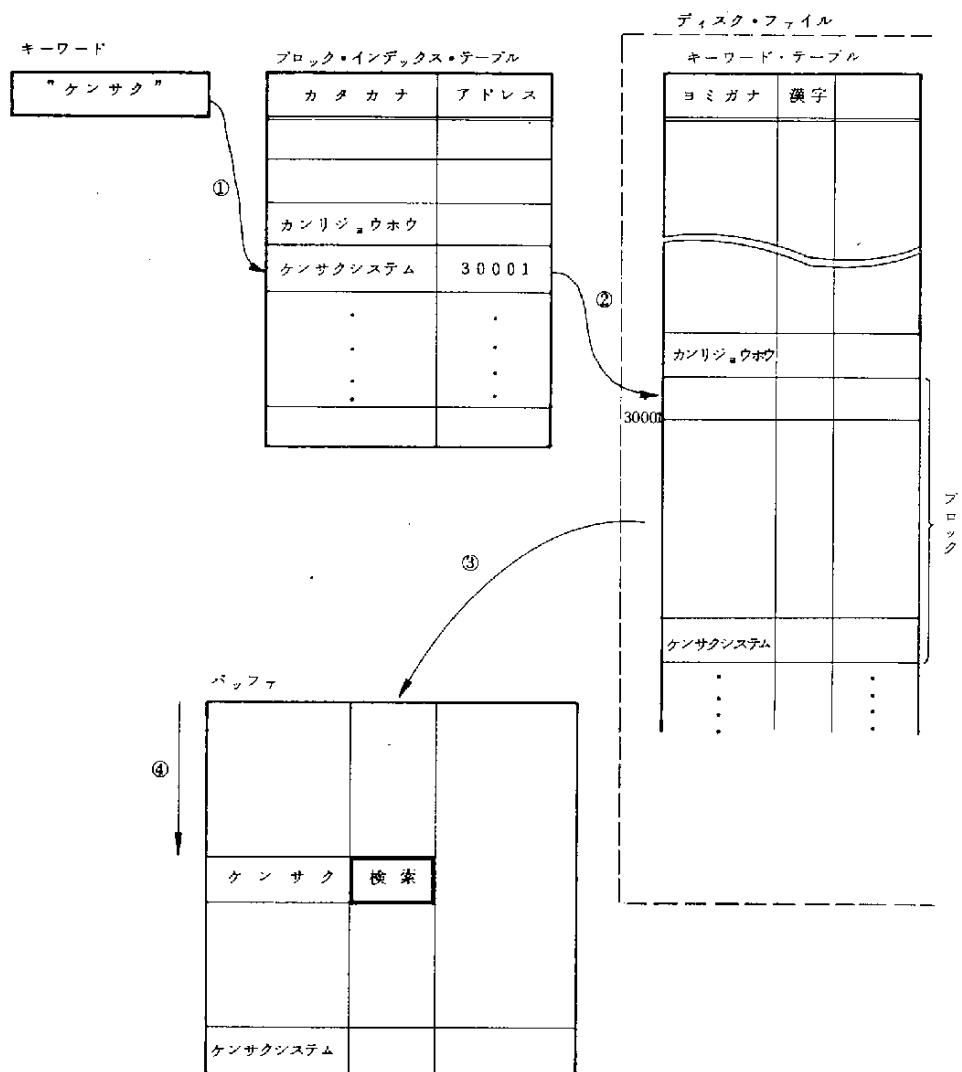


図 4 - 1 7 キーワード漢字変換の流れ

ロック・インデックス・テーブルのカタカナ項目と大小比較を行なって、該当する位置を探す。

- ② ブロック・インデックス・テーブルのアドレス項目はキーワード・テーブル上の該当ブロックを指しており、カタカナは該当ブロックの最終レコードのヨミカナ（頭12文字）である。
- ③ キーワード・テーブルの該当ブロックをインコアに読込む。
- ④ 読込んだブロックの先頭からサーチして対応する漢字キーワードを探し出す。

●同音異義語の処理

カタカナで検索する際に次の例にみられるような、同音異義語の問題が生じる。

シュウオンキキ	収音機器
	集音機器
ジョウホウキカン	情報機関
	情報帰還

このような場合、検索サービスでは、どちらをとるか端末の利用者に問合せて選択させる。

●前方一致サービス

該当するキーワード（漢字）の前方2字について前方一致をとり、2字が共通するキーワードを関連語とみなして検索表示する。

たとえば、キーワードとして“情報”を与えた場合、次のようなキーワードが検索サービスされる。

情報管理
情報検索
情報検索システム
情報サービス
情報システム
⋮
etc.

前方一致処理の流れを図4-18に示す。

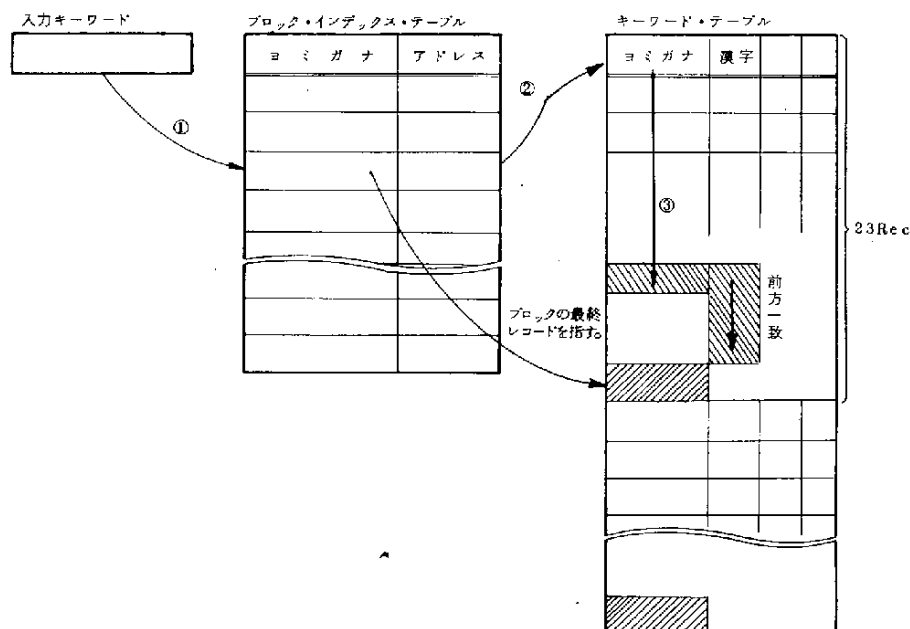


図4-18 前方一致処理の流れ

注) 前方一致処理では非ディスクリプタは一致の対象としないが、USE関係にある語を検索して表示する。

(例) IR

IRシステム

USE

IRIAC

IRS

情報

情報検索システム

この処理で注意することは、該当ブロック外にも、前方一致となる語が存在する場合がある。したがって該当ブロックの前後にあるブロックについても検索チェックを行なう必要がある。

● シソーラス・サービス

シソーラスの処理は、自由な言葉の発想により、整理された語を検索することであり、当システムでは下位語に相当するキーワードをサービスする。

たとえば、キーワードとして“情報検索”を与えた場合、次のものが検索される。

情報検索	
事項検索	} 下位語
そ及検索	
文献検索	

このうち“事項検索”に対して、さらに下位語をたどると

事項検索	↓ 下位語
データ検索	
化学構造検索	

が得られる。

シソーラス処理の流れを図4-19に示す。

シソーラス・リンク情報には、シソーラス関係語のアドレスが格納されており、各キーワードの下位語あるいはディスクリプタがリンクされている。

下位語の場合には、複数個の下位語が対応するために下位リンク・サブテーブルの格納アドレスをたどって間接的に求める方法になっている。非ディスクリプタからディスクリプタへのリンクはキーワード・テーブル上のアドレスを持つ。

● キーワードの保存

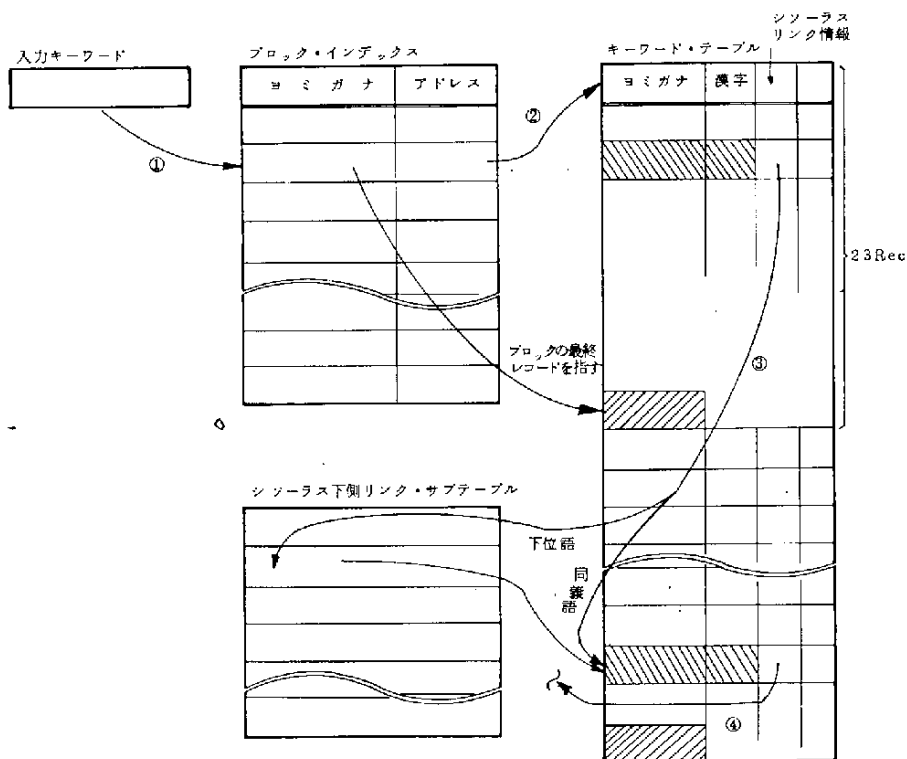


図 4 - 1 9 シソーラス処理の流れ

キーワード入力によって検索された関連キーワードは、前方一致、シソーラスの処理にかかわらず、すべてキーワード・エリアに順次格納されるが、キーワード保存の指定ができるのは、このエリアに格納されているものに限られる。

保存語の指定がなされたキーワードには、検索時の語番号とは別に新たに保存語番号が付加される。

キーワード保存処理の流れを図 4 - 2 0 に示す。

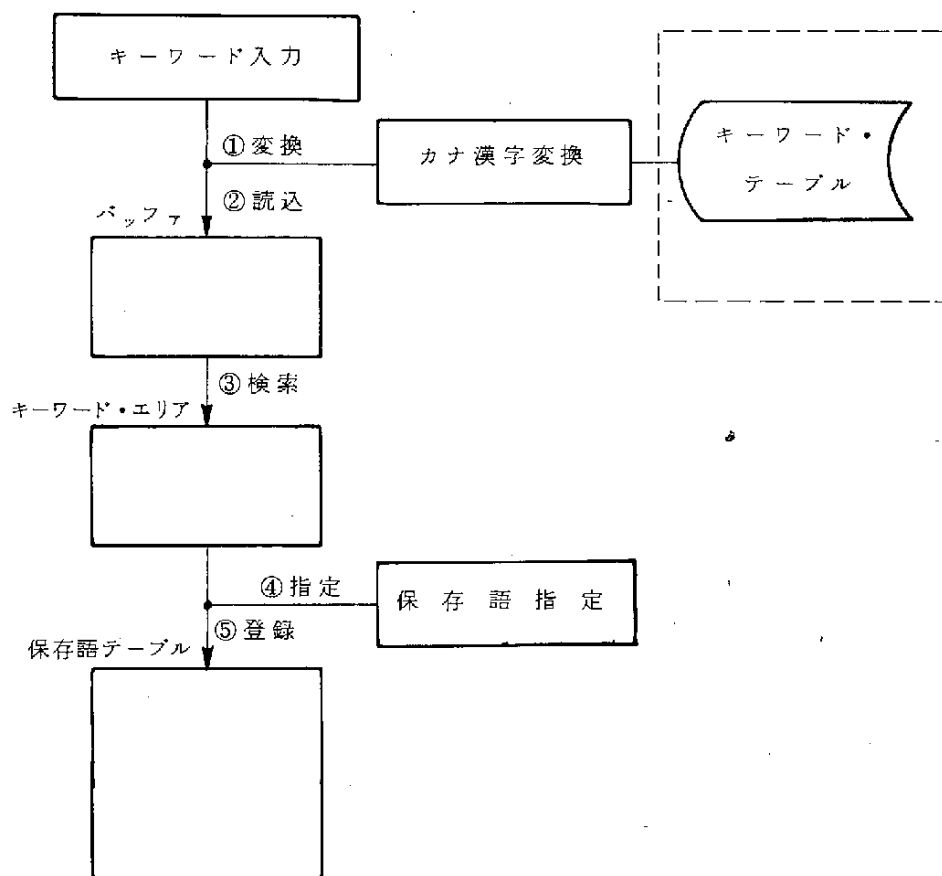


図 4 - 20 キーワード保存処理の流れ

検索式処理ブロック

(1) 機能

- 入力された検索論理式のチェックを行なう。
- 正常な論理式からオペレーション・テーブルを作成する。
- オペレーション・テーブルにより，論理検索を行なう。

(2) 概要

検索式処理ブロックは，“検索式”コマンドが入力されたとき動作する処理ブロックであり，その処理概要を図4-21に示す。

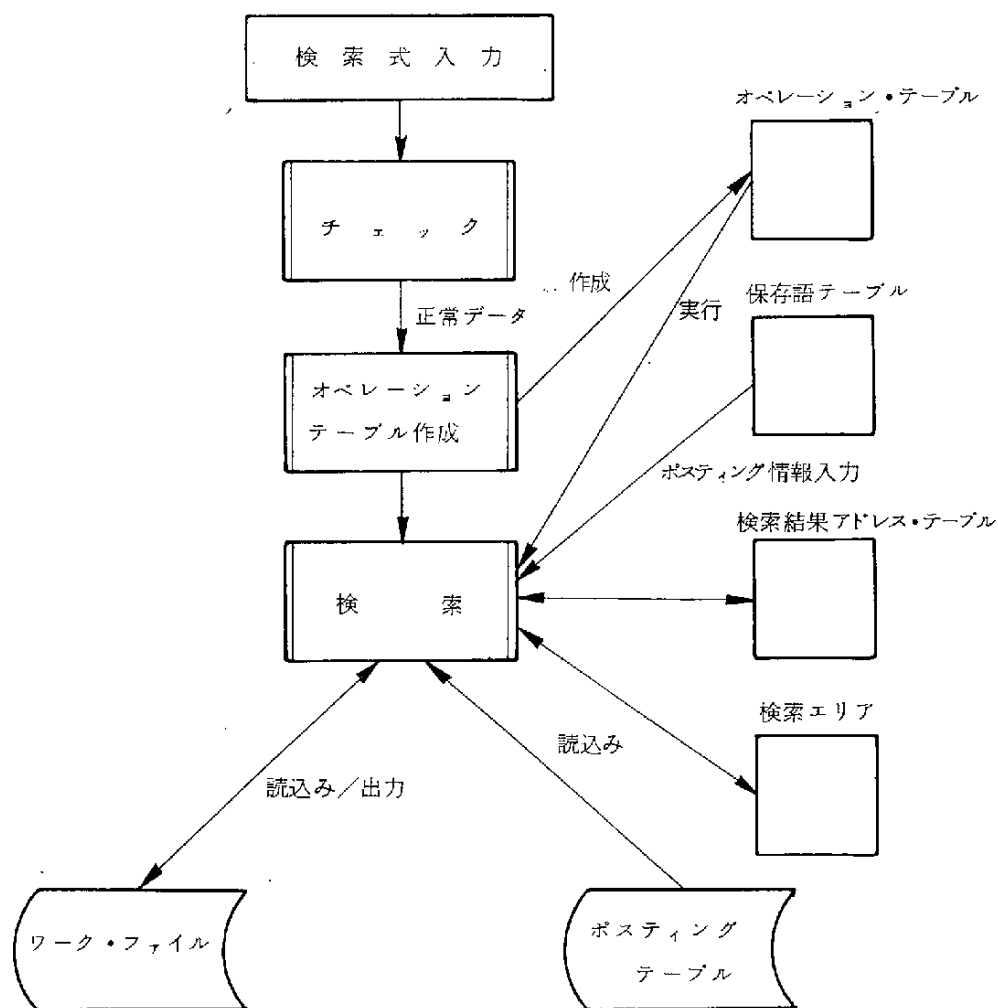


図4-21 検索式処理概要

保存語番号と論理演算子などを組合せて入力される検索式はデータ・チェックが行なわれ、検索式にエラーが検出されると、処理ブロック管理表のエラー・ステータスに、エラー内容を通知して主ルーチンに制御が戻る。

正しく入力された検索式は、オペレーション・テーブルを作成するため、項目（論理式中の各項目で保存語番号等）の照会などが行なわれる。

オペレーション・テーブルは検索式を一定の規則で変形したものであり、これによって実際の検索が行なわれる。検索された中間結果は検索エリアに格納され、最終結果が得られた時点で、検索結果データ・ファイルに出力される。

(3) 主な処理の説明

●検索式のチェック

検索式は、選定したキーワードでより確かな文献を検索するために作成するものであり、この式自体が間違っていたのでは検索は行なえない。

検索式のチェックは次の3つについて行なう。

- 文字種類
- 式
- 項目

●オペレーション・テーブルの作成

オペレーション・テーブル作成の目的は、論理演算を簡単化することと検索手順を明確にすることであり、検索式を変形して作成される。

検索式の各項目は、保存語番号あるいは式番号で構成されており、これを演算子で結んでいる。

オペレーション・テーブル作成の流れは図4-22のとおりである。

正常に入力された検索式は、一定の規則で分解されてオペレーション・テーブルが作成される。規則とは主として、演算（検索）の順序関係であり、その関係を表4-6に示す。

この表を使って分解した式の演算順位が求まったら、オペレーション

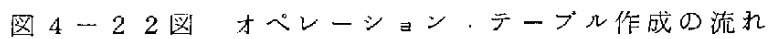


表 4 - 6 演算子の順序関係

順 位	演 算 子	重 み
1	N O T (-)	3
2	A N D (*)	2
3	O R (+)	1

(カッコは最優先し、重みは3とする)

テーブルに、次の規則で各項目を格納する。

- 演算順位の高いものから格納する。
- 検索式の各項目に使われた保存語番号はそのままの形で格納する。
- 演算項目に使われる式番号とは検索結果アドレス・テーブルの格納エリアに相当するアドレスであり、式保存番号に変換したものを格納する。
- 検索結果格納エリアの決定は検索エリア管理テーブルにより、空エリアを探して格納する。

例をあげてオペレーション・テーブルを作成してみると以下のようになる。

検索式

$$K 0 1 . (0 1 + 0 2) * (0 3 + 0 4)$$

この式を分解して演算順位を並べかえる処理は次のように行なわれる。

$$\begin{array}{rcccl}
 & (0 1 + 0 2) & * & (0 3 + 0 4) & \\
 \text{重み付加} & 3 + 1 & & 2 & 3 + 1 \\
 & \downarrow & & \downarrow & \downarrow \\
 \text{順位得点} & 4 & & 2 & 4
 \end{array}$$

この結果から順位得点が高いほど演算順位は高くなり、検索優先権が与えられる。これにしたがい、オペレーション・テーブ

ルの各項目がセットされる。

検索結果を格納する検索エリアの決定は次のような手順で行なわれる。

- 検索エリア管理テーブルのステータスをみて、空エリアを探す。
- 空エリアに付られている検索エリア番号が検索結果格納エリアとなる。
- その検索エリア番号に対応するステータスにフラグを立てる。
- オペレーション・テーブルのオペランドに検索エリア番号を格納したら、そのステータスをリセットする。

よって上記の検索式のオペレーション・テーブルは表 4-7 のようになる。

表 4-7 オペレーション・テーブル作成例

検 索 優先権	検 索 結 果 格 納 エ リ ア	演 算	オ ペ ラ ン ド	
			第 1	第 2
4	4 1	+	0 1	0 2
4	4 2	+	0 3	0 4
2	4 3	*	4 1	4 2
0	0	0	0	0

● 検索方法

論理検索はオペレーション・テーブルの検索優先順位に従って順次行なわれる。

論理演算は、文献アドレスを用いて比較するので、必要なデータを該当するテーブルから得ながら処理を行なう。

検索処理（論理演算）の流れを図 4-23 に示す。

- ① オペレーション・テーブルの第 1 オペランドをチェックし、保存語番号が得られる。

表 4-8 テーブル関連表

オペランド項目	データを得るためのテーブル名
保存語番号	保存語テーブル
式保存番号	検索結果アドレス・テーブル
検索エリア番号	検索エリア管理テーブル

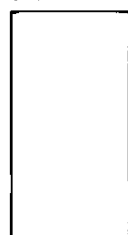
オペレーション・テーブル (概略)

検索結果 格納エリア	オペランド	
	第 1	第 2
41	01	05
⋮	⋮	⋮

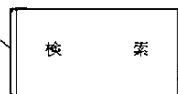
保存語テーブル (概略)

保存語番号	アドレス
01	01011
05	02000
⋮	⋮

検索エリア-I



検 索



ポスティング・テーブル

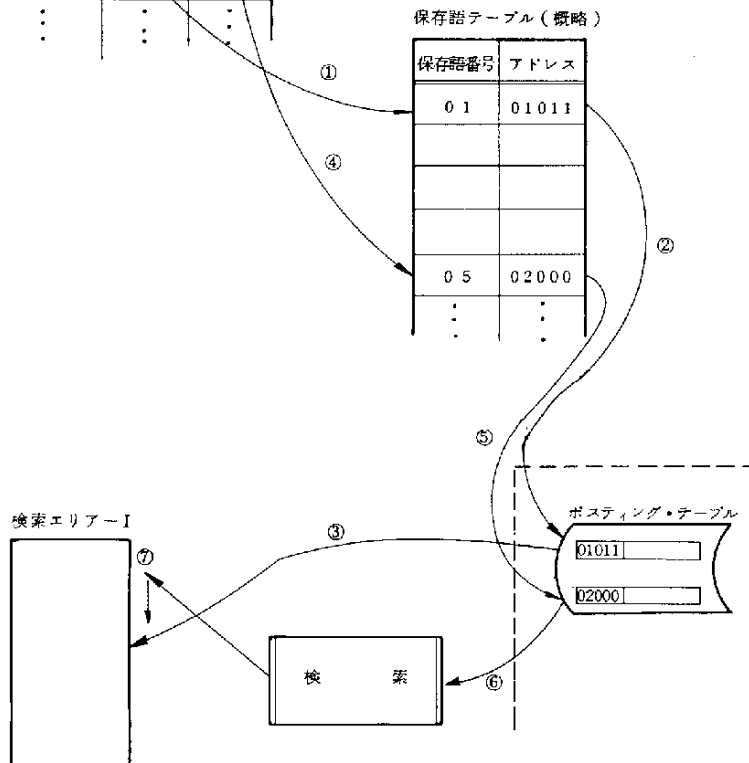
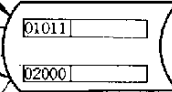


図 4-23 検索処理の流れ

- ② 保存語テーブルを参照して、ポスティング・テーブルのアドレスを知る。
- ③ ポスティング・テーブルの該当するレコードを、検索エリアに読込む。
- ④, ⑤ 第2オペランドについても①, ②の処理を行なう。
- ⑥, ⑦ 第2オペランドから得られたポスティング・テーブルのレコード内容と、検索エリアの内容とで演算子にしたがって論理演算を行なう。

条件式処理ブロック

(1) 機能

- 入力された条件論理式のチェックを行なう。
- 正常な論理式からオペレーション・テーブルを作成する。
- オペレーション・テーブルにより、論理検索を行なう。

(2) 概要

条件式処理ブロックでは、検索式が条件式に変わるほかは、ほぼ検索式処理ブロックと同様の処理が行なわれるが、条件項目による検索処理が追加される。

(3) 主な処理の説明

●条件項目による検索

条件項目による検索は、論理式によって論理検索をしたあと、その結果に対して条件項目を比較する。

条件項目は次の4種である。

- 言語コード
- 発行国コード
- 発行年
- 記事区分

これらの条件項目は、ポスティング・テーブルに格納されているので、データベースまで検索しなくても比較できるようになっている。

文献出力編集処理ブロック

(1) 機能

- 検索した文献データを、漢字ディスプレイ装置に表示するために画面編集を行なう。
- 文献データを、漢字プリンタ装置に出力するための帳票編集を行なう。
- 利用者からの出力形式の指定にしたがい、文献データを SYSOUT ファイル（システム出力用のファイル）に出力する。

(2) 概要

文献出力編集処理ブロックは、“出力”コマンドが入力されたとき動作する処理ブロックであり、その処理概要を図 4-24 に示す。

端末から入力された出力情報（出力形式の指示情報）をチェックして、エラーが検索されると、処理ブロック管理表にエラー内容を通知して、制御を主ルーチンに戻す。

正しく入力された出力情報は、出力する文献データを取得するために、出力様式の入力で指示された式番号について検索結果アドレス・テーブルの内容が照合され、検索結果データ・ファイルを読み込む。その結果、ポスティング情報が得られ、データベースから文献データを読み込むことができる。検索された文献データは、出力形態に応じた編集が行なわれたうえで出力／表示される。

漢字ディスプレイには、原則として画面に 1 文献分のデータを編集して表示する。

1 画面内に収まらないときは、最初の画面に許されるところまでを第 1 画面として編集し、“継続”コマンドが入力されたら、残りの部分を第 2 画面として編集する。

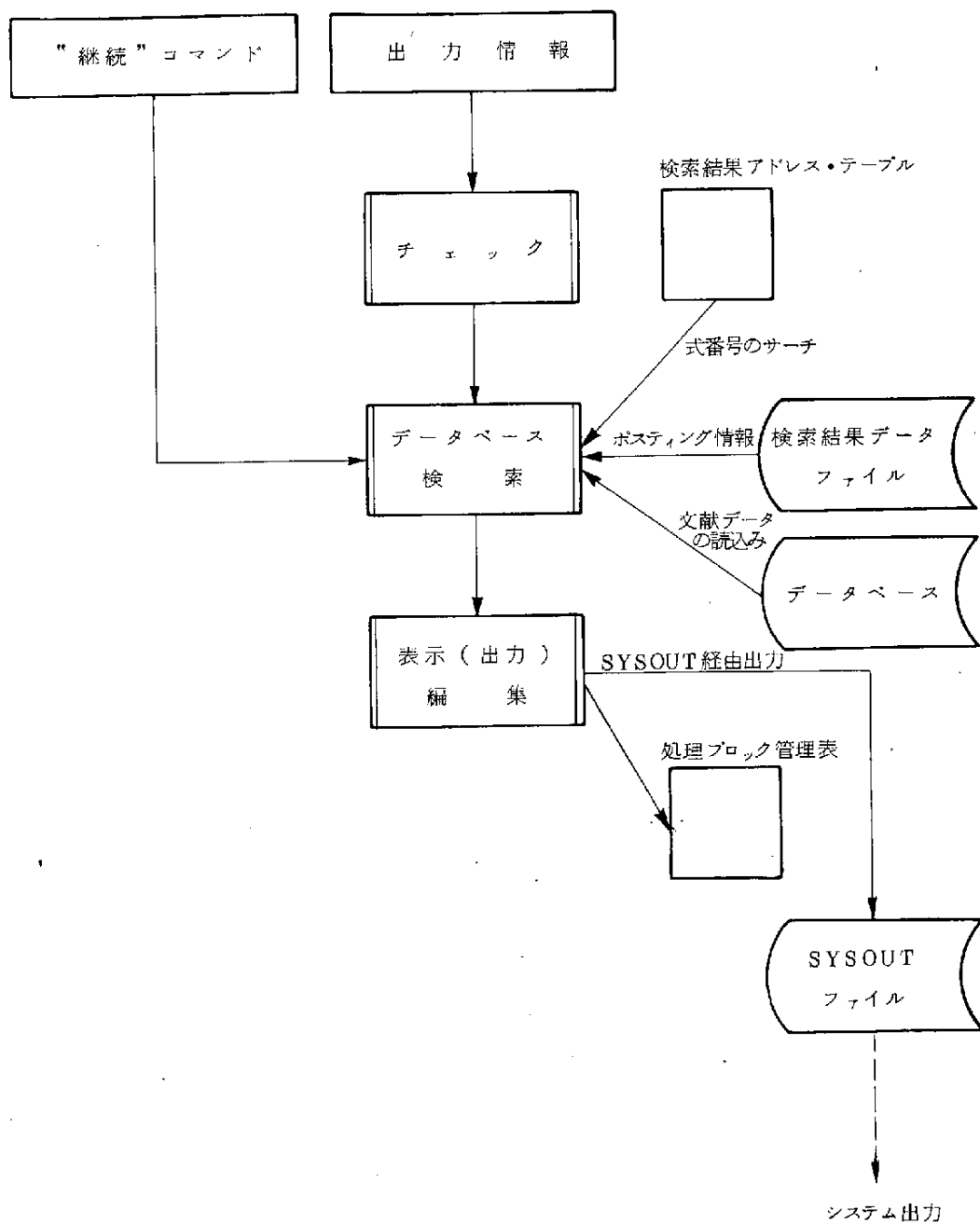


図 4 - 2 4 文献出力編集処理概要

また、“ハード・コピー”コマンドが入力されたならば、その時点で表示している文献データを帳票編集して漢字プリンタに出力する。

漢字ディスプレイに表示することなく、漢字プリンタへ直接出力する場合には、検索した文献データを帳票編集して出力する。

(3) 主な処理の説明

●データベースの検索

検索式や条件式によって適合した文献データの格納アドレスは、検索結果データ・ファイルのポスティング情報によって示されている。文献データの検索手順を図4-25に示す。

- ① 検索結果アドレス・テーブルから式番号をサーチして、ポスティング情報が格納されているアドレスを得る。
- ② 検索結果データ・ファイルからポスティング情報を検索する。
- ③ 文献データの格納アドレスを順に取出す。
- ④ データベースから文献データを読み込み、画面エリアあるいは帳票エリアに編集する。

●継続処理

“継続”コマンドが入力されたときの文献データの編集処理は図4-26に示すとおりである。

継続処理では、次のことを行なう。

- 文献データが継続する場合は、画面の最左端に継続文字を表示する。
- 検索した全文献データの表示が終わったら、表示ポイントを全文献データの先頭にセットしなおす。

●表示（出力）時変換項目

次にあげる項目は、データベースのレコード格納モードがBINARYあるいはEBODIC表示（出力）時に該当するテーブルを用いて、漢字モードに変換する。各テーブルは付録を参照。

