記事情報検索のための
データ・マネジメント

昭 和 47 年 3 月

JIPDEC

財団法人　日本情報処理開発センター

この事業は，日本自転車振興会の機械工業振興資金による「昭和46年度情報処理に関する調査研究補助事業のうち「データ・マネジメント・システムの研究」の一部として実施したものであります。

序

大規模な情報サービスのシステムを効果的に運用するためには，関連する複数のデータをベースとした情報の利用とか，利用者とコンピュータ・システム間のインタラクティブな情報の交換といった，新しい情報処理技術の確立が必要であり，データ間の相互関係を考慮した，いわゆるデータ・マネジメント，情報検索等の技術開発が内外のメーカ，ユーザ，研究機関等において活発に行なわれております。

当財団においても，コンピュータによるデータ・マネジメントのソフトウエア技術を確立するため，情報処理提供サービスという観点から統計データあるいは記事情報のマネジメント・システムに関する研究を行なっておりますが，昭和46年度には，その一環として，内外の主要メーカがこれまでに開発した各種汎用データマネジメント・システムの中から代表的と思われるものをいくつか選び出し，当財団設置のコンピュータを用いてその処理能力と経済性，応答性と安定性といった相対する特性を実験的に考察するとともに，記事情報をできるだけ理解しやすい形で提供する手段の１つとして，漢字・カナ文字混り日本語文のコンピュータの処理に関する技術的問題をとりあげ，複数の漢字を専門分野別の用語として扱うという新しい概念に基づく漢字・カナ文字の変換方式の事例研究を実施しました。とくに，この漢字・カナ文字変換に関する研究の成果は，今後，出版界等において，広く利用されるものと確信しております。

ここに，今年度の研究実施にご尽力ならびにご支援を賜わった，島野滋雄(トピー工業)，杉本太郎(千代田化工建設)，後藤榛男(学習研究社)，村岡洋右(三菱電機)の各氏および関係各位に心から感謝の意を表しますとともに，この報告書が各方面に活用され，わ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に存じます。

昭和 47 年 3 月

財団法人　日本情報処理開発センター
会長　難波捷吾

記事情報検索のためのデータ・マネジメント

目　　次

総　　論

1. 総　　論

1.1　研究の意義

人間社会で取扱われている情報を形態的にとらえると，定形的な情報として記号情報，数値情報，不定形な情報として記事（文章）情報，図形情報，音声情報などがある。

今日までの情報処理は，事務処理，科学計算，自動制御などにおける定形的な情報処理の取扱いが主体であった。情報検索に限定しても記号情報，数値情報を取扱ったものがほとんどで，生産における部品管理，オンライン預金サービス，旅行予約サービス，クレジット・サービスなどは，その代表的な検索システムである。これらのシステムが実現するにあたって処理言語，ファイル編成，アクセス方式などの検索技法がこれまでに数多く開発されている。

ところが最近，通信回線の自由化もあいまって情報提供サービスがクローズ・アップされてきており，その中でも取扱いが面倒であるために放置されていた不定形な情報の利用が注目されるに至って，その管理技法の向上が問題になってきている。情報提供サービスには，広域性，即応性が要求されているが，それぞれの利用目的に応じて特色を生かした資料配布サービス，テレフォン・サービス，オンライン・リアルタイム・サービスが今日までに実現をみている。

これらのサービスのうち，コンピュータを使用するシステムはオンライン・リアルタイム・サービスであるが，そこで取扱う情報は，科学計算，簡易事務，財務データ，発券などの数値情報が大部分を占めている。数値や記号は，文章や言葉より感覚的に受入れにくいため，利用者からは一般に敬遠されがちであるが，反面このような定形化，規格化された情報は，これまでコンピュータ処理の上で最も取扱い易く，管理し易かったために処理データの主流を占めてきた。しかし，文字のもつ表現力，解説力，連想力は，記事情報や図形情報の方が記号情報や数値情報よりはるかに優れている。そこで，これら不定形な取扱いにくい情報を如何にしてコンピュータに管理させるかという研究が，情報処理の高度利用という観点から無視できない状況になってきている。

不定形な情報の記憶としては，現在使われているコンピュータと周辺記憶装置とは別なマイクロ・フィルムやホログラフィなどのパターン記憶がいくつかの機関で研究されており，その成果が実用に近い段階にきているので，大型コンピュータと連動した大量情報管理もまもなく可能になろうとしている。

また，現在のコンピュータ・システムを使って不定形な情報を管理する地道な研究も進められており，文章で記された記事情報や製図などの図形情報を記号化，定形化し，既存の方法を活用して不定形な情報のコンピュータ処理を可能にしようとする努力も払われている。本報告書第9章の記事情報検索事例研究：ＭＡＳＣＯＴにおいては，雑誌，文献，ニュースなど不定形な文章で表現された記事情報をできるだけ定形化し，現在までに開発されてきた検索技法を十分活用して，情報検索の分野でいまだ例をみない記事情報検索の可能性を実証しようとしている。

記事情報とは，広義には書物の中に記された情報全般を指すが，ＭＡＳＣＯＴでは，とくにコンピュータに関する一連の記事をとりあげ，不定形な情報を取扱った検索システムを作りあげており，技術面，利用面から収集した記事情報を，一定の規則で分類整理してデータ・ベースで管理することにより，利用者に必要な情報を検索サービスしようとするものである。

不定形でしかも長文の文章で表現された生情報を加工せずに大量に記憶することは，現在の主記憶装置や周辺記憶装置では不可能であるので，これらの情報の内容を損うことなく要約して，セレクトした情報に作り直す必要がある。この作業を行なうことは，単位当りの情報価値を高めることにもなり，大量不定形情報の記憶を可能にする第１歩である。

記憶情報を分類し整理して検索が行なえるようにするためには，文章の中からその内容を表現するいくつかのキーワードを抽出しなければならない。このキーワードを決める作業は，記事情報を取扱う上での要となるもので，利用面で十分満足ゆくものにする努力が肝要である。ＭＡＳＣＯＴでは，これらのキーワードを決める作業を専門家の手に任せることにしている。キーワードを体系的に整理したものをシソーラスというが，定形化した情報のシソーラスは一般に階層木構造をとることが多い。人間の手でシソーラス造りをする労力と時間を省くために，コンピュータ・プログラムにこの作業を任せるシソーラスの自動編成や，関連情報を自動的にうまく整理して記憶する連想記憶の研究が一部で進められているが，まだ実用化の段階には至っていない。

このように記事情報検索では，データの定形化やキーワードの抽出などデータ・ベースへ蓄積する以前の処理において，数値情報などの定形情報の取扱いよりも手間がかかる。ここで記事情報収集から運用管理までの作業手順を示すと次の通りである。

記事情報の収集 ── 内容の分析 ── 分類 ── 抄録化 ── キーワードの抽出 ── シソーラスの作成 ── データ変換 ── データ・ベースへ格納 ── 運用管理

また，検索面からみた，記事情報検索サービスのための一般的なステップは，およそ次の通りである。

検索指示 ── キーワードの突合わせ ── 論理検索 ── 加工 ── 報告書作成 ── 提供 ── 評価

記事情報は，データを収集してから利用されるまでの間に情報の整理，統合，変換，加工など比較的多くのプロセスを経た後に提供されるため，情報の真実性を失い易いのでこの点十分注意を払わなければならない。また，検索された結果が利用者の要求にできるだけ合致するよう，利用目的を限定した情報提供サービスにしなければならない。記事情報提供サービスが十分利用されるまでには，実験段階においてデータの修正，シソーラスの再編成を試行錯誤的に何度か行なう必要があろう。度重なるシステムの変更と改良の結果，はじめて利用者の要求を満足させるに至るまで情報価値を高めることができるのであろう。

一方，メーカやソフトウェア会社から提供されている情報管理のための道具として注目されているソフトウェアに，汎用データ・マネジメント・システム（Generalized Data Base Management System）がある。このソフトウェアは，データ構造に合った格納，データ・ベー

スの管理，論理検索，加工・編集などの機能をもっており，これらの面から記事情報を管理するためには，有効な手段となり得ると思われる。ＭＡＳＣＯＴでは，これらの機能を確証するために汎用データ・マネジメント・システム：ＲＡＰＩＤ（Ｒｅｔｒｉｅｖａｌ　Ａｎｄ　Ｐｒｏｄｕｃｔｉｏｎ　ｆｏｒ　Ｉｎｔｅｇｒａｔｅｄ　Ｄａｔａ）をベースにして記事情報検索システムの実用化試験を行なっている。

現在，コンピュータとその周辺機器で取扱える文字は，英数字とカタカナそれに幾つかの特殊文字に限られており，これらの文字だけで取扱う日本語の文章は，非常に読みにくいといった欠点がある。日本語特有の問題として，カタカナ，ひらがな，漢字の混合文が，日本人に最もよく理解され利用されているという事実から，できることなら記事情報を現在のコンピュータの管理限界を越えた漢字混合文で表現したいという希望が強くなってきている。

本来ならば，直接漢字でインプットし，処理して，漢字でアウトプットする方法が望ましいが，現段階では漢字処理を行なう場合，漢字編集システムとコンピュータ・システムとがハードウェア的に連動した形のものが作られておらず，各々が別個のシステムとして存在しているため，このような状況で日本語処理を行なうには，コンピュータ・システムから得られた検索結果を一度テープに出力し，カタカナ漢字変換を行なったうえで，電算写植編集システムを利用して漢字混合の日本語文を作りあげる方法が現実的と思われる。ＭＡＳＣＯＴでは，カタカナ漢字変換プログラムを作成してカナ・漢字混合の日本語文を作成するための実験を，この考え方に従って行なっている。

現在の日本は，生産中心の時代からサービス中心の時代に移り変りつつあるといわれているが，その中で情報提供サービスの広域利用がとくに叫ばれており，コンピュータを中核とする不定形な情報検索サービスの要望が強まっている。中でも日本語情報処理の問題は，その必要性から最も重要視されている。このような背景から情報提供サービスを，より高度化するための基盤として，記事情報検索サービスの取扱いを管理面，検索面，サービス面から実験的に考察し，データの収集，キーワードの抽出とデータ・リンケージ，ファイリング，検索，メンテナンス，日本語出力などの処理方法について，その実用化の方向を探ることは意義あることと思われる。

1.2　報告書の要旨

コンピュータを主体とするデータ・マネジメントの研究は，昭和４３年度以来一貫して進めている経営情報システムを，側面から技術的にサポートする一つの方法として，４４年度より実施しているものである。すなわち，４４年度は情報処理におけるデータ・マネジメントの重要性を中心に報告書４４－Ｓ００３「中堅企業のＭＩＳ構造」をとりまとめた。４５年度は定形情報の取扱いとして統計情報をとりあげ，その管理方法を検討して報告書４５－Ｓ００１「経営予測のためのデータ・マネジメント」を作成した。そこで，４６年度は今日取扱いが問題になっている不定形な情報の管理のうち，とくに記事情報をとりあげ，その管理方法を検討して報告書「記事情報検索のためのデータ・マネジメント」をまとめることになった。

本研究の主要目的は，記事情報を管理運用するために汎用データ・マネジメント・システムを活用すること，ならびに記事情報をできるだけ利用者に受入れ易くするための漢字処理の重要な問題であ

るカナ漢字変換システムを作成することであり，各々の有用性を実験的に検証することである。

第1章「総論」では，これまでのコンピュータによる情報処理の主体は定形情報であったが，利用者の立場からみると必らずしもこのような情報の形態が好ましいとは限らない。むしろ不定形な図形，音声，記事（文章）などの情報の方が情報量も多いし受入れ易いという長所があることを述べ，そのための管理方法として従来の処理方式に合せた定形化や，パターン情報処理などがあることを解説している。

第2章「記事情報提供サービスのための検索システム」では，記事情報提供サービスを行なうためのコンピュータ検索システムのメリットについて活用面，管理面，技術面，採算面などから検討するとともに，記事情報検索システムが今後活発に利用されるための諸条件について議論を展開している。また，現在までに開発された情報提供サービスのための検索システムのいくつかを紹介し，記事情報検索システムを設計するうえでの参考にしている。

第3章「記事情報検索における汎用データ・マネジメント・システムの活用法」では，情報処理における検索システムの特異性について述べ，記事情報検索の難しさを指摘している。また，このような取扱いにくい情報を効果的に管理するために汎用データ・マネジメント・システムの機能を有用することが考えられるが，その場合のシステム設計上の問題点を提示し，さらに検討を加えている。

第4章「データ収集の方法」では，データの収集作業は，情報の利用目的を明確にしたうえで，優れたライブラリアンにより十分な時間と労力と費用をかけて行なわれなければならない，また，データの良し悪しが情報価値の有無を決定することから，収集のための選定基準をはっきり規定しておかなければならないなどを解説している。

第5章「データ・リンケージの方法」では，データ・リンケージの重要性を中心に述べており，とくにキーワードの抽出法とシソーラス（thesaurus）の作成法がシステムの有効性を決定する大きな鍵となっていることを強調している。

第6章「ファイリングの方法」では，データの構造，データの量，処理方式の各面からシステムの検討を加え，データ・マネジメントの機能を十分に生かしたファイル編成を採用することが望ましいと述べている。

第7章「ファイル・メンテナンスの方法」では，記事情報を管理するうえで新しい情報を追加したり，情報の一部を修正したり，不要な情報を削除するなどのファイル・メンテナンスは頻繁に起るので，それをやり易くしたファイル編成が必要であり，また，シソーラスの変更にともなうデータ・リンケージの補修にも十分対処できるシステムを予め編成しておくことが肝要であることを強調している。

第8章「検索の方法」では，検索システムを受動的な方法をとるものと，能動的な方法をとるものとに分けて解説し，検索の質問様式や処理方法を決定するために，システムの用途，範囲，サービスの形態などの明確化が必要であることを述べている。

第9章「事例研究：コンピュータ情報・検索システム（ＭＡＳＣＯＴ）」では，記事情報としてコンピュータに関する一連の情報をとりあげ，これを提供サービスするための検索システムの運用法を

汎用データ・マネジメント・システム（ＧＤＢＭＳ）を活用して管理面，検索面から実験的に検証している。コンピュータ情報・検索システム：ＭＡＳＣＯＴは，コンピュータ用語を解説するＭＡＳＣＯＧ，コンピュータに関するニュースを紹介するＭＡＳＣＯＡ，コンピュータに関する雑誌文献を紹介するＭＡＳＣＯＬの３つのサブシステムから構成されており，その各々について詳しい説明を行なっている。ＭＡＳＣＯＡでは，ＧＤＢＭＳと比較検討するため，とくにデータ・ベースに独自のインバーテッド・ファイル（ｉｎｖｅｒｔｅｄ　ｆｉｌｅ）をもつ専用システムを作成している。

第１０章「汎用データ・マネジメント・システム活用上の問題点」では，汎用データ・マネジメントシステム（ＧＤＢＭＳ）がシステム機能として，またデータ管理機能として勝れた特長をもちながらも，ユーザ・サイドであまり活用されていない事実を述べるとともに，事例研究の実験結果から記事情報検索システムを作成するうえでのＧＤＢＭＳの有効な活用方法について解説している。

第１１章「カタカナ漢字変換システム」では，日本語処理におけるひらがな・漢字表現の重要性を述べ，現在コンピュータで取扱かわれているカタカナ情報を漢字情報に直すためのカタカナ漢字変換を中心に，漢字処理の問題を解説している。ここで行なった実験は，ＭＡＳＣＯＴの出力を電算写植編集システムを介して漢字情報として提供サービスする際の，カナ漢字変換システムの作成と変換率を向上させるための方法を検証することである。

記事情報提供サービスのための検索システム

2. 記事情報提供サービスのための検索システム

2.1 記事情報提供サービスの必要性

記事情報検索システムとは，書籍・新聞・雑誌などの記事を体系づけて分類し，ファイルしておき，それらの内容について問合せを受けたさい，タイムリーに応ずるシステムである。一見簡単そうだが，実際に運用を行なう上では，きわめてむづかしい問題が山積している。

たとえば，記事が簡単に廉価に収集できるか，その記事を体系的に一貫した思想のもとに分類・加工・蓄積できるものか，それを要望に応じて，自由に検索することが可能であるか，ざっと考えただけでもこれらの問題点がある。

さらに，経済的採算性を考えた場合，より困難な問題に逢着する。アポロで名高いＮＡＳＡのＩＲシステムや，医学文献を収録してあるMEDLARS，生物化学文献のCASなど有名な情報検索システムは，そのほとんどが国家・軍事・学術などのような公益に関するものであり，採算性をそれほど配慮しなくても済む。また，情報自体も一般的に価値の高いものである。

しかし，ここで扱う記事情報検索システムはどの程度公益たりうるものか，または，検索される情報に果して，市場価値のようなものがあるのだろうか，現段階では未知のものである。

情報検索システムの動向をみると，技術的には定着化の兆が見え，その対象も単なる研究・遊びの段階から，ＭＥＤＬＡＲＳのような実用価値をもつものの出現もあり，一歩すすんだものとなってきた。記事情報検索システムがこのうちのどの段階まですすんでいるのか，以下概観してみよう。

2.1.1 記事情報検索システムの価値

記事情報検索システムが必要であるためには，

第１に，検索される情報自体の価値が必要である。通常，情報の価値は新鮮で，稀少であるほど高い。内容の面では学術・研究のようなものは高く評価される。しかし，記事情報のように，一般に公表される情報は時間とともに著るしくその価値を喪失して行く。ほとんどは捨てても支障のないものとなる。蓄積された情報が価値をもつには，たとえば，利害得失のきびしい法律適用のような場合，その事実証明のための情報は価値が高いが，一般の記事情報はほとんど興味本位の検索効果しかないのが普通である。

第２に，情報が価値をもつには，情報の量が膨大で，その処理が人間ではできない場合であろう。記憶を正確によみがえらせることができない場合がこれに当る。人間はある限度の複雑な情報の中から行間の意味を抽象し，論理的な判断を加えて，情報の意味を分析することに向いているが，単純に量のみが天文学的に多く，ただそのまま検索することにはむしろ，機械の方がはるかにすぐれている。このような，機械としてのコンピュータの特徴を活かした情報については，処理の過程で価値をもつことがある。

第３に，複雑な論理構造をもっているため，第２の膨大な量と併せて，それをコンピュータで処理

することによって，価値をもつことがある。たとえば，言語の解析をするさい，ある長い文章の中で個々の単語の出現頻度，前後の関連語，著者による偏向のような処理で得る情報は人間ではとうてい取扱えないので，コンピュータ処理による価値をもつ。

第4に，情報の発生が広域に亘っており，それを即時に処理することによって，地域間の格差が是正できる場合がある。

たとえば，A地で発生した情報によって，B地のものが行動を起す場合，郵便その他の伝達方法によっていては，地域が限定され伝達がおそく，手続が煩しくなる場合などがこれに該当しよう。この場合のおくれは，価値を失なわせるものとなるので，それをコンピュータまたは通信回線を使って補えば，それだけ価値の肯定ができることになろう。

以上のような場合，情報は価値をもつことになろう。実際に，日本で動いている情報検索システムは特許情報，不動産取引情報，雇用情報などであり，上にあげたいづれかの価値をもつ場合に該当しよう。

しかも，このいづれもがかなり，特殊な分野のものであることに注目したい。記事情報のように，相当一般性の強い情報，たとえ，それを特殊な分野に限定してみても，上にあげた各システムとは違うとみるのが正しいであろう。従って，記事情報検索システムはこれらに較べて，その成立は相当困難であることが予想される。

その理由を具体的に述べると，

① 記事情報には実用的・市場的価値が薄く，極端にいえばなくても済むことが多い。記事情報の対象である日常生活および商取引で情報によって左右される場合もあるが，ほとんどの情報は金銭的に取引されるほどの価値をもつことは稀れである。ただし，情報を精密に分析し，効果あるように加工されている場合は別である。これはコンピュータによる検索によるものではなく，人間の頭脳による調査といえるであろう。

② 記事情報は，わざわざ情報検索システムを使わなくても，他の方法で通常の場合は，十分に目的を達することができる。たとえば，図書館的索引，新聞社の記録係，図書相談員，綜合文献索引書などで十分なことが多い。これらに携わるすべての機関が大同団結して，センターでも創設する場合は，コンピュータによる検索も有効になるにちがいない。しかし，一般にはその機運はまだない。

③ 記事情報検索システムの開発・維持には相当の費用と優れた要員とを必要とする。それに見合うメリットを生むことは容易ではない。ＭＥＤＬＡＲＳでもハネウエル８００という大型のコンピュータで処理していることを考えれば，それをサポートする要員および費用も相当膨大になることが考えられる。コンピュータ・サイドの費用だけでも膨大なのに，情報の原資料の購入費用，記事をデータに加工するライブラリアンなどの賃金などを加えれば，さらに，膨大な費用となる。

常識的に考えて，記事情報検索システムの成立は上記のように相当難かしい。

それでもなお，記事情報検索システムが必要性をもつには，その提供者と利用者とのいづれか，または，双方の利益の合致が成り立たなければならない。

2.1.2 記事情報検索システムの必要度

記事情報検索システムが必要であるとすれば，それは

① 記事情報検索システムが何らかの価値をもち，提供者，利用者のいづれかが利益をうける。できれば経済性の面から満足できること。実際にどういう分野があるか見当がつけにくいが，その有効な分野を探り当てたとき，記事情報検索システムは新たな展開を見せることになろう。情報そのものが取引である不動産取引の紹介などはその一つであろう。

② たとえ，直接の利益はえられなくとも，そのシステムを作るさい，取得できる手法，ソフトウエア，または，ハードウエアの開発が一般に役立つこと。

③ 当面，直接的間接的な現実の利益がなくとも，公益または研究のために，このようなシステムの開発・提供・管理が必要であること。

④ ②と③とを含めて，記事情報検索システムの手法およびシステムをそのまま，または，多少の改善を加えることによって，他の情報検索システムに切換・応用できること。

以上，いづれかに該当するとき，記事情報検索システムは存在理由をもつと評してよいであろう。

ここで進める研究は上記中②③④のいづれかに該当すると思われる。今後，このシステムでの経験を活かして，記事情報検索システムに向く分野を探して行くことになるが，権利とか，義務とか，または，危険が情報の如何にかかってくるようなものが，分野として有望であろう。

今まで述べてきたのは現時点のことであるが，今後の記事情報検索システムの必要性を占うものは，この他に，

① 経済的社会的要請の変化

② 経済的社会的環境の変化

③ 価値観の変化

などが考えられる。たとえば，コンピュータの使用料がきわめて低廉となったときなど，検索される情報の価値は低くとも，相当頻繁に使われるようになろう。または，技術が向上し人間の能力ではそれを記憶しえなくなったとき，広域かつ精密な情報を何らかの形で処理しなければならなくなったとき，などはこれに該当しよう。

そのような必要性は今後相当広い範囲で起ってくる可能性がある。記事情報検索システムがその必要性を十分確保するために，克服しなければならない条件がある。これについて次に述べる。

2.1.3 記事情報検索システムの成立条件

提供者と利用者との各々の側から記事情報の有効性を認識させる条件を調べてみる。

① 提供者側の条件

a．利用者が多数いて，提供する使命を十分理解していること。利用者が少ないことは，システムを成立する大事な条件が欠けていることを意味する。その利用者に情報を提供する使命を十分理解していないと，折角のシステムも活かせなくなってしまう。

b．記事情報の収集体制が確立しており，ルーチン化が可能であること。たとえば，収集が断続的，遅延，重複，権威喪失，偏向などがなく，しかも，収集に費用および手間のかからないことが望

ましい。

c．システムを運営するにふさわしいライブラリアン，プログラマ，オペレイタなどがいること。とくに，記事情報のように，記事そのものに内容や水準のちがうものが入り混っているものについては，ライブラリアンの質量ともにシステムの成立条件の中ではとくに重要なものである。

d．データ・ベースの作成更新が簡単であって，運営する上で，支障のない時間内で処理できること。たとえば，常にデータ・ベースを変更していては費用の上でも負担が大きい。また，他のシステムが優先で，検索システムが常に後順位で，深夜でしか利用できないという状況や方針では問題がある。

e．記事情報の入力変換が容易で，キーワードなどの一般化がすすんでおり，利用上の煩わしさがないこと。利用者の誰にでも，すぐ利用できることが大切な条件である。特殊な使用法の解っている人のみしか使えないとなると，システムの利用は衰退する。もっとも，これは秘密保護とは別である。

f．リアル・タイムおよびバッチのいづれの処理によっても検索が可能であること。経済的要因もあって一概にいえないが，リアル・タイムで処理できることがのぞましい。

g．ユーザの要求に絶えず注意し，満足の行く提供サービスができること。

② 利用者側の条件

a．必要な記事を迅速，かつ，正確に検索提供ができること。当り前のことであるが，分類技術，キーワード作成技術，データ・ベース作成技術などに大きく関連する。

b．操作が簡単で，判かりやすいこと。この条件は，利用者層を広めるよい方法である。

c．出力の表現が容易に正確に解読でき，また，不明瞭な場合，再問合せ，または，用語検索がすぐできること。出力の表現については，漢字印刷のできることが日本語の場合には望ましい。

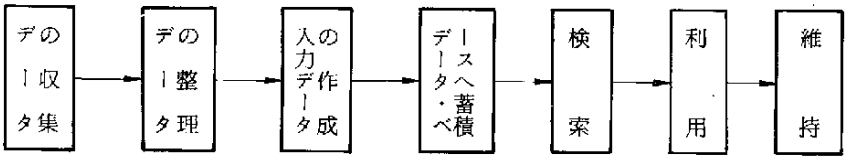
d．若干複雑な条件をつけて，検索可能なこと。単に，辞書的に出力するのではなく，コンピュータの中で比較検討，さらに加工して出力できることが望ましい。

e．記事の内容が利用者の満足できるものであること。再現性，正確性が高いものが望ましい。そのため，収集段階から，維持保守の段階まで，常に点検し，水準を向上させるよう努力しなければならない。

以上で述べたように，提供者側と利用者側の両方に受け入れ易い体制を作ることはシステムを有効に活かすために必要な条件である。この他，利用者はきわめて要求に忠実であり，たった数秒出力が遅いだけでも拒絶反応を示すことすらある。まして，情報の内容についての要求はきびしいものと考えなければならない。この点に提供者側は十分留意する必要がある。