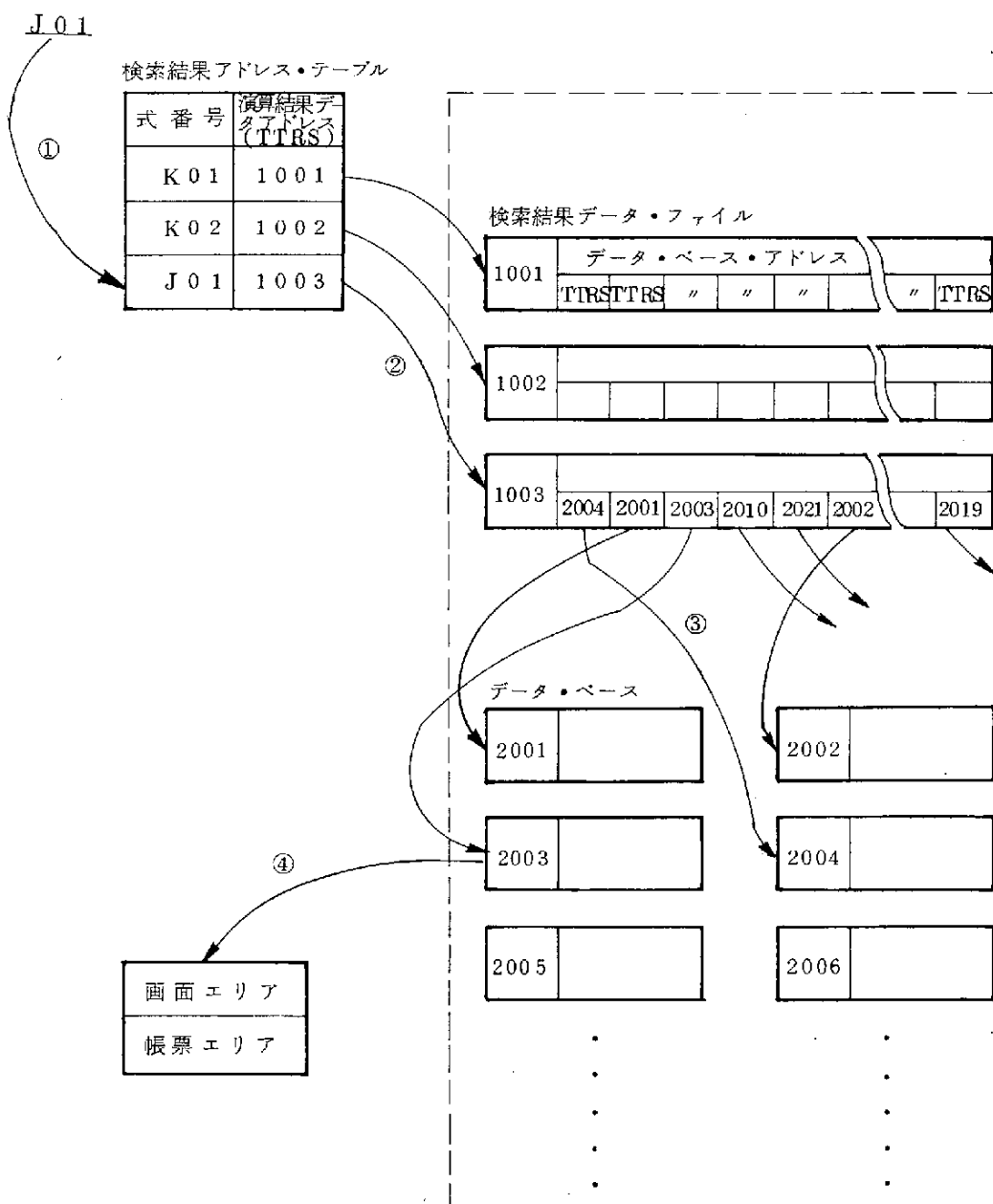


図 4 - 2 5 文献データの検索手順

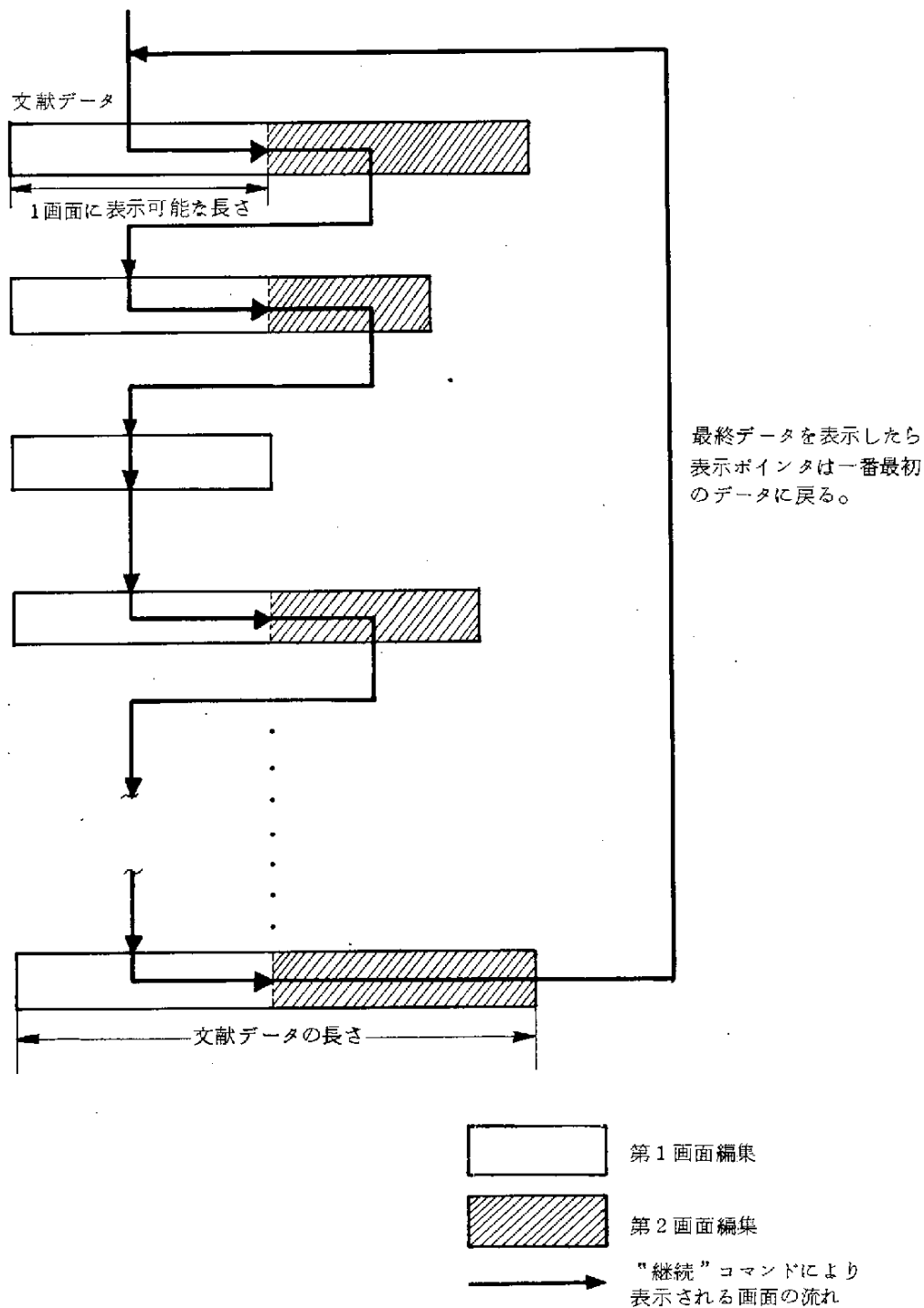


図 4 - 2 6 文献データの継続処理

4.2.4 ファイル概要

(1) 当サブシステムで扱うファイル

当サブシステムで扱うファイルは、表4-9のように、キーワード処理関係、検索出力処理関係およびメッセージ処理関係とに分けられる。

処 理	フ ァ イ ル 名
キーワード処理	ブロック・インデックス・テーブル
	キーワード・テーブル
	シソーラス下位リンク・サブテーブル
検索出力処理	ポスティング・テーブル
	データベース
	検索結果データ・ファイル
	SYSOUT ファイル
メッセージ処理	エラー・メッセージ・ファイル

表 4 - 9 ファイルの分類

これらのファイルはすべて磁気ディスク・バック上に割成される。

(2) ファイルの説明

- ブロック・インデックス・テーブル

キーワード・テーブルのブロック・インデックスが格納されている。詳細は「キーワード管理サブシステム」を参照。

- キーワード・テーブル

ヨミカナと漢字のキーワードが対応したテーブルであり、キーワードごとにポスティング・テーブルのアドレスが格納されている。詳細は「キーワード管理サブシステム」を参照。

- シソーラス下位リンク・サブテーブル

キーワード入力によるシソーラス・サービスで使用され、キーワード・テーブルにおける下位語のアドレスが格納されている。詳細は「キーワー

ド管理システム」を参照。

- ポスティング・テーブル

キーワード単位にデータベースにおける文献データのアドレスが格納されている。詳細は「データベース管理サブシステム」を参照。

- データベース

検索サービスする文献データ（情報処理関連文献）が格納されている。詳細は「データベース管理サブシステム」を参照。

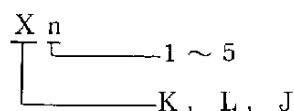
- 検索結果データ・ファイル

検索結果データ・ファイルは、検索式および条件式で検索された結果（ポスティング情報）を格納するファイルである。ファイルのレイアウトは次のようになっている。

式 番 号	件 数	ポインター 1						ポインター 2		ポインター n
		文献 アドレス TTRS	言語 コード	発行 年	発行 国 コード	記事 区分				
2	2	4	1	2	1	2	2			

- 式番号

検索結果の識別番号を次の形式で入れる。



- 件 数

検索の結果、該当文献が何件かという件数がBINARYで格納される。

- ポインタ（ポスティング情報）

適合した文献データのデータベースにおけるアドレスおよび条件項

目からなる。(ポスティング・テーブルから得られる)

- SYSOUT ファイル

検索した文献情報(標題・抄録文等)を検索サービス終了後、漢字プリンタに出力させるために、文献情報を一時的に格納しておくファイルである。このファイルのレイアウトは次のようになっており、各レコードは漢字プリンタの1行分に相当する。

チ ェ ッ ク ・ コ ー ド	文 献 情 報	空 白	C P L F
(2)	(64)	(4)	(2)

- チェック・コード

このコードは漢字プリンタ出力におけるデータの桁ずれ(ピッチ)を検出するためのものであり、桁ずれが発生しても、このコードによってデータは正しいピッチに補正される。

- 文献情報

データベースから得られた文献データであり、漢字プリンタに出力する際の1行分(32文字)に編集してある。32文字に満たない場合は空白が詰められる。

- 空 白

ワード・バウンダリのための空白である。

- CPLF

漢字プリンタの改行コードであり、ハード的に定められたステップ(行間隔)で紙送りを行なう。

- エラー・メッセージ・ファイル

検索サービス・サブシステムでは、端末から入力される各種の電文(デ

ータ、コード)をチェックし、誤りが検出されたら適切なメッセージを端末へ表示する。このメッセージを集めたのがこのファイルであり、レイアウトは次のようになっている。

メ ッ セ ー ジ 番 号	メ ッ セ ー ジ 本 文	空 白
(4)	(64)	(4)

- ・メッセージ番号

メッセージに付けられた固有の番号である。

- ・メッセージ本文

漢字まじりのエラー・メッセージ本文である。文の長さは漢字ディスプレイ画面の1行分(32文字)である。

- ・空白

ワード・バウンダリのための空白である。

4.2.5 検索対象項目

検索サービス・サブシステムで文献検索の対象とする項目は表4-10のとおりである。

表 4 - 1 0 検索対象項目

検索項目	収 納 フ ァ イ ル		表 示 オ プ シ ョ ン	条 件 項 目
	ポスティング・ テ ー ブ ル	デ ー タ ベ ー ス		
記 事 番 号		○		
言 語 コ ー ド	○	○		○
発 行 国 コ ー ド	○	○		○
発 行 年	○	○		○
記 事 区 分	○	○		○
著 者 名		○		
和 文 標 題		○		
原 文 標 題		○	○	
巻		○	○	
号		○	○	
資 料 名		○	○	
抄 録 文		○		
検索文献番号		○		

このうち、表示オプションは文献情報を表示・出力する際に要・不要を選択できる項目である。

また、条件項目は文献を検索するとき、条件として指定できる項目である（条件式で指示）。

4.2.6. 検索コマンドとパラメータ

(1) 検索コマンド

システムに対して何らかのサービスを受けようとするとき、利用者の意思を伝える手段が必要である。当サブシステムではこの手段のことを検索コマンドという。

一般にコマンドは次のような形で入力されることが多い。

コマン ド	空白	パラメータ
-------	----	-------

しかしながら、当サブシステムの検索コマンドは、文字や記号でコマンドを入力する代りにキーボード・アダプタに付いているファンクション・キーと検索コマンドを対応付け、そのキーを押すことにより同じ意味をもつようにした。その理由としては、次のことがあげられる。

- ・漢字キーボードでは、1文字入力するのに2ストローク必要であり、わずらわしい。
- ・漢字キーボードに配列されている文字数が多い(5,376字)ために、文字を拾うことが利用者にとって負担となる。

(2) パラメータ

当サブシステムでは、検索情報を主としてパラメータ形式で入力する。

パラメータの入力は、できるだけ利用者に負担をかけないように簡易な方法で入力できることが望まれる。したがって漢字キーボードの操作ではわずらわしいので1ストロークで入力できるテン・キーを使って入力する方法を採用する。しかしながら、キーワード(カタカナ)や論理式の演算子等の入力については、テン・キーだけでは不十分であり、必要最少限にとどめてキーボードからの入力を許すことにした。

(3) 検索コマンドとファンクション・キーの対応

検索コマンドの配置は、会話の流れに合った合理的なものが好しい。その要素をあげると、一般に

- ・利用者にとって憶えやすいこと。
- ・使用頻度の多いものは操作しやすい所に配置する。
- ・相反する検索コマンドは離す。

などが考えられる。このような点を考慮して当サブシステムの検索コマンドはキーボード・アダプタのプログラム・ファンクション・キー上に図4-27のように配置した。

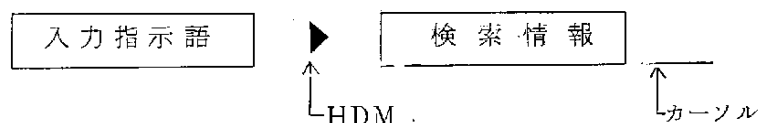
会 話 開 始			
キーワード 入 力	保 存 語 指 定	検 索 式	条 件 式
出 力	ヘー ド コ ピー		
継 続			会 話 終 了

図 4 - 2 7 検索コマンドの配置

4.2.7. 検索情報の入力形式

(1) 入力形式

検索情報（パラメータ）の入力はすべて次の形式で行なわれる。



入力指示語は検索情報の種類を表わす語であり，検索コマンドに対応する▶印はHDM（ヘッディング・マーク）で，このあとに検索情報を入力する。入力指示語および▶は，検索コマンドを押したとき，サブシステムから表示される。

(2) セパレータ

セパレータは，検索情報として入力するパラメータが複数あるとき，パラメータとパラメータの間に入れる記号である。

セパレータは入力操作，および処理のしやすさという面から1つの記号に統一した方が好しく，当サブシステムではキーボード・アダプタにあるテン・キーの[.]キーを用いる。

例

K 0 1 . 0 1 * 0 2
 ↑
 セパレータ

(3) 検索情報の種類

● 検索キーワード ▶ [キーワード [. 選択子]]

これは文献検索を行なうために検索キーとなる語（キーワード）を指定する。キーワードは、カタカナによる入力で最大入力文字数は40字である。1回の操作で入力できるキーワード数は1語である。選択子はキーワード関連情報のサービスを受けようとするときの方法を指定する。

0：前方一致

1：シソーラス

● 保存語番号 ▶ [指示子 [. 語番号]]

この検索情報はキーワードの保存指定を行なうものである。

指示子はキーワードの扱いを指示する。

これには、D、S、Aの3種がある。

D：DELETE（消去）で、指定された語番号以外のものを保存する。

S：SAVE（保存）で、指定された語番号のものを保存する。

A：ALL（全部）で、全ての語番号のものをテーブルの制限数内で保存する。制限数は30個である。

● 検索式 ▶ [検索式番号 . 論理式]

この検索情報は保存した語をもとに論理式を作成するものであり、これによって論理演算（検索）が行なわれる。

検索式番号は論理式ごとに付けられる式番号である。検索式番号はK01～K05、L01～L05のいずれかである。

論理式の演算子にはAND：*、OR：+、NOT：-、を用い、項目は式番号によって次のように識別する。

式 番 号	項 目
K	保存語番号
1.	K01～K05, 保存語番号

●条件式 ▶〔条件式番号・論理式〔・条件項目〕〕

この検索情報は検索式によって適合した文献データに対し、さらに論理式および条件項目で範囲を限定するためのものである。

条件式番号は論理式ごとに付けられる式番号であり、J01～J05のいずれかである。

論理式の項目は検索式番号である。また条件項目は次の4種の特定項目について指定する。

発行年：数字で2桁（西暦年下2桁）

発行国：数字で2桁（発行国コード表）

記述言語：数字で2桁（言語コード表）

記事区分：英数字で2桁（記事区分分類表）

●出力様式 ▶〔出力形態・式番号〔・NNNN〔・N'N'N'N'〔・選択子〕〕〕〕

この検索情報は検索した文献データを漢字ディスプレイあるいは漢字プリンタに表示出力するときの形式を指示するためのものである。

出力形態は出力する装置を指定する。

0：漢字ディスプレイ装置

1：漢字プリンタ装置

式番号は検索式番号あるいは条件式番号である。

NNNNは最大4桁の数字で、何件目のデータから出力させるかを指定し、N'N'N'N'は最大4桁の数字で何件出力するのかを指定する。

選択子は一部の項目について出力の要・不要を指定する。

0：原文標題，資料名，巻，号は表示・出力しない。

1：原文標題，資料名，卷，号を表示・出力する。

●出力方法 ▶ 出力方法

この検索情報は漢字プリンタ出力の方法を指定する。

0：SYSOUTによる出力

1：直接出力

●その他の情報

次の4つの検索コマンドについては，特に入力する情報としては何もない。

- ・ 会話開始
- ・ ハード・コピー
- ・ 継続
- ・ 会話終了

それぞれの機能については表4-1を参照。

4.2.8 表示・出力のレイアウト

(1) 漢字ディスプレイ画面の表示様式

漢字ディスプレイ画面のレイアウトは，図4-28に示すように表示画面と操作画面とに分けられる。

表示画面は検索過程における検索結果およびエラー・メッセージ等の表示を行なう部分であり，操作画面は検索手順の指示と検索情報の入力表示が行なわれる部分である。表示画面および操作画面の検索手順部分は画面保護のためプロテクト・フィールドにする。プロテクトがかからない部分はオペレーション・フィールドと称し，検索情報の入力操作が行なわれて送信の対象となるフィールドである。

画面に表示する場合，各行の始まりは通常第2文字位置とする。ただし次の場合は利用者の注意を促す意味で第1文字位置から表示する。

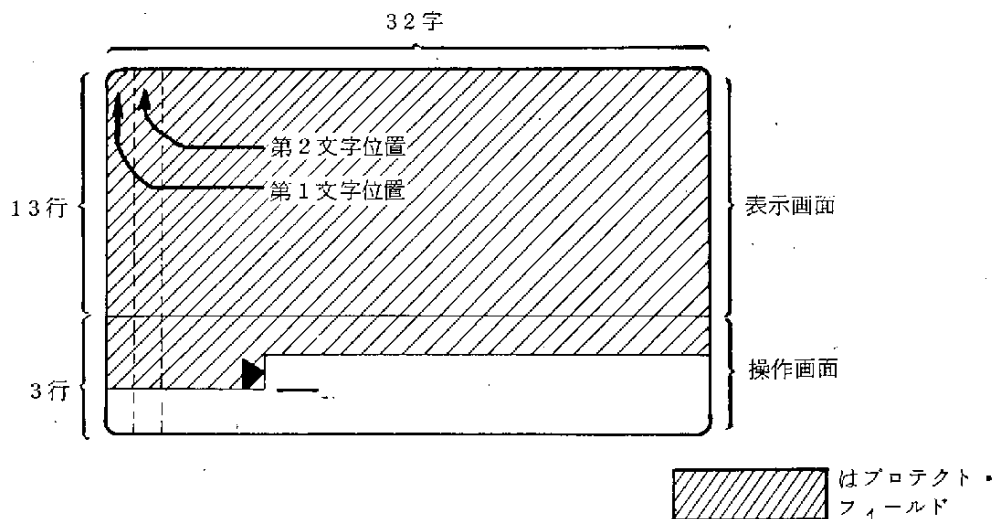


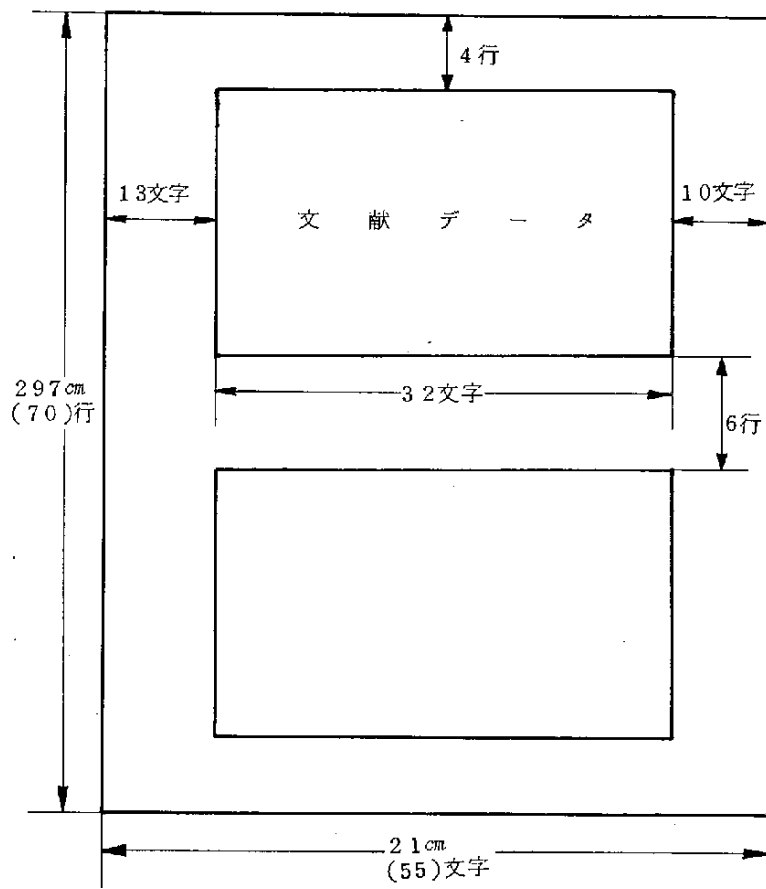
図 4 - 2 8 画面レイアウト

- ・継続（“ 継 ”を表示）
- ・エラーの場合（“ 誤 ”を表示）
- ・操作画面の検索手順（“ ▷ ”を表示）

また、例外として文献情報の記事番号は第 1 文字位置から表示する。

(2) 漢字プリンタの出力様式

漢字プリンタの出力様式を図 4 - 2 9 に示す。



サイズ : A 4

行 幅 : 6 行 / インチ

図 4 - 2 9 出力レイアウト

4.2.9 エラー・メッセージ

(1) エラー・メッセージの形式

当サブシステムでは，端末の漢字ディスプレイ画面に表示するエラー・メッセージの設計ポイントを次の2点に置いた。

- ・漢字まじりデータなので，長いメッセージは必要ない。
- ・会話型処理なので，メッセージは検索情報として入力されるデータに限って応答を出す。

以上のことからエラー・メッセージは次のような内容にした。

項 目	内 容
メッセージ長	画面1行分(32文字)
メッセージ数	20種
本 文	漢字まじり文

また，エラー・メッセージの形式は次のとおりである。

エ ラ ー 記 号	空 白	メ ッ セ ー ジ 番 号	空 白	メ ッ セ ー ジ	本 文
1	1	4	1	25	

—文字数

- ・エラー記号

このメッセージがエラー・メッセージであることを表わす1文字の記号である。

- ・メッセージ番号

個々のエラー・メッセージに付けられた4桁の番号である。

- ・メッセージ本文

漢字まじりの文章で，エラーの内容を表わしたものである。

(2) 表示方法

エラー・メッセージの表示方法としては、一般に、

- ① エラーは無条件に出力する。
- ② エラーの出力は利用者にまかせる。
- ③ エラー・メッセージを簡易化して出力する。

が考えられるが、当サブシステムでは①の方法を採用した。したがって、端末からの入力データに対してのエラーは常に表示できる部分とし、画面の一部を表示欄として確保する。この位置は表示画面の最下行（13行目）とする。

4.2.10 EBCDICコードの扱い

(1) EBCDICコードの入力方法

当システムのハードウェア構成のうち、漢字ディスプレイ装置のキーボードの一部、およびアダプタのテン・キーの文字についてはEBCDICコードでも、入力することが可能である。

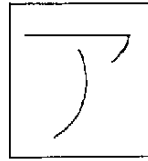
EBCDICコードの入力方法と文字種は表4-11のとおりである。

表4-11 EBCDICコード文字と入力法

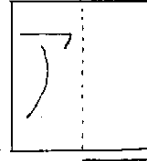
入 力 方 法	文 字
キーボード・アダプタ のテン・キーを押す。	0 ～ 9 ・
画面制御キーの[A]キー を押し、キーボードの 文字を入れる。	0 ～ 9 A ～ Z ア ～ ン 特殊記号

このいずれかの方法で入力したEBCDICコードの文字は、漢字コードで入力した場合とは画面の表示が次のように異なる。

漢字コード



EBCDICコード(半ビッチ)



カーソル

(2) コード変換

当サブシステムでは、利用者自らの操作による入力データは内部処理のしやすさ、および主記憶容量等から、EBCDICコード(半ビッチ)で行なうが、端末の漢字ディスプレイおよび漢字プリンタへの表示出力の際はすべて漢字コードで行なう。したがって、EBCDICコードから漢字コードへの変換が必要である。

4.3 データベース管理サブシステム

4.3.1 処理概要

データベース管理サブシステムは、KIND システムにおける文献データの情報管理を総合的に行なうサブシステムであり、データベースに蓄積する文献情報の収集・加工およびデータベースの作成・管理を合理的に行なう。

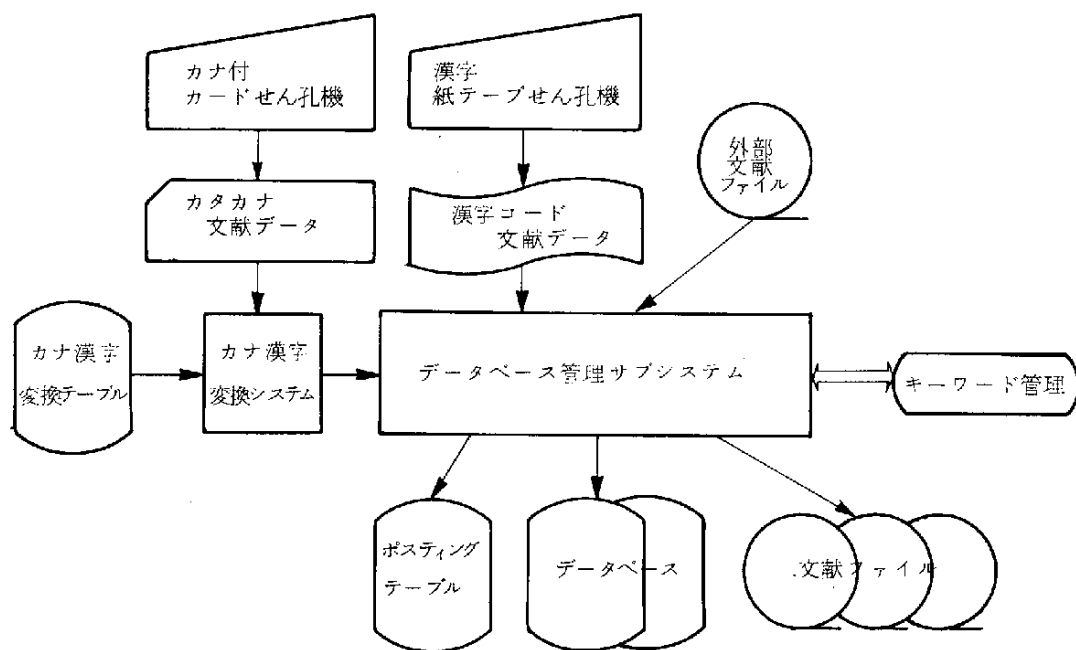


図 4-30 データベース管理サブシステムの構成図

KIND システムで扱う文献情報の適用分野は科学技術文献のうち情報処理に関連する技術文献である。よってデータベースに蓄積する文献情報の収集にあたっては、その手段として、外部の情報提供サービス機関で収集し作成した漢字モードの文献ファイル入手して使う場合と、当財団で収集した文献情報からデータを作成して使う場合と、2とおりの方法が考えられる。

漢字モードのデータを作成するには、一般に漢字入力装置を使ってデータせん孔し、モニタ・リストをとって校正を行ない、訂正データのせん孔、訂

正処理という工程を経なければならない。このデータ作成工程は、アルファ・ニューメリック・データ作成に比較して、作成コストも高く、また作成時間を多く要するなどデメリットな面が多い。このようなことから、データの収集と作成を効果的に行なうために、外部からの文献ファイルの導入も考慮して、データベースの管理を行なうことにする。

当サブシステムでは、上記の2とおりの方法いずれからでもデータ作成が行なえるようにした。

当財団で収集した文献情報については、データの量や内容などによってデータ作成方法を変えることができる。第1の方法は、漢字モード・データ作成手順に従い、第2の方法は、カナ・モード・データ作成手順に従って、最終的に漢字モードの文献ファイルを作成する。なお、カナ・モード・データ作成から漢字モード文献ファイルへ変換処理する過程に当財団で開発したカナ漢字変換システムを利用する。

外部からの文献ファイル導入については、KINDシステムの文献検索サービスに必要な項目を編集・加工したうえで利用しなければならないが、一般に、外部機関の漢字コード体系がKINDシステムの漢字コード体系（富士通標準漢字コード）と異なることが多いために漢字コード変換を必要とする。漢字コード体系は、現状では標準文字数／標準コードが設定されていないために、ハードウェア・メーカーあるいはユーザのアプリケーションによって異なることが多い。そのため一義的に同一コードが同一漢字と対応するとは限らない。よって外部機関で作成し提供される文献ファイルについては、漢字モードを当システムの文字種および文字コードにすべて変換して使用する。

いっぽう、当サブシステムにおけるデータベースは、集団磁気ディスク装置に持ち、記憶領域の管理を行なって、KINDシステムを効果的に運用する。データベースは、記憶容量の制限等から随時メンテナンスを行ない、古くなった情報の削除と新しい情報の追加を一括して処理する。