2.2 記事情報検索システムの在り方

一般的に記事情報検索システムは，次の図のような過程で示される。



データの収集から維持にいたる間に日常の表現形式を電子的表現形式に転換することによって，有効にコンピュータを使うことがこのシステムの狙いである。それだけに，この転換は単なるデータ処理とはちがういろいろな問題を解決しなければならない。これを広義にとらえれば，いわゆる情報検索（ＩＲ；Information Retrieval）の問題であるし，その中を細分化すれば，データ構造なり，シーソラスの問題になる。これらはまだ開発の途上であり，ハードウエアの発展とともに今後大いに発展することが期待されている分野である。従って，このシステムのあり方は今後いちじるしく変って行くにちがいない。ここでは現在の一般的な水準において論じてみよう。

ここで問題にしているのは，一般的なデータによる情報検索システムではなく，記事という特異な分野での情報検索システムである。その大きな特徴は，一般のデータが確立した形式構造で処理されるのに対し，記事情報はその範囲と量とを深め，価値あらしめようとすれば，するほど，確立した形式や構造の中に押し込むことがむづかしいことである。

たとえば，データの収集について，記事の源泉に新聞をとりあげてみよう。一般大衆紙Ａと経済専門紙Ｂとでは同じニュース源から得た情報でありながら扱い方，見方がちがう。質量ともに大巾な差を生ずることが多い。これに対して通常の販売データなどは，どんなに複雑で項目が多くなっても，共通のパターンをとり易い。

記事情報についての，この特異性はその情報検索システムのあらゆる過程で見い出せるものであり，それをふまえて，記事情報検索システムのあり方を考慮しなければならない。

2.2.1　データの収集

記事情報データの源泉となるものは，たとえば，

① 新　　聞
② 専 門 紙
③ 論 文 紙
④ 文　　献
⑤ 官 公 報
⑥ 広 告 紙
⑦ 雑　　誌
⑧ マニュアル

などがある。

新聞といっても国内国外の別，大衆紙と業界新聞との別など各種ある。上にあげたほとんどは活字であるが，ラジオ・電話による音声のものやテレビ，映画などの視覚に訴えるものもあろう。記事情報データの源泉は限りなく広大な領域に存在する。

これらの源泉からとりうる記事情報はどんな分野のものでもまさしく洪水のように巨大な量になる。この中から必要な記事情報を入手しようとすれば，ハードウエアが無限でないかぎり，制限せざるをえない。このためには，情報源とそれからとり出される記事情報データの両方の選定に，明確な基準を設定することが不可欠である。

第1に，情報検索システム一般についてもいえることであるが，システムの目的が明確でなければ，検索そのものが十分目的を果せないものになってしまう。記事情報検索システムは，MEDLARSのような医学専門の情報検索システムに較べ，その対象が多次元多種類多量の情報検索システムなので，目的も単一でないことが多い。この目的を明らかにしないと，情報源の選択基準を設定することができない。

第2に，システムの目的に合致した情報源の範囲が決まると，情報源自体の価値を考察しなければならない。これは客観的基準による情報源の格付けで整理する。

第3に，収集の経済性も問題である。情報検索を行なう価値のないデータを格納する必要はない。収集活動はハードウエアおよびソフトウエア上の経済性とデータの価値性との関連から決定されよう。

第4に，維持・保全というメンテナンス上の面から収集が限定されてくる。収集活動はメンテナンスも含めて，容易に経済的に行なえるものが望ましい。

2.2.2 データの整理

ここでいうデータの整理とは情報源の素材をコンピュータの処理に便利なように加工・転換することである。たとえば，新聞の記事を所定のカード・ファイルに納めることがこれである。原データといっても記事情報検索システムは単なる数値表現の情報ばかりでなく，言語表現の情報も含まれており整理の段階では主観的な要素が入ってくるので，相当むづかしい。

第1に，言語選定の基準を設定しなければならない。とくに，キーワード，抄録になるものは明確にしておく。しかし，将来，格納されている記事そのものの中から，キーワードでない言語をキーワードの代わりに使うことも考えられるので，このような場合の取扱いにとくに厳格な基準の設定を必要とする。

第2に，整理して行く過程で，情報源のデータは選定され，取捨選択される。それは将来の価値も考えて，選別されて行く。この選別の基準も明確にすることが望ましい。

第3に，データの整理は機械操作的にできることが望ましい。人によって，情報の原始データの加工法が違っては困る。

第4に，この過程の作業はハードウエアおよびソフトウエアと深く関係する。それによって，情報の入力データの形式は変ってこよう。

データの収集と整理で述べてきた，情報源から記事情報の原始データを作成するまでの作業は人手による。この段階ではコンピュータ特有の性質や性格は表面にでてきていない。

しかし，コンピュータで情報検索する以上，それに必要な考え方やデータの構造などは無視できない。その制約の中で，開発されたいくつかの手法が記事情報検索システムのあり方として選択される。そのうちのとくに重要な項目として，次のものがあげられる。

第1に，シーソラスとキーワード

第2に，データ構造

第3に，システム全体のマネジメント

以下，これらについて触れてみよう。

2.2.3 シーソラスとキーワード

必要な情報を検索するためには，予め必要とするデータを収集し，必要とするときに，迅速，正確に取り出せるように加工した上で，蓄積しておかなければならない。その中には普通，見出し語（キーワード）の集まりと抄録（アブストラクト），それに目録とが含まれている。

検索のさいには，同義語辞典といわれるシーソラスと照合しないと完全な検索はできない。シーソラスには同意語ばかりでなく，反意義，類似語，上位・下位語，関連語などが体系的に整理されている。

検索の効率はこのキーワードとシーソラスの作成にかかっている。キーワードとシーソラスの作成は通常の情報検索システムでさえ，かなり困難な作業である。これが記事情報検索システムになると，記事の多様性，量の多さ，次元の差の大きさ，情報価値の変化の激しさなどのため，一層困難であるといわなければならない。記事情報検索システムでのキーワードとシーソラス作成については，次の諸点を考慮に入れる必要がある。

第1に，新しく発生する記事情報やそのデータなどの変化に十分即応できること。

第2に，ハードウエアや関連ソフトウエアの発展に順応できること。

第3に，将来のあり方に正しい予測を行なえるようにすること。

このような観点から，シーソラスとキーワードの作成には十分時間をかけて完全なものにすることが望ましい。そのために，

第1に，シーソラスとキーワードの論理構造を安定した，強固なものにすること。

第2に，コンピュータが人間面の不完全性をカバーすること。

第3に，コンピュータに意味づけの機能を付加すること。

などが考えられる。また，現在の意味中心の考え方から，「パターン認識」による新しい検索方法にすすむことも考えられる。

2.2.4 データ構造

検索の正確性と効率化とを保証するために，データを蓄積する形式とデータの関連づけの方法とを確立しておかなければならない。

前者の場合のファイル形式には次のようなものがある。

1) 順次編成ファイル（シークエンシャル・ファイル）

2) 分割編成ファイル（パーティショント・ファイル）

3) 索引順次編成ファイル（インデックスド・シークエンシャル・ファイル）

4) 直接編成ファイル（ランダム・アクセス・ファイル）

これらの各々は特徴をもっており，ハードウエアの特性と使用法とを考えて，最も適合するファイ

ルをとるのがよい。

また，データ構造も代表的な４つの構造を示すことができる。

1) 線型構造（リニア・ストラクチャ）

2) 樹木構造（トリー・ストラクチャ）

3) 階層構造（ハイラキー・ストラクチャ）

4) ネットワーク構造（ネットワーク・ストラクチャ）

各構造は各々使用方法によって特徴も違うし，効果も違う。このうち，情報検索でよく使われるのは階層構造と樹木構造，それにこの２つの組み合せ型である。技術的にはネットワーク構造をはじめ難かしい，高度な手法が開発されている。

一般的に，データ量が膨大な規模の処理には，階層構造としてとらえれば検索処理時間が短かくて済む。反面，メンテナンスやデータのローディングに時間がかかる欠点がある。

どの構造を採用するかは使用目的とシステムの難易によって決まろうが，その他に使うコンピュータのハードウエアやソフトウエアによっても異なる。これを整理すると，

1) 検索機能の範囲

2) データのボリュウム

3) O／S，ハードウエアの性能

4) ユーザの利用形態

などを考慮した上でデータの構造やファイル編成を決めることが最適であろう。

情報検索システムでのデータ量は当然膨大なものである。最適のデータ量は使用目的やデータ構造やファイルの蓄積形式，さらに，ハードウエアの容量，とくにメモリ容量などから決められよう。記事情報のデータ量は膨大なので，場合によっては文脈の解釈ルーチンを設けたパック変換なども考慮に入れておく必要がある。データ量は検索の迅速性，正確性に関連があり，強いては経済性にも関連がある。うらを返せば経済性もデータの選択基準になる。

2.2.5 システム全体のマネジメント

システム全体のマネジメントには，コンピュータの運用上の問題とシステムの維持運用上の問題がある。

第１のコンピュータ運用上の問題とは入力から出力にいたる間での処理を指し，広義のシステムの維持運用の部分である。

入出力からみると，通常のバッチ処理では記事情報検索システムのサービス形態からいって間に合わないであろう。オンライン・リアルタイム処理でなければ，情報の新鮮さは失われよう。場合によっては，会話型問合せの可能なＴＳＳ処理，さらにディスプレイなどを使った図形表示も有効となろう。日本語の取扱いではとくに，漢字やひらがなの使用が可能でないと，効果が薄れる。音声表示は実用的価値を当分もたないと思うが，研究課題にはなろう。

注） データの構造，ファイル編成，処理方式については，４５Ｓ－００１「経営予測のためのデータマネージメント」第５章を参照のこと。

情報提供サービスのあり方としては経済性実用性の観点から決められよう。経済的価値からみると，ディスプレイ付ＴＳＳのようなものが果して引合うのかどうか疑問である。しかし，情報提供サービスを行なう場合には次のような作動方式が不可欠であろう。

1) 大容量のメモリを必要とする。

2) 問合せ方式をとる場合，コンカレントなオペレイションが可能である。

3) 記事情報検索用などページライタなどの機能が完備している。

4) データ収集，加工，編集，レポート出力，更新などの機能はバッチ処理で十分であるが，記事情報のレポート出力として日本語を取扱う場合には，漢字用プリンタの必要性が高い。

データ処理においては，パフォーマンスを無視できない。アクセス方式としては，直接アクセス方式，シーケンシャル・アクセス方式があり，ファイル編成と関連させて適当な処理方式が選択されよう。

とくに，記事情報検索システムではある属性をもった項目を関係づけるために，インバーテッド・インデックスを付加することにより，目的用途を拡張したシステムを形成するという特徴がある。この方式は，データ量の節約，および検索のスピード向上には効果的であるが，記録の追加・削除の頻度が多いシステムにはあまり向いていない。

記事情報検索システムでは削除は少ないが，追加は常に行なわれるので，上の長短からみて，インバーテッド・インデックスは特定の項目だけに限定しておいた方がよい。

ホスト言語については，誰でも容易に使えることが前提である。従って，プログラマでなくとも簡易に書ける言語がよい。しかし，汎用性を失ってはならない。検索システムには，システム独自の専用言語をもつこともできるが，記事情報検索システムにどの言語が適しているかは，利用者，利用機能，ハードウエアおよびソフトウエアによってうまく使い分ける必要があろう。

第２のシステムの維持運用上の問題点とは，より人間サイドの範囲の問題を指す。機密保護はシステム運用上の最も大きな問題である。普通，機密保護の上で考えられることには，他のユーザに対して機密を保護することと，ユーザの機密をコンピュータ・センタに対して保護することがある。一般には，機密キーを設けることにより，特定のユーザのみに利用させるなどの工夫で機密性を守って行く。この方法では，各レベルでパスワードを付加することができる。データ・ベース，ファイル，エリア，プロパーティ，セットなど。

記事情報検索システムは恐らく，個々の記事情報に機密価値をおくのではなく，加工し，綜合し，検索された情報に価値をもつことが多い。

機密保護と並んで回復機能も重要である。システムの信頼性を高める条件として，障害対策をどの程度サポートするかが重要である。回復機能では人間サイドで講ずる対策とコンピュータ・サイドで解決する対策とがある。後者ではシステムの種類やファイルの編成によってその方法も異なる。前者では，システム間またはシステム内での機能転換，たとえば，オンラインが使用不能になれば，バッチ処理に切換えるなどの方法があろう。

システムの維持には常に最新の記事情報を収集し，加工し，蓄積できるようにすることが必要であ

る。そのためには継続的かつ権威ある収集体制と，定期的にデータの更新を行なう努力が必要である。また，システムは常に拡大し，他の情報との連繋が保たれていなければならない。このためファイルの交換やデータなどの互換性が確立されていなければならない。

以上の諸点は，記事情報検索システムを設計するさいに注意しなければならないことがらである。

2.3 記事情報検索システムの動向

2.3.1 情報検索の歴史

情報検索はコンピュータの初期の段階頃から問題にされ，ＩＢＭのLuhnが今から15年位前にＫＷＩＣというキーワードを中心とした検索の考え方を明らかにして一躍その基礎が固まったのは衆知のとおりである。その後のハードウエアの発展に伴い，キーワードから，シーソラスへと発展し，規模の上でもデータ・リンケージの取扱いは大きくなっていった。