また検索方法にキーワード検索を適用するため、インバーテッド・ファイルの一部であるポスティング・テーブル（文献データの格納アドレスと検索条件に使用する項目から成る）の管理も当サブシステムで行ない、キーワード管理サブシステムと連系をとる。

なお、データベースおよびポスティング・テーブルが確保する最大記憶容量は、100メガ・バイトの磁気ディスク・パック2台とし、合計200メガ・バイトについて当サブシステムが管理する。この最大記憶容量は、文献データ約8万件分に相当する。

4.3.2 処理フロー

(1) 文献ファイル作成

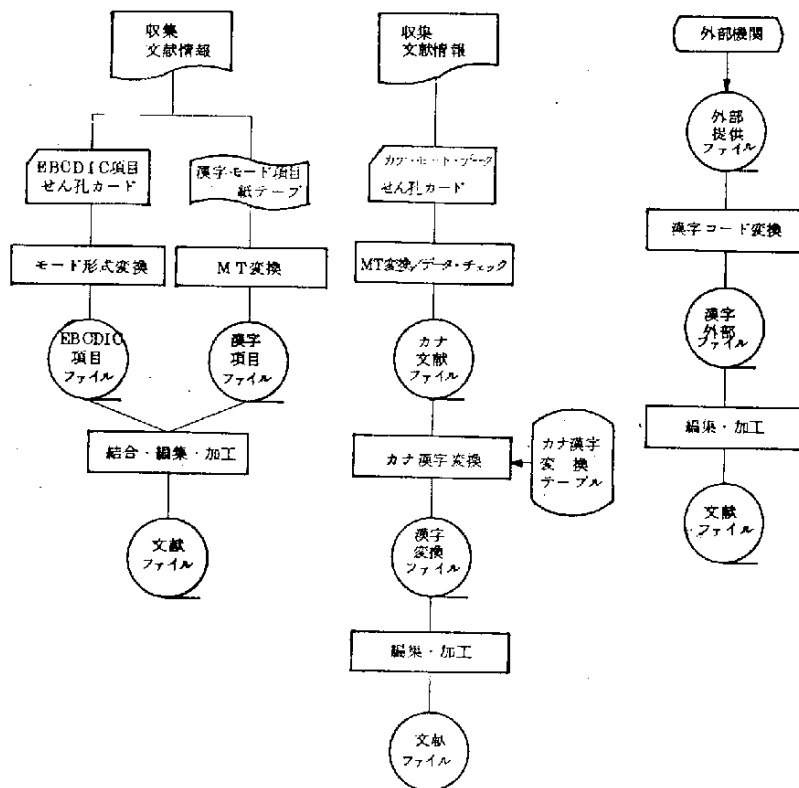


図4-31 文献ファイル作成フロー

(2) データベース格納

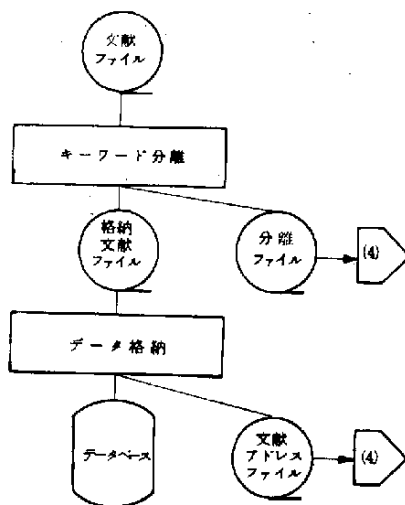


図4-32 データベース格納フロー

(3) データベース・メンテナンス

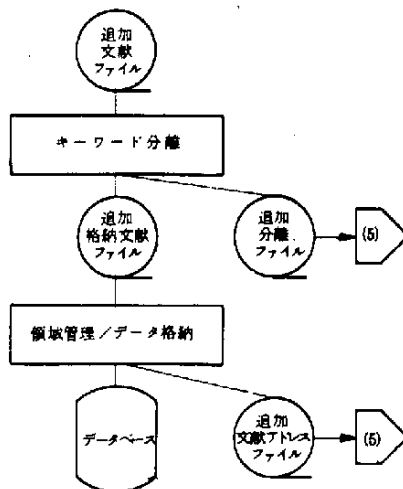


図4-33 データベース・メンテナンス・フロー

(4) ポスティング・テーブル作成

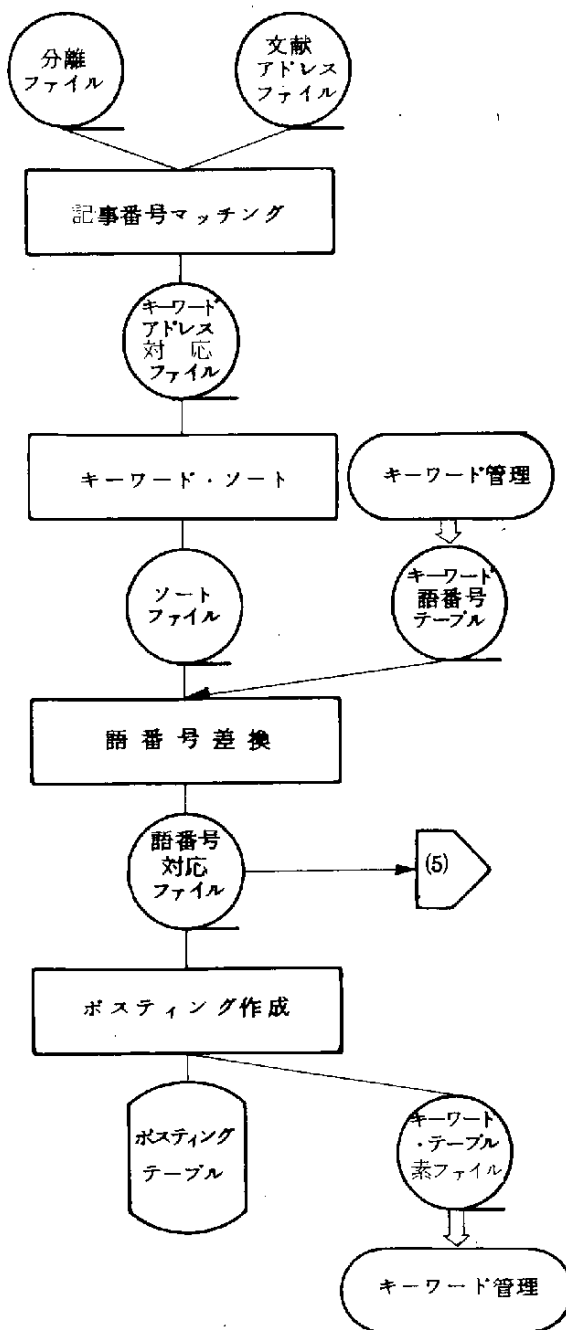


図4-34 ポスティング・テーブル作成フロー

(5) ポスティング・テーブル・メンテナンス

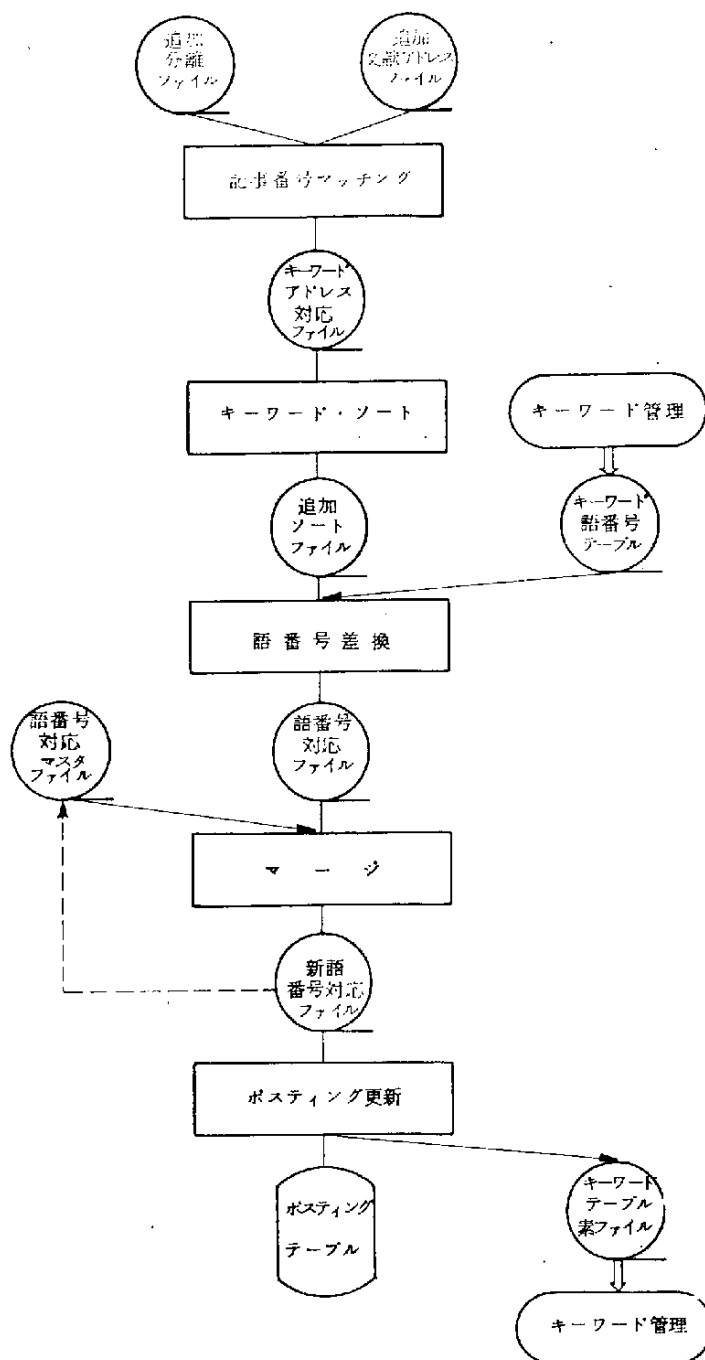


図 4-35 ポスティング・テーブル・メンテナンス・フロー

4.3.3 データ作成

(1) 文献データ作成方法

当財団では、国内・国外を問わず各種の情報処理技術文献を収集し、その技術動向等を調査する部門があることから、そこで収集した文献情報について新規にデータを作成する。

データの作成方法については、最終的には漢字モードの文献ファイルの創成であり、その処理過程として、漢字入力装置を使った漢字モード・データの作成方法と、カナ付きカードせん孔機を使ったカナ・モード・データの作成方法との2とおりの方法いずれかによってデータを作成する。

漢字モード・データ作成には、当財団設置の漢字入力装置(F6801A)を使って(あるいは外注)漢字モード・データのせん孔を行ない、モニタープリント(校正ゲラ・リスト)は漢字プリンタ(F6521A)に出力される。またデータ項目中、EBCDIC/BINARYモード項目部分についてはカード・データを作成し、漢字モード項目と結合して文献ファイルを完成する。

カナ・モード・データ作成では、漢字モード項目・EBCDIC/BINARYモード項目ともカード・データを作成する。このうち漢字モード項目については、当財団で開発したカナ漢字変換システムのデータ作成基準(付録参照)に従ってデータをせん孔し、内部でEBCDICカナ・モードから漢字モードへ変換処理をして漢字モードの文献ファイルを作成する。

以下に文献データを構成する各項目についてその内容を概述する。なお、各項目の()内は文献ファイルにおける最終モードを漢字・EBCDIC・BINARYで識別し、数字は、各項目の最大文字数を現わしている。ただし、漢字モードは文字数(1文字=2バイト)、EBCDIC/BINARYモードは桁数(1桁=1バイト)で表現している。

記事番号(EBCDIC, 9)

各文献ごとに付けられる唯一の番号。

分野区分	作成年	作成月	一連番号
1	2	2	4

分野区分：情報処理文献は，管理システム技術に関するものはB，電気に関するものはEとする。

作成年：データの作成年を西暦年下2ケタで表わす。

作成月：データの作成月（01～12）

一連番号：作成月単位の一連番号（0001～9999）

資料略名（漢字，150）

各文献を掲載している資料の略名。資料が「逐次刊行物所蔵目録」（JICST刊）にあるものについては，資料番号付きの略名。

(資料番号)	スペース	略名
5 or 7	1	150

資料番号：5桁ないし7桁のコードから成る。詳細は上記所蔵目録を参照のこと。

巻（EBCDIC，30）

雑誌・会議資料（定期刊行）の巻番号，あるいは会議資料（不定期刊行）の会議回次。

号（EBCDIC，30）

雑誌・会議資料（定期刊行）の号番号

発行国コード（BINARY，1）

資料が発行された国の国名コード。コード表は付録参照。

発行年 (BINARY, 2)

資料が発行された西暦年の下2ケタ。発行が2年以上にまたがる合併号等の場合、前年を第1バイト目に、後年を第2バイト目に持つ。他の場合は第2バイト目は0とし、発行年不明の場合は両バイトとも0とする。

記事区分 (EBCDIC, 2)

記事の形式を区別するコード。コード表は付録参照。

言語コード (BINARY, 1)

それぞれの文献を記述している言語のコード。コード表は付録参照。

著者名 (漢字, 120)

文献の著者名、あるいは特許の発明者。個人著者名、複数著者のうち第1著者名、団体名、のいずれかの名称。

著者名フリガナ (EBCDIC, 30)

日本語の個人著者名、あるいは複数著者のうち第1著者名のフリガナ。

所属機関名 (漢字, 120)

文献の著者あるいは特許の発明者が所属する機関名。

原文標題 (漢字, 300)

外国語で書かれた文献の標題。ただし、英字で表現できない文献は省略。

和文標題 (漢字, 150)

日本語で書かれた文献の標題、または外国語で書かれた文献の和訳標題。

[一般形式]

主標題 主標題の報告番号 副標題 副標題の報告番号

キーワード (漢字, 40×15個=600)

文献から抽出された主要概念を表わす用語。キーワード管理サブシステムで管理するキーワードから選定。

抄録文（漢字，500）

文献の内容について要点をぬき出して短かくまとめた文章。

表4-12 データ作成モード対応表

モード 処理内容	漢字モード・データ作成		カナ・モード・データ作成
データ・シート 記入	漢字モード項目用データ・シート	EBCDIC/BINARY 項目用データ・シート	カナ・モード・データ・シート
データせん孔	漢字入力装置 （紙テープせん孔）	カナ付カードせん孔機 （紙カードせん孔）	カナ付カードせん孔機 （紙カードせん孔）
データ・チェック	漢字プリンタ （モニタ・プリント）	ラインプリンタ （チェック・リスト）	ラインプリンタ （チェック・リスト）
データ訂正	グラ校正 （訂正紙テープせん孔）	訂正カードせん孔	訂正カードせん孔
コード変換	漢字→EBCDIC変換	EBCDIC→BINARY 変換	カナ漢字変換 （EBCDIC→漢字） （EBCDIC→BINARY）
編集・加工	漢字・EBCDIC項目結合→文献ファイル形式編集		文献ファイル形式編集

(2) 外部ファイルの利用方法

外部の情報提供サービス機関で作成された漢字モードの文献ファイル（以降外部ファイルと呼ぶ）を利用するにあたっては、KINDシステムで扱うコード体系との相異，文字種の違い等の問題があり，単純に適用することはできない。そこで，外部ファイルの入手以前に，これらの問題を解決すべく，コード変換表の作成，文字の対応表作成等を行なっておく必要がある。

KINDシステムの漢字コード体系は，当財団設置の漢字入出力装置が富士通製であることから富士通標準漢字コードで構成される。このコード体系における文字種は次のように割当てられている。

漢 字……………4,290種

ひらがな……………240種

カタカナ…………… 2 4 0 種

英 字…………… 2 6 8 種 (英記号も含む)

数 字…………… 1 4 8 種 (数記号も含む, 漢数字も一部含む)

その他記号…………… 1 9 0 種

計 5, 3 7 6 種

(これらの文字およびコードについては付録を参照。)

外部機関で使用している文字種と, 当システムの文字種の違いは,

- ① 外部機関の文字にあって, 当システムの文字にないものは, 読み仮名等でおきかえる。
- ② 外部機関の文字になく仮名になっているものは, 当システムでもその仮名を使用する。

という方法で処理する。

その他, KIND システムに外部ファイルを適用するには, 主に次のような処理を行なう必要がある。

- ① 外部機関の漢字コードから, 当システムの漢字コードへ変換する。
- ② 当システムに必要なデータ項目を編集する。
- ③ 編集した項目について一部加工し, モード変換を行なう。

漢字コードから EBCDIC コードへ

漢字コードから BINARY コードへ

EBCDIC コードから BINARY コードへ

外部ファイルの適用例として, JICST 文献検索磁気テープを採用した場合の処理例を以下に述べる。

JICST で 扱っている漢字種は約 2,000 字であることから, 当システムの 4,290 字に大部分は含まれている。(当システムにない文字は「纒」と「苦」の 2 字で, それぞれの読み仮名 “ラン” と “トマ” で代用する) しかしながら, JICST で処理している文献が科学技術文献という特徴から, 数学記号・化学記号等の特殊な記号類と, 外国文献の範囲が

世界各国に及ぶために、ギリシャ文字／ロシア文字／ドイツ文字といった特殊な文字類が出現する可能性がある。これらの記号・文字類は当システムにはないために、次のような方法をとる。

- ① 記号類については、他の記号で代用する。
- ② 文字類については、これらの文字類の文献は希少であることから、出現すると思われるデータ項目の原文標題，資料略名，抄録文の各項目を省略することがある。

表4-13は、JICSTテープから当システムの文献ファイルへ編集する場合の対応表である。

表4-13 外部ファイルからの編集例

項 目	JICSTテープ			KINDシステム文献ファイル			備 考
	モード	最大文字数	個数	モード	最大文字数	個数	
記 事 番 号	R	9	1	E	9	1	○ (特○印) 必ずある項目、△印 無いこともある項目、※印 不要となる項目)
資 料 種 類 コー ド	E	1	1				○
I S S N	E	9	1				○
C O D E N	E	6	1				○
資 料 番 号	漢	7	1	漢	150	1	○
資 料 略 名	漢	150	1				○
資 料 名 フリガナ	E	150	1				○
巻 号	漢	30	1	E	30	1	○
発 行 国 コー ド	漢	30	1	E	30	1	○
発 行 年 号	漢	2	2	B	1	1	○
抄 録 番 号	漢	7	1	B	2	1	○
レ ー ー ジ	漢	30	1				○
ベ ー ジ	漢	60	1				○
記 事 区 分	漢	2	1	E	2	1	○
語 コー ド	漢	2	3	B	1	1	○
写 真 図	漢	2	1				○
表 数	漢	2	1				○
参 考 文 献	漢	3	1				○
著 者 名	漢	30	9	漢	30	1	○
通 体 著 者 名	漢	120	9	漢	120	1	○
著 者 名 フリガナ	E	30	9	E	30	1	○
通 体 著 者 名 フリガナ	E	12	9				○
所 属 機 関 名	漢	120	9	漢	120	1	○
所 属 機 関 名 フリガナ	E	120	9				○
原 文 標 題	漢	300	1	漢	300	1	○
和 文 標 題	漢	150	1	漢	150	1	○
カナ・タイトル	E	300	1				○
J I C S T 統一分類コード	漢	8	9				○
U D C	漢	30	9				○
キーワード	漢	40	15	漢	40	15	○
キーワード・フリガナ	E	40	15				○
自動追加キーワード	漢	40	120				○
自動追加キーワード・フリガナ	E	40	120				○
抄 録 文	漢	500	1	漢	500	1	○

4.3.4 データ格納

漢字モードの文献ファイルが作成されたあと、各文献データをデータベースに格納する処理を行なう。データベースは集団磁気ディスク装置（F477K）の2台のディスク・パックに創成され、その規模は200メガ・バイトの記憶容量におよそ8万件の文献が格納できる。

データベースのレコード構成は、文献ファイルのレコード項目から、キーワード項目を取除いたものから成る。レコード形態は文献ファイルと同様に可変長形式をとり、各項目の他、レコード長識別区分（4バイト＝1ワード）とディレクトリ区分で構成される。可変長形式をとる理由は、磁気ディスク・メモリを節約するためであり、文献データの特徴から、各項目が長短まちまちであること、しかも最大文字数の大きい項目が可変項目であるなど、可変長形式の効果が十分得られる。また検索サービス・サブシステムにおいて、会話応答の時間をスピード・アップさせる必要があることから、1ブロック＝1レコードの形態をとり、各レコードごとに磁気ディスク上の格納アドレス（TTRS）が得られるように格納する。データベース内のファイル編成は、順編成で格納処理を行ない、同時に格納アドレスを確保して、検索時にはダイレクトにアクセスする方法がとられる。

磁気ディスク・メモリの使い方としては、2台のディスク・パックで計200メガ・バイトがデータベースの規模となり、平均2,500バイトのレコードで約8万件の文献が格納できる算定になる。この8万件の文献という数字は、1ヶ月おおよそ1,500～2,000件の文献を格納するという想定から、ほぼ4年分の文献になる。よってデータベースを相対的に4分割し、かつ12ヶ月分の領域管理ができるようにして、データベースの作成・更新を行なう。

データ格納の処理前に行なうキーワード分離は、各文献レコードに唯一に付られた記事番号とともにキーワードをすべて取りはずし分離ファイルを作成する。そして、データ格納時には、記事番号と格納アドレスおよびポスティング項目（検索条件に使用するデータ項目）の言語コード／発行国コード

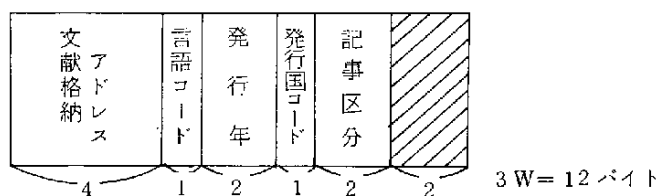
／発行年／記事区分とを組合せて文献アドレス・ファイルを作成する。この2つのファイルは次のポスティング・テーブル作成処理で使用される。

4.3.5 ポスティング・テーブルの作成

ポスティング・テーブルは、インバーテッド・ファイル形態をとるキーワード・テーブルとデータベースとの中間ファイルであり、データベースに蓄積した文献データとのリンク情報を保持する目的で作られる。

ポスティング・テーブルの各レコードは、それぞれに対応するキーワードごとに文献格納アドレスを集合したもので構成される。リンク情報とは、この文献格納アドレスのことであり、アドレスをたどって文献データをダイレクトに検索することが可能になる。

レコードの構成要素は、ポインタ項目として文献格納アドレスの他に、検索条件の要素をこのポスティング・テーブルのレベルで判別できるように、言語コード／発行国コード／発行年／記事区分を保有する。1レコードには、



このポインタ項目を最大1,085個もつことを可能にし、レコード・サイズは最大長13,024バイトである。

ポスティング・テーブルの記憶容量を算出すると、文献データ数8万件に対し、それぞれの文献にキーワードが平均5個付られているとして、約40万件のポインタ項目が必要となる。単純にポインタ項目12バイトを乗算すると4.8メガ・バイトになり、ブロック間隔(1RG)などを含めて、約5.4メガ・バイトが最大容量となる。(1レコード=1ブロック当りの平均ポインタ数を100として計算)

ポスティング・テーブルの作成処理は、データ格納処理で作成した分離フ

ファイル(記事番号/キーワード)と文献アドレス・ファイル(記事番号, ポインタ項目)とを使って, 記事番号でマッチングをとり, キーワード・アドレス対応ファイルを作成し準備する。さらにこの対応ファイルをキーワードをキーにしてソートしておく。ここで, キーワードの長さが可変長で大きい(最大40文字=80バイト)ことから, キーワードを語番号(キーワード個々に付られた番号でバイナリ4バイト)に置換える処理をする。そしてこの語番号ごとにポインタ項目を集計しながらポスティング・テーブルを創成していく。また同時に, キーワード・テーブルに必要なポスティング情報として, 語番号/記事番号/ポスティング・スタート・アドレス/ポインタ件数/ポスティング・レコード数を確保して, キーワード・テーブル素ファイルを作成する。この素ファイルは, キーワード管理サブシステムへ送られて, キーワード・テーブルのポスティング・アドレス格納処理で使用される。

図4-3-6は, ポスティング・テーブルのリンク情報の関連性を例示した図である。

4.3.6 データベース・メンテナンス

データベースのメンテナンスは, 当システムでは文献データの追加と削除の処理に限って行なう。これは, 文献ファイルの作成工程において十分な訂正処理が行なわれることから, 文献データの個々についての修正・更新は無いものと判断したためである。

文献情報の収集は月ごとにまとめられてデータ作成が行なわれるため, 原則として月1回のペースで追加の処理を行なう。また追加処理が重ねられると, データベースの記憶領域がオーバーフローを生じるので, 古くなった情報から削除を行ない, 常に新しい情報から約8万件をデータベースに蓄積しておくようにする。

データベースの記憶領域の管理方法については, 月単位の文献格納領域の管理を中心に行なう。これは前述のごとく, 文献ファイルが月ごとに発生す

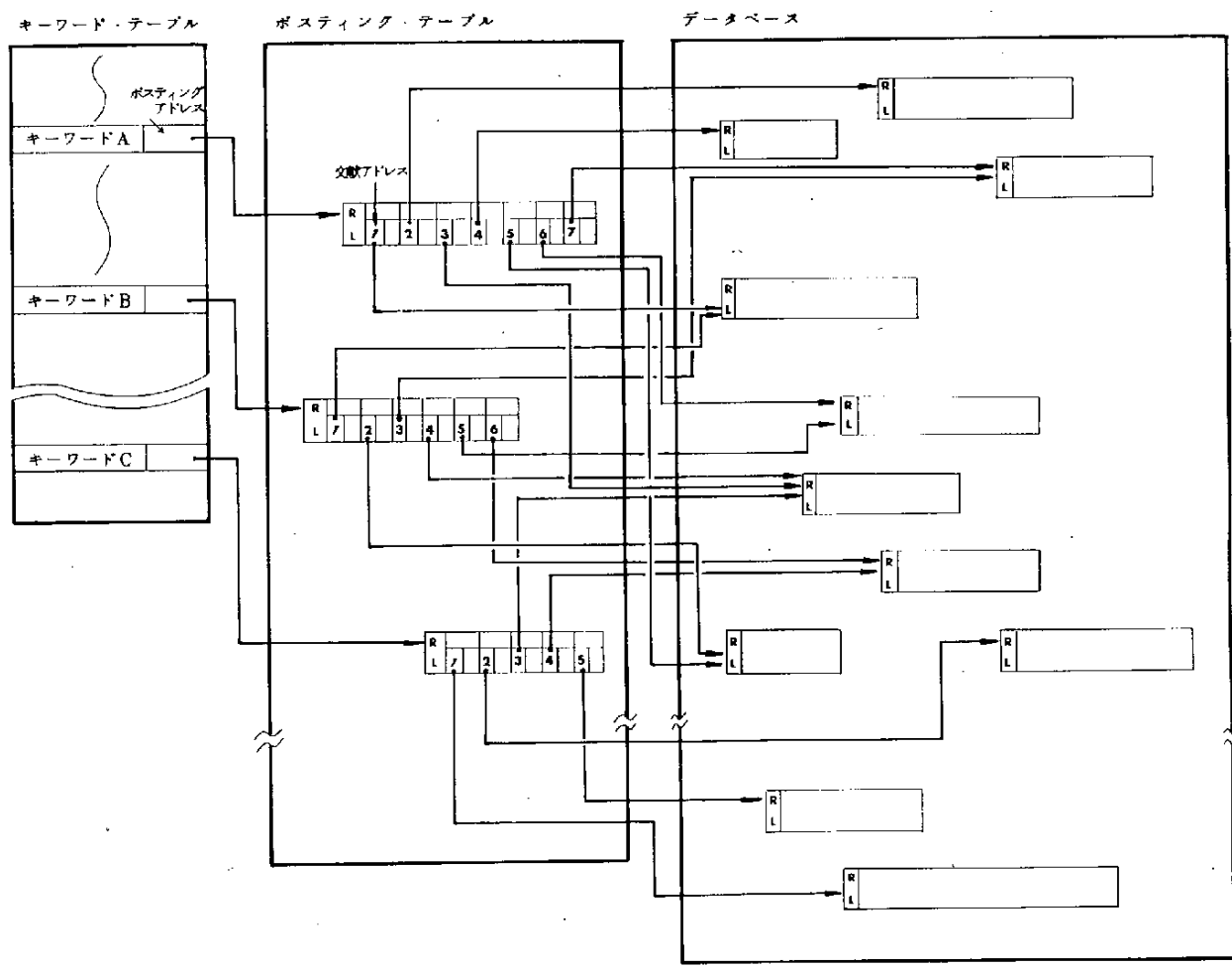
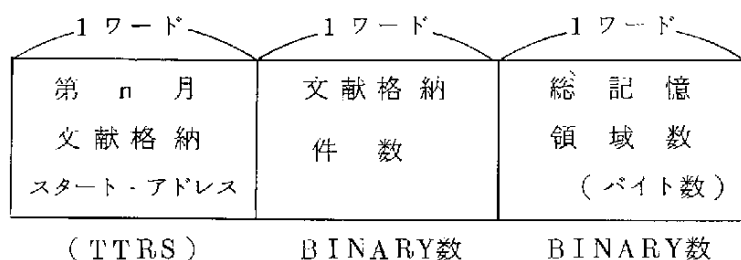


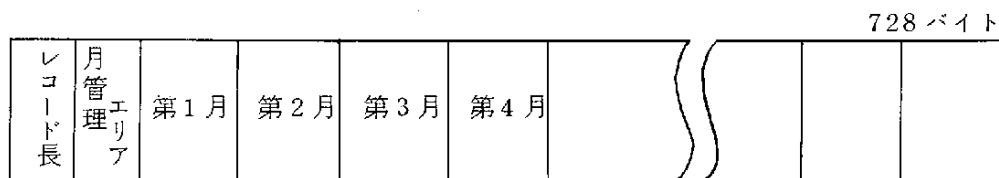
図 4 - 3 6 関連ファイル間のリンク

ることに起因している。

文献格納領域の管理ブロックは、データベースの先頭ブロックとし、その構成内容は次のようになっている。この形式で第1月から第48月までの管



理項目と、予備項目として12ヶ月分、合計60ヶ月の管理ができる。ブロック長は、12バイト×60ヶ月+4バイト(レコード長エリア)+4バイト(月管理エリア)=728バイトになる。なお、ブロック構成のうち、総記憶領域数を全月分集計することにより、実記憶領域と空領域とが算出でき、この空領域から予備項目の管理のもとにデータベースの記憶効率は向上する。



(月管理エリアは、前回のデータ格納月を記憶する)

以上のような管理のもとに、データベースのメンテナンスは行なわれるが、オーバーフロー後の追加は、古い文献データの格納してある領域を重ね書きによって処理する方法をとる。また重ね書きと同時に管理項目の書き換えも行なり。よって当システムの削除は追加処理の延長で自動的に行なわれる。

なお、消去されてしまった古い情報については、バックアップ用の文献ファイルを必要に応じて再格納すれば利用可能である。

4.3.7 ポスティング・テーブル・メンテナンス

KIND システムでは、キーワード検索によるインバーテッド・ファイル形式をとっているため、データベースのメンテナンスが生じるたびにポスティング・テーブルおよびキーワード・テーブルのポスティング情報を修正しなければならない。よってポスティング・テーブルのメンテナンス・サイクルもデータベースと同様に原則として月1回のペースで行なうことになる。

ポスティング・テーブルのメンテナンス処理は、データベースのメンテナンス処理で作成した追加分離ファイルと追加文献アドレス・ファイルをもとにして、記事番号マッチング、キーワード・ソート、語番号差換、のそれぞれの処理をポスティング・テーブル作成処理と同様の処理を行なって語番号対応ファイルを作成する。

語番号対応ファイルは、ポスティング・テーブル作成およびメンテナンスの処理を行なうごとに保存し、前回までの語番号対応ファイルを毎回マージして新語番号対応ファイル＝語番号対応マスタ・ファイルを作成しておく。これはポスティング・テーブルのメンテナンスを全面書き換え形式で行なうために、不可欠なファイルである。

新語番号対応ファイルが作成されると、ポスティング作成処理と同様の処理で、ポスティング更新処理を行なう。またキーワード・テーブル素ファイルも前処理と同様である。

なお、キーワード管理サブシステムで、キーワード・テーブルのメンテナンスが行なわれた場合、語番号差換で参照しているキーワード語番号テーブルが更新されるので、語番号対応ファイルの語番号を新語番号に直したうえで、ポスティング・テーブルをメンテナンスしなければならない。

4.3.8 ファイル概要

(1) 文献ファイル

1 5						21										61												ff																																																													
レ コ ー ド 長 (B)						記 事 番 号 (E)						言 語 コ ー ド (B)						発 行 年 (B)						発 行 国 コ ー ド (B)						記 事 区 分 (E)						ディレクトリ (B)												著 者 名 (E) フリガナ 名 (E) 所 属 機 関 名 (E) 和 文 標 題 名 (E) 原 文 標 題 名 (E) 資 料 略 名 (E) 巻 号 抄 録 文 キ ー ワ ー ド キ ー ワ ー ド 文 録																																									
4						9						1						2						1						2						1						4						4						4						4						4						4						4						4					
固定長項目						ディレクトリ項目						可変長項目																																																																													

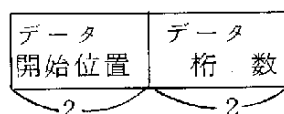
← 固定長項目 → ← ディレクトリ項目 → ← 可変長項目 →

a. 固定長項目

記事番号(9バイト)、記事区分(2バイト)はEBCDICコード。
言語コード(1バイト)、発行年(2バイト)、発行国コード(1バイト)はBINARYコード。発行年は西暦年下2桁をBINARYでもち、
第1バイトは前年、第2バイトは後年のエリアで、未使用のときは'0'。

b. ディレクトリ項目

可変長項目に対するディレクトリで、各エントリは



の形式1ワード(4バイト)からなる。

データ項目が無いエントリは'0'。

c. 可変長項目

フリガナ、巻、号はEBCDICコードで、他はすべて漢字コード。各項目とも可変長であるため、処理上ワード・バウンダリにして区切る。
キーワード項目については、先頭2バイトにキーワードの個数をBINARYで持ち、各キーワード間のセパレータはX'FF'とする。

各項目の最大長は、著者名240バイト、フリガナ30バイト、所属機関名240バイト、和文標題300バイト、原文標題600バイト、資料略名300バイト、巻30バイト、号30バイト、抄録文1,000

バイト，キーワード1,200バイト。

d 最大レコード長：4,000バイト（1,000ワード）

(2) データベース

1 5		21										57																			
レ コ ー ド 長 (B)		記 事 番 号 (E)		言 語 コ ー ド (B)		発 行 年 (B)		発 行 国 コ ー ド (B)		記 事 区 分 (E)		ディレクトリ (B)										著 者 名 (E) 所 属 機 関 名 (E) 和 文 標 題 略 名 (E) 原 文 標 題 略 名 (E) 資 料 略 号 (E) 巻 号 (E) 号 (E) 抄 録 文 (E)									
												著 者 名 (E) 所 属 機 関 名 (E) 和 文 標 題 略 名 (E) 原 文 標 題 略 名 (E) 資 料 略 号 (E) 巻 号 (E) 号 (E) 抄 録 文 (E)																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2		1		2		1																			
4		9		1		2																									

← 固定長項目 → ← ディレクトリ項目 → ← 可変長項目 →

a. 固定長項目：文献ファイルと同様

b. ディレクトリ項目：文献ファイルと同様の形式で，キーワードエントリを除いたもの。

c. 可変長項目：文献ファイルと同様の形式で，キーワード項目を除いたもの。

d. 最大レコード長：2,800バイト（700ワード）

(3) 語番号対応ファイル

1 5 9		15					24							
語 番 号 (B)	文 献 格 納 ア ド レ ス (B)	言 語 コ ー ド (B)	発 行 年 (B)	発 行 国 コ ー ド (B)	記 事 区 分 (E)	記 事 番 号								
4	4	1	2	1	2	9								1

a. 語番号：キーワード個々に付られた番号をBINARY 4バイトでもつ。

b. 文献格納アドレス：データベースに格納した文献データの磁気ディスク上のアドレス（TTRS）。

c. その他の項目は文献ファイルからとったもので同内容。

d. レコード長：24 バイト (6 ワード)

(4) ポスティング・テーブル

1 5		17												
レ コ ー ド 長 (B)	ポインタ 1						ポインタ 2	ポインタ 3				ポインタ n		
	文 献 格 納 ス (B)	言 語 コ ー ド (B)	発 行 年 (B)	発 行 国 コ ー ド (B)	記 事 区 分 (B)									
	4	4	1	2	1	2	2							

a. ポインタ内の各項目は語番号対応ファイルと同内容。

b. ポインタ数は1キーワードに対応する文献数にあたり、1レコード内に最大1,085のポインタが作成される。文献数がこれを越えると、それに応じて継続レコードが作成される。

c. 最大レコード長：13,024 バイト

(5) キーワード・テーブル素ファイル

1 5		9		13		24	
語 番 号 (B)	ポ ス テ ィ ン グ (B)	カウント情報		レ コ ー ド 数 (B)	ポ ィ ン タ 数 (B)	記 事 番 号 (B)	
		ス タ ー ト ・ ア ド レ ス					
4	4	1		3		9	3

a. ポスティング・スタート・アドレス

語番号 (キーワード) に対応するポスティング・レコードの開始アドレスで、磁気ディスク上の相対アドレス (TTRS)。

b. カウント

レコード数は、ポスティング・レコードの数であり、継続レコードの有無を識別するために使われる。

ポインタ数は、対応するポスティング・レコードの総ポインタ数。

c. 語番号、記事番号は語番号対応ファイルと同内容。

4.4. キーワード管理サブシステム

4.4.1. 処理概要

KINDシステムでは、情報検索の技法としてキーワードによる検索方式を採用しており、このキーワードを管理するのがキーワード管理サブシステムである。データベースに蓄積する情報が、情報処理に関連する技術文献であ

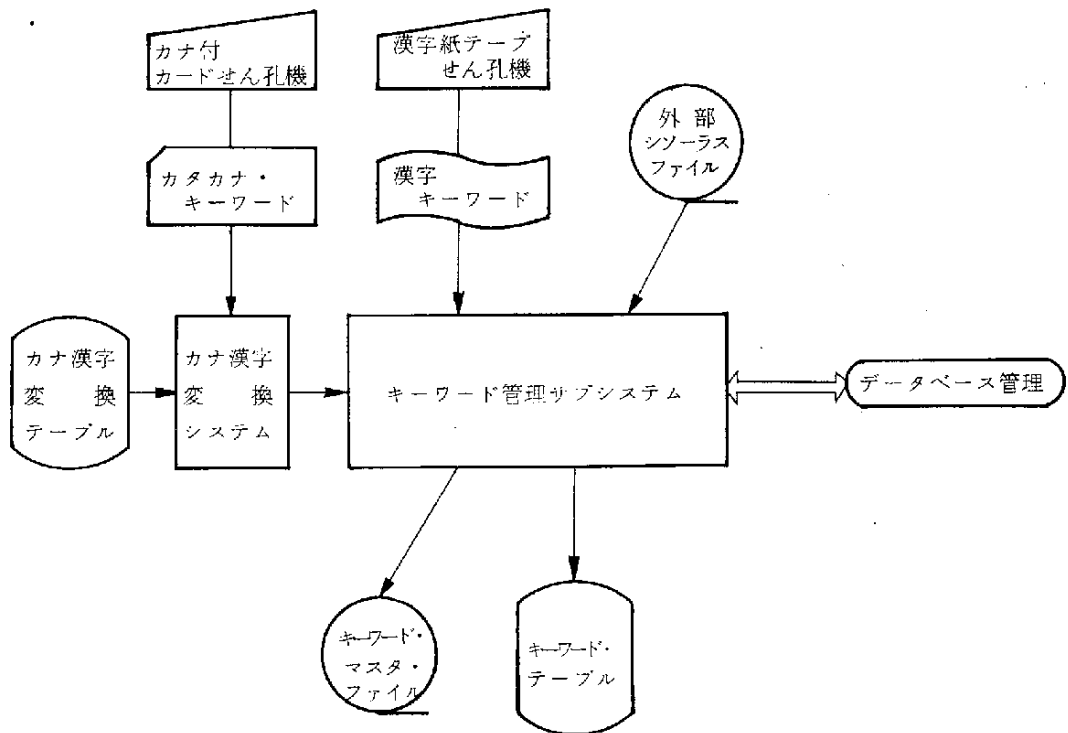


図4-37 キーワード管理サブシステムの構成図

ることから、当サブシステムで管理するキーワードも情報処理に関連した用語を選定して、キーワード・テーブルに登録する。

キーワードを選択する方法としては、文献中出现する用語をすべてキーワードにする方法なども考えられるが、当システムでは外部の研究機関で作成されているキーワードを利用するとともに、当システム用に情報処理文献や用

語集の中から重要と思われる用語を選定して、併合整備したキーワード全体がキーワード・テーブルに登録される。

キーワード・テーブルは、単なるキーワードの順編成ではなく、シソーラスの概念として下位語あるいは同義語の関係リンク情報を持たせる。これは検索サービスにおいて、利用者の要求に応じてこれらのシソーラス・サービスを行なうためである。その他、キーワード・テーブルには、各々のキーワードに対応した文献データの格納アドレスで構成されるポスティング情報を参照するために、ポスティング・テーブル上の格納アドレスと件数などを保持させる。

キーワード・テーブルのメンテナンスは、キーワードの追加と削除の処理を行なうが、ともにキーワード・マスタ・ファイルに対して処理を行ない、キーワード登録処理によってキーワード・テーブルを全面書き換えする方法をとる。

この他、キーワード管理サブシステムではキーワード登録処理を行ないながら、ブロック・インデックス・テーブルを作成して、検索時のアクセス・スピードを向上させる。

4.4.2 処理フロー

(1) キーワード・ファイル作成

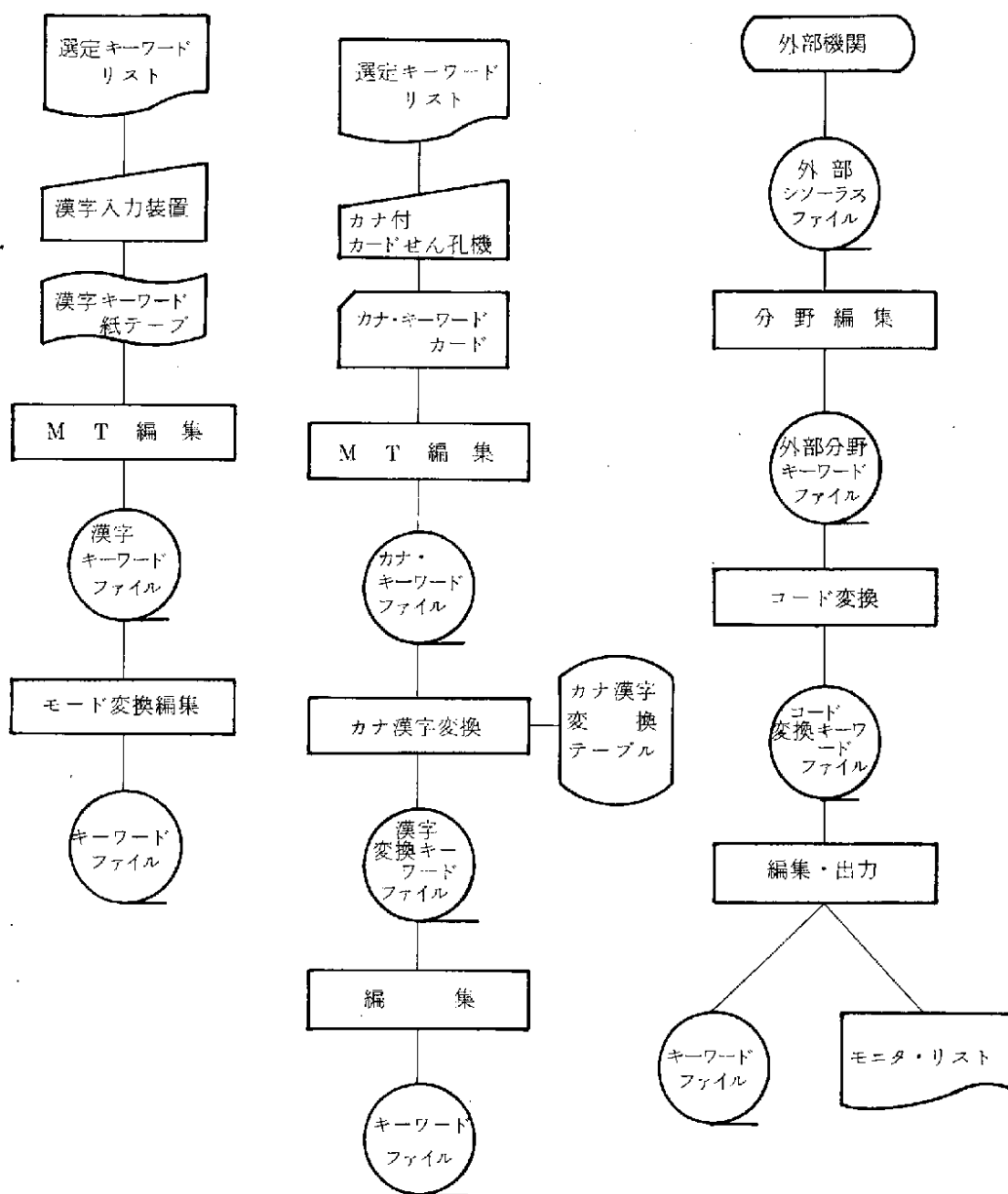


図 4 - 38 キーワード・ファイル作成フロー

(2) キーワード・テーブル登録

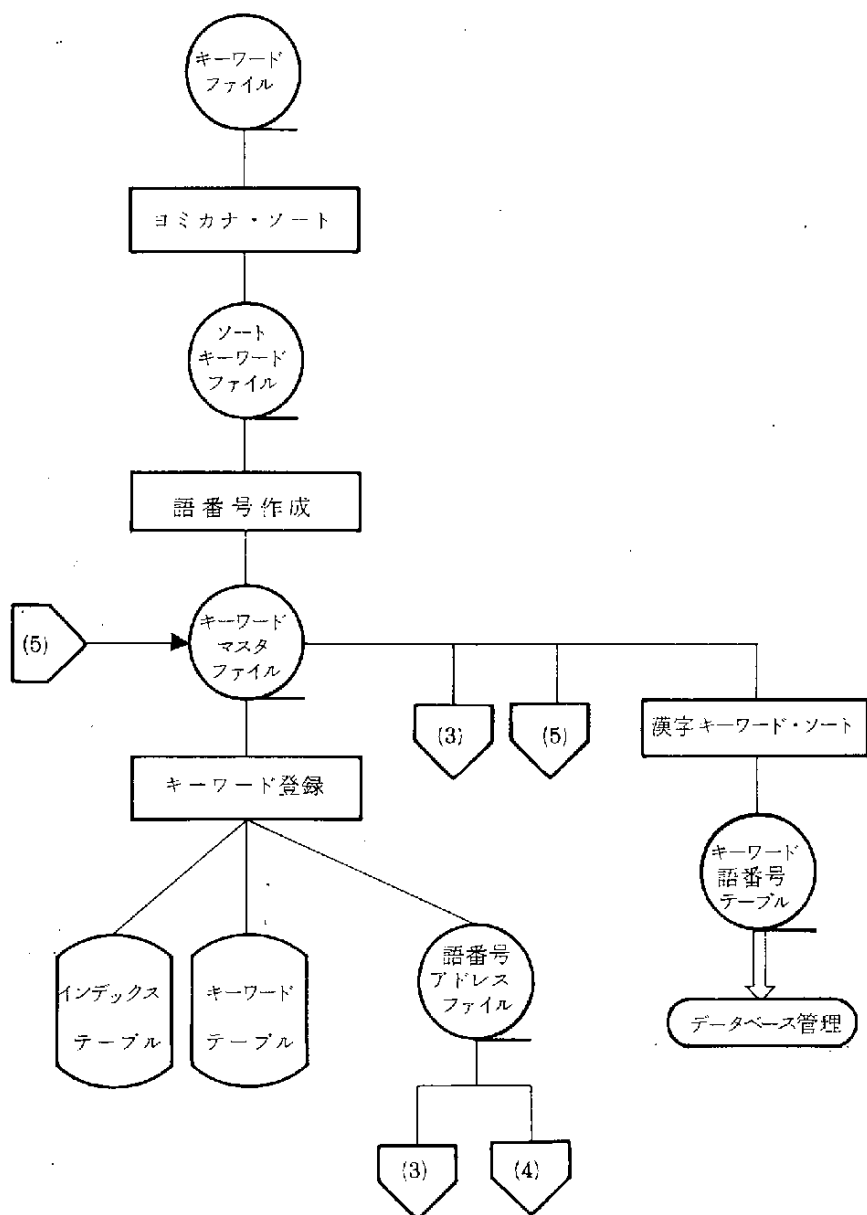


図 4 - 39 キーワード・テーブル登録フロー

(3) 関係リンク作成

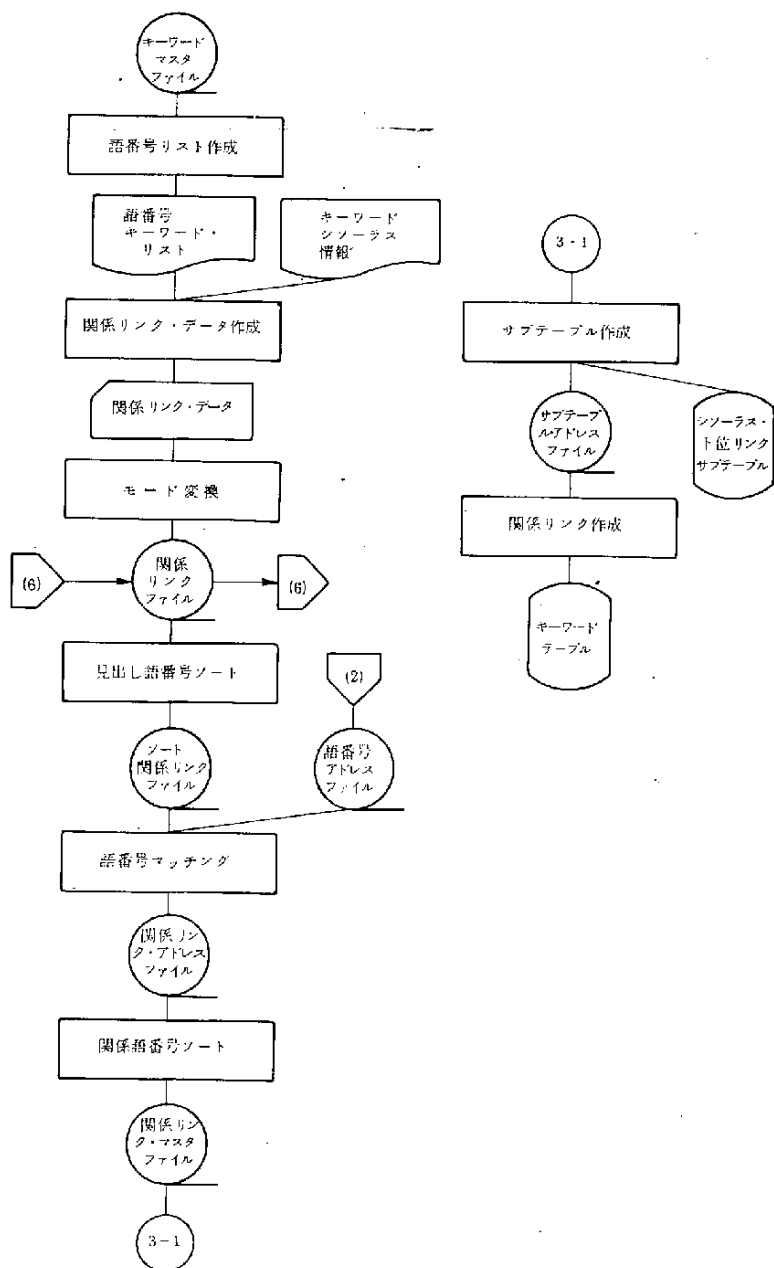


図 4 - 40 関係リンク作成フロー

(4) ポスティング・アドレス格納

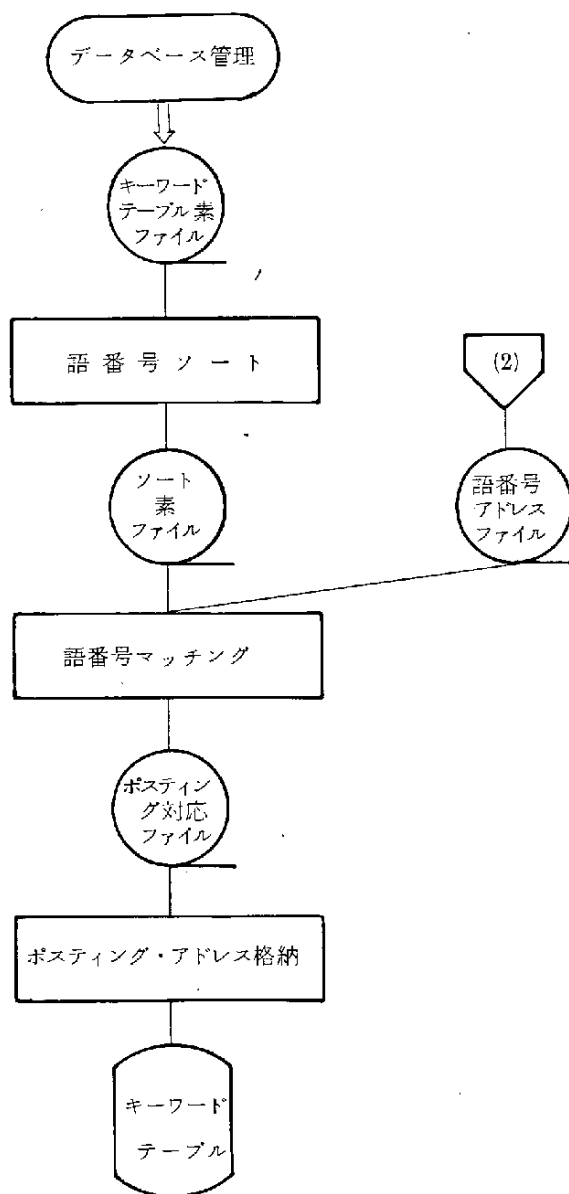


図 4 - 41 ポスティング・アドレス格納フロー

(5) キーワード・テーブル・メンテナンス

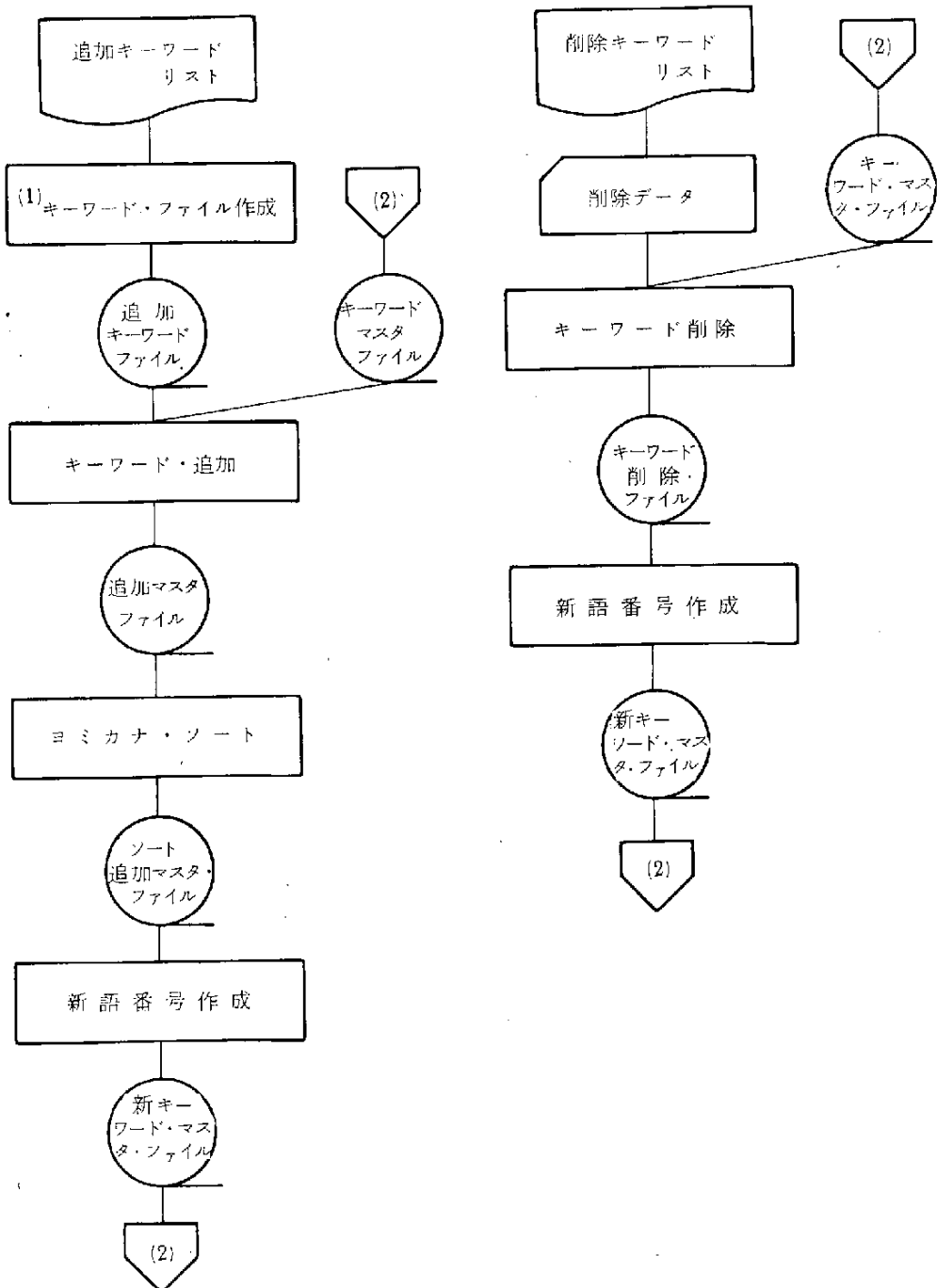


図4-42 キーワード・テーブル・メンテナンス・フロー

(6) 関係リンク修正

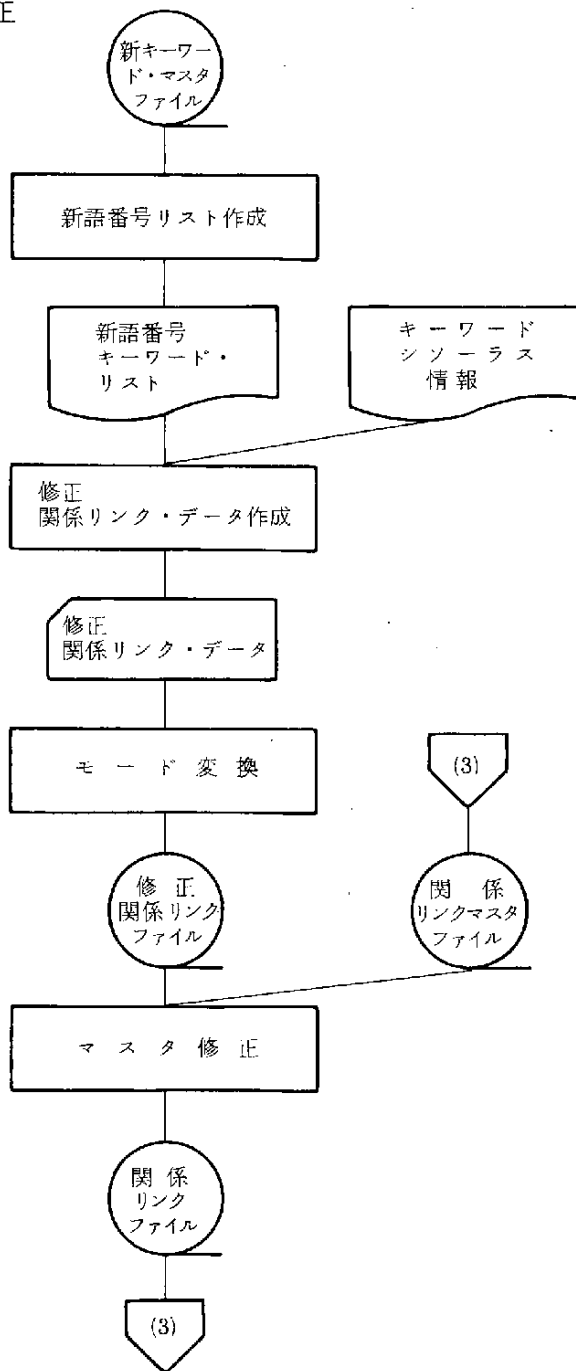


図 4 - 4 3 関係リンク修正フロー

4.4.3 キーワード・テーブルの作成

キーワード・テーブルの作成工程として、キーワード・ファイルの作成－キーワードの登録－関係リンクの作成－ポスティング・アドレスの格納、という4つの処理が行なわれる。

キーワード・ファイルの作成では、まず最初にキーワード・テーブルに登録にするキーワードの選択を行なわなければならない。KINDシステムで取扱う文献情報は情報処理に関連する技術文献であり、情報処理の分野で特徴ある用語をキーワードとして選定する必要がある。この場合、専門書や専門用語集などに出現する用語について特徴あるものを選び出し、文献の内容を表現できる代表的な用語をキーワードにする方法がとれる。しかしながら、キーワードの選定は容易なものではない。そのため、当システムでは外部の研究機関などですでに整備されているキーワードやシソーラスの利用を検討し、併用したキーワード・テーブルの構成になっている。