当時は試験的段階にすぎなかったものが，今日では単なる試験から発展して実用的段階に達したものも数多くみられる。

最近はデータ・ベースの確立とともに，検索も含む汎用データ・マネジメント・システムが数多くつくられ，ＭＩＳの確立に尽している。内容的にも単なる情報から，抄録へ，さらに記事内容そのものへと当然関心が向けられて行こう。

2.3.2 情報検索サービスの実状

世界的にもっとも有名なのは，医学雑誌・文献の情報を蓄積・検索するＭＥＤＬＡＲＳであり，すでに提供サービスも行なわれている。

この他に，化学の分野ではＣＡＳなど学術研究的な情報検索提供サービスが行なわれている。

日本でも最近ようやく，官公庁，公益団体などを中心に情報サービスのための検索システムの研究が行なわれ始めた。そのいくつかを概観してみよう。

① 労働省の「職業紹介システム」

労働省ではリアルタイム方式で全国の職業安定所とオンラインで結んでいる。

求人側と求職側の両方の要求を蓄積情報にもとづいて照合し，適合するものを選び出すものである。全国的なリアルタイム方式であるところに大きな特徴がある。

② 日本特許情報センターの「特許情報システム」

技術の発展とともに特許の情報は厖大な量に達している。特許は一つの法律による権利であり，権利保護のため，公告日とか請求範囲とか，外国での特許とか貴重な情報が沢山織り込まれている。

一つの特許を開発するに厖大な時間と費用とがかかるので，重複投資や埋没特許の再利用のためにも，特許情報サービスは必要である。

日本特許情報センターの検索サービスは特許公報にもとづき，書誌事項の第１検索サービスと技術詳細の第２検索サービスとを行なうことになっている。

処理はバッチ方式による。特徴は単なる数値や記号の処理でなく，文章を加工し，コンピュータの処理のし易いようにしている点記事情報検索システムの参考になる。

③　日本科学技術情報センターの「技術文献検索システム」

日本科学技術情報センターは，内外の科学技術文献で工学および理学にいたる領域を中心に経営管理も含めた文献データを収集し，重要な論文や記事を抽出し，索引を行ない易いように加工している。ここのサービスの特徴は漢字を始め，各種文字による出力，さらに一過程おいたマイクロシステムに結びつけている点である。

サービスの形態は文献のコピーと磁気テープの貸し出しである。

④　総理府統計局の研究

総理府統計局では統計データ・サービスを供給する目的で統計データ・バンクの設立を考えている。その研究の一環として，米国センサス局のセンサス・ユース・スタディを中心に研究しているが，これは情報検索の意欲的な発展の方向を暗示している。

このセンサス・ユース・スタディの内容を要約すれば，都市の小地域についてのいろいろなしかも厖大な情報をコンピュータ向に加工・蓄積し，必要に応じ検索しようとするものである。この特徴の一つは地理的な2次元表現をコンピュータの情報として蓄積することである。出力するさいにはグラフィック化も必要となる。もう一つの特徴は，この情報を他の機関の情報データと関連させ，加工しようとするものである。

第3には，これらの計画との関連で都市計画のような新しい創造的な情報へと発展する可能性が開かれているということである。

記事情報検索システムも，このような発展的構想を参考にして検討することによって一段と現実的な方向に進むであろう。

2.3.3　情報検索の基本的技術の動向

情報検索の有効な使い方は，その基本的な技法の開発によって，飛躍的に発展して行くものである。その動向をとらえるために，2，3の例をあげてみよう。

第1に，KWIC（Key Word In Context），KWOC（Key Word Out of Context）がある。この有名な検索技法は文章中の語そのものをキーワードとして，索引言語とする。このキーワードをアルファベット順に配列すれば，索引表が作成できる。キーワードを中心に整理し，その側に前後の文を示すことにより，文脈（Context）の中で占めているキーワードの役割が解かるようになっている。KWICを使った索引は，原文献，抄録，標題のいづれからでもできるが，一般には標題のみを対象にしている。現在，随所でみられる索引分類や情報検索は，キーワードの管理が比較的容易にできることからKWOC方式が圧倒的に多い。また，コンピュータを用いて行なうKWIC検索は，人手でやるのにくらべ適合率は低いが，現在実用的にはまずまずの段階である。

記事情報のような場合には，文章の自動索引は有効であるが，問題は日本語のKWICがない点である。日本語は漢字・平仮名・カタ仮名があり，また専門分野とくに，コンピュータの分野では外来語が多く用語が乱れているので，自動索引は容易でない。これらの点については，将来解決すべき問題であろう。

第2に，SDI（Selective Dissemination of Information）がある。SD

Ｉは有効な情報を各部門および関連部門へ，新鮮な情報としてサービスするシステムである。これ自体は情報検索ではなく，情報検索と並行して有効な効果を生みだすものである。これには３つの情報として，通知用，抄録用，応答用がある。その利用の程度によって，抄録で済む場合は抄録で，原文が必要なときはコピーまたはマイクロ・フィルムで提供することができる。肝心なことは，利用者の関心の度合を常に管理して，ファイルを取捨選択できるところにある。このシステムはかなり実用化されており，ＮＡＳＡやＡＳＣＡやＩＢＭで稼動中である。

ＳＤＩサービスを本格的に行なうためには，厖大な作業量と巨額のコストがかかるので，商業ベースですすめることはむづかしい。しかし，ＳＤＩでの情報のサイクリングや重要度による利用方法の考え方は記事情報検索システムにも取り入れられるべきであろう。

2.3.4　記事情報検索システムの今後

文章による記録を中心とする記事情報検索システムの今後の発展を展望するために，利用者の動向なり要求なりをみて行くことにしよう。

第１に，オンライン処理が容易にできること。情報には２種類あって，過去の記録に価値を見つけるものと，最新の情報に価値をもつものとがある。ほとんどのものはバッチ処理で間に合うはづであるが，人間はより便利なものを志向する欲望があるので，オンラインリアルタイム処理を目指すようになろう。

第２に，ＴＳＳのように他方面から多くの人が利用できること。記事情報を利用する層と数にもよるが，問い合せ方式やマイクロ・フィルム索引なども含めた，情報の段階的利用のためにＴＳＳは当然志向されよう。

第３に，漢字やひらがな，それに文字の大小のような一般のコンピュータ出力にないものが求められる。とくに，日本語の場合，漢字変換は検索の効果を一段高めるが，変換の技術が，問い合せ方式やマイクロ・フィルムによる索引技術なども含めた精密な情報を索引できるような段階へと進むであろう。とくに，記事情報という以上，日本語が中心の特殊なものであり，漢字変換は検索の効果を一段と高める。難点は処理にコストに負担がかかることで，果して採算に合うかどうかである。

第４に，図形処理ができること。

利用する情報として，今後ますます視覚に訴えるものが多くなる。設計図，絵画，図形など情報として，文章よりも，綜合的で，かつ，正確に表現できるものの処理も考慮しなければならない。これらの導入は運営コストおよび技術との関係で決定されよう。

第５に，便利な言語を使用できること。

言語は少くともＣＯＢＯＬ程度の理解で十分使いこなせるものが要求される。

ＴＳＳを使えば，会話型の言語も必要であるし，特殊装置をつければ，専用言語との結びつきも考えなければならない。他の機械および将来のことも配慮して，互換性および継続性の高い言語で，しかも，やさしい言語が望ましい。

第６に，上記と関連して，次のような周辺機器に簡単に接続できること。

1)　マイクロ・フィシュ

2) EVR

3) キャラクタ・ディスプレイ

4) 音声入出力装置

などである。これらの周辺装置はすぐ必要といえないが，今後ともその性能を充実して行く傾向にあるので，常に，これらのシステムと継続性および互換性がきくように準備しておくことが必要である。

一方，提供側の方から，その動向なり，要求なりを探ってみると，

第1に，ハードウエアおよびソフトウエアの使い方が容易であること。

たとえば，バッチ処理はできるが，オンライン・リアル・タイム処理はできないというソフトウエアより，両方できるソフトウエアの方が評価は高い。また，機械によって限界があるよりも，拡張・転換が可能である方が望ましい。

第2に，分析・抽出が容易にできること。

KWICのように一部自動化の実例もあるが，そこまですすまなくても，論理構造が堅固で正しい分析・抽出が容易にでき，しかも情報検索の理にも適う方法があればよい。当然，こういう面に今後の研究が向けられて行こう。

第3に，すぐれたライブラリアンの確保・養成を行なえること。

情報検索システムは結局，人間が索引方法を決定し，そのシーソラス作成能力を駆使して，つくり上げるものである。従って，その作業を効率的に行なえるライブラリアンの確保とその養成とは緊急事の重要な問題である。まして，記事情報のように，多元的で，複雑で，情報量の多い分野では，とくにその良し悪しはライブラリアンにかかっている。

第4に，情報の収集が容易で，コストが低廉であり，しかも，権威ある情報を大量に管理できること。各機関が協力して，重複のない収集・維持体制を築くことが肝要である。さらに個別の収集機関だけではなく，いくつかの機関が将来結合できるように体制づくりしておくことも必要である。

以上，利用者側および提供者側の動向なり，要求なりをみてきた。今後の記事情報検索システムは，こうした双方の要求を汲み入れながら，発展して行くものと思われる。

その中で，とくに記事情報検索システムとして整えるべき条件を若干あげてみることにする。

① 文章分析の発展

文献などの情報の主体は文章であり，これを分析し，評価を行ない，分類や抽出を行なって抄録化して行く作業が記事情報検索システムの最初にして，最大の重要課題である。

これを支障なく行うためには，優れたライブラリアンを確保しているか，あるいはすでに一連の作業基準が確立しているか，いづれかであろう。後者の場合にはその一環として，シーソラス作成の自動化も日程にのぼってこよう。しかし，現状ではまだ人間作業に頼るところが大きい。文章分析を高めるためには，言語学，記号学，論理学などの基礎的な専門分野での発展の拡がりに俟つところ大である。

② 漢字印刷など周辺装置の発展

記事情報検索システムでは文章を中心とする以上，日本の特殊性として，漢字による入出力は当

然必要になってくる。この研究でもそれを実験的に試みている。記事情報検索システムでは漢字印刷，さらには近い将来，漢字のパターン認識も含めて，欠くことのできないものとなろう。

通常の検索結果は印刷出力であるが，将来は単なる印刷では済まなくなり，前にも述べたように，グラフィック化，または，場合によっては音声による入出力機器を使って検索する必要も起こってこよう。方式もバッチ方式から，リアル・タイム方式，次いで応答可能な会話型ＴＳＳ方式が使われるようになろう。記事情報検索システムが実用化される段階で欠くことのできない条件として，このようなものがあろう。

③ メモリの大容量化とハードウエアの進歩

情報検索の基本的前提条件は大容量のメモリが必要であろう。記事情報検索システムは単なるデータの情報検索システムよりも，構造的にも複雑であるし，文章的にも長い。集録される情報の量は厖大なので，当然大容量のメモリが必要である。単に，大容量ばかりではなく，経済的にも割安で技術的にも簡単でなければならない。このような容量の拡大と，それに伴うハードウエアおよびソフトウエア上の発展は蓄積され利用される情報の量を拡張して行くことになる。これは前に述べたように長文の記事情報検索システムにとって，大きな価値をもつ。また，処理の上でも，加工やチェック機能など，利用面を充実する各種機能も付加されることになろう。これらは，記事情報検索システムにとって基本的条件といえよう。

④ 他システムとの関連（国際化も含む）

センサス・ユース・スタディにみられるように，他の機関，または，他の情報の利用が今後ますます生じてこよう。すでに，日本科学情報技術センターと米国のＭＥＤＬＡＲＳとの情報交換協定計画のような実例も生れている。このようなテープ交換による国際的情報交換サービスは今後ますますふえよう。そして，蓄積された情報データが総合的に利用され，創造的な計画策定の方面にも大いに活用されて行こう。こうした傾向を助成する基本的な条件は，標準化，互換性などであろう。記事情報検索システムを整えるべき基本的条件の一つとして他システムとの関連は，それだけ情報検索の価値を高めることになろう。

記事情報検索システムの今後の発展の動向を握る鍵は，利用者および提供者の動向や整えるべき条件で述べたものにつきる。しかし，最も大事なことは，記事情報の価値感という問題であろう。現在は，記事情報検索に対する要求は大きいが，といって記事情報の中から有効な価値を見出すことはむづかしい状態である。これは初期的現象で新聞が近代価値をもってきたと同様，やがて記事情報検索システムは誰からもその重要性を認識され，その価値を認められるようになるであろう。記事情報検索システムが発展するかどうかの最大の鍵は，以下のような項目に集約できよう。

第１に，記事情報の価値の増大。

第２に，情報検索技術の向上。

第３に，編集ライブラリアンの能力向上。

第４に，権威ある情報を大量に廉価に収集維持できる体制の確立。

第５に，国際的記事情報との結合。

現在のところコンピュータに関する記事情報提供サービスは，それほど市場価値をもっているとは思えない。記事情報に価値があるかどうかは，記事情報検索システムの発展の最大の鍵である。政治的，経済的，社会的，環境の変化に対する利用者の必要性と記事情報検索システムの内容自体が充実されていったとき，記事情報検索提供サービスは重要な位置を占めるであろう。

事実，環境は情報化社会として，次第に，記事情報検索システムの発展を促す方角に向って行きつつある。記事情報検索システム自体も質量ともに充実する方向に向っているので，それが提供者側からみて，買手市場から売手市場に変わって行く機運に乗ったとき，記事情報検索システムの真価が発揮できるにちがいない。その利用分野は情報の価値からみて，専門的な分野から徐々に拡がって行くであろう。たとえば，権利義務を明確にする判例・特許，公共的な行政・経済・犯罪，学問的・技術的なものでは生物学・医学・考古学・歴史，商業取引の不動産取引・信用取引などがあろう。ここで，コンピュータを中心とした記事情報検索システムがすぐにこれらを受入れられるかどうか断定しがたいが，この研究で試みた実験は，記事情報提供サービスを行なうために必要な，検索および管理システムの有効性を検討するうえで重要な資料となろう。

記事情報検索における汎用データ・マネジメント・システムの活用法

3. 記事情報検索における汎用データ・マネジメント・システムの活用法

3.1 情報検索システムの問題点

大量に蓄積されたデータから，必要なデータを，必要な時に，必要な形で選び出して利用しようという情報検索システム（Information Retrieval System）の概念が発表されてからすでに十数年の年月が経っているので，多少のニュアンスの違いがあるにせよ，その考え方がどんなものであるかということは，一般に理解されている。

情報検索は，その概念が多くの利用分野に普及しており，コンピュータを利用した情報検索システムの必要性が叫ばれていながら，あまり実用化されていないことも事実である。

なぜ情報検索システムは，その必要性が要求されていながら，実用的なシステムができないのであろうか。これには，いろいろな原因があげられるが，それらを下記の5つに要約することができる。

1) 情報検索のシステムとしての本質的な難しさ。
2) シソーラス作成の難しさ。
3) ハードウエアの問題。
4) ソフトウエア（特にデータ・ベース）の問題。
5) データ通信活用の問題。

3.1.1 情報検索システムとしての本質的な難しさ

情報検索システムには，他のアプリケーション・システムと異なった難しさがある。それは情報検索システムが，システムを作成し維持する側とシステムを利用する側とで，システムに対する考え方で合意をみることは難しく，往々にしてシステムの作成者は，利用者と無関係にシステムを作成してしまうことが多い。また情報検索システムは，システムを利用する対象者が不特定であるため，利用者の範囲もまちまちであり，システムを設計する場合の対象やレベルが不明確になりやすい。この他システムの利用頻度，データの更新やメンテナンス費用，情報の検索コストの決定，検索コストとメンテナンス・コストのバランスの調整などわかりにくい問題が多い。これは情報検索システムが，他のアプリケーション・システムと異なる本質的な特徴であり，このため他のアプリケーション・システムにない難しさがある。

3.1.2 シソーラス作成の難しさ

情報検索システムの中でも特に記事情報検索システムは，利用者が不特定多数であるため検索情報の問い合せに使用する言葉を統一することが難しい。またシステムとしては，ファイルに同一データをいろいろな形で蓄積することはファイルや検索の効率を低下させることになるので避けなければならない。このため蓄積されるデータは唯一種類であり，これを検索する言葉（キーワード）も唯一であることが要求される。このようなシステムの設計者と利用者の相反する条件を解決するために，利

用者からのいろいろな形の問い合せを整理し，唯一の言葉に翻訳してシステムに渡すためには辞書が必要となってくる。設計者と利用者の間のインタ・フェースをはかるものとして辞書を作成するが，これがシソーラス（見出し語辞典）である。より良いシステムの活用を計るためには，良い辞書（シソーラス）を作成しなければならない。それでは良いシソーラスとは，どのようなものであろうか。この問題は，利用者の立場で考えるならば，

1) インプットが簡単であること。

2) 日常会話に利用しているような自然言語であること。

3) 問い合せの順序やキーワードの長さが自由であること。

といったことであるし，システム設計者の立場で考えるならば，

1) コンピュータが取り扱える言語や記号であること。

2) 同一の事象を表示する言語は，一通りしかないこと。

3) 問い合せの順序やキーワードの個数，組み合せの方式などが決まっていること。

4) キーワードの長さが限定していること。

などがあげられる。このようにシステムの設計者と利用者とでは全く異なった要求が起こってくる。このためよいシソーラスを作成することは，この両者のギャップをいかに少なく，効率良く連結するかということになり，言語形態として自然言語にするか，索引言語にするか，キーワードの構造をどうするかなど難しい問題があり，検索システムの実用化を遅らせている原因となっている。

3.1.3 ハードウエアの問題

情報検索という言葉が発表された当時のコンピュータ・システムは，紙テープや磁気テープを中心としたシステムであり，コンピュータのアクセス・タイムも２～１０ｍｓと云う遅いものであった。当時は必要なデータを検索しようとすると，人意的なファイルの選択やファイル間の照合などの作業に手間がかかり，検索時間も数十分～数時間必要となるため，検索システムの必要性が要求されながら実用化ができない状態にあった。しかし，今日ではコンピュータの検索時間も速く，大容量の記憶装置の開発とランダム・アクセス方式の採用，端末機器や通信回線の発達などにより，十分といえないまでも，ハードウエアの問題が，いくつか解決できるようになってきた。このため最近になって各分野において実用化をめざした情報検索システムの開発が着手されている。

3.1.4 ソフトウエアの問題

ここ数年来問題となっているＭＩＳ（経営情報システム）の実現をはかるために最近データ・ベース・システムとか，データ・マネジメント・システムの開発が行なわれている。この思想は従来のファイリング・システムと異なり，多目的なシステムに対応できるようにプログラムとデータを独立させることによりシステム環境の変化への追従を容易にさせるものである。また，データの更新やプログラムのメンテナンスも簡単に行なえるようになり，メンテナンス・コストを引き下げられると考えられている。このため多目的なデータ・マネジメント・システムは，コンピュータ・メーカを中心として開発が進められ，着々とその成果が発表されてきている。この考え方は，大量のデータを取り扱う情報検索システムにとっても，システム効率を向上させるうえに要求されることであり，ＭＩＳを

実現するために必要なニーズとなっている。さらに，従来の情報検索システムの実用化に対するソフトウエア上のネックとされていたファイリングの問題を解決するための糸口となるものである。しかしハードウエアに比べ新しいデータ・マネジメント・システムは，より多くの解決しなければならない問題（技術的な問題や使用制限に関する問題など）を含んでおり，その活用法の研究もいまだ不慣れなため，１日も早くデータ・マネジメント・システムにおけるファイリングの応用法に対する研究がなされなければならない。

3.1.5　データ通信活用の問題

通信回線や端末機器の発達は，コンピュータのオンライン化を可能とし端末を通してコンピュータと自由に会話を行なうことができるようになった。オンライン処理は，従来のバッチ処理に比べて応答速度が速く，しかも，複数の仕事を同時に処理してしまうという点で便利になったということができるし，情報検索システムを活用した情報提供サービスの分野においても要求する情報を即時に提供することができるため，利用者のシステムに対する期待を増大させ，より広い需要者の獲得を可能にした。しかしデータ通信を利用した検索システムが開発されたのは，ここ数年前であり，通信回線の認可の制度や範囲，費用などにも問題があったため，いまだ特定の専用システムを除いては，あまり実用化されていない。

この問題は，情報提供サービスの分野の発展を阻害する原因ともなり，開発されたものの多くは，バッチ処理によるものであったため，情報提供のタイミング，提供の仕方などで利用者を満足させることができなかった。しかし，最近の通信回線の開放や端末機器の発達は，人間とコンピュータとの間において直接コミュニケーション（マン・マシン・コミュニケーション）を可能にしたので，情報検索を主体とした情報提供サービスにおける通信回線の利用度は極めて大きいことから，今後ますます情報検索の利用分野は拡大されて行くであろう。

また，情報検索では大量のデータ作成を行なうが，最近の傾向としてパンチ・カードや紙テープなどの入力媒体を使用してインプットする方法から，キーツー磁気テープやディスプレイ装置などのように直接入力する方法に進みつつあり，そのために手軽に使用できて，安価な装置の開発も進んでおり，ますますデータの取扱いは便利になってきたと思われる。

3.2　記事情報検索とデータ・マネジメント・システム

前述のように情報検索システムは，他のアプリケーション・システムと異なった問題を含んでいるが，なかでも記事情報検索システムでは，

1）大量のデータの作成。
2）必要なデータの選択と取り出し。
3）情報の分析と加工。
4）その結果の表示。
5）データやシソーラスの更新。
6）ファイルの保守。

などの処理を行なってゆかなければならない。

また，現代のような伸展のめざましい情報化社会における検索システムは，データの変化や環境・ニーズの変化に追従して行かなければ利用価値のないものになってしまう。このためシステムを常に活用できる状態に保守しなければならないが，磁気テープ・ベースのシステムのように，データの更新のたびにファイルを作り直したり，環境変化にともなってその都度関連するプログラムを訂正したのでは，メンテナンス・コストが厖大なものとなり，システムの運用が成り立たなくなってしまう。このような不合理を解消するため，これらのデータや環境の変化に対し最小限の変更で済ませることのできるデータ・マネジメント・システムが要求され，同時にデータ・ベースと端末装置をより効率良く利用するために，より良いオペレーティング・システムの開発が要求される。また記事情報サービスを効果的に運営するためには，記事情報のとらえ方，環境変化にともなう体系づけの変更など弾力的な処理の可能なシステムにしていかなければならない。このようなわけで記事情報システムを運用するためには，従来の磁気テープを利用したファイリング・システムのような融通性の乏しいものでなく，多能性のあるデータ・マネジメント・システムが要求される。

3.3　汎用データ・マネジメント・システムの現状

記事情報検索システムに要求されるようなデータ・マネジメント・システムやオペレーティング・システムを開発することは，情報検索システムを製作するうえで不可欠な要素である。しかし，これらのソフトウエアをユーザの手で開発することは，容易なことではない。コンピュータ・メーカやソフト・ウエア会社でない応用分野においてこれらのソフトウエアを開発することは，能力・工数・費用の点からも非常にむつかしいので，一般には，コンピュータ・メーカやソフトウエア会社で開発した，汎用システムをそのまま利用して応用分野の検索システムを製作するか，あるいはメーカの保有する汎用マネジメント・システムを検索システムにあわせて改良するかを考える必要がある。

現在発表されている汎用データ・マネジメント・システムは，オペレーティング・システムのもとで，ファイリングするデータ・ベースの管理，データや入出力に関するデータの管理，オンラインでの対話を可能にするためのコマンド・プログラムや機密保護など，いろいろな機能を持っている。これらの機能を遂行させるためにはシステムとしての多くの制約条件があるので，ユーザが，汎用データ・マネジメント・システムを利用したシステムを設計する場合には，適用範囲の制限や機能上の問題でその制限にぶつかることが多い。（ＳＤＣ社のＴＤＭＳは，システムの中にオペレーティング・システムを持っている。）

このように活用面では，現行の汎用データ・マネジメント・システムはまだまだ不完全であり，これを利用してデータ・ベースを設計したり，アプリケーション・システムを設計する場合は，従来のファイル設計以上に十分な検討と研究が必要となり，この検討が不十分であったために，設計できなかったり，設計できても効率を極端に低下させてしまい実用に供さない例が少なくない。

3.4 汎用データ・マネジメント・システムの活用法

現在開発されている汎用データ・マネジメント・システムは，適用範囲や機能上の制限が多いので，これを活用するためには，今迄のシステム設計以上に十分な研究が必要であることは前節で述べたが，この他，従来取扱われてきたような単一目的のファイル設計と異なり，多目的ファイルとしてのデータ・ベースを使用するので，データ・ベース設計の良し，悪しがシステム全体の効率を左右することになる。このように，汎用データ・マネジメント・システムを利用したシステム設計においては，ファイルをどう設計するかということが，一番大切な問題となる。

汎用データ・マネジメント・システムを活用してファイルを設計する場合には，つぎの点を考慮しなければならない。

1) データの構造をどうするか。

2) データ・ベースにレコードをどう定義するか（１つのレコードのうち，キーを何にするか，レコード中のフィールドのレベル分けをどうするか）。

3) ファイルの編成と処理方式をどうするか（シークエンシャルか，ランダムか）。

4) バッファ・サイズをどの位にするか。

これらの問題は，ファイル中のデータの性質，処理のタイミング，使用のされ方によって決められる。すなわち従来のファイル・システムと異なり，単一の目的のためにファイルを作成するのではなく，多くの目的に共通に使用できるファイルとして，データ・ベースを作成することになる。システムを効果的に運用するためには，それに関連する各システムの利用頻度や検索のされ方，データ・ベースの大きさ（ファイル・サイズ），ファイルへのレコードの追加・修正の頻度などを考慮した上で設計されなければならない。一般には，データ・ベースを取り巻くシステムのうち，使用頻度や更新頻度の高いものに合わせてレコードのレイアウトやファイルの編成を考え，使用頻度の低いものについては，そのレコードのレイアウトとどう接続するかを考えることにより設計を行なう。汎用データ・マネジメント・システムを活用した具体例としては，９章で富士通提供のＲＡＰＩＤ（Retrieval And Production for Integrated Data）を利用したコンピュータ情報・検索システム（ＭＡＳＣＯＴ；MAnagement and retrieval System for COmpuTer's information）についてその応用実験の結果を述べる。