キーワード・ファイルは、各レコードとも漢字コードのキーワードと、EBCDICコードのカタカナで表現したヨミカナとがペアになって構成される。

キーワード・テーブルには、キーワード・ファイルのヨミカナをソーティング（昇順）したうえで、ブロック化した順編成ファイルとして作成される。またこのブロック化のインデックス情報として、ブロック・インデックス・テーブルを同時に作成する。

関係リンクの作成では、キーワードの個々に付られた語番号を使って、キーワード間のシソーラス関係を表わす関係子とともに関係リンク・データを作成し、キーワード・テーブルのシソーラス関係アドレスの項目に、リンク情報を付加する。またキーワード・テーブルのサブファイルとして、リンク情報の間接的な役割を果たすシソーラス下位リンク・サブテーブルを作成する。

ポスティング・アドレスの格納は、データベース管理サブシステムで作成したキーワード・テーブル素ファイルと、キーワード登録時に作成される語番号アドレス・ファイルとからポスティング情報を得て、それぞれを各キー

図 4-44 キーワード・テーブルの構成例

キーワード・ヨミカナ	漢字キーワード	関係子／関係キーワード（U：同義語，N：下位語）
ジョウホウカツドウ	情報活動	
ジョウホウカンリ	情報管理	U - ドキュメンテーション
ジョウホウカンリシャ	情報管理者	U - ドキュメンタリスト
ジョウホウキカン	情報帰環	N - 再送訂正，判定帰環
ジョウホウキカン	情報機関	N - 情報センタ，情報分析センタ，データセンタ
ジョウホウケンサク	情報検索	N - 事項検索，そ及検索，文献検索
ジョウホウケンサクシステム	情報検索システム	N - SMARTシステム
ジョウホウゲン	情報源	
ジョウホウゲンフゴウカ	情報源符号化	
ジョウホウゲンモデル	情報源モデル	
ジョウホウサービス	情報サービス	N - SDI，コンテンツサービス，図書館サービス，……
ジョウホウシステム	情報システム	N - INIS，INSPEC，情報検索システム，……
ジョウホウシュウシュウ	情報収集	N - 資料収集
デンシ	電子	N - Auger 電子，熱い電子，エクゾ電子，……
デンシギジュツ	電子技術	
デンシケイサンキ	電子計算機	N - アナログ計算機，ディジタル計算機，……
デンシシャシン	電子写真	
デンシセン	電子線	U - 電子ビーム

ワードに付加する処理を行なう。

以上の処理を完了すると、インバーテッド・ファイルはすべて完成し、検索サービス・サブシステムの実行が可能になる。

4.4.4 インバーテッド・ファイルの構成

KINDシステムにおけるインバーテッド・ファイルの構成は、キーワード・テーブル - ポスティング・テーブル - データベース、という磁気ディスク上に作成される3つのファイルが、それぞれのデータ格納アドレスによってリンクがとられる構造になっている。

キーワード・テーブルは、さらにブロック単位のインデックスをとった、ブロック・インデックス・テーブルとリンクされる。

ファイルの関連図を図4-45に示す。このうち、ポスティング・テーブルとデータベースはデータベース管理サブシステムで管理している。

(1) キーワード・テーブルの構造

a. キーワード数

外部のキーワード採用	約 5,000 語
新規のキーワード作成	約 5,000 語
全キーワード数	10,000 語

外部のキーワード数5,000という数字は、事例としてJICSTシソーラスに登録されているキーワードのうち、情報処理に関連すると思われる「管理システム技術編」と「電気編」のキーワードを合計した数が約4,500語であることから想定している。

当サブシステムで管理するキーワードの最大数は10,000語と定義する。

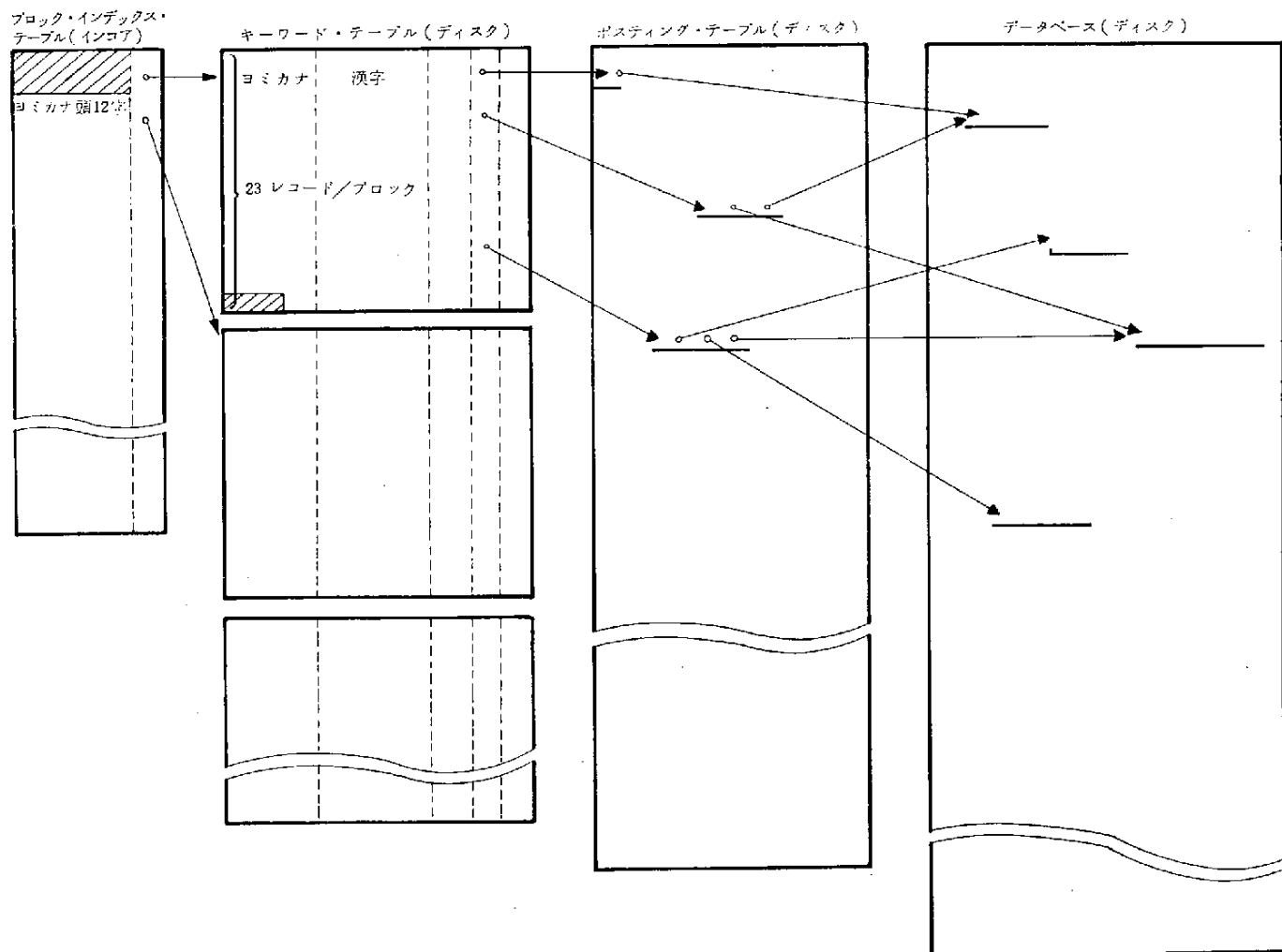
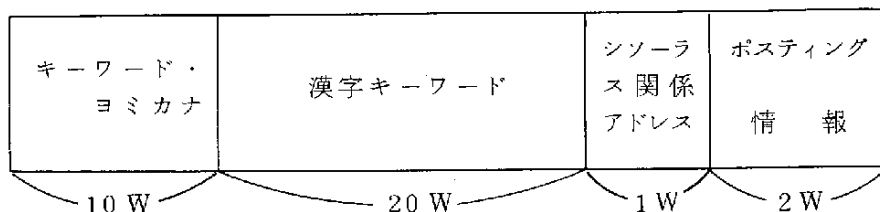


図 4-45 インバーテッド・ファイルの関連図

b.. レコード構成



・ キーワード

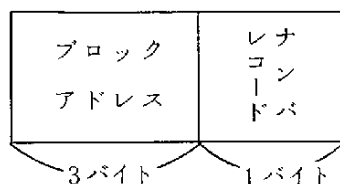
ヨミカナ：EBCDICコード 40 バイト

漢 字：漢字コード 40 文字 = 80 バイト

・ シソーラス関係アドレス

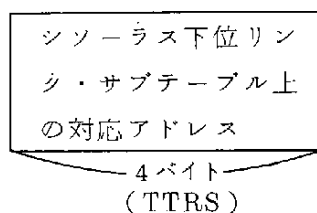
シソーラスの概念として、下位語および同義語の関係をリンクさせるための項目エリアであり、同義語の関係をもつキーワードのアドレスあるいは下位語の関係を間接的にもったシソーラス下位リンク・サブテーブルの対応アドレスが保持される。同義語・下位語の識別は、次のように行なう。

同義語のアドレス



（ブロック・アドレス：キーワード・テーブルのブロッ
キング・アドレス (TTR)
レコード・ナンバ：ブロック内のレコード・シーケ
ンス・ナンバ

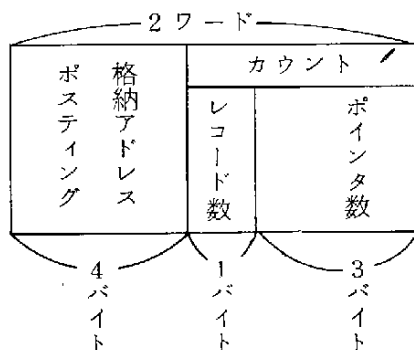
下位リンクのアドレス



関係リンクのない場合はこの項目はオール“0”とする。

・ ポスティング情報

データベース管理サブシステムで作成されるポスティング・テーブルについて、各キーワードに対応するポスティング・レコードの格納アドレス（複数レコードの場合先頭アドレス）と、レコード数・ポイント数のカウント情報から成る。



c. ブロック構成

キーワード・テーブルを格納する集団磁気ディスク装置（F477K）の1トラックの容量が13,165バイトであることから、

$$13,165 \leq N (136 \times L + 135)$$

という計算式が得られる。

- N : トラック当りのブロック数
- L : ブロック当りのレコード数
- 136 : レコード長
- 135 : レコード制御バイト数

この式に対してN = 1 から 5 までの計算結果は表 4 - 14 のようになる。

表 4 - 14 ブロック構成算出表

N	L	トラック当り レコード数	ブロック当り バ イ ト 数
1	95	95	12,920
2	47	94	6,392
3	31	93	4,216
4	23	92	3,128
5	18	90	2,448

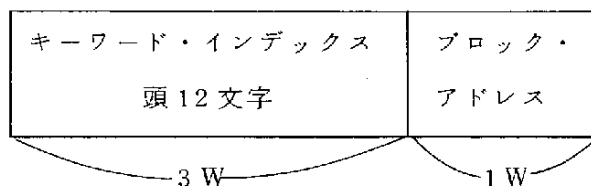
以上の結果から、検索サービス実行時にインコアに必要なバッファ・サイズを考慮してN=4を採用する。よってキーワードのブロック化は23レコード単位に行なわれ、最大レコード総数10,000語に対して計435ブロック必要になる。

なおヨミカナ・キーワードで発生する同音異義キーワードについては、同音異義のキーワードが2つのブロックにまたがることを禁止する。その処置方法は、各ブロックの最終レコードと次のブロックの先頭レコードをチェックして、同音異義のキーワードであった場合、前ブロックの同音異義キーワードをすべて次のブロックにもっていく。この場合前ブロックの同音異義キーワードのエリアはゼロ・クリアし、縮小ブロック(1ブロック<23レコード)の形式をとる。

(2) ブロック・インデックス・テーブルの構造

キーワード・テーブルに対して、ブロックごとのインデックスをもったのがブロック・インデックス・テーブルである。このテーブルは、当サブシステムのバッチ処理では外部記憶装置に作成され、検索サービス・サブシステムにおいては、インコアにイニシャル・ローディングして使用される。

a. フィールド構成



・ キーワード・インデックス

キーワード・テーブルのブロック化に対して、ブロックの最終レコードのキーワード・ヨミカナ40バイトの頭12バイトをインデックスにとる。

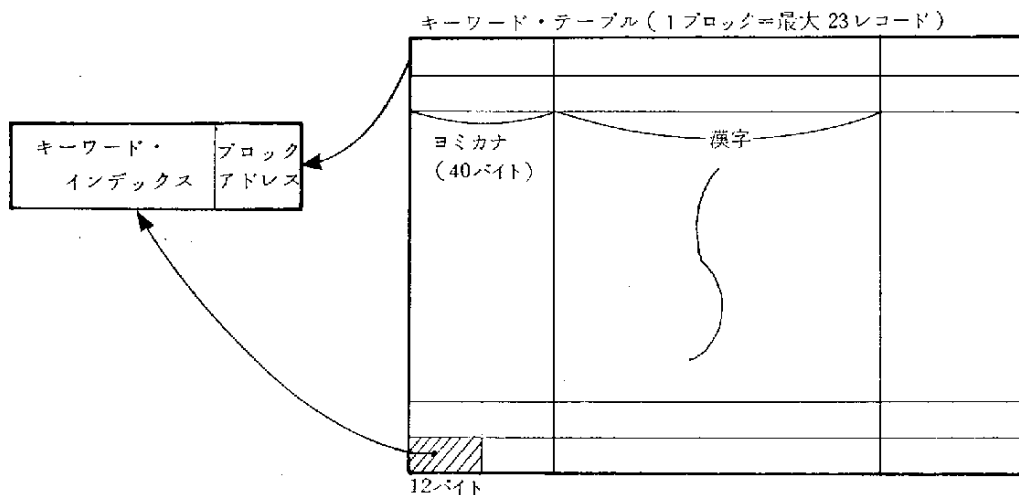


図 4-46 ブロック・インデックス関連図

- ・ ブロック・アドレス

キーワード・テーブルの各ブロックを磁気ディスクに作成して得られる相対アドレス (TTRS)。

b. テーブル・サイズ

キーワード10,000語に対して、435ブロックが作成される。よって
テーブルの最大容量は、

$$16 \text{ バイト (4 ワード)} \times 435 = 6,960 \text{ バイト (1,740 ワード)}$$

になる。

4.4.5 キーワード・テーブル・メンテナンス

キーワード・テーブルのメンテナンスは、追加および削除についての処理を行なう。

技術文献の特徴から、新しく発表される研究論文等には、従来使用されたこともない用語が出現することがある。また「ことば」の移り変りともいえる流行を帯びた用語の出現や、一方では長期間にわたって一度も使われることもなくすでに陳腐化してしまった用語になることもある。このようなことから当サブシステムの管理では、システムの複雑さをさけるとともにメンテナンス処理の簡易化を考慮して追加と削除の処理に限定して行なう。

追加キーワードの作成方法は、キーワード・ファイルの作成と同様の処理を行なって追加キーワード・ファイルを作成する。そして前回までにキーワード・テーブルに登録してあるキーワード・マスタ・ファイルと照合を行ない、新たな語番号をふりなおして新キーワード・マスタ・ファイルを作成する。

キーワードの削除方法は、前回までのキーワード・マスタ・ファイルに付られている各キーワードの語番号を使って、削除データ・カードをせん孔し、語番号をキーにしたマッチングにより削除の処理を行なう。このあと新たに語番号をふりなおして新キーワード・マスタ・ファイルを作成する。

追加あるいは削除によって作成された新キーワード・マスタ・ファイルは、キーワード登録処理によって、キーワード・テーブルとブロック・インデックス・テーブルを全面更新し、インバーテッド・ファイルを作りかえる。またキーワードのメンテナンスとともに、関係リンクの更新も行なう。

4.4.6 ファイル概要

(1) キーワード・ファイル

1	41	120
キーワード・ヨミガナ	漢字キーワード	
(E)	(K)	
40	80	

a. キーワード・ヨミガナ

キーワードのヨミガナをEBCDICコードのカタカナで40バイト以内で表現。残りは“0”。

b. 漢字キーワード

キーワードを漢字コードで40文字＝80バイト以内で表現。残りは“0”。

c. レコード長：120バイト

(2) キーワード・マスタ・ファイル

1	5	9	49	128
語番号	旧語番号	キーワード・ヨミガナ	漢字キーワード	
(B)	(B)	(E)	(K)	
4	4	40	80	

a. 語番号

キーワード・ヨミガナのソート順（昇順）に付られたシーケンス番号をBINARYで表現。

b. 旧番号

キーワードのメンテナンスによって語番号が付け加えられた場合の、

前回までに使われた語番号。

c. キーワード・ヨミガナ／漢字キーワード

キーワード・ファイルと同様。

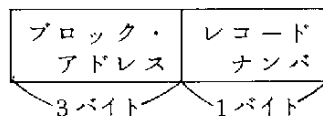
d. レコード長：128 バイト

(3) 語番号アドレス・ファイル

1	5	9	13	53	132
語 番 号	旧 語 番 号	キ 格 納 ワ ア ド レ ス	キ ー ワ ー ド ・ ヨ ミ ガ ナ	漢 字 キ ー ワ ー ド	
(B)	(B)	(B)	(E)	(K)	
4	4	4	40	80	

a. キーワード格納アドレス

キーワード・テーブル上のアドレスで、ブロック・アドレスとレコード・ナンバが



の形式になっている。ブロック

・アドレスは磁気ディスク・ファイルの相対アドレス (TTRS) の TTR である。レコード・ナンバは、ブロック化されたキーワード・レコードのブロック内のシーケンス番号である。

b. 語番号／旧語番号

キーワード・マスタ・ファイルと同様

c. キーワード・フリガナ／漢字キーワード

キーワード・ファイルと同様

d. レコード長：132 バイト

(4) キーワード語番号テーブル

1	5	9	49	128
語 番 号 (B)	旧 語 番 号 (B)	キーワード・ヨミガナ	漢 字 キ ー ワ ー ド	
4	4			

a. 各項目の内容はすべてキーワード・マスタ・ファイルと同様であるが、このテーブルでは漢字キーワードのソート順（漢字コードの小さい順）に並びかえてある。

b. レコード長：128 バイト

(5) 関係リンク（マスタ）ファイル

1	5	9	12	20
見語 出番 し号 (B)	関 係 子 (E)		関係語 番号 (B)	
4	1	3	4	8

（最小レコードが20
バイトに制限されて
いる。）

a. 見出し語番号／関係語番号

いずれの語番号もキーワード・マスタ・ファイルの語番号と同様の意味であるが、見出し語と関係語の関連性は主体となるのが見出し語で、それに対して関係語はどういう関係かを関係子で識別する。

b. 関係子

EBCDIC コードで識別。「B」のときBT関係（Broader Term；上位語），「U」のときUF関係（Used For；同義の非ディスクリプタ）を意味する。

c. レコード長：20 バイト

(6) 関係リンク・アドレス・ファイル

1	5	9	13	20	
見出し語番号 (B)	格納アドレス (B)	関 係 子 (E)		関係語番号	
4	4	1	3	4	4

a. 格納アドレス

見出し語番号に対応するキーワードの格納アドレスで、語番号アドレス・ファイルのキーワード格納アドレスの項目からとったもの。

b. その他の項目は、関係リンク・ファイルと同様。

c. レコード長：20 バイト

(7) シソーラス下位リンク・サブテーブル

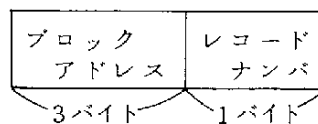
1	5	下 位 リ ン ク (B)									
レコード長 (B)	格納アドレス										
4	4	4	4	—	—	—	—	—	—	—	4

可変長レコード

a. 下位リンク

シソーラス関係のうち下位に対応するキーワードの格納アドレスをリンクとしてもつ。

格納アドレスは



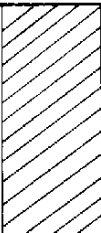
の形式でキーワード・テ

ブル上の相対アドレスである。

1 キーワードに対して最大 200 個の下位リンクを確保する。

b. 最大レコード長：808 バイト

(8) サブテーブル・アドレス・ファイル

9			20	
関係語番号 (B)	関係子 (E)		シソーラス関係アドレス (B)	
8	1	3	4	8

a. 関係子／関係語番号

関係リンク・アドレス・ファイルの見出し語と関係語の関連を逆転した形式で関係子をもつ。

EBCDIC コードで「N」のとき NT 関係 (Narrower Term ; 下位語), 「S」のとき USE 関係 (Use ; 同義ディスクリプタ参照) を意味する。

b. シソーラス関係アドレス

関係語を主体にしたシソーラス関係 (下位語, 同義語) のキーワードのアドレスを保有する。

下位語に対する関係アドレスは, シソーラス下位リンク・サブテーブル上のアドレス (TTRS) を, 同義語に対する関係アドレスは, キーワード・テーブル上のアドレスをもつ。

c. レコード長：20 バイト

4.5 索引誌作成管理サブシステム

4.5.1 処理概要

KINDシステムでは、オンライン会話型による文献情報の検索が主目的であり、検索サービス・サブシステムをメインにして、データベース管理サブシステムおよびキーワード管理システムにおいて各種のファイルが作成・管理され、情報検索が行なわれる。これら3つのサブシステムに比べて特異な、独立したサブシステムとして索引誌作成管理サブシステムがある。

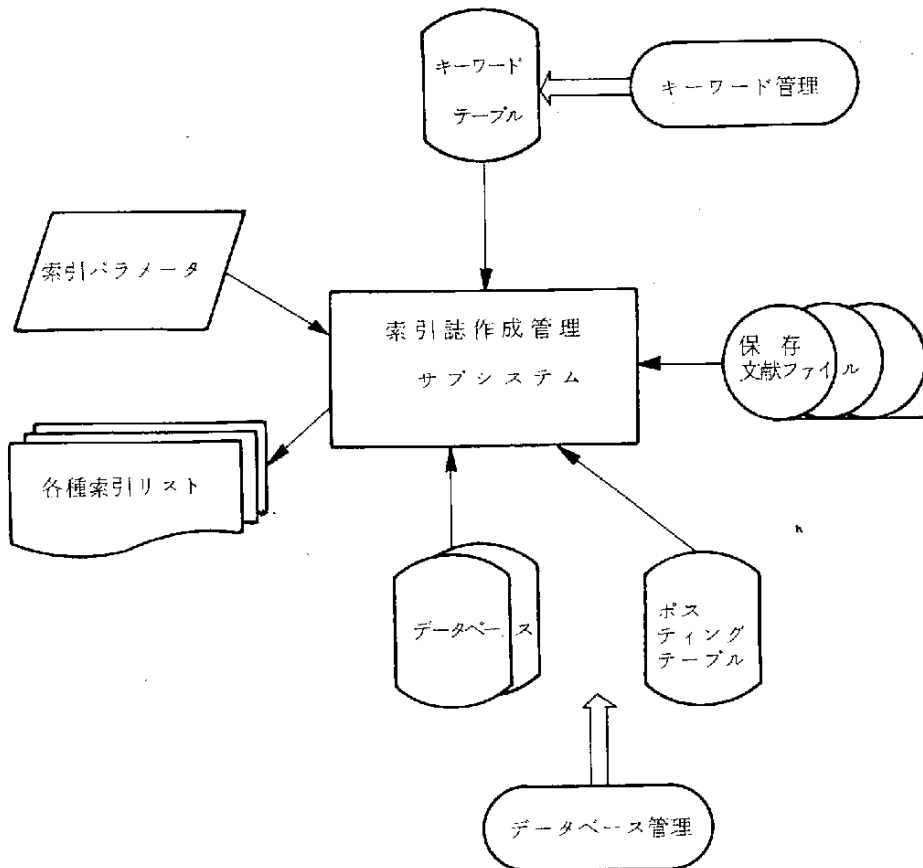


図 4 - 47 索引誌作成管理サブシステムの構成図

索引誌作成管理サブシステムは、すでに作成されているデータベースとキーワード・テーブル、およびこれらをリンクするポスティング・テーブルとを使って、各種索引リストや統計リストを作成するためのサブシステムである。

当サブシステムでは、主に次のような索引リストおよび統計リストを作成する。

- ① 著者名索引リスト
- ② キーワード索引リスト
- ③ キーワード統計リスト

著者名の索引リストの作成は、文献それぞれの著者に対して一連の研究論文等を調査する目的で行なうものであり、文献データの著者名および著者名フリガナの項目を取り出して集計を行ない、索引リストを作成する。

キーワード索引リストは、キーワード・テーブルに登録されている各キーワードに対して、文献データがどのような分布で蓄積されているかを知る手がかりを与えるための索引リストであり、各キーワードに文献の記事番号を対応させた形式で提供される。

キーワード統計リストは、キーワード・テーブルの各キーワードについて、年度別の使用頻度分布をとって、キーワードの使われ方を調査する目的で行なうものであり、キーワード・テーブルのメンテナンスの判断資料にも使われる。

4.5.2 処理フロー

(1) 著者名索引作成

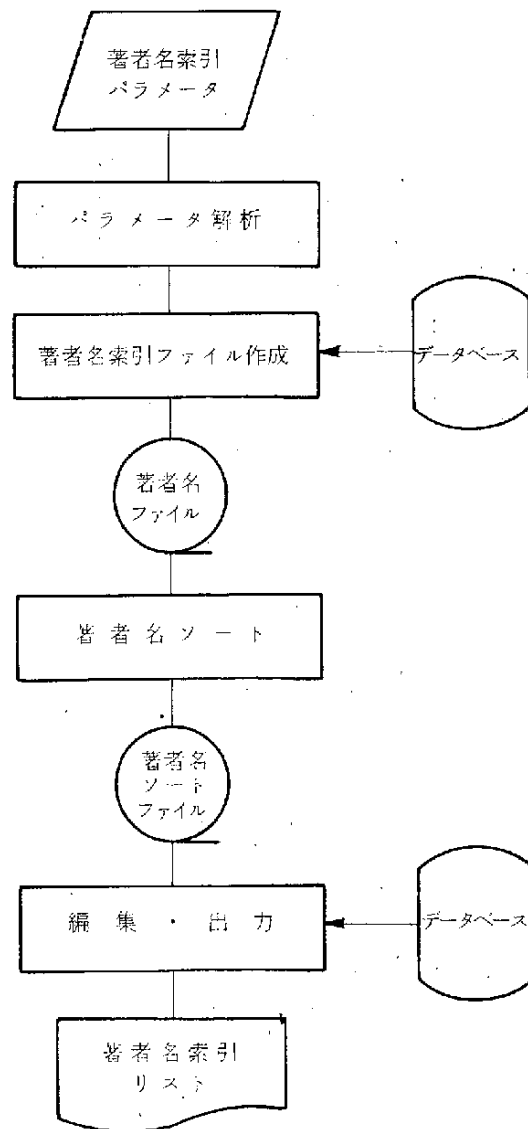


図4-48 著者名索引作成フロー

(2) キーワード索引作成

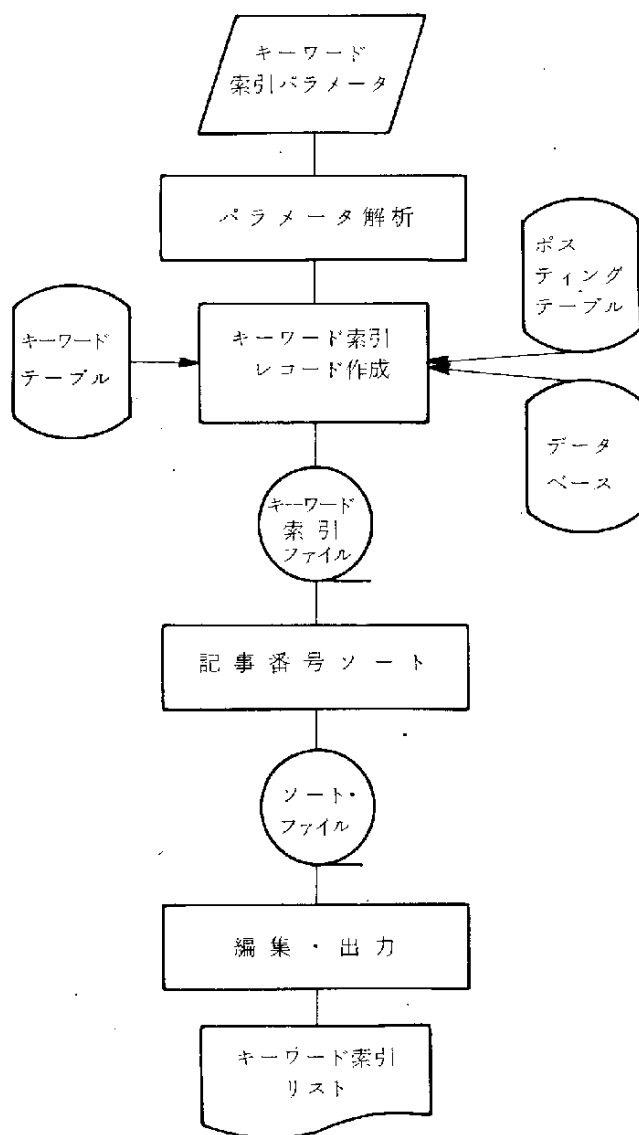


図 4 - 49 キーワード索引作成フロー

(3) キーワード統計リスト作成

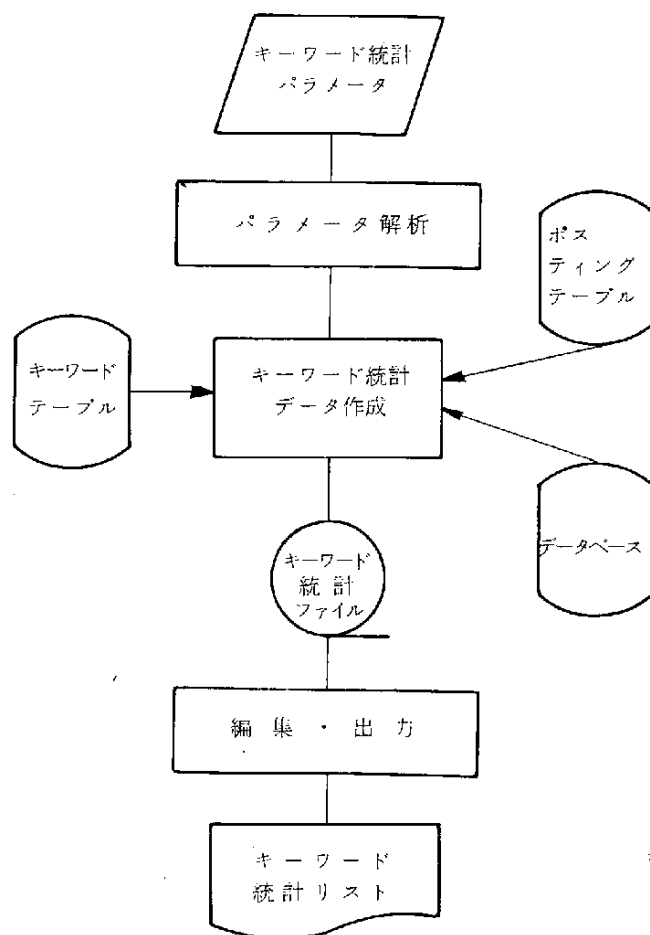


図4-50 キーワード統計リスト作成フロー

4.5.3 著者名索引

著者名の索引は、それぞれの著者について一連の研究発表論文を追跡して成果等を調査したり、どんな著者がどういった研究を行なっているか等を調査する目的で作成するものである。