データ収集の方法

4. データ収集の方法

4.1 記事情報の収集法

記事情報検索システムの設計を成功させる第1歩は，記事情報の収集をいかにうまく行なうかということである。この段階でいくつかの考慮すべき点があるが，それらについて以下に述べてみよう。

第1に，情報の利用目的である。無尽蔵にある情報を加工し，蓄積し，検索するにはそれに価する利用目的があって，それに基づく選定基準を明確にしなければならない。次元のちがう情報に対する選定基準を設定することは容易でない。ある程度，実際の収集の過程で個々につくられた基準を蓄積して決められることになろう。

この利用には，過去の記録のうち現在高い価値をもっている情報ばかりではなく，将来の質問に応えて，今後価値をもつかも知れない情報の芽も配慮しなければならない。

第2に，情報の利用者である。情報の利用者が誰であるかが，情報の収集源および収集方法を限定する。利用者が高度の専門家であるときは，学術的領域から情報が収集されよう。利用者が一般大衆であれば，学術的領域からの情報は不要である。記事情報の特徴は利用範囲がきわめて広く，利用する層もそれに従って広いことにある。この中で，どの層に焦点をあてるか，まずその利用レベルを決定しなければならない。その選択を誤かにすると，検索システムそのものの存在価値が問われることになる。利用者は大きく分けて，学術研究家，メーカ，ユーザ，関係者，一般大衆などに分けることができる。

利用者の範囲が決まれば，情報源が自動的に定まってくる。情報源としては内容的にみて，論文，説明，発表，議事録，論評，公報，カタログ，写真，設計図，談話，予想記事，統計，辞書，電信文などがある。形式的にみれば，学術研究書，新聞雑誌，官公報，文献，辞典などが代表的なものであろう。

これらの情報源の中から，各々ちがう選定基準により，蓄積すべき情報が選定される。ここで情報の評価が問題となってくる。類似の情報であっても，記事情報は必ずしも同一の表現をとっていないことが多い。たとえば，同一の対象の情報であっても，実際に記事として表現されたものの内容はちがうことがある。その中のどれを選定するか判断がつかない場合もある。情報源については，コンピュータやシステム理論などの発展分野では，当初予想もされないものが記事情報として価値をもつことがある。たとえば，言語学などもコンピュータの関連学問となってきたし，グラフの理論も情報検索で実用的価値をもってきている。今後ともこのような記事情報の対象範囲は拡大して行こうが，利用目的は逆に限定されて行こう。ただ，いくつかの目的を満足するために，目的別ファイル間にはデータ・リンケージがもたらされるであろう。

情報検索におけるメンテナンスの重要性は衆知のとおりであるが，それに関連して，システムを運用して行くためには情報源，それにもとづく情報選定，情報の加工方法などの諸面で継続性をもたな

ければならない。

第3に，情報の収集方法である。情報収集には，論文や雑誌のように一方的な伝達による場合と問合せや調査のように，交互伝達による場合とがある。そのいづれの情報か，または両方の情報かによって収集結果は異なってくる。とくに，収集の体制に大きく関係する。普通は前者の一方的伝達の情報に限定される。

収集の方法はハードウエアの発展によって地域的に拡大する。通信回線の利用が可能になれば，情報の収集範囲も変わってくる。それに伴って，利用方法も，情報源も変ってこよう。一方的伝達方法による情報ばかりではなく，調査のような能動的な交互伝達の情報も技術的・経済的に導入可能になる。

第4に，情報の形態による収集方法の相違がある。情報には文字，記号，数値，図形，音声などがあり，図形には，写真，設計図，絵画などもある。文字にもカタカナ，ひらがな，漢字，英字，その他特殊文字，外国文字などがあり，文字を使った単語，語句，文章などもある。これらの使い方はハードウエアの在り方にも関連する。とくに，周辺装置の，たとえば，ＸＹプロッタ，ディスプレイ，漢字プリンタ，カナ文字プリンタ，などのどれがどのような働きをするかによって，情報の収集する範囲が限定されてくる。これは記事情報が利用者および利用方法に深い関連をもつためである。

4.2　記事情報の加工法

記事情報は記事をつくる人によって，その表現方法が異なるので，定形的表現をとらない。ほとんどが不定形情報である。これをコンピュータ処理に向く定形情報にするのが情報の加工である。必要とする定形情報はコンピュータの機能に合せて，設計しなければならない。

第1に，定形情報として，標準形式を決めておく必要がある。文章にはいわゆる５Ｗ１Ｈの要素，すなわち，Ｗｈａｔ（何が），Ｗｈｅｒｅ（どこで），Ｗｈｏ（誰が），Ｗｈｅｎ（何時），Ｗｈｙ（何故），Ｈｏｗ（どのように）がある。文章の中でこれらの要素を見出し，抽出し要約して定形化を行なう必要がある。

ファイルの大きさや形式は，情報のボリュウムやハードウエアおよびソフトウエアのもつ技術によって決められるが，さきの５Ｗ１Ｈの要素が欠けていれば，完璧な情報の蓄積とはいえないであろう。記事情報を情報検索システムとして，完成するためには，５Ｗ１Ｈの要素を含んだ標準形を設定し，それに向って記事情報を加工することが望ましい。

この標準形はコンピュータによる情報検索の規則に適合するのが条件である。普通，情報検索では，その抽出の度合によって，文章→主題→抄録（Ａｂｓｔｒａｃｔ）→標題（Ｔｉｔｌｅ）→キーワード→索引（Ｔａｇ）の順にすすめられる。

この過程で文章を分析し主題発見，主題抽出，索引作成が行なわれる。この各々を行なうためにいくつかの手法がある。

手法の一つは，分類方法である。５Ｗ１Ｈもこの一種である。記事情報を作成するにはこの方法が最も良いと考えられる。また，学術的分類もある。図書館分類法も有益であろう。その他，属性とその

属性のウエイトづけによる分類もある。あるいはまた，特殊な学問分類もある。記事情報のサービスの目的が何であるかによって有益な方法がどれか決められよう。

記事情報の特性から，発行者名，著者名，引用文献，引用個所，引用日時などは共通の項目として，主題と並んで重要なものである。主題分析で明らかにされた部分から，コンピュータ処理に必要な主題が抽出される。主題の抽出は，主題の数にも関連するが，そのもつ意味が後にのべるシーソラスに位置づけられ，それが文章全体の内容を表現することになる。主題のもつ意味の豊かさによって，検索効果に差がでてくる。さらに，主題の選び方によって他の情報との関連がでてくる。ここにデータ・リンケージの重要性がある。

今までのべたことは，ほとんど人間が行なうものであるが，これを自動的に行なうことが実験的に試みられている。論理および言語の厳密な科学技術の分野では適しているが，複雑かつ不安定なデータ・リンケージをもつ記事情報では厳密な表現が確立していないだけに，自動化はむづかしいのではなかろうか。

第2に，索引作成とシーソラス編成の問題である。情報の中から主題を分析し，主題を抽出し，これを検索しやすいように，キーワードや索引（Tag）を作る過程である。

索引作成は正確には情報収集とは別に取扱かわれるべきであるかも知れないが，データ収集方法を広く解訳すれば，この範疇に含めて考えても問題はないであろう。一般にはシーソラス作成も含めて，この段階の作業は人間が行なうので，ここに扱った方がよいと思われる。

索引作成は3つの点に留意しなければならない。一つは，はっきりした概念を主題の中から求めて記号に変換し，索引やキーワードを作成することである。その場合には，概念そのものを他の情報と区別できるよう明確かつ厳密に抽出しなければならない。

次に，同意語や反対語などの体系をつくることが必要である。これについてはシーソラスの説明として後で述べる。さらに，シーソラスに関連するが，キーワードの上位・下位および同位の関係を明らかにする必要がある。この階層化および位置づけによって，検索システムでの索引効果は倍増される。

これらの索引作成とシーソラス作成は実務的にむづかしい作業である。記事情報では高度な専門技術書からデータを収集できるとは限らないので，上にのべた3つの点をどの資料からも判然することは困難である。

しかし，これらの索引編成が記事情報検索システムでもっともおくれており，その使命を制するポイントとなっている。ここでシーソラスをみてみよう。

シーソラスとはいわゆる見出し語辞典ともいわれており，キーワード（見出し語）の位置づけを体系化したものである。見出し語を同意語，類語，反対語，上位語，下位語，関連語，語源などの面から体系的に明らかにしたものである。

シーソラスの作成は専門的知識を必要とする厖大な作業なくしてはできない。しかし，一度できあがれば，コンピュータの利用によって，飛躍的にその検索効果を発揮する。シーソラスを正確迅速につくるためには，完成された，権威の高い辞典類が必要である。この研究で問題としているコンピュ

ータの記事情報ではこのような辞典がいまだ作られていないという面から他の情報検索システムにおけるシーソラス作りよりむづかしいといえる。たとえば，用語一つをとってもメーカによって，表現が違うし，しかも，コンピュータはハードウエアおよびソフトウエアともに加速度的に発展をつづけており，それに従ってシーソラスも変えて行かなければならないためメンテナンスの繁雑さが問題になってくる。

一部では，シーソラス作成の自動化も試みられているが，まだ実用にまでは至っていない。

シーソラスをつくるさいのアプローチには２つの方法が考えられる。一つは語源的なアプローチで，他は利用上のアプローチである。

記事情報検索システムのシーソラスは厳密性に欠けるので，利用上のアプローチによるシーソラスの方がよいのではなかろうか。

索引作成およびシーソラス作成の内容の決め方は，データ構造に深い関連をもつ。データ構造は他の情報のデータ構造とリンケージすることによって，大きくその情報のネットワークを拡げて行くことができる。

そこで注意しなければならないのは，ソフトウエア技術によって，使えるデータ構造も変ってくるし，それに伴い，索引作成やシーソラス作成も修正しなければならなくなってくる。これらの索引編成はソフトウエアによって，とくに，プログラムによって制限される。また，処理能率もシステムの編成によって変ってくるが，情報をいかに加工するかは，日常言語と機械処理との接点の問題といえよう。

4.3　収集上の問題点

データ収集方法の方法についてこれまで述べてきたが，そこにはいろいろな問題点がある。そのいくつかをあげてみる。

第１に，コストの問題である。厖大な記事情報を収集するには，厖大な作業量と技術力を必要とする。文献を集めるだけでも容易ではない。それを加工し，蓄積することはさらに，コストを増加させることになる。記事情報の検索利用価値との相対的関係でコストの評価がされようが，いづれにせよ，収集のコストだけでも相当の費用を要する。これをどう考えるかが第１の問題である。

第２に，技術力がある。単なる収集ならそれほどむづかしくはないが，収集した資料を死蔵化することなく，情報検索システムの中で使えるように加工することは実は大へんむずかしい作業である。すなわち，言語表現を転換する技術，情報検索システムへ加工する技術，ソフトウエアを適用する技術の３位一体の綜合的技術が必要である。しかし，主題分析やシーソラス作成でみられるようにここでも技術力の画一化を示すことができないのが現状である。さらに，データ収集の自動化も含めて，これらの技術の問題は今後の大きな課題であろう。

第３に，対象外の情報との関連性である。システム化が社会・科学のすべての分野において今後すすんでこようし，別に国際化の要請もある。そういう面でデータ・リンケージを解決するためには，データとしての記事情報の収集と，データ構造の面で他のデータ構造との共通化，標準化をすすめて

行くことが必要であろう。

情報収集活動にもはっきりしたシステム化ができれば，情報源を分散して，メーカはメーカから，官公庁は官公庁から分担してデータを収集する体制もできようし，多角的な活用も可能になろう。

データ・リンケージの方法

5. データ・リンケージの方法

5.1 データ・リンケージの重要性

利用者の要求する情報と，コンピュータを通してデータ・ファイルの中に蓄積されている情報とを結びつけるデータ・リンケージの問題は，情報検索システムの活用面や処理面において非常に重要な問題である。すなわち情報検索システムの設計に際して一番頭をいためるのは，これら検索要求と蓄積されたデータをいかに結びつけるかというデータ・リンケージの問題であり，この良し悪しによってシステム全体の評価がなされると考えても過言でない。このため，データ・リンケージは，データの蓄積面からも検索面からも，一番適したものにしなければならない。

この場合，人間同志であれば，常識的な概念で簡単に判断する情報も，コンピュータを使用した情報処理の分野では，コンピュータが直感的な判断が下せないために，プログラムによる論理的な方法をとって行なわれなければならない。このため，コンピュータの処理できる判断機能（照合・比較・論理演算など）を使用して，コンピュータが理解できる言語で記述した処理方法が必要になる。データのリンケージは簡単で，間違いにくいものでなければならない。蓄積や検索に用いられる情報は，その内容をよく分析したうえで情報の特性を表わす言葉（キーワード）を選択し，コンピュータ処理に適した索引コードに変換される。一般に処理方法や効率の面から，索引コードと情報内容とは，各々の連結をつけながら別々にファイルに記憶・蓄積される。この様なキーワードと索引コードの関係やキーワードと標準的なキーワードとの関係を示したファイルをシソーラス・ファイルとよび，蓄積や検索処理における辞書の役割をはたしている。