文献ファイルおよびデータベースに蓄積される著者名項目と著者名フリガナ項目の特徴として、日本語で表記された著者名（日本人あるいは日本語で表記可能な外国人）と外国語で表記された著者名（外国人あるいは外国語で表記した日本人）、および団体著者名（日本語表記に限る）という3種類の著者がある。また著者名フリガナは、EBODIC コードのカタカナで表わすために日本語の著者名に限ってフリガナが付られている。しかも文献の著者名には姓だけのもの、あるいは名があっても読み方が不明なもの、などがあるために姓の一般的な読み方で著者名フリガナは構成されている。

このようなことから、著者名について索引をとるには、これら著者名の種類を識別して索引をとる必要がある。よって著者名索引リスト作成処理では、これらの識別をパラメータで指示・選択して行なわれる。

処理手順としては、まずパラメータを解析して指定された著者名に関する文献をデータベースから抜取り、著者別に整理してリストが作成される。索引リストには著者別に、和文標題／記事区分／資料略名／発行年／巻／号、および記事番号とが漢字コード編集されて出力される。

・著者名索引パラメータ形式

$$\begin{pmatrix} \text{NAME} \\ \text{NAME2} \end{pmatrix}, \begin{pmatrix} \text{JI} \\ \text{JG} \\ \text{F} \end{pmatrix}, \text{文献発行年指定}, (\text{姓フリガナ } n_1, n_2, n_3, \dots, n_{10})$$

出力形式

NAME : 和文標題, 記事区分, 資料略名, 発行年, 巻, 号, 記事番号
NAME2 : 記事番号

著者名記述形態指定

J : 日本語の著者名 (個人, 団体とも)

JI : 日本語の個人著者名

JG : 日本語の団体著者名

F : 外国語の著者名

省略: 全著者名

文献発行年数

YY : 年指定

YY - Y'Y' : 年範囲指定

省略: 全年指定 (YY, Y'Y' : 西暦年下2ケタ)

特定姓指定

著者名記述形態で「JI」指定のとき有効

著者の姓をフリガナ (カタカナ) 12字以内で指定

最大10名のフリガナ指定可

ただし, フリガナ姓のため同音姓も混同する。

・索引リスト出力形式

NAME 指定

著者名							
記事番号	和	文	標	題	記事区分		
	資	料	略	名	発行年	巻	号

NAME 2 指定

著者名				
記事番号 1	記事番号 2	記事番号 3	記事番号 4	記事番号 5

NAME 指定のときの指力例を図4-51に示す。

	1	2	3	4	5	10	20	30	40	50	55
5 行 改 行	1										
	2										
	3										
	4										
	5										
	6	石井 一郎	B75051305	雪害・土砂害に関する道路情報のシステム化について〔b②〕							
			F433A	道路建設('74)[323]							
	10	石本 新	B75051193	論理学と自然言語Ⅲ 内包的論理学による自然語の分析〔b②〕							
			F251A	オペレーションズ リサーチ('74)19[12]							
			B75051195	OR入門 論理学と自然言語 4-述語論理の翻訳定理〔b②〕							
			F251A	オペレーションズ リサーチ('75)20[1]							
		小山田英一	B75050945	中小企業経営と職業安定行政〔b①〕							
			G786A	職業研究('75)29[1]							
		春日 敬三	B75052071	新交通システムの安全基準等検討委員会の答申について〔b②〕							
			S079A	J R E A('74)17[11/12]							
	20	坂井利ゆき	B75051525	インハウス・コンピュータ・ネットワークとHOSTコンピュータ〔a①〕							
			G427A	情報処理('74)15[12]							
		高木 尙一	B75050938	調停に現われた病院の労使問題〔b①〕							
			G395A	労働の科学('74)29[11]							
		高木 和男	B75050939	欧州産業給食見学旅行2〔b②〕							
			G395A	労働の科学('74)29[11]							
	30										
		西川 広	B75052710	工業高校における造船教育の一考察〔b②〕							
			F166A	造船界('74)[66]							
	65										
	66										
5 行 改 行	70										

4.5.4 キーワード索引リスト

キーワード索引は、キーワード・テーブルに登録されている全キーワードについて、各キーワードに文献がどのように対応しているかを記事番号（各文献ごとに付られた唯一の番号で、分野区分／作成年月／一連番号から成る）を対応させて出力作成したものである。

キーワードは、文献中に記述されている情報の内容から、情報検索上重要な事柄を「ことば」として表現したものであり、このキーワード索引リストを参照することによって利用者は文献検索サービスを受ける際に大きな手がかかりとなる。とくに検索キーワードを入力する時には、キーワードに対応する文献の有無が明確に得られ無駄な検索をさけることができる。

キーワード索引の作成にあたっては、キーワード・テーブルの登録順（EBCDICコードのヨミカナ順）にポスティング・テーブルを参照しながら文献データのデータベース内の格納アドレスを求め、そのアドレスから文献データを探索して、対応する記事番号を得る。この記事番号は、蓄積モードがEBCDICコードであるため、漢字プリンタに出力するには漢字コードに変換され、出力様式に編集されてキーワード索引リストが作成される。

キーワード索引リスト作成のパラメータ形式は次のようになる。第2項目の文献、発行年の指定は、全文献データについての索引を行なうか、あるいは一部指定年に行なうかを指示する項目であり、文献データ中の発行年を参照して識別する。

- ・ キーワード索引パラメータ形式

KEYWORD, 文献発行年指定

文献発行年指定

YY : 年指定

YY - Y'Y' : 年範囲指定

省略 : 全年指定 (YY, Y'Y' : 西暦年下2ケタ)

$$\frac{7-1}{2}$$

出力例を図4-52図に示す。

図 4-52 キーワード索引リスト出力形式

4.5.5 キーワード統計リスト

キーワード統計リストは、キーワード・テーブルに登録されている全キーワードについて、各キーワードが年とともにどのように使われているかの頻度統計をとったリストである。

キーワードの意味することは前述のごとく文献の内容を代表する重要な「ことば」である。このキーワードについての分布を年別にとってみると、あるキーワードが高い頻度で出現している年ならば、その年にはこのキーワードに関連した研究論文や文献が多く発表・出版されたという判断のもとに、キーワードの流行、換言すればそのキーワードの意味するような内容の研究熱が高まった年であると推察することができる。また、あるキーワードの出現頻度が年とともに下降状態になり、ついには出現頻度がゼロになって数年の経過が続いたならば、もはやそのキーワードは陳腐化したか、あるいは別の新しい「ことば」へと変化してしまったという予測も可能である。その他にも、キーワードのメンテナンスに役立つ資料としてもまた有効な統計と思われる。

キーワード統計リスト作成は、パラメータ方式で行ない、年別のキーワード出現頻度を集計して漢字プリンタに出力する。

(なお、統計リストの種類は、現時点では1種であるが、将来他の統計リストの作成も行なう予定である)

・キーワード統計パラメータ形式

TOKEI, KEYWORD, 年指定

年指定 (文献発行年)

YY : 年指定
(YY -- Y'Y' : 年範囲指定 (範囲 ≤ 15 年))
省略 : 現在の年から過去 15 年

(YY, Y'Y' : 西暦年下 2 ケタ)

・索引リスト出力形式

キーワード				
年…	頻度	年…	頻度	年… 頻度

出力例を図4-53に示す。

	1	2	3	4	5	6	10	20	30	40	50	55
5行改行	1	2	3	4	5	6						
	6						経営管理					(320)
							'69 ...	35	'70 ...	101	'71 ...	80
							'72 ...	45	'73 ...	31	'74 ...	15
							'75 ...	13				
10							経営情報システム					(522)
							'69 ...	150	'70 ...	77	'71 ...	70
							'72 ...	60	'73 ...	55	'74 ...	50
							'75 ...	60				
							計算機アルゴリズム					(168)
							'69 ...	15	'70 ...	30	'71 ...	25
							'72 ...	21	'73 ...	35	'74 ...	18
							'75 ...	19				
							計算機システム					(711)
							'69 ...	190	'70 ...	130	'71 ...	95
							'72 ...	68	'73 ...	132	'74 ...	89
							'75 ...	102				
							計算機シミュレーション					(181)
							'69 ...	50	'70 ...	20	'71 ...	18
							'72 ...	33	'73 ...	15	'74 ...	21
							'75 ...	16				
							計算機プログラミング					(459)
							'69 ...	95	'70 ...	103	'71 ...	83
							'72 ...	54	'73 ...	29	'74 ...	58
							'75 ...	32				
30							ゲーム理論					(183)
							'69 ...	35	'70 ...	29	'71 ...	83
							'72 ...	19	'73 ...	22	'74 ...	15
							'75 ...	25				
5行改行	65											
	66											
5行改行	70											

図4-53 キーワード統計リスト出力形式

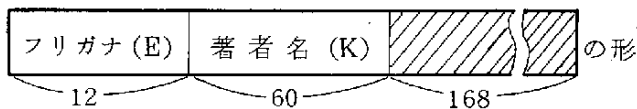
4.5.6 ファイル概要

(1) 著者名ファイル

1	241	248
著者名(フリガナ)	該格 納 文 ア ド レ ス	
(K)		
240	4	

a. 著者名

「J1」指定のとき



式をとり、その他の指定は著者名のみ。

b. 該当文献格納アドレス

データベースに格納してある該当文献データのアドレス (TTRS)

c. レコード長：248 バイト

(2) キーワード索引ファイル

1	5	9	13							
レ コ ー ド 長 (B)	語 番 号 (B)	記 事 番 号 件 数 (B)	記事番号	1	2	3	4	5	6	n
4	4	4	9	9	9	9	9	9	9	9

a. 語番号

キーワード・ヨミカナのソート順(昇順)に付られたシーケンス番号。

キーワード・テーブルのレコード格納順と一致する。

b. 記事番号件数

キーワード索引によって得られた記事番号の件数。

1レコードには最大100件の記事番号まで記録でき、それをこえる場合は100件以内のレコードを継続させて、総件数を第1レコードに入れる。継続レコードにはゼロを入れる。

c. 記事番号

文献データごとに付られた番号で、EBCDICコード9桁から成る。

d. 最大レコード長：912 バイト

(3) キーワード統計ファイル

語 番 号 (B)	発行年カウント・エリア (B)														
	第 1 年 (B)	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	4	4	4	4											

a. 語番号

キーワード・ヨミカナのソート順（昇順）に付られたシーケンス番号。キーワード・テーブルのレコード格納順と一致する。

b. 発行年カウント・エリア

指定された基準年（古い年）をエリアー1にセットし、以降最大15年間のカウントを入れる。指定年が15年未満のとき、不要なエリアにX'FF'を入れる。

4.6 システムの利用法

4.6.1 端末の構成

検索サービスサブシステムで使用する端末の構成を図4-54に示す。

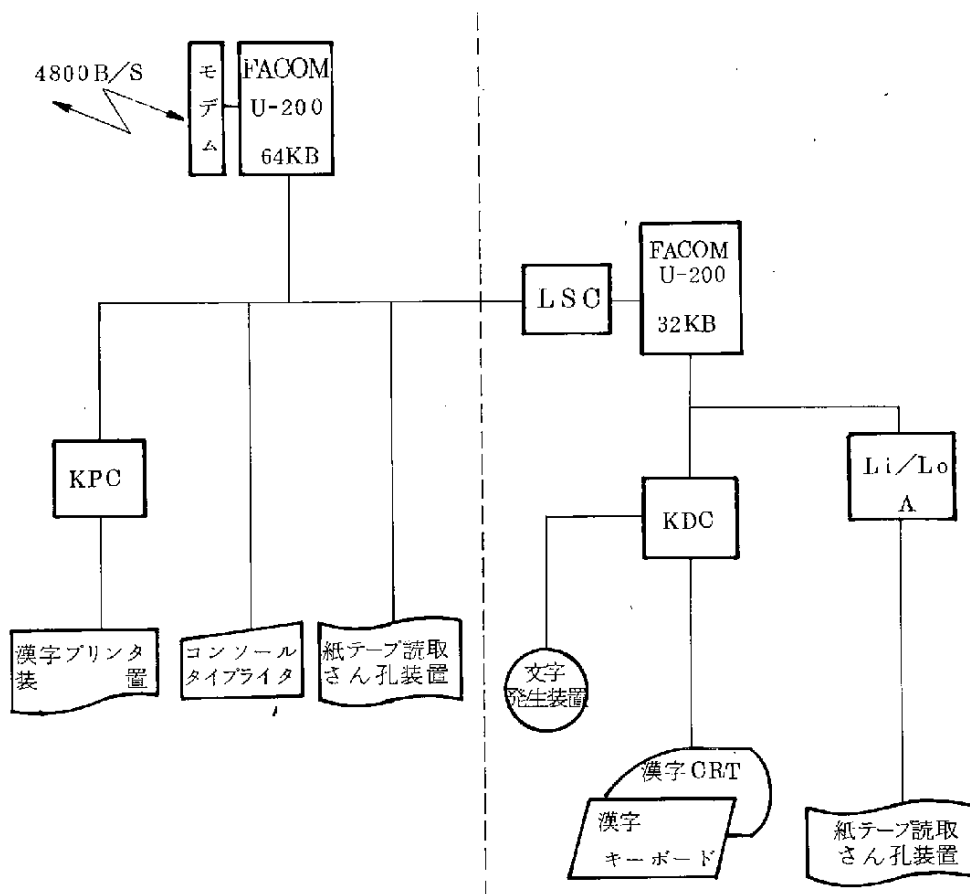


図4-54 端末装置構成図

4.6.2 漢字ディスプレイ装置の盤面配列

(1) キーボードの配列

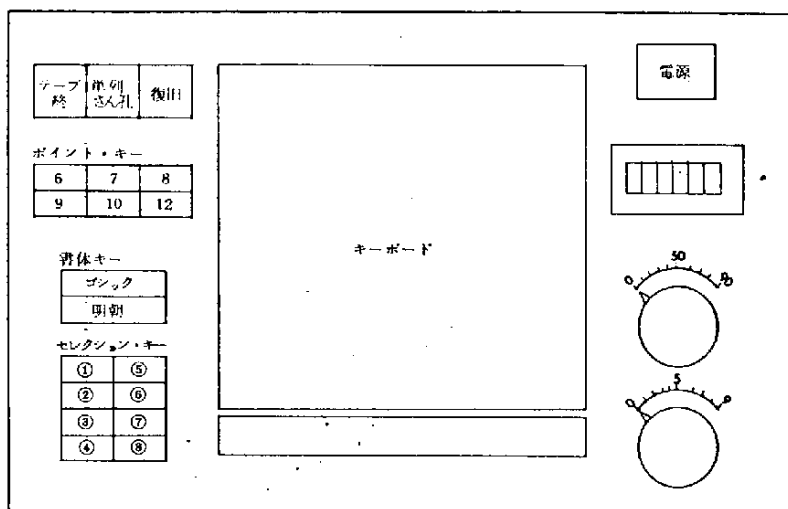


図 4 - 5 5 漢字キーボードの配列

(2) キーボード・アダプタの配列

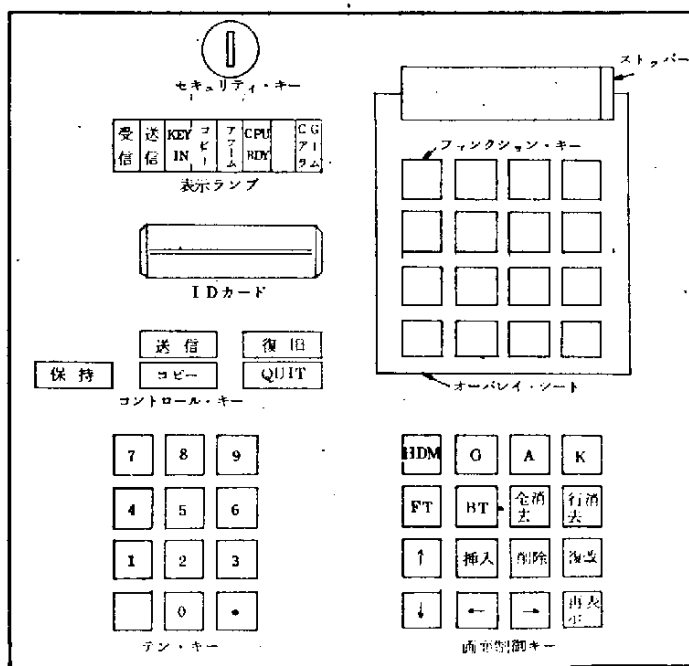


図 4 - 5 6 キーボード・アダプタの配列

4. 6. 3 各種キーの説明

文献検索時において、システムとの会話に必要な各種のキーを表4-15に概説する。これらのキーは、キーボード・アダプタに配置されている。

表4-15 各種キーの説明

キー種別	操 作 キー	説 明
ファンクション・キー	検索コマンドが指定されているキー	オーバーレイ・シートを併用して、検索コマンドとして使用する。配置は、「検索コマンド」の項を参照。
テン・キー	 キーを除くすべてのキー	このキーを押すことにより、数字およびピリオドが入力できる。
コントロール・キー		このキーは、画面に表示した入力データをセンター側に送るとき使用する
画面制御・キー		データをEBCDICコードで入力する。当システムでは、データを入力する前には必ず押す。
	   	これらのキーを押すことにより、カーソルの移動を行なう。これは、入力データの修正などに使う。
	再表示	現在表示した画面を再び表示させる。

4.6.4 表示・出力の様式

検索に関する諸情報は、漢字ディスプレイ装置、および漢字プリンタ装置を使って表示・出力される。

(1) 漢字ディスプレイの画面

漢字ディスプレイ画面には、検索過程における各種の情報や、検索の結果である文献情報を表示する。

画面の様式は図4-57のとおりである。

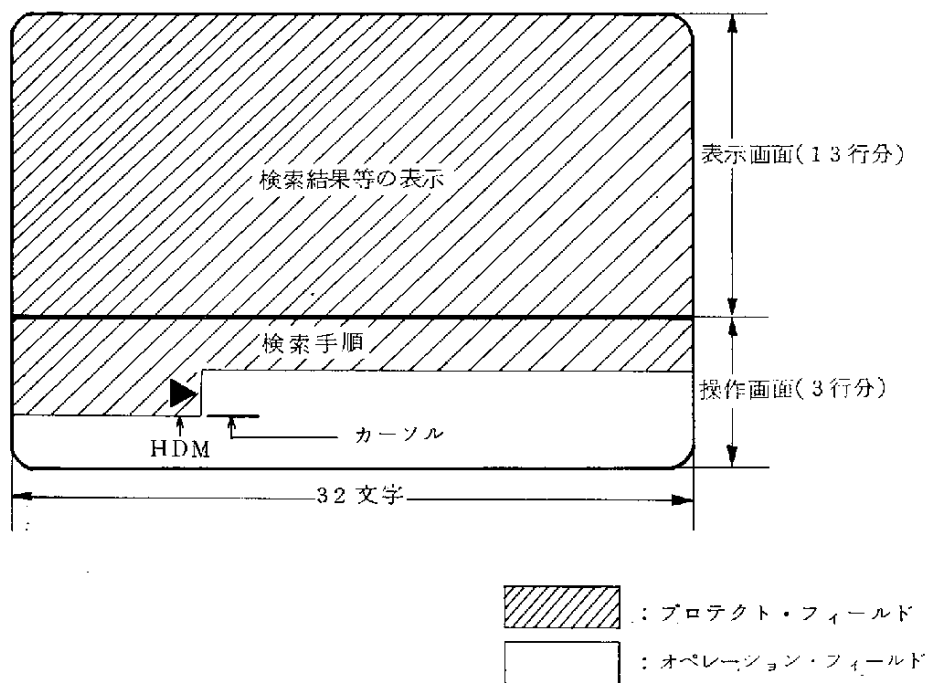


図4-57 画面の様式

。表示画面

この部分は、検索の中間情報や文献情報の表示を行なう部分である。
また、エラー・メッセージもこの部分の最下行(13行目)に表示される。

。操作画面

この部分は、主として検索に関するデータを入力するための画面である。なお、この始めの部分には、検索手順が示される。

ディスプレイ画面のうちプロテクト・フィールドが指定されている部分には、文字の入力操作はできず、オペレーション・フィールドに限って可能である。

(2) 漢字プリンタ用紙

漢字プリンタには、検索した文献情報を出力する。

出力用紙は、A4サイズの大きさに、ステップ間隔は6行/インチである。原則として用紙1枚に2文献収めるが、前の文献が長い場合は1文献のこともある。

出力様式は図4-58のとおりである。

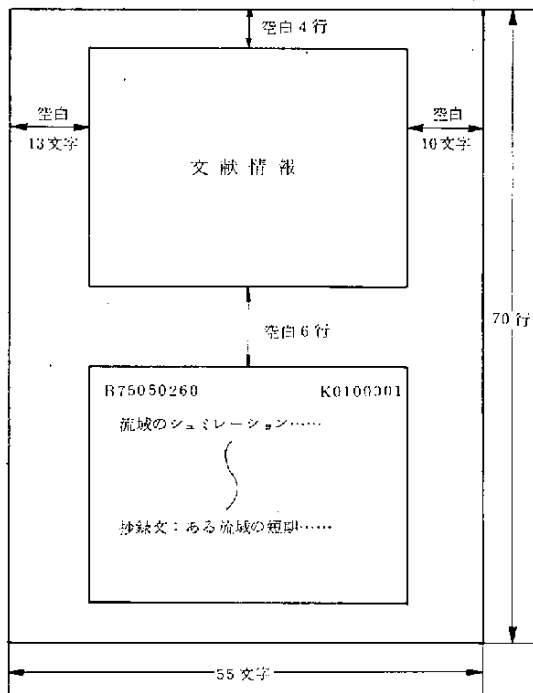


図4-58 出力用紙の様式

4.6.5 検索コマンド

(1) 検索コマンドの種類

検索コマンドは全部で9個用意されており、その種類と使用法を表4-16に示す。

表4-16 検索コマンドの使用法

処理	検索コマンド名	入力指示語	説明
	会 話 開 始		検索サービスの開始をサブシステムに知らせる
キーワード処理	キーワード入力	検索キーワード	検索に必要なキーワードを入力する。
	保 存 語 指 定	保 存 語 番 号	キーワード群から保存したいものを指定する。
検索処理	検 索 式	検 索 式	選択したキーワードを使い論理式を作成する。
	条 件 式	条 件 式	検索式で得た結果に条件を与える。
出力処理	出 力	出 力 様 式	検索された文献を表示（出力）する。
	ハード・コピー		ディスプレイ装置に表示されている文献を漢字プリンタに出力する。
	会 話 終 了		検索サービスの終了をサブシステムに知らせる。
	継 続		ディスプレイ画面の続きを表示する。
		出 力 方 法	出力、ハード・コピーのサブ・コマンドである。

(2) 検索コマンドの配列

検索コマンドはキーボード・アダプタのプログラム・ファンクション・キー上にあり、配列位置は図4-59のとおりである。

会 話 開 始			
キーワー ド入力	保存語 指 定	検 索 式	条 件 式
出 力	ハ ー ド コ ピ ー		
継 続			会 話 終 了

図 4 - 5 9 検索コマンドの配列

(3) 検索コマンドの入力方法

検索コマンドの入力方法は、キーボード・アダプタ上に配置したプログラム・ファンクション・キーを押すことにより行なわれる。また、システムがパラメータを要求する場合は、操作画面上に入力指示語，HDM(▶印)，カーソル()が示される。

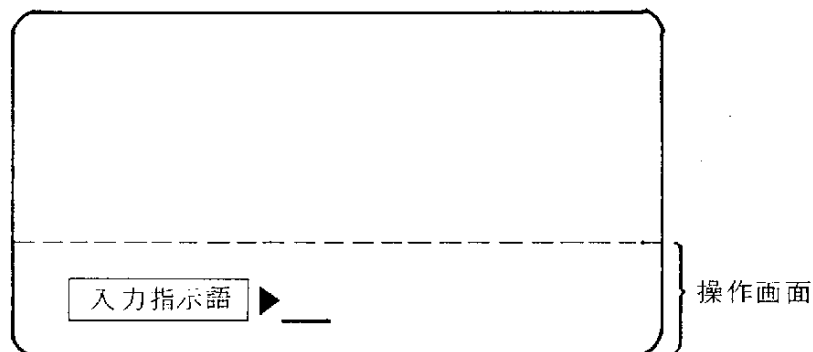


図 4 - 6 0 検索情報の入力位置

4.6.6 データ入力方法

(1) 入力文字種

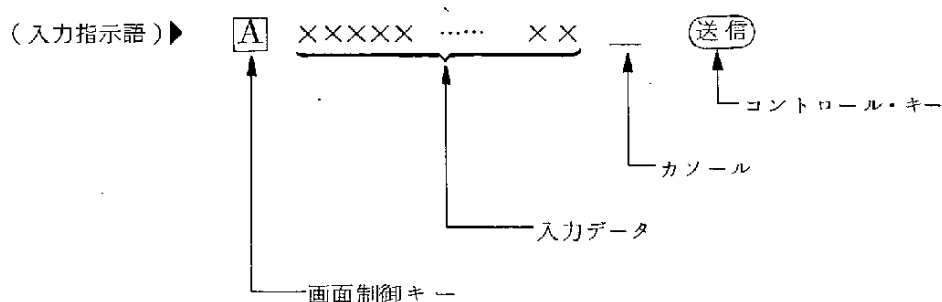
利用者が文献検索サービスを受ける時に、端末のキーボードを使ってデータ入力が行なえる文字種を、入力操作するキー別に示したのが表4-17である。

表4-17 データ入力文字種

入力キー	文 字	備 考
テン・キー	0～9 ・	数字およびピリオドはテン・キー入力が望しい。
キーボード	ア～ン, ^, , ° 0～9 +, -, *, .	キーボード・アダプタにある画面制御キーの A キーを押してから入力する。

(2) データ入力形式

入力データは、次の形式で入力する。



(3) キーボード文字の入力方法

キーボード上の文字は、2ストローク方式によって入力される。すなわち、キーボード上の入力したい文字がある鍵盤を押すと同時に、セレクション・キーの、文字位置に対応する番号を押す。操作例を図4-61に示す。

セレクトジョン・キー

①	⑤
②	⑥
③	⑦
④	⑧

「エ」
押す 押す

鍵盤

ア	オ
イ	カ
ウ	キ
エ	ク

図 4 - 6 1 キーボード文字の操作

(4) 入力における注意事項

- 文字キーをはじめ、各種キーを押すときは、

KEY IN

 ランプが点灯していなければならない。
- 書体キーは

明朝

 とする。
- 入力データの送信は、画面制御キーの

送信

 キーを用いる。
- 入力文字の訂正は、カソール移動キー

↑

↓

←

→

 で、訂正したい文字の下までカソールを動かし、正しい文字を入力する。
- チェック・コード間隔ダイヤルは2つとも0にセットしておく

漢字キーボード

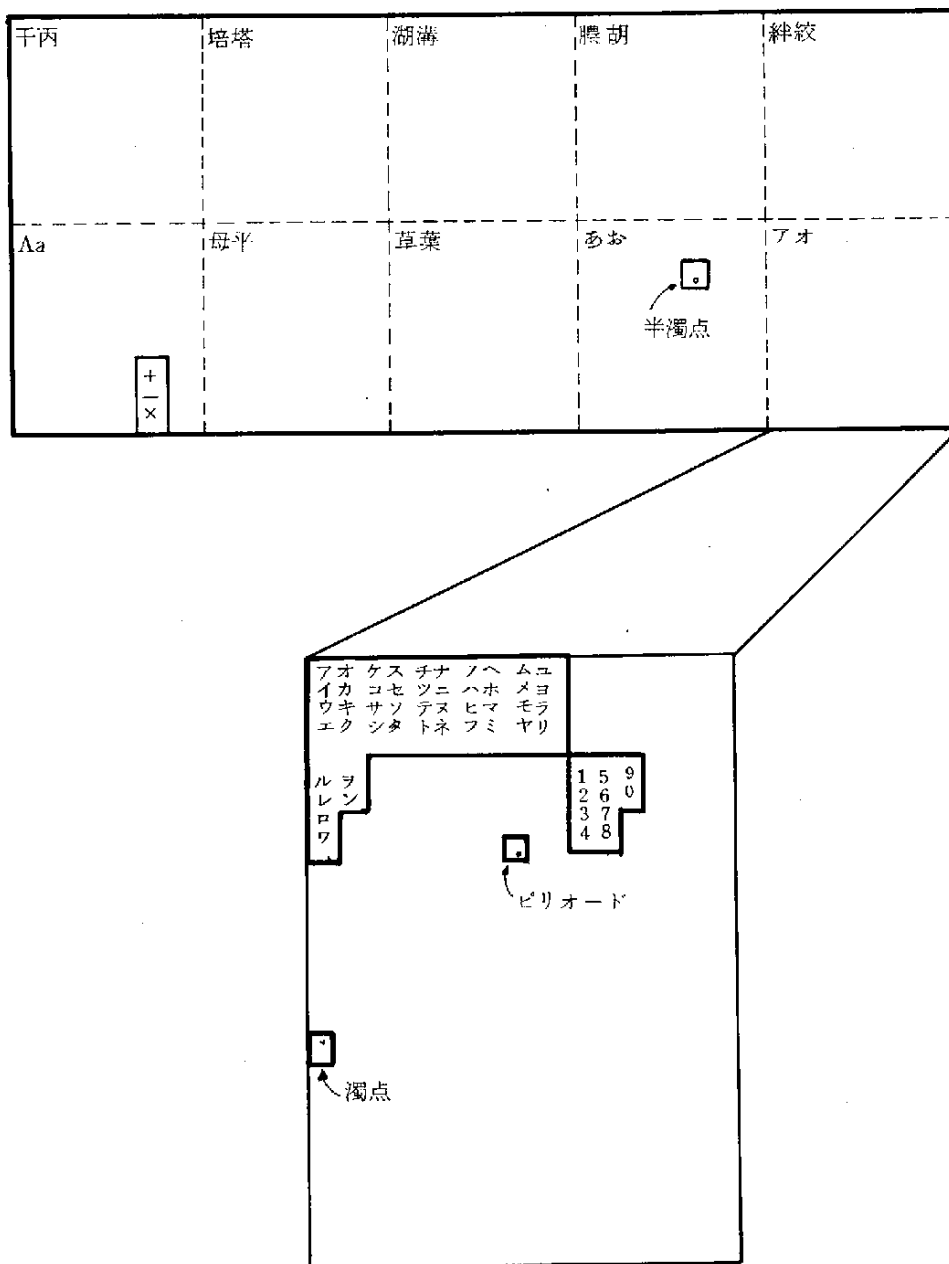


図4-62 漢字キーボードの文字位置

4.6.7 検索の方法

(1) 会話の流れ

検索作業におけるシステムと利用者との会話は、次の図4-63のように一定の流れに従って行なわれる。検索処理の流れの中で、その主流からはずれるようなときは、検索コマンドを押すことにより行なわれ、通常のやりとりは、ディスプレイ画面に表示された検索手順によって会話が進められる。