シソーラス・ファイルは，辞書の役割を果さなければならないので，その構成は体系的な分類にもとずく階層的な構成か，あるいはキーワードをアルファベットや五十音のようなルールに従って順序付けたキーワード・タイプの構成をとることになり，下記に示すような機能が要求される。

1） 一般的なものから特殊なものにいたる幅広い組織的な構成であること。
2） 情報検索システムの範囲内にある全ての分野を完全に包括していること。
3） いろいろな観点から分類できる柔軟性をもつこと。
4） 思想の連続を示す論理性を有すること。
5） 一般的な表示法（普遍的な表示法）が可能であること。
6） 明示性があること。
7） 便宜性があること。
8） 簡単な表示法ができること。
9） システムの拡張にともなう表示法の柔軟性があること。
10） 標準的な項目の規定と，その項目に関連する項目の連結が可能なこと。
11） コンピュータの取扱い易い索引形式であること。

12）　利用面から扱い易く検索効率が高い索引形式であること。

これらの要求される機能を全て備えたシソーラス・ファイルを作成することは，現実にはかなりむずかしい。そのためには，これらの機能をできるだけ多く活かせるようなファイルを設計する必要があるし，この機能の選択の方法や重点の置き方によって，シソーラス・ファイルの構成を活用面からの体系的な分類タイプにするか，語源的なキーワード構成タイプにするかを選ぶことができる。

シソーラス・ファイルは，いずれのファイル構成をとるにせよ，その基礎となっている分類の概念や形態がはっきりしており論理的に体系化できるものであるならば，少々の環境変化に対しても追従していくことができるが，分類の概念や形態が，表面的かつ不安定なものでは，多少の環境の変化にも常に注意を払いシソーラス・ファイルを修正しなくてはならない。しかし，シソーラス・ファイルにおける分類の概念や形態は，理論だけで組立てられるものではなく，情報検索システムをより効率的に運営するための手段であることから，実用上の可能が高く利用し易いものであり，情報検索システム全体からみて有益性のあるものでなければならない。

5.2　データ・リンケージの作成法

新しく発生し，収集されたデータを分析してファイルに蓄積したり，ファイルから検索するためにデータとコンピュータ・ファイルとのリンケージをはかるデータ・リンケージは，どんな方法で行ったならばよいであろうか。この問題は，シソーラス・ファイルの構造やキーワードの取扱い方などによっていろいろと変化するので断定的な表現はできないが，一般的なデータ・リンケージの手順を述べると次のようになる。

1）新しく発生したデータを収集する。

2）収集したデータを分析し，その情報内容を代表するキーワード（見出し語）を選択する。

3）選択されたキーワードを分類体系にもとずく索引ファイルの標準用語（標準的なキーワード）と照合し，そのキーワードと標準用語との関係をはっきりさせる。

4）インプットされたキーワードを索引ファイルの標準用語に変換する。

5）変換した標準用語にキーワードのデータ内容が入っているデータ・ファイルのアドレス情報を登録し，索引ファイルとデータ・ファイルを作成する。

このうち最も難かしいのは，２）のキーワードの選択と３）の標準的な用語との関連づけであろう。これは，キーワードをどう分類し，整理すればよいかということが，はっきりさせにくい点に起因している。そこでキーワードを分類する場合には，次のような事項に注意する必要がある。

1）大項目から中項目，小項目と細分化して行く場合，体系的な視野に立って整理して行かなければならない。

2）細分化の過程や細分化された項目の重複がおきないように，各レベルにおけるグループ化を常に検討していかなければならない。

3）将来，分類項目やレベルの拡張が起こっても，吸収できるようなカテゴリの決定と表示法の研究をしておかなければならない。

4) 分類方法が理論的な面からだけでなく，実際的な面からも実行可能なものであることを確めておく必要がある。

このように，データ・リンケージを考えるうえには，キーワードの分類とシソーラスの作成が最も重要な点である。次節ではこのシソーラスについて，もう少し掘り下げて考えて行く。

5.3 シソーラスの方法

シソーラスは，情報検索システムにおける辞書であることは前にも述べたが，この他情報検索システムにおける用語の内容を管理する上でも重要である。すなわちいろいろな表現を取るキーワードとシソーラス・ファイルの標準的なキーワードの結びつけを通して，標準的な用語の表現を知ることや，用語とそれに関連する内容を知ることができる。このように標準的な用語と他の用語との関係を関係子という記号を用いて同義語，関連語，上位語，下位語などの形で表現することもできる。

このようにシソーラスでは情報検索システムを効率よく活用するために，索引ファイルや検索式に用いられる用語の管理を行なっており，データ・リンケージやデータのファイリング，検索処理の各段階において有効な手段である。

また，シソーラス・ファイルの構成は，前にも述べたように活用面から体系的に分類したタイプのものと語源的に体系づけたキーワード・タイプのものがある。

活用面から体系的に分類したタイプのものの構成は，十進分類法にみられるように上位概念から下位概念に段々と細分化して行く"木構造"のタイプであり，この長所は，利用者が知りたい概念を探す時に上位または下位の概念から手がかりを得ることができることにある。しかし情報の正確な位置付け（分類体系のどのレベルのどこに位置付けるか）が難しく，新しい環境の変化に対応することが弱いという欠点を持っている。これに対して語源的に体系づけたキーワード・タイプのものの構成は，個々の情報の特性を代表するキーとなる言葉を選択し，その情報のキーワードとして検索の手がかりとしている。これにより複数個のキーワードの選択やその組合せにより概念の組合せが可能となるので多元的な観点にもとずく検索ができるという長所をもっている。しかしこのための用語の整理および統合がむずかしく，1つの概念にいくつかの概念があったり，1つの言葉がいくつかの意味に取られたりして混乱を招くこともある。この問題を解決するためには用語の範囲，限定句，関連語などを規定しなければならない。

また，1つの用語をいくつかの用語の合成と考え，個々の用語に分解し（1つの体系でなく，種々の面に対してそれぞれ異なった体系で分解する），これら1つ1つの用語にそれぞれの特徴を表わしたキーワードを与える方法がある。この方法により，1つの用語をその特性を表わす別の用語で表現することが可能になる。このような分類方法をファセット（面）分類と呼ぶが，9.2 コンピュータニュース・検索システム（MASCOA）の事例研究において，この形をみることができる。

ファイリングの方法

6. ファイリングの方法

6.1 データ・マネジメントの機能

一般にコンピュータを使用して検索システムを設計する場合には，データを蓄積しているファイルとしての外部記憶装置（主として磁気ディスク，磁気ドラム）に，いかに効率よくデータを蓄積し，修正し，検索できるかが問題となる。より良いデータ・マネジメント・システムを設計する場合には，そのシステムの用途や範囲によって最も適合したファイリング・システムを採用することが大切である。しかし，ファイルとどのような形でデータのリンケージをとりながら蓄積・修正・検索を行なえば良いかということは，システムの利用者に情報を提供する方法として，即時処理によるリアル・タイム方式にするか，一括処理によるバッチ方式にするかという処理方法によって異なる。すなわち，データの利用のされ方に適合した形にファイル中のデータを登録・編成・配置・管理して行かなければならない。

そこで，この章では，データを効率良くファイルするために要求される要素や機能について考ゆく。外部記憶装置としてのファイルにデータを記憶させる場合，まず要求される事項は，インプットとして入力されたデータをファイルにどう記憶するか，そのために必要なデータの構造はどうあるべきかということであろう。

一般的なデータ構造の種類には，

線型構造
階層構造
木 構 造
階層木構造
グラフ（ネット・ワーク）構造

などがあり，このデータ構造に対して

シークエンシャルなファイル処理
ランダムなファイル処理

の組合せが考えられる。

最も簡単な組合せの例としては，磁気テープ・ファイルに見られるような線型構造のファイルをシークエンシャルに処理する方法であり，複雑な例としては，グラフ構造のファイルをシークエンシャルとランダムの両方の処理ができるような方法である。

このように，データの構造とファイル処理方式を任意に組み合わせると幾通りかのファイリング・システムが形成できる。現在コンピュータ・メーカやソフト・ウエア会社で発表されている汎用マネジメント・システムは，このような幾通りかのデータ構造とファイル処理方式を組合わせたファイリング・システムを採用している。この他ファイリングに関連して，単なるファイル編成や処理方式だ

けでなく，システムへの適用性・プログラムの簡易性・インプットの簡易性，言語の汎用性，リカバリ（修復）の機能などについても同時に考慮しなければならない。

現在発表されている汎用データ・マネジメント・システムの多くは，データ構造やファイル処理方式の他，

(1) データを記憶・蓄積する機能 — ＳＴＯＲＥ —

(2) データを検索する機能 — ＲＥＴＲＩＥＶＥ —

(3) データを更新する機能 — ＵＰ　ＤＡＴＥ —

などデータ・マネジシントとしての基本的な機能について簡単なパラメータや言語（ホスト言語）を用いて，直接処理できる工夫がなされている。またこれら基本的機能の他，データ・マネジメント・システムをより実用的なものとするために下記に示す付加機能のいくつかが準備されている。

(4) データ・チェック，エンコード，デコードの機能

(5) ガーベージ・コレクションの機能

(6) 端末機からのオンライン問合せの機能

(7) セキュリティや機密保護の機能

(8) ファイル構造のディスクリプタ

(9) ファイルのリカバリ（修復）機能

(10) チェック・ポイント，リスタートの機能

(11) データの加工機能

(12) 作表の機能

(13) アカウンティングの機能

6.2　データ・マネジメント・システムの形態

ファイルの形態をデータ・マネジメントの立場で分類すると表6.2.1のようになる。

表6.2.1　　データ・マネジメントの分類

第１分類	少種類		多種類	
第2分類 / 第3分類	単能的	多能的	単能的	多能的
小量	ケース１	ケース２	ケース５	ケース６
大量	ケース３	ケース４	ケース７	ケース８

第１分類：データ・マネジメント・システムで取扱うことのできるデータの種類（少種類，多種類）

第２分類：データの使用形態（単能的，多能的）

第３分類：データの量（小量，大量）

表6.2.1でもわかるように，ファイルの形態は，８つのケースに分類され，ケース１からケース８に移るにしたがってデータ・マネジメントは複雑な機能を要求される。

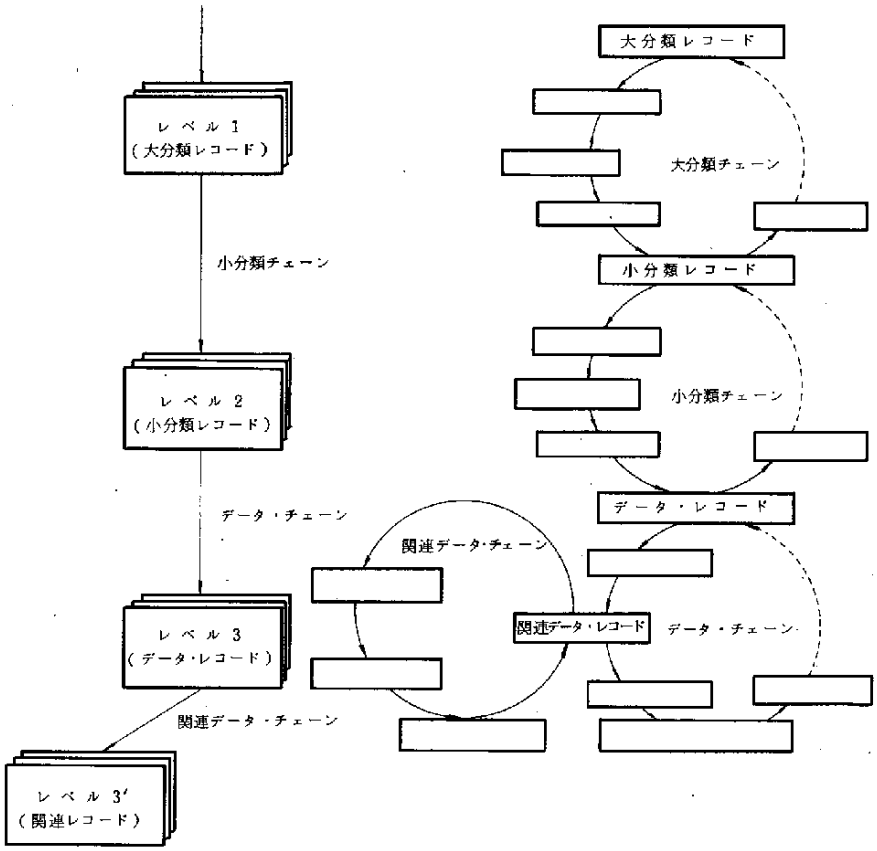


図6.2.1　簡単なファイル・システム

ケース1は，少種類のデータを小量かつ単能的に使う場合のデータ・マネジメント・システムであり，従来のテープ・システムに代表されるような線型構造でシークエンシャルな処理を取扱う場合である。ケース8は，多種類のデータを大量かつ多能的に使用する場合のデータ・マネジメント・システムであり，システム論的には設計できても，実用化の段階でかなりの制約や技術的問題が残る可能性が大きい。このような複雑なケースでは，次のような問題が解決されなければならない。