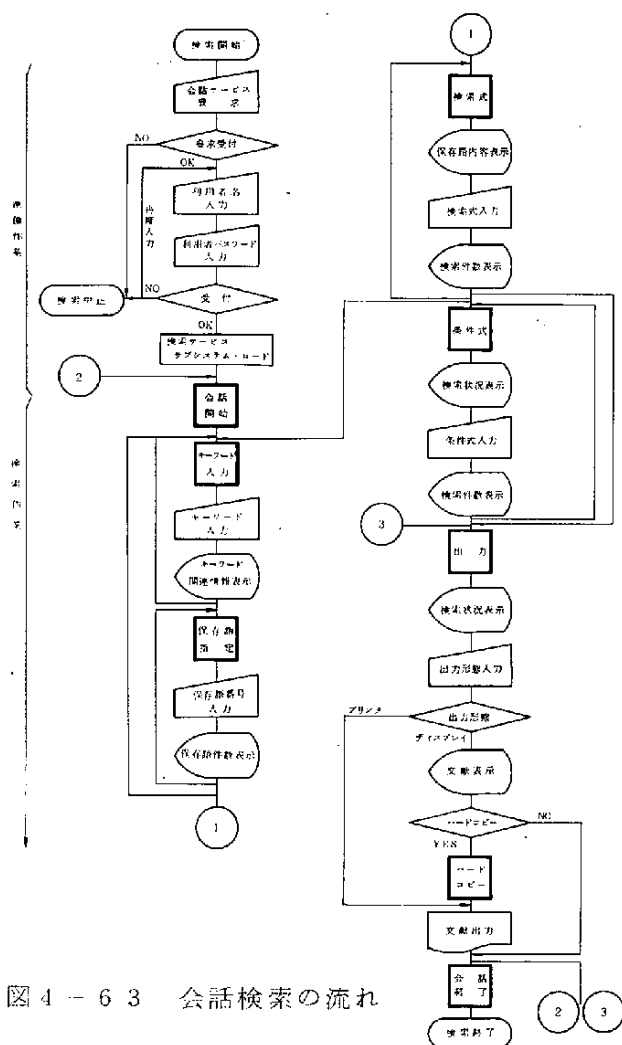


図4-63 会話検索の流れ

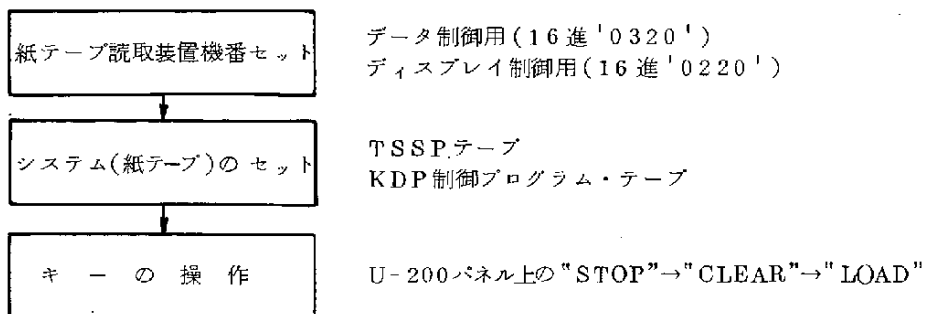
(2) 検索サービス・サブシステムの起動

① U-200のシステム・ロード

検索サービス・サブシステムの起動に先立ち、端末処理装置であるU-200に、制御システム（紙テープ提供）をローディングする必要がある。端末制御用のシステムは、データ制御用に「TSSP」を、漢字ディスプレイ制御用に「KDP制御プログラム」をローディングする。

。ローディング手順

端末制御用のシステムのローディングは次の手順で行なう。



注) 各システム・テープをロードする前に、U-200のメモリを
"PCLEAR"テープでイニシャル・セットする。手順は上と
同じ。

② 検索サービス・サブシステムの呼出し

検索サービス・サブシステムの呼出しは、端末側のコンソール・タイ
プライタから行なう。

操作手順は図4-64のとおりである。

利用 者 入 力	シ ス テ ム 応 答
REQ キー押下	***
" /C " を入力	ENTERランプ点灯
" ¥¥CPS " を入力	メッセージ出力
" 利用者名 " を入力	メッセージ出力
" パスワード " を入力	メッセージ出力
	" # " 出力後、入力待ちとなる
" CALL△・P/KINDS " を入力	メッセージ出力

図 4 - 6 4 検索システムの呼出し手順

(3) 検索コマンドの使い方

① 会話開始コマンド

。機 能

このコマンドを押すことにより、検索サービス・サブシステムの検索処理が開始され、各種のデータやテーブル類の初期設定および各ファイルの開設が行なわれる。

。パラメータの説明

な し

。画面出力例

付録を参照。(以下のコマンドについても画面出力例は付録参照)

画面番号：①

② キーワード入力コマンド

入力指示語	HDM	パラメータ
検索キーワード	▶	{ キーワード $\left[\begin{array}{c} \cdot 0 \\ \cdot 1 \end{array} \right]$ }

。機 能

文献データの検索キーワードを入力することにより、関連するキー

ワード情報を前方一致あるいはシソーラスにより検索する。

。パラメータの説明

キーワード：文献検索を行なうためのキーとなる語を指定する。

カタカナで最大40文字である。

1回の操作で入力できるキーワードの数は1語である。

- 。0：前方一致によるキーワード検索を行なう。
- 。1：シソーラスによるキーワード検索を行なう。

。入力例

検索キーワード▶ジョウハウショリ・1  ←コントロール・キー
↑
カーソル

。画面出力例

画面番号：②③④

③ 保存語指定コマンド

入力指示語	HDM	パラメータ
保存語番号	▶	$\left\{ \begin{array}{l} \text{D. 語番号-1 [. 語番号-2 [. ……]]} \\ \text{S. 語番号-1 [. 語番号-2 [. ……]]} \\ \text{A} \end{array} \right\}$

。機能

検索キーワードおよび関連キーワードの中から、指定されたキーワードを保存語テーブルに登録する。

。パラメータの説明

D：DはDELETE(消去)の意味で、指定した語番号以外のキーワードを保存する。

S：SはSAVE(保存)の意味で、指定した語番号のキーワードを保存する。

A：AはALL(全部)の意味で、表示されたすべてのキーワードを制限数内で保存する。

語番号：ここでいう語番号とは検索表示されたキーワードにシークン

なお、保存語テーブルの保存制限数は30語である。

・入力例

○画面出力例

画面番号：⑤ ⑥ ⑦

④ 検索式コマンド

入力指示語	HDM	パラメータ
検 索 式	▶	$\left\{ \begin{array}{l} K_{nn} \cdot \text{論理式} \\ L_{nn} \cdot \text{論理式} \end{array} \right\}$

機能

保存語番号と演算子によって作成された論理式を解析し、それによつて文献を検索する。

○パラメータの説明

Knn : 保存語番号を使って論理式を作成する場合の検索式番号である。nn は 01 ~ 05 である。

Lnn : 保存語番号およびKnn の検索式を組合せて論理式を作成する場合の検索式番号である。nn は01～05である。

論理式：論理演算子（AND：* OR：+ NOT：-）と論理演算項目で構成されている。

Knn の論理演算項目は保存語番号である。

L_{nn} の論理演算項目は検索式 K_{nn} と保存語番号である。

○ 入力例

。画面出力例

画面番号：⑧ ⑨

⑤ 条件式コマンド

入力指示語	HDM	パラメータ
条 件 式	▶	{ J nn . 論理式 [. 発行年 [. 発行国 [. 記述言語 [. 記事区分]]] }

。機 能

検索式番号を使って作成された論理式を解析し、それによって文献を検索し、さらに条件項目の指定があれば、条件下による検索を行なう。

。パラメータの説明

J nn : 条件式番号である。 nn は 0 1 ~ 0 3 である。

論理式: 論理演算子 (AND : * OR : + NOT : -) と論理演算項目で構成されている。

J nn の論理演算項目は検索式番号 (K nn , L nn) である。

発行年: 文献の発行年で、西暦の下 2 桁である。(7 6 → 1 9 7 6 年)

発行国: 文献の発行された国で、英字で 3 桁のコードである。

(BEL → ベルギー)

記述言語: 文献が書れている言語で、英字で 2 桁のコードである。

(EN → 英語)

記事区分: 文献の種類であり、英数字 2 桁で分類される。

(A 1 → 論文)

。入力例

条件式 ▶ J 0 1 . K 0 1 * K 0 2 . . . EN 送信 ← コントロール・キー

カーソル

発行年, 発行国, 記事区分を省略した例。

省略の場合はその分ピリオド (.) を送る。

。画面出力例

画面番号：⑩ ⑪

⑥ 出力コマンド

入力指示語	HDM	パラメータ
出力様式	▶	[出力形態・式番号 [．nnnn [．n'n'n'n' $\left[\begin{smallmatrix} \cdot 0 \\ \cdot 1 \end{smallmatrix} \right]]]]$
出力方法	▶	出力方法

。機能

検索された文献を指定された装置上に表示（出力）する。このコマンドはサブコマンドを有する。

。パラメータの説明

出力形態：文献を表示（出力）する装置を指定する。

0 → 漢字ディスプレイ装置

1 → 漢字プリンタ装置

式 番 号：検索式番号または条件式番号である。

nnnn：最大4桁の数字で何件目の文献から出力させるかを指定する。指定しなかったときは最初から行なう。

n'n'n'n'：最大4桁の数字で何件出力させるかを指定する。

指定しなかったときは、nnnn以降すべての文献を出力する。

●0： 原文標題，資料名，巻，号 は表示しない。

●1： 原文標題，資料名，巻，号 を含めて表示する。

出力方法：0 → SYSOUT経由による方法

1 → 直接出力

。入力例

出力様式 ▶ 0 . J 0 1 . . 1 0 (送信) ← コントロール・キー
 ↑
 カースル
 省略した場合、その分ピリオド(.)を送る。

この例は J 0 1 に検索されているデータの先頭から 1 0 件を漢字ディスプレイ装置に表示させる例である。

。画面出力例

画面番号：⑫ ⑬ ⑭ ⑰

⑦ ハード・コピー・コマンド

入力指示語	HDM	パラメータ
出力方法	▶	出力方法

。機能

漢字ディスプレイ装置に表示されている文献を出力方法に従って出力する。このコマンドはサブコマンドを有する。

。パラメータの説明

出力方法：0 → SYSOUT 経由による方法

1 → 直接出力

。入力例

出力方法 ▶ 1 (送信) ← コントロール・キー
 ↑
 カースル

。画面出力例

画面番号：⑮ ⑯

⑧ 継続コマンド

。機能

現在表示されているディスプレイ画面の続きを表示するコマンドであり、コマンドが入力された時点で画面が変わる。

。パラメータの説明

な し

⑨ 会話終了コマンド

。機 能

このコマンドを押すことにより，検索サービスが終了し，使用した
各種ファイルがクローズされる。

。パラメータの説明

な し

。画面出力例

画面番号：⑱

4.6.8 エラー・メッセージ一覧表

表 4-18 エラー・メッセージ一覧表

検 索 コマンド	メッセージ 番 号	メ ッ セ ー ジ	処 置
		検索コマンドの操作に誤りがある。	操作手順に指定された検索コマンドを押す。
キ ー ワ ー ド 入 力		入力したキーワードの文字に誤りがある。	カタカナで入力する。
		入力したキーワードが登録されていない。	他のキーワードを入力する。
		入力したキーワードの選択子に誤りがある。	正しい選択子を指定する。
保 指 存 定		入力した語番号に誤りがある。	正しい語番号を入力する。
		キーワードが保存できない。	保存語数が規定値になった。
検 索 式		検索式の右辺に誤りがある。	正しい式を入力する。
		検索式の左辺に誤りがある。	正しい検索式番号を入力する。
		検索式の右辺に使った番号がおかしい。	番号が登録されていないので正しい番号を入れる。
条 件 式		条件式の右辺に誤りがある。	正しい式を入力する。
		条件式の左辺に誤りがある。	正しい条件式番号を入力する。
		条件式の右辺に使った番号がおかしい。	番号が登録されていないので正しい番号を入れる。
		条件項目に誤りがある。	正しい条件項目を指定する。
出 力		出力形態の指定に誤りがある。	正しい出力形態を指定する。
		式番号に誤りがある。	正しい式番号を指定する。
		出力件数指定に誤りがある。	正しい出力位置、件数を指定する。
		出力方法に誤りがある。	正しい出力方法を指定する。
		指定された式番号の文献件数が0である。	出力する文献情報がない。
		システムにエラーがあるので処理を中止する。	システム作成者に連絡する。
		入力の際、“A” キーを押さなかった。	画面制御の <u>A</u> キーを押して再度入力する。

4.7 サービス・プログラムの概要

4.7.1 TSSP

TSSP (Terminal Station System Program) はFACOM U-200の端末に用意された端末制御プログラムで、センターと端末間のデータ・リンクを確立する。その主たる用途はリモート・バッチ処理、デマンド処理である。

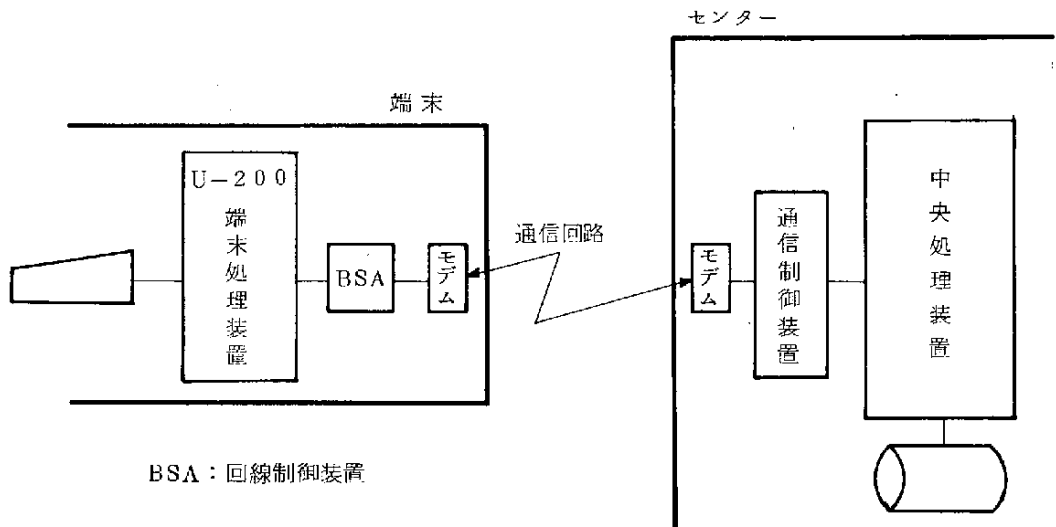


図4-65 端末とセンターのつながり

TSSPの役割りは、センターから送られてきたデータを受信し、端末の各装置に出力することと、端末の入力装置から入力されたデータを、センターに送信することである。

送受信されるデータはメッセージとして一定の形式を持っている。これは、ヘッディング情報およびテキストから成っており、ヘッディング情報には装置割当て、データ・モード、ブロック情報等が含まれており、テキストには入力したデータに伝送制御文字を付加した形で構成される。

4.7.2 ETAM

ETAM (Extended Telecommunication Access Method) は、富士通 OS (M-6/7) のサブモニタの 1 つである。ETAM はセンターにつながっている端末に対して、各種のサービスを行なうことができる。

主な特徴としては、

- 端末とのデータ送受信は、順編成アクセス・マクロと同一インタフェースで行なうことができる。
- 入力待ちのオンライン処理プログラムをロールアウトすることにより、主記憶の有効利用をはかることができる。
- 端末とオンライン処理プログラムは ETAM で管理するメッセージ・ファイルを介して結合される。
- COBOL 言語でも複数の端末を同時並行処理するようなオンライン処理プログラムを作ることができる。

などがあげられる。

オンライン処理プログラムは、メッセージ・ファイルに対してリード・ライト命令をかけることにより、センターと端末間のデータの送受信ができる。

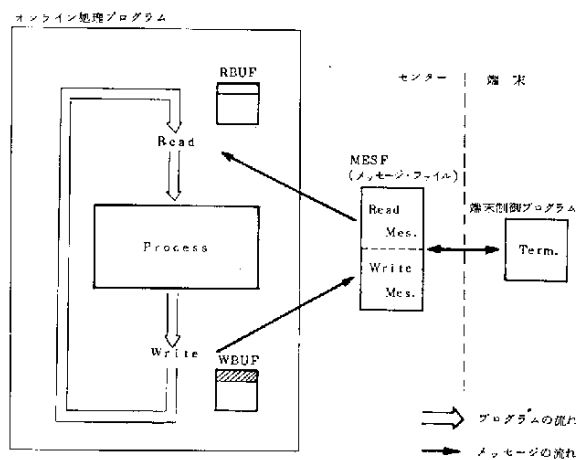


図 4-66 センター・端末間の送受信

端末から入力されたメッセージ（データ）は、一度メッセージ・ファイルにスタックされる。このメッセージはオンライン処理プログラムからリード命令が出されたとき、バッファに転送される。また、出力メッセージはプログラムのライト命令によりメッセージ・ファイルにスタックされ、そのあと端末に出力される。

4.7.3 KDP制御プログラム

KDP制御プログラムは、漢字ディスプレイ制御用処理装置FACOM U-200にローディングされ、上位の端末制御用処理装置U-200と漢字ディスプレイ間のデータ送受信の制御を行なうが、本来漢字ディスプレイ・キーボードとともにハードウェアの一部としてファームウェアとみなされるものである。

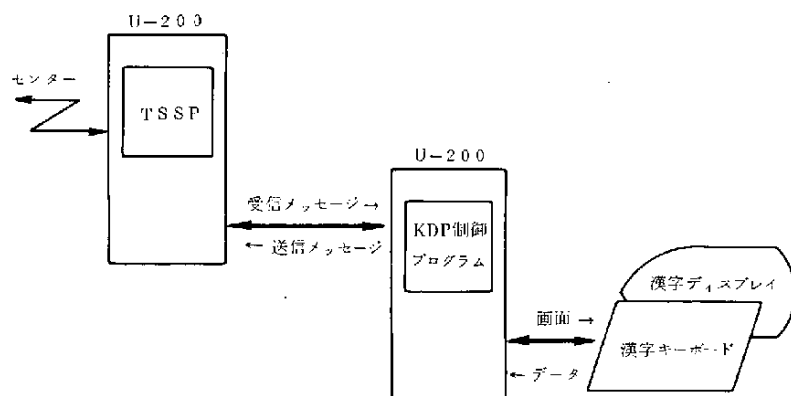


図 4-67 漢字ディスプレイのデータの流れ

端末制御プログラム（TSSP）で受けたセンターからのメッセージはヘッダ情報をもとに漢字ディスプレイ用のデータ・フォーマットに変換されてKDP制御プログラムに送られ、漢字ディスプレイに表示される。また、漢字キーボードから入力されたデータは、KDP制御プログラムによって、TSSPで扱えるデータ・フォーマットに変換したうえで、上位に送られる。

4.8 付 録

(1) 表示画面形式一覽

① 会話開始釦を押したとき

[illegible]

② キーワード入力釦を押したとき

キーワードを40字以内のカタカナで入力して下さい。

検索キーワード▶

③ キーワードを入力したあと

[illegible]

④ 継続釦を押したとき

[illegible]

⑤ 保存語指定釦を押したとき

[illegible]

⑥ 継続釦を押したとき

[illegible]

⑦ 保存語番号を指定したとき

保存語は 件登録されました。

⑧ 検索式作成釦を押したとき

[illegible]

⑨ 検索式を入力したとき

[illegible]

⑩ 条件式釦を押したとき

[illegible]

(11) 条件式を入力したとき

[illegible]

(12) 出力釦を押したとき

[illegible]

(13) 出力形態にディスプレイを指定したとき

B75050260	領域のシミュレーションモデルの開発のため、の字際的アプローチ	K01000001
流	Ant interdisciplinary water shed simulation development models.	
	記事区分：解説的記事②	
	使用言語：英語	
	発行者国名：米国	
	著者資	
	資料名：WALTERS, T. C. H. Vol. 6	
	：H572A, Tech. 6	
	Soc. Change	
	【3】	
	(発行年：1974)	
継		
▷	継続・ヘードコピ一鉛のいずれかを押して下されい。な、おし、中、途、打、切	
り	をす・と、は、会、話、開、始、の、鉛、の、会、話、を、了、押、し、の、い、を、押、お、し、下、さ、い、	

⑭ 継続釦を押したとき

B	7抄学の経。がブにル	5録のよ濟こ必の対開	0文シう学の要成す発	5：ミな的よと果るの	0ユモなうなとしつ一般	2あレデ予名を。ての一般的	6ある一測目的を得シア	0流シハ含めにころミア		域ヨマ含にころミア	のングめはでれュロ	短モノ包多はたレー	期デジ括くカ大一チ	おルメ的に研大洋論	よガンに研究の岸ンじ	び使ト扱究の岸ンじ	長わ単り分多北モら	期れ位と野分西デれ	のるとこの野部ルテ	フコしいう知のとかい	ロとして知識研ケ絶る。	一が一かを究ベ介。	バ多流ら総者ッさ	Kーなを全合にけれ	タク域は台にけれ	オンっ生くにけるゾと	1のて物不活ワエと	0調い学完用一一も	0査る的全すクムに	0にがおでるンス、	0水、よあこヨ灣モ	0理こびるとン域デ
▷	継続する	鈕す	・	ハと	一き	ド会	ピ話	一開	鈕始	の鈕	いの	ず会	れ話	か終	了	押鈕	しの	下す	され	いか	を押	をおし	中下	途さ	打い。	切						

(15) ハードコピー鉤あるいは出力形態としてプリンタを指定したとき

▷出力方法を指示して下さい。	出力 = 0	直接出力 = 1
出力方法 ▶		

(16) 直接出力を指定したとき

ただいま文献抄録を漢字プリンタへ出力しています。

次の会話までしばらくお待ち下さい。

(17) 指定された文献がすべて出力されたとき

文献出力を終わります。

出力文献数 件

▷ 出力釦・会話開始釦・会話終了釦のいずれかを押して下さい。

⑬ 会話終了釦を押したとき

[illegible]

(2) 保存語テーブル

保 存 語 番 号 (2) E		頻 度 (2) B		キ ー ワ ー ド (漢 字) (80)40 文字 B	ポ ス テ ィ ン グ ・ テ ー ブ ル ア ド レ ス (T T R S) (4) B			
30 {				}}				

保存語番号 (2 バイト)

保存されたキーワードに付けられたシーケンス番号。

EBCDIC コード 01~30

頻度 (2 バイト)

このキーワードの対象となる文献数。

BINARY コードにより記録されているので、表示時には

2 進 → 10 進 → 漢字

の変換が必要である。

キーワード (80 バイト)

保存の指定を受けたキーワード (漢字)。

最大 40 文字。

ポスティング・テーブル・アドレス (TTRS) (4 バイト)

このキーワードの対象となる文献のアドレスを持つ、ポスティング・
テーブルの位置を TTRS の形で表わす。

テーブル・サイズ

$$(2 \text{ バイト} + 2 \text{ バイト} + 80 \text{ バイト} + 4 \text{ バイト}) \times 30 \text{ 個} = 2,640$$

バ イ ト

(3) オペレーション・テーブル

検 索 優 先 権 (1) B	検 索 結 果 格 納 エ リ ア (2) E	演 算 (1) E	オペランド			
			第 (2)	1 E	第 (2)	2 E

検索優先権(1バイト)

論理演算式を解説し、各演算子に重みを付けて算出したもの。

大 $\rightarrow$ 小の順に並ぶ。

検索結果格納エリア(2バイト)

検索された結果がどの検索エリアに格納されているかを示す。

検索結果用に用意された検索エリアは3つある。

演算 (1 バイト)

オペランドに対してどのような論理演算（論理積，論理和，論理否定）

を行なうかその演算子 ($*$, $+$, $-$) が格納されている。

オペランド

検索の対象となる各種番号（保存語番号，検索エリア番号，式保存番号）が格納されている。

検索は第1オペランドを基にして行なう。

テーブル・サイズ

(1 バイト + 2 バイト + 1 バイト + 2 バイト + 2 バイト) × 1 0 個 =
8 0 バイト

(4) 検索エリア管理テーブル

検 索 エ リ ア 番 号	空 白	ス テ ー タ ス
(2) E	(1) E	(1) B

検索エリア番号 (2 バイト)

検索エリアに付けられた番号。

4 1 ~ 4 3

空白 (1 バイト)

未使用

ステータス (1 バイト)

使用中か否かを表わす。

使用中 : 1

否 : 0

テーブル・サイズ

4 バイト × 3 個 = 1 2 バイト

(5) 検索エリア (Ⅰ, Ⅱ, Ⅲ)

250

(検索エリアⅠ, Ⅱ, Ⅲ, とも同形式)

文献アドレス(4 バイト)

データベースにおける文献データのアドレスを TTRS の形で表わす。

言語コード (1 バイト)

文献が記述されている言語を表わす。

発行年(2 バイト)

文献が発行された年を西暦年で表わす。

発行国 (1 バイト)

文献が発行された国を表わす。

記事区分(2 バイト)

記事内容を表わす。

空白 (2 バイト)

未使用

テーブル・サイズ

$$12 \text{ バイト} \times 250 \text{ 件} \times 6 \text{ (最大論理和個数)} = 18,000 \text{ バイト}$$
$$18,000 \text{ バイト} \times 3 \text{ 個 (エリア数)} = 54,000 \text{ バイト}$$

(6) 検索結果アドレス・テーブル

式保存番	件数	式番号	空白 (i) B	検索結果データ アドレス (TTRS)
(2) E	(2) B	(3) E	B	(4) B

式保存番号(2バイト)

保存した式番号（検索式，条件式）に付けられる番号で，この番号は検索の対象となるとき，オペランドとして使われる。

$$\bar{5} \quad 1 \sim 5 \quad 5$$

6 1 ~ 6 5

$$7 \ 1 \sim 7 \ 3$$

件数 (2 バイト)

検索により適合した文献の件数である。

式番号 (3 バイト)

検索式，条件式に付けられた番号である。

K 0 1 ~ K 0 5 J 0 1 ~ J 0 3 L 0 1 ~ L 0 5

空白 (1 バイト)

未使用

呼出し状況 (1 ビット)

あるコマンド内において同じコマンドが押されたときの呼出し状況である。

エラー通知 (1 ビット)

実行中の処理ブロック内で、エラーが発見されるとセットされる。

継続有無 (1 ビット)

画面表示で継続があるかないかの区分である。

プログラム・コード (4 ビット)

実行中の処理ブロック内プログラム番号である。

エラー・ステータス (1 バイト)

エラー内容がセットされる。

画面情報 (4 バイト)

文字アドレス・ポインタ

画面上のどの位置から出力するか指定 (2 バイト) である。

メッセージ長

メッセージのレコード長 (2 バイト) である。

テーブル・サイズ

8 バイト × 9 個 = 72 バイト

(8) 国名コード表

バイナリ シーケン スコード	国 名	J I C S T 国名コード	バイナリ シーケン スコード	国 名	J I C S T 国名コード
0 1	アイスランド	(I S L)	3 4	キューバ	C U B
0 2	アイボリーコースト	(C I V)	3 5	ギリシャ	G R C
0 3	アイルランド	(I R L)	3 6	ケニア	K E N
0 4	アフガニスタン	A F G	3 7	コスタリカ	C O T
0 5	アメリカ合衆国	U S A	3 8	コロンビア	C O L
0 6	アラブ共和国連邦	(E G Y)	3 9	コンゴ人民共和国	C O G
0 7	アルジェリア	(D Z A)	4 0	サウジアラビア	S A U
0 8	アルゼンチン	A R G	4 1	シェラレオネ	S L E
0 9	アルバニア	A L B	4 2	ジャマイカ	J A M
1 0	アンゴラ	(A G O)	4 3	シリア	S Y R
1 1	イエメン	Y E N	4 4	シンガポール	S G P
1 2	イギリス	G B R	4 5	スイス	(C H E)
1 3	イスラエル	(I S R)	4 6	スウェーデン	S W E
1 4	イタリア	I T A	4 7	スーダン	S D N
1 5	イラク	I R Q	4 8	スペイン	E S P
1 6	イラン	I R N	4 9	スリランカ	(L K A)
1 7	インド	I N D	5 0	セネガル	(S L K)
1 8	インドネシア	I D N	5 1	ソビエト連邦	S E N
1 9	ウガンダ	U G A	5 2	タイ	S U N
2 0	ウルグアイ	(U R Y)	5 3	台湾	T H A
2 1	エクアドル	(U R U)	5 4	タンザニア	T W N
2 2	エチオピア	E C U	5 5	チェコスロバキア	(T Z A)
2 3	エルサルバドル	E T H	5 6	中国	(T A N)
2 4	オーストラリア	S L V	5 7	チェニジア	C S K
2 5	オーストリア	A U S	5 8	チリ	C H N
2 6	オランダ	(A U T)	5 9	デンマーク	T U N
2 7	ガーナ	N L D	6 0	ドミニカ	C H L
2 8	カナダ	G H A	6 1	ドミニカ	D N K
2 9	韓国	C A N	6 2	トルコ	(D M A)
3 0	カンボジア共和国	(K O R)	6 3	ニカラグア	(D O M)
3 1	北朝鮮	(K R S)	6 4	西ドイツ	(T T O)
3 2	北ベトナム	(P R K)	6 5	日本	(T D T)
3 3	キプロス	(V N N)	6 6	ニュージーランド	T U R
		C Y P			N I C
					D E U
					J P N
					N Z L

バイナリ シーケンス スコード	国 名	J I C S T 国名コード
6 7	ノルウェー	NOR
6 8	パキスタン	PAK
6 9	パナマ	PAN
7 0	ハンガリー	HUN
7 1	東ドイツ	DDR
7 2	ビルマ	BUR
7 3	フィリピン	PHI
7 4	フィンランド	FIN
7 5	プエルトリコ	(P R I P T R)
7 6	ブラジル	BRA
7 7	フランス	FRA
7 8	ブルガリア	BGR
7 9	ベネズエラ	VEN
8 0	ペルー	PER
8 1	ベルギー	BEL
8 2	ポーランド	POL
8 3	ポルトガル	PRT
8 4	ホンコン	HKG
8 5	ホンデュラス	HND
8 6	マレーシア	MYS
8 7	南アフリカ連邦	ZAF
8 8	南ベトナム	(V N M V N S)
8 9	メキシコ	MEX
9 0	モナコ	MCO
9 1	モロッコ	MAR
9 2	ユーゴスラビア	YUG
9 3	ルーマニア	ROM
9 4	ルクセンブルグ	LUX
9 5	レバノン	LBN
9 6	ローデシア	RHO
9 7	複数国表示	INT
9 8	その他	ZZZ

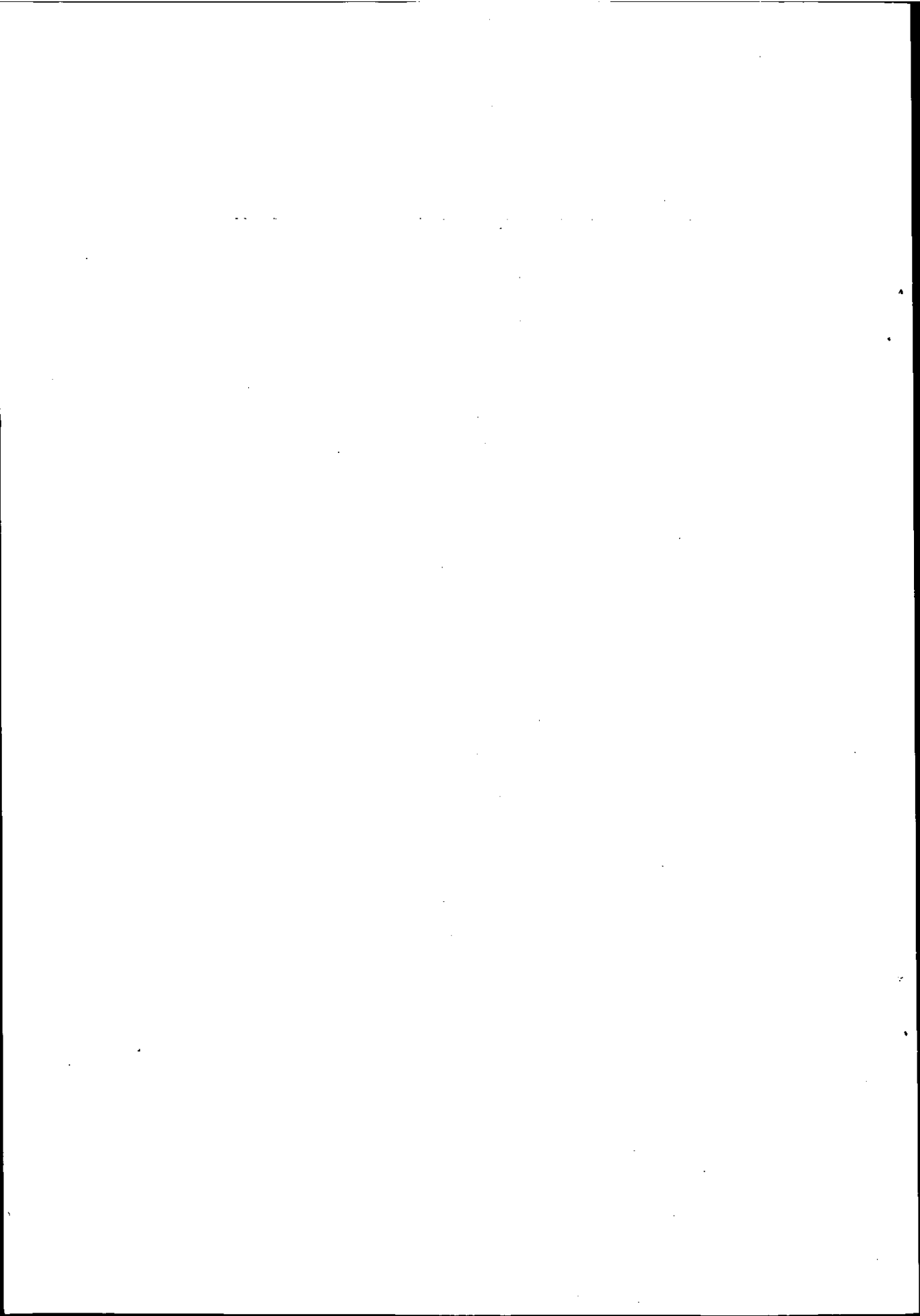
(注) 準拠する標準が、ISO/DIS
3166からISO 3166に変わった際、
コード体系が変更になるもののみに、
()内にISO/DIS 3166のコ
ードを示した。

(9) 言語コード表

バイナリ シーケン スコード	言 語	J I C S T 言語コード	バイナリ シーケン スコード	言 語	J I C S T 言語コード
0 1	アイスランド語	I O	3 1	デンマーク語	D A
0 2	アゼルバイジャン語	A Z	3 2	ドイツ語	D E
0 3	アフリカーン語	A F	3 3	トルコ語	T R
0 4	アラビア語	A R	3 4	日本語	J A
0 5	アルバニア語	A B	3 5	ノルウェー語	N O
0 6	イタリア語	I T	3 6	白ロシア語	B E
0 7	インターリングア語	I A	3 7	ハンガリア語	H U
0 8	インドネシア語	I N	3 8	パンジャブ語	P A
0 9	ウクライナ語	U K	3 9	ヒンディ語	H I
1 0	ウルドゥー語	U R	4 0	フィン語	F I
1 1	英 語	E N	4 1	フェレーズ語	F A
1 2	エストニア語	E S	4 2	フランス語	F R
1 3	エスペラント語	E O	4 3	ブルガリア語	B G
1 4	オランダ語	N L	4 4	ベトナム語	V I
1 5	ギリシャ語	G R	4 5	ヘブライ語	H E
1 6	グジャラチ語	G U	4 6	ベルシャ語	P E
1 7	グルジア語	G E	4 7	ベンガル語	B N
1 8	クロアチア語	C R	4 8	ポーランド語	P L
1 9	ゲール語	G A	4 9	ポルトガル語	P T
2 0	サンスクリット語	S A	5 0	マケドニア語	L U
2 1	スエーデン語	S V	5 1	マレー語	M A
2 2	スペイン語	S P	5 2	蒙古語	M G
2 3	スロバク語	S K	5 3	モルタビア語	M O
2 4	スロベニア語	S N	5 4	ラテン語	L A
2 5	セルビア語	S E	5 5	リトアニア語	L I
2 6	セルボクロアート語	S H	5 6	ルーマニア語	R O
2 7	タイ語	T H	5 7	ロシア語	R U
2 8	チェコ語	C S			
2 9	中国語	C H	5 8	多言語	M L
3 0	朝鮮語	K O	5 9	その他	Z Z

(10) 記事区分分類表 (JICST使用の区分)

記事区分コード	内 容
a	原著的文章
a ①	論文 発表形式が論文形式のもの
a ②	短報 論文形式をもたずに、編集者あての通信等簡略化した形式のもの
b	解説的文章
b ①	文献レビュー 関連文献資料等にもとづき、解説または評論した記事
b ②	解説 上記以外の解説的文章
c	実用技術資料
c ①	データシート、フローシート、ノモグラム等
c ②	安全基準、規格、仕様書、作業規準等
d	一般的記事
d ①	科学技術に関する方策
d ②	学会予稿集等
d ③	紹介的文章
p	特許
p ①	日本公告特許
p ②	特許制度



(12) カナ・モード・データ・カード形式

記事番号	年・月・日	地域区分	発着時刻	発着時刻		記事番号	電	セパレート	電	最終セパレート
				前	後					
100	(2)	(6)	(1)	(2)	(2)	(2)	(max 3.0)	(2)	(max 2.0)	(2)

一連番	作反月	作反如	分型区
-----	-----	-----	-----

主 取組区分はデ 4(※各ごとにシーケンス番号(001～999)を付与)

記事番号	年 月 日	記 述 区 分	セ レ ク ト	最 終 セ レ ク ト
101	12月	101	101	101

0.2	——	著者名	フリガナ	所属機関名
0.3	——	和文標題		
0.4	——	原文標題		
0.5	——	資料題名		
0.6	——	抄録文		
0.7	——	キーワード		

(13) 漢字モード・データ・カード形式 (EBCDIC/BINARY項目)

記事番号	発行国	雑誌名	巻数	発行年	発行部	発行部	記事区分	港 セパレータ シム 号	号	付録セパレータ	
000	01	02	03	04	05	06	07	(0000-10)	08	(00000-30)	09

11	2	2	1
分野区	作成年	作成月	速需具

※林区分はデ、*区分(011)内は、カラム右足の場合 ケース番号(001~999)を付して記述デ、*とする

(14) 漢字モード・データ・紙テープ形式 (漢字モード項目)

記 事 欄	第 一 分 冊	著 者 氏 名	ナ リ ノ チ	所 属 機 関 名	和 文 協 会	第 二 分 冊
00	2					3

第 一 分 部	第 二 分 部	第 三 分 部	第 四 分 部	第 五 分 部	第 六 分 部	第 七 分 部	第 八 分 部	第 九 分 部	第 十 分 部	第 十一 分 部	第 十二 分 部	第 十三 分 部	第 十四 分 部	第 十五 分 部	第 十六 分 部	第 十七 分 部	第 十八 分 部	第 十九 分 部	第 二十 分 部	第 二十一 分 部	第 二十二 分 部	第 二十三 分 部	第 二十四 分 部	第 二十五 分 部	第 二十六 分 部	第 二十七 分 部	第 二十八 分 部	第 二十九 分 部	第 三十 分 部	第 三十一 分 部	第 三十二 分 部	第 三十三 分 部	第 三十四 分 部	第 三十五 分 部	第 三十六 分 部	第 三十七 分 部	第 三十八 分 部	第 三十九 分 部	第 四十 分 部	第 四十一 分 部	第 四十二 分 部	第 四十三 分 部	第 四十四 分 部	第 四十五 分 部	第 四十六 分 部	第 四十七 分 部	第 四十八 分 部	第 四十九 分 部	第 五十 分 部	第 五十一 分 部	第 五十二 分 部	第 五十三 分 部	第 五十四 分 部	第 五十五 分 部	第 五十六 分 部	第 五十七 分 部	第 五十八 分 部	第 五十九 分 部	第 六十 分 部	第 六十一 分 部	第 六十二 分 部	第 六十三 分 部	第 六十四 分 部	第 六十五 分 部	第 六十六 分 部	第 六十七 分 部	第 六十八 分 部	第 六十九 分 部	第 七十 分 部	第 七十一 分 部	第 七十二 分 部	第 七十三 分 部	第 七十四 分 部	第 七十五 分 部	第 七十六 分 部	第 七十七 分 部	第 七十八 分 部	第 七十九 分 部	第 八十 分 部	第 八十一 分 部	第 八十二 分 部	第 八十三 分 部	第 八十四 分 部	第 八十五 分 部	第 八十六 分 部	第 八十七 分 部	第 八十八 分 部	第 八十九 分 部	第 九十 分 部	第 九十一 分 部	第 九十二 分 部	第 九十三 分 部	第 九十四 分 部	第 九十五 分 部	第 九十六 分 部	第 九十七 分 部	第 九十八 分 部	第 九十九 分 部	第 一百 分 部
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

	$p = 12 \quad f = 1$	$p = 17 \quad f = 2$	$p = 19 \quad f = 3$	$p = 23 \quad f = 4$	$p = 29 \quad f = 5$	$p = 37 \quad f = 6$	$p = 47 \quad f = 7$	$p = 59 \quad f = 8$	$p = 71 \quad f = 9$	$p = 83 \quad f = 10$	$p = 97 \quad f = 11$	$p = 113 \quad f = 12$	$p = 131 \quad f = 13$	$p = 151 \quad f = 14$	$p = 173 \quad f = 15$	$p = 197 \quad f = 16$	$p = 223 \quad f = 17$	$p = 251 \quad f = 18$	$p = 281 \quad f = 19$	$p = 313 \quad f = 20$	$p = 347 \quad f = 21$	$p = 383 \quad f = 22$	$p = 421 \quad f = 23$	$p = 461 \quad f = 24$	$p = 503 \quad f = 25$	$p = 547 \quad f = 26$	$p = 593 \quad f = 27$	$p = 641 \quad f = 28$	$p = 691 \quad f = 29$	$p = 743 \quad f = 30$	$p = 797 \quad f = 31$	$p = 853 \quad f = 32$	$p = 911 \quad f = 33$	$p = 971 \quad f = 34$	$p = 1033 \quad f = 35$	$p = 1097 \quad f = 36$	$p = 1163 \quad f = 37$	$p = 1231 \quad f = 38$	$p = 1301 \quad f = 39$	$p = 1373 \quad f = 40$	$p = 1447 \quad f = 41$	$p = 1523 \quad f = 42$	$p = 1601 \quad f = 43$	$p = 1681 \quad f = 44$	$p = 1763 \quad f = 45$	$p = 1847 \quad f = 46$	$p = 1933 \quad f = 47$	$p = 2021 \quad f = 48$	$p = 2111 \quad f = 49$	$p = 2203 \quad f = 50$	$p = 2297 \quad f = 51$	$p = 2393 \quad f = 52$	$p = 2491 \quad f = 53$	$p = 2591 \quad f = 54$	$p = 2693 \quad f = 55$	$p = 2797 \quad f = 56$	$p = 2903 \quad f = 57$	$p = 3011 \quad f = 58$	$p = 3121 \quad f = 59$	$p = 3233 \quad f = 60$	$p = 3347 \quad f = 61$	$p = 3463 \quad f = 62$	$p = 3581 \quad f = 63$	$p = 3701 \quad f = 64$	$p = 3823 \quad f = 65$	$p = 3947 \quad f = 66$	$p = 4073 \quad f = 67$	$p = 4201 \quad f = 68$	$p = 4331 \quad f = 69$	$p = 4463 \quad f = 70$	$p = 4597 \quad f = 71$	$p = 4733 \quad f = 72$	$p = 4871 \quad f = 73$	$p = 5011 \quad f = 74$	$p = 5153 \quad f = 75$	$p = 5297 \quad f = 76$	$p = 5443 \quad f = 77$	$p = 5591 \quad f = 78$	$p = 5741 \quad f = 79$	$p = 5893 \quad f = 80$	$p = 6047 \quad f = 81$	$p = 6203 \quad f = 82$	$p = 6361 \quad f = 83$	$p = 6521 \quad f = 84$	$p = 6683 \quad f = 85$	$p = 6847 \quad f = 86$	$p = 7013 \quad f = 87$	$p = 7181 \quad f = 88$	$p = 7351 \quad f = 89$	$p = 7523 \quad f = 90$	$p = 7697 \quad f = 91$	$p = 7873 \quad f = 92$	$p = 8051 \quad f = 93$	$p = 8231 \quad f = 94$	$p = 8413 \quad f = 95$	$p = 8597 \quad f = 96$	$p = 8783 \quad f = 97$	$p = 8971 \quad f = 98$	$p = 9161 \quad f = 99$	$p = 9353 \quad f = 100$	$p = 9547 \quad f = 101$	$p = 9743 \quad f = 102$	$p = 9941 \quad f = 103$	$p = 10141 \quad f = 104$	$p = 10343 \quad f = 105$	$p = 10547 \quad f = 106$	$p = 10753 \quad f = 107$	$p = 10961 \quad f = 108$	$p = 11171 \quad f = 109$	$p = 11383 \quad f = 110$	$p = 11597 \quad f = 111$	$p = 11813 \quad f = 112$	$p = 12031 \quad f = 113$	$p = 12251 \quad f = 114$	$p = 12473 \quad f = 115$	$p = 12697 \quad f = 116$	$p = 12923 \quad f = 117$	$p = 13151 \quad f = 118$	$p = 13381 \quad f = 119$	$p = 13613 \quad f = 120$	$p = 13847 \quad f = 121$	$p = 14083 \quad f = 122$	$p = 14321 \quad f = 123$	$p = 14561 \quad f = 124$	$p = 14803 \quad f = 125$	$p = 15047 \quad f = 126$	$p = 15293 \quad f = 127$	$p = 15541 \quad f = 128$	$p = 15791 \quad f = 129$	$p = 16043 \quad f = 130$	$p = 16297 \quad f = 131$	$p = 165$
--	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	-----------------------	-----------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	-------------------------	--------------------------	--------------------------	--------------------------	--------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	---------------------------	-----------

(例: 4区画 2, 3, 4, 5, 6, 7,) は、以下でも標準 K B のアランツァンキ を使用)

(15) カナ・モード・キーワード形式

[illegible]

クハ・セ ド・キニリ ド！キ リ 0331 カト右 小幡 学電機システム内学 9作成基準に

と、それについて、

地域区分：市、町、村、市界内であるがフロンティアや、流入、分岐存在

とによって1粒の力 ドで4尾のとき1粒シーケンスを付ける。

(16) 漢字モード・キーワード紙テープ形式

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524
--

★リード・カタカナ：漢字・ワ・下のカタカナを全化（後の処理でEBCDIC変換される）

漢字モ　ワ　ド：漢字かな混りキーワードをサス引。

$\varphi = \frac{1}{2} \left(\frac{1}{\sqrt{2}} \right)$ かつ $\varphi = \frac{1}{2} \left(\frac{1}{\sqrt{2}} \right)$ を (2) 分 (ソリッド) とし、 $\varphi = \frac{1}{2} \left(\frac{1}{\sqrt{2}} \right)$ (X⁺ 8881⁺)

12 漢字キ ワードを区分 (フナノクシ ヲノコ・ド 区 分 X'84H1')

(17) キーワード削除データ形式

シェンヌ ド		前 除 ホ ー フ ー ド 出 部 号												
		1	2	3	4	5	6	7	8	9	10	11	12	13

(18) 関係リンク・データ形式

[illegible]

※ 見出し語番号・関係語番号：語番号はキーワード・ヨミカタ中に付られたシ ンケイ番号。

見出し語を主体にして関連するキーワードを関数番号で対応させる。

関係子：見出し語に対する関係語の関連性を識別する。

上位群「R」，同義群「U」のいふ部分がそれだ。

(19) 関係リンク修正データ形式

1	2	3	4	5	6
正 見 出 止 分	国 際 航 空 航 路 公 司	運 送 船 隻 数 計			

修正区分: 1 A のとき追加す - *

| D ! のとき $\Delta F_{\text{デ}} = 0$

1. M_1 のとて閉包-子集止。 $\bullet$

カ ドごとに修正が 夕は定む

(20) カナ漢字変換システムの利用方法

1. データ作成基準

- (1) 読みカナで入力する。
- (2) 漢字熟語（複合語・合成語）は原則として2字単位の単語に分ち書きする。
- (3) 分ち書きした読みカナはEBCDICコードのカタカナにして12文字以内とする。
- (4) 漢字単語の読みカナ中には長音・拗促音を禁止し、正規の読みあるいは大文字読みで入力する。
- (5) 分野用漢字単語には「井」を、固有名詞には「%」を、カタカナ単語には「*」を、それぞれの読みカナの先頭にファンクション・キーとして付加する。
- (6) カタカナ単語中には、「・（中黒）」、「ー（長音）」、「／（拗促音ファンクション・キー）」が使用できる。
- (7) ひらがなはカタカナにして入力する。
- (8) 英数字は特殊変換を除いてそのまま入力する。
- (9) ブランクを各単語間、文字記号間等のセパレータに用いる。
- (10) 特殊記号はファンクション・キーに使用されているもの以外はそのまま入力する。
- (11) ファンクション・キーに使われる記号は、原則として記号の前後にブランクを挿入する。その他は次のように使用すれば記号とみなす。

井, %, *		次の文字がカタカナ以外
/		次がカタカナ, (,) 以外
!		次が「オ」以外
\$		次が英字以外
>		次が数字以外

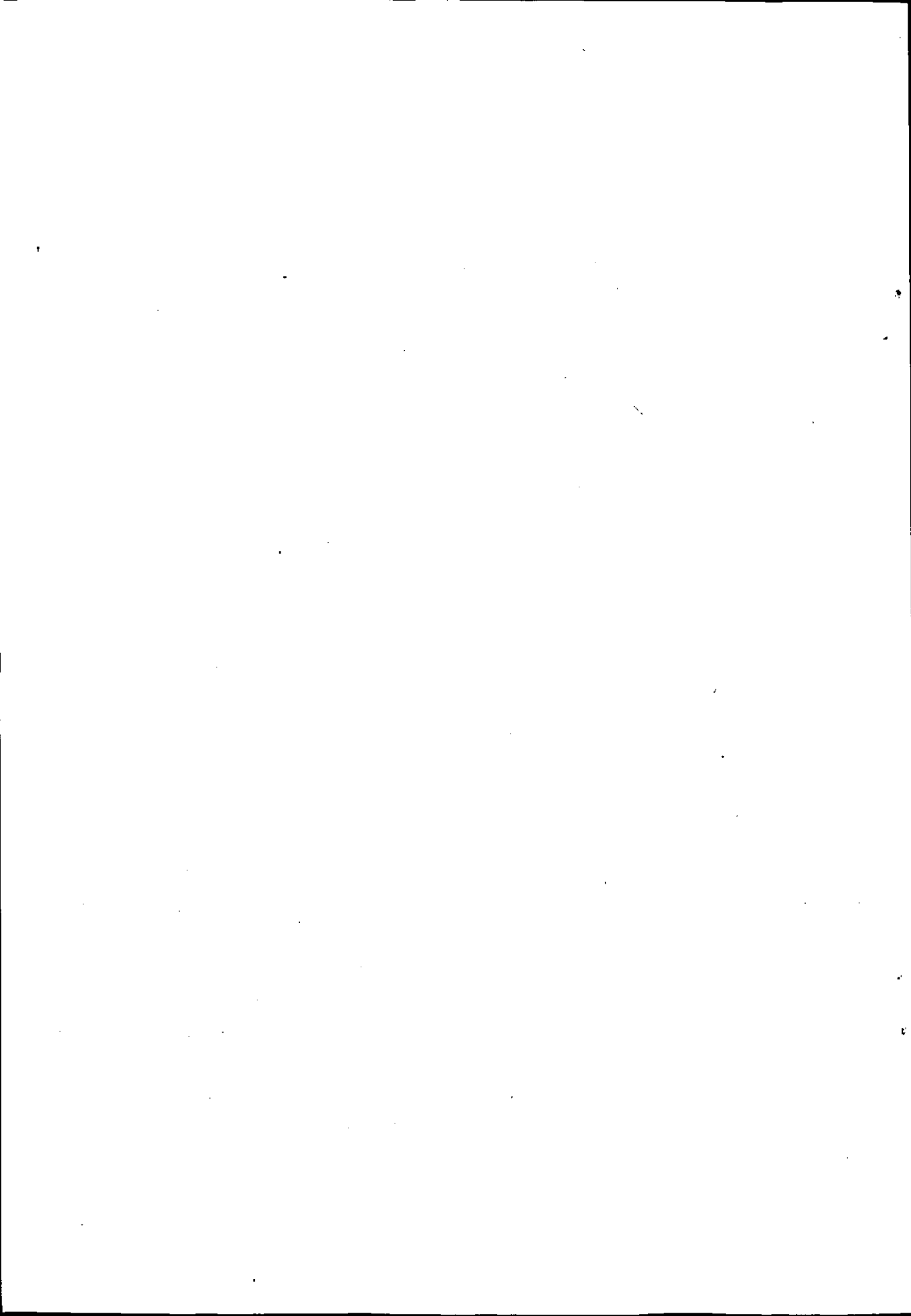
"、&、@ …… 原則を適用せずに必ず記号の前に：をつける。

他， $n + 1$ 個のブランクで n 個のスペース・シフトを行なう。

3. 同音異義語の選択

漢字単語および固有名詞を読みカナで入力するために同音異義語が発生する。この場合，データ作成時に同音異義語表を参照しながら，同音異義語選択番号を読みカナの直後に付加する。

(例) 井ジョウホウ 1 *システム



5. 今後の課題

5. 今後の課題

コンピュータの適用分野が広がり数値情報だけでなく文書情報の取り扱いが増すにつれて当然のことながら日本語情報の処理が問題となってくる。これは単に日本だけの問題ではなく、アルファベット以外の文字で表現される言語を使用する国々における共通の問題といえるであろう。我が国では従来からこの種の情報はローマ字やカタカナの表音文字を用いて処理してきたが、読みにくい、意味にあいまいさがあるなどのために、コンピュータからの出力情報も価値の低下や理解力の低下をまねき、最近では日本語本来の表現形式である漢字かな混りに対する需要が高まってきている。このため各メーカから相ついで各種の漢字入出力装置が開発され実用に供するようになってきた。一方、コンピュータからの出力情報を漢字かな混りにしようとするソフトウェアの開発等も活発化しつつあり、日本語情報（和文情報）をコンピュータで処理する際の問題点もいろいろ提起されている。その中でも漢字の種類が非常に多いことが最大の障害となっている。一般に、日本語の文書で取り扱われる漢字の種類は当用漢字1,850字をベースにした2,000～5,000字といわれており、姓名、地名などの固有名詞をカバーしようとするれば10,000字以上が必要となってくる。しかも、適用分野によって使用する字種が大幅に異なるため、日本語情報処理を行なう企業、機関で独自の字種を選定し、漢字入力装置の鍵盤上の配列も文字の使用頻度などを考慮して、独自の配置を行なっているため、字種・配列だけでなくデータ・コードも異っているのが現状である。このため、漢字コードで表現されたデータに関しては全く互換性が無く、必要に応じてコード・コンバートを行なって処理しているが、いわばハードウェアに起因する障害をソフトウェアでカバーし、不便をしのいでいるということができよう。そこで、漢字の種類やコードの標準化の問題が起ってくる。

漢字入出力装置を早くから導入した新聞業界では昭和34年にCO-59コードを規定し、現在では業界内でこのコードに統一している。

また、行政機関においても日本語情報処理を行なう省庁が増加しつつあることから、行政管理庁が中心になって昭和48年度に漢字コードの標準化に着手し、昭和50年度に一応「行政情報処理用漢字コード」としてその標準化を行なった。

更に、工業技術院では特定業界だけでなく広く一般に適用できる漢字コードの標準化として昭和49年度より「漢字符号の標準化に関する調査研究」に着手し、字種の選定とその符号化を行ない、JIS原案としてまとめていることから、近い将来発表されるものと思われる。

この原案では使用頻度の高い各分野に共通性のある基本漢字群として約3,000字を選定し、これを第1水準漢字群として代表的音訓順に配列している。また、それ以外の必要漢字を第2水準漢字群として選定し、部首画数順に配列し、合計約8,000字となる予定である。

国立国語研究所で行なった新聞の全国紙3紙についてのサンプリング調査によれば、漢字は約2,900字しか現われないので第1水準漢字群約3,000字でほぼ100%処理可能となるであろう。また、漢字の選定には人名、地名などの固有名詞についても十分の配慮が払われていることから、著者名や地名を含めて文献検索などにも十分満足すべきものとなるであろう。この標準化は今後、日本語情報処理を行なっていく上で1つの方向づけとなり、大きな意味をもつものと考えられる。

更に、コンピュータで日本語情報処理を行なう際の問題点として、日本語情報の入力をあげることができる。

この問題は漢字入力装置の操作性、コスト、オペレータの育成、定着率など種々の問題を含むことから必ず論議されるところであるが、これも漢字の種類が非常に多いことに起因することはいうまでもない。

現在最も多く利用されている漢字入力装置としては漢字テレタイプライタをあげることができるが約3,000字もある鍵盤上から必要とする文字を探すことは専任のオペレータでない限り非常に困難な作業であり、素人に適用することとはその操作性からも無理な部分が多い。このため各メーカーや研究機関で種々

の研究が行なわれ和文タイプライタの改良やライトペンを用いた漢字入力装置などが開発されているが、まだ日本語情報を合理的に入力するための有効な入力装置が出現しているとはいいがたい現状である。

また、従来フルキーボード方式の入力装置を用いて日本語情報を入力することが困難であることから、これをソフトウェアでカバーするための方法としてカナ漢字変換についての研究が行なわれている。

つまり、漢字にくらべてはるかに文字数の少ないカナ鍵盤を用いて漢字の読みを入力し、機械との対話によって同音異義語を処理し、漢字かな混り文へ変換する方法である。

当財団で研究開発した（昭和48、49年度）カナ漢字変換システムもその1つであるが、大学、研究機関で早くから研究されており、その変換率も70%～95%が得られている。

この場合は同音異義語の処理を完全に行なうことが難しく、また、100%の変換率を求めることも非常に困難であるためどの分野にでも直に適用することは好ましくないが、出力情報に対する理解力の向上をめざすシステムには大きな効果が期待できるであろう。いずれにしても、カナ漢字変換は本質的には日本語の機械翻訳の問題であるため日本語の構文分析や文章の意味処理などの技術が相当進歩しなければ解決できない問題を含んでいるようである。

その他にも漢字の配列順序が一定でないこと、漢字の字型が複雑であること、外字処理が必要となることなど数値情報とは本質的に異なる問題があり、解決しなければならない問題が山積されている状態である。

この様な現状の中で、当財団ではマン・マシン・インタフェースの面から、特に必要と考えられる日本語情報処理の適用として文献検索をとりあげ、日本語情報を用いた検索システムの設計を行なった。詳細については第4章で述べているが、日本語情報文献検索システム（KINDシステム）の設計にあたっては必要十分な調査研究期間もなく、またハードウェア上の制限などもあって検索システムとして完璧かつ満足されるものではないと思われる。当面、この

KINDシステムに機能拡充が望まれるであろう課題として、次の5つがあげられよう。

- ① ジャーナル管理および統計処理
- ② 検索コマンドの充実
- ③ データ・セキュリティ
- ④ シソーラス関係の拡張
- ⑤ データベースの拡充

①の機能によって、システム運用上で得られるデータ、とりわけ利用者の情報源に対するアプローチの実際や、検索コマンドの使われ方など適切な資料が得られよう。②については、当初会話型検索を中心にしたコマンドの選定を行なっていることから、ガイダンス機能などが網羅されていないので、拡充が望まれよう。③は、当システムの本質的目的からそれるため除いてあるが、データベースにパーソナル・ファイルの登録をも可能にすれば必要な機能となろう。④は、現在、同義語と下位語のサービスしか行なっていないものを、さらに上位語、関連語の順に追加して、シソーラスの有効な利用が望まれる。⑤は、システムの利用者に対するサービスの1つとして、KINDシステム用の外部記憶装置、いわゆる磁気ディスク・メモリを拡充して、異なる技術分野のデータ収録や、収録件数の増加をはかるもので、この点は主としてハードウェアの課題である。

以上のような機能拡充は望まれるところであるが、日本語情報を扱った検索システムのモデル・ケースとして、当システム開発は意義あるものと確信している。

今後、当システムの開発および実験研究を通して、他の企業機関において日本語情報を扱った検索システムを構築する上で、有用な種々のデータが提供できれば幸いであり、また逐次使用体験と機能の充実を繰返し行なって、より価値あるシステムに発展させることが望まれる。

禁 無 断 転 載

昭和51年3月 発行

発行所 財団法人 日本情報処理開発協会
(旧財団法人 日本情報処理開発センター)

東京都港区芝公園3-5-8

機械振興会館内

TEL(434) 8211(代表)

印刷所 東京都小金井市前原町1-15-19

(有) 秀 嶺 社

50-S001