(1) 多種類のデータを多能的に使用できるようなデータ構造をもたなければならない（グラフ構造の採用）

(2) 多能的に使用するため，シークエンシャル処理やランダム処理が任意に指定できなければならない。

(3) 平均検索時間をできるだけ短かくするようにデータ構造・記憶・処理方式を考慮しなければならない。（右回りの検索，左回りの検索，順序関係，逆順序関係，ページングなど）

(4) 大容量のディスク・ファイル装置を必要とする。

(5) ファイルの保護，データの機密性を考慮しなければならない。

現在これらの問題について，メーカを中心に開発が進められており，序々にではあるが解決の方向へと向っている。

6.3 ファイリングとシソーラス

記事情報システムは，他のアプリケーション・システムと異なり，シソーラスをどうするかが重要なシステム設計上の問題である。シソーラスの作成は，

(1) 記事情報の内容を分析し，それを代表するキーワード（見出し語）の選択

(2) キーワードの表現方法や構成の決定

(3) キーワードをコンピュータ・ファイルに登録するための索引言語（標準的なキーワード）への変換

(4) 索引言語をファイルに記憶する時の形態の決定

(5) 索引言語とデータの連結方法の決定

などの問題を解決しなければならない。

このため起事情報のシソーラスをコンピュータに記憶させるためには，

a) 記事情報を代表するキーワード（見出し語）は，1つで良いのか，2つ以上必要なのか。

b) 2つ以上必要な場合は，その組合せをどうするか。

c) キーワードは，単一語で表現できるか，それとも複合語を含めて表現するか。

d) キーワードを標準的なキーワードにおき変えられるか。

e) キーワードと標準的なキーワードのカテゴリをどう設定するか。

f) カテゴリの分類体系をどうするか。

g) 標準的キーワードを索引コードにするのか，別に考えるか，その場合の標準的なキーワードから索引コードへの変換をどう行なうか。

h) 索引コードをファイルに記憶するのにどんな方法を選択するか。（テーブル・ルックアップ方式にするか，インデックス方式にするか，それとも索引コードそのものを特殊コードにして記憶するか）

i) 索引コードとデータとの接続方法（ポイントの順序づけ）をどうするか。

などの事項について検討を行なわなければならない。

このように記事情報検索システムを設計する場合には，記事情報のデータ本体のファイリングの方法とデータを蓄積したり，検索したりするのに必要なシソーラスのファイリングとを考慮して，システムを設計する必要がある。

ファイル・メンテナンスの方法

7. ファイル・メンテナンスの方法

7.1 記事情報検索システムにおけるファイル・メンテナンスの考え方

情報検索システムに限らず，コンピュータを利用したアプリケーション・システムにおいて，ファイル・メンテナンスの問題は，システムの良否を決定する上に重要な要素となっている。記事情報検索システムにおいても，新しく発生した記事情報をファイルに蓄積する場合，情報のファイルへの挿入や情報と検索キーを連結するためのシソーラスの追加，シソーラスやデータのファイル中でのアドレスの変換など，新しく発生した情報をコンピュータ・ファイルに蓄積するための一連のオペレーションがなされる。この他データ・ファイル中の情報の変更・削除など，ファイルを正しく保持していくためには，頻繁にファイルの更新がなされなければならない。また検索・蓄積などオペレーションの途中でファイルやコンピュータに欠陥が生じて，それ以降のオペレーションができなくなることもある。このような場合もできるだけ速くファイルを修復し，以降のオペレーションを可能にするようにしなければならない。このようにシステムにとって，メンテナンスの問題を度外視して設計を行なうことはできないので，メンテナンス効率を上げることが重要な問題となってくる。

また，ケースによっては，検索効率を上げるよりもメンテナンス効率を良くし，メンテナンス・コストを低減することに重点を置いてシステムを設計する場合もある。このようにファイル・メンテナンスは，システム設計上の重容な要素であるが，次にどのような時にファイル・メンテナンスが必要になるかを整理してみよう。

ファイル・メンテナンスは，大きくデータの更新と，コンピュタやファイルの故障にともなうファイルの修復とに別けられる。データの更新は，

(1) 新しく発生した情報をファイルに蓄積するためのレコードの追加
(2) 不要な情報をファイルから削除するためのレコードの削除
(3) 蓄積された情報の値を変更するためのレコード中のアイテムの値の変更
(4) 蓄積されている情報のレコード構造の変更
(5) 蓄積されている情報の接続の変更
(6) 蓄積されている情報ファイル媒体の変更

などがあげられる。

また，ファイルの修復は，下記のような場合に必要となってくる。

(1) ハード・ウエアとしてのファイル媒体（磁気テープ，磁気ディスク，磁気ドラムなど）そのものの欠陥によるファイルの故障
(2) コンピュータ本体やファイル装置，入出力装置など，コンピュータやファイルを動かす機構の故障によるファイルの破壊
(3) データ処理の方法（プログラムやオペレーション）の間違いによるファイルの破壊

これらの問題にどう対処していけば常にファイルを完全な状態に保持することができるか，また，それを保証できるかということが，ファイル・メンテナンスの重要な点であり，これらの処理をいかに効率良く行なうことができるかがファイル・メンテナンスの設計目的である。

7.2 ファイル編成とデータの更新

新しく発生した情報を追加（挿入）したり，不要な情報を削除したり，データの内容を修正するなどデータの更新処理は，環境やニーズの変化，時間的な状況の変化にともなって頻繁に発生する。このためにデータの更新をいかに効率良く処理するかということが問題となるが，これは情報を蓄積するファイルの編成方式により大きな相違がある。そこで，これらのファイル編成方式によりデータの更新がどう異なるかを次に考えてみよう。

(1) シークエンシャル・ファイルにおけるデータの更新

シークエンシャル・ファイルにおけるデータの更新を考えた場合，レコードに追加・削除がない場合とある場合の2通りが考えられる。

外部記憶装置として磁気テープを使用している場合は，いずれの場合でも更新にともなって全く別のファイルを新製しなければならないが磁気ディスクを使用している場合は，追加・削除のない場合と追加・削除のある場合とでは処理の方法が異なる。前者の場合は，更新データをファイルのシークエンスにあわせてあらかじめ分類しておき，前者のマッチキー（照合キー）によりデータの抽出を行なって該当するレコードだけを修正する。また，後者の場合には，新しいレコードが追加されるためにファイル自身も新製しなくてはならない。これをさけるために，オーバー・フロー・エリアを準備して置くことも考えられるが，純粋な意味でのシークエンシャル処理では通常オーバ・フローの処理機能はもたない。

(2) インデックスド・シークエンシャル・ファイルにおけるデータの更新

インデックスド・シークエンシャル・ファイルにおけるデータの更新は，オーバ・フロー・エリアをファイルの中に持っているので，シークエンシャル・ファイルのようにレコードの追加・削除によるファイルの書き替え（新製）は行なわない。

この場合は，更新データのキーとファイルのインデックス・エリアのキーをシークエンスまたはランダムに照合することにより，直接プライム・エリアの該当レコードを取り出し，データの更新が行なわれる。もし更新データにレコードの追加がある場合は，プライム・エリアやオーバ・フロー・エリアに追加される。

(3) ダイレクト・ファイルにおけるデータの更新

ダイレクト・ファイルにおけるデータの更新は，更新データのキーとＤＡＳＤ（Direct Access Storage Device）上のレコードのアドレスをランダマイジングによって関連づけているので更新データのレコードをダイレクトに取り出して処理する。このためレコードの追加・削除についてもファイルを書き直す必要はなく，直接行なわれるので処理時間は速い。しかし，アドレスの関連づけをランダムイジング・テクニックなどを用いて行なっているので，シノニムの

取り扱いやレコードの更新にともなうアドレス変更など，全てプログラム上で補わなければならないので，プログラミング作業がむづかしいという欠点がある。

このようにデータの更新の場合，データが記憶されているファイルの編成方式により更新の仕方が異なってくる。そこで，ファイル編成の選択を行なう場合には，次の点を考えて選択するとよい。

(1) シークエンシャル・ファイル

シークエンシャル・ファイルの更新は，ファイルに対するレコードの追加・削除の頻度が少なく，ファイル全体のデータ量に対して更新レコードの量が多いものに適し，逆の場合には，不適である。

(2) インデックスド・シークエンシャル・ファイル

インデックスド・シークエンシャル・ファイルの更新は，レコードの追加・削除の頻度，更新レコードの量の大小はそれほど問題とならないが，更新データを探り出す場合にインデックス・エリアとデータ・エリアの2回のアクセスを必要とするので，即時性の強いシステムでは問題である。しかし，ファイルの大きさが大きい，即時性の弱いオンライン・システムには，ダイレクト・ファイルの場合のシノニムの処理などがないので適している。

(3) ダイレクト・ファイル

ダイレクト・ファイルにおけるファイルの更新は，ランダム処理の利点を用いて，1回のオペレーションにおける更新データの量が，ファイルの全体量に対して少ない場合や特定のレコードに頻繁に起こる場合に適する。また該当するレコードを取り出す場合，ランダマイジング・テクニックなどを用いるので，あまり大きなファイルではシノニム処理が大変になることから比較的小さい，即答性の強いリアルタイム・システムに適する。また，ダイレクト・ファイルでは，シノニムの取扱いなどが発生するので，ファイル・スペースを１００％利用することはほとんど不可能であり，通常は７０～８０％位利用される。これはランダム処理においてはファイル・スペースよりもアクセス・タイムに重点がおかれていることを示している。

このようにファイル更新に対しては，ファイルの性質・ファイルの編成方式，ファイルの処理方式（オンラインか，バッチか）などを十分考慮したうえで，システムを考えなければならない。

この他ランダム処理のファイルでは，レコードの追加く削除が直接行なわれるものの更新が度重なると，ファイル上に使用されないエリアやレコード間のギャップが生じたり，頻繁に検索されるレコードがオーバ・フロー・エリアに記入されていたりして，ファイルや検索効率を落すことがある。このため定期的にガーベージ・コレクションを行ないファイルの中味を効率よく管理できるように整理をしておく必要がある。

7.3 データの正当性と信頼性

情報検索においては，データ処理の迅速性と正確性は重要な要素である。そこでこの正確性について考えてみる。

正確性を論ずる前提としては，インプット・データ，データを処理するためのプログラムおよびオペレーション，それにハード・ウエアなど，全てが正しく保持されてなくてはならない。このために

は,

a) エラーを発生させない。

b) エラーを迅速に発見する。

c) 発見されたエラーをメンテナンスする。

などの対策がなされなければならない。このため具体的な方策として，データのチェック，プログラムのデバックやテスト，オペレーション・マニアルの整備，教育，コンピュータの定期検査などが必要となるが，ここではインプット・データのチェックに的をしぼって述べることにする。

データ・チェックの方法には,

(1) 人間の目やモニタによる方法。（読み合せ，パンチ・ベリファイの併用など）

(2) プログラムによる方法。（シークエンス・チェック，マッチング・チェック，フォーマット・チェック，トータル・チェック，検証コードによるチェック）

(3) ハード・ウエアによる方法。（Ｉ／Ｏ命令やチャネルによる再試行）

などが考えられる。チェックを行なう段階も，インプット作成段階，キー・パンチの段階，コンピュータへ投入時の段階など種々の段階が考えられる。

また，オンラインでデータ処理を行なう場合，通信回線が回線雑音や瞬断などにより誤りが発生する場合がある。このためデータの正当性を確保するために誤りを検出し訂正しなければならない。この誤りを検出する方法として最も一般的に用いられるものとして，パリティチェックがある。これには，垂直パリティチェックと水平・垂直パリティチェックがあり，前者に比べると後者は検出能力が高い。

なお，チェック方式については，表7.3.1に示す。

表7.3.1　誤り制御方式

伝を送も情た報せにる記方号法	誤り訂正符号の使用	ハミング符号
		サイクリック符号
	誤り検出符号を使用	垂直パリティチェック
		水平パリティチェック
		垂直・水平パリティチェック
		群計数パリティチェック
		定マーク・定スペース方式
伝に送よる方方式法	返送方式	返送照合方式
		チェック符号返送方式
	連送方式	並列伝送方式
		反復伝送方式

検索の方法

8. 検索の方法

8.1 検索システム設計上の考え方

これまでコンピュータを利用した記事情報検索システムの設計に関する種々の機能のうち

データの収集

データ・リンケージ

ファイリング

ファイル・メンテナンス

について述べてきた。そこで本章では，検索についての考え方や方法について述べることにする。

ここでもう一度"検索"という言葉について考えてみたい。今までわれわれは，検索という言葉を"探しだす"という行為としてとらえてきた。この"探し出す"という行為は，"情報の所在を探す（明らかにする）"という行為と"探した情報を再び取り出す"という行為が含まれているので"探す"行為と"再び取り出す"行為の間に実行される必要な過程が検索であるということになる。しかし，システムによっては単に質問に従って情報をファイルから抽出するというだけにとどまらず，利用者の要求によって取り出した情報を加工・編集して提供したり，さらに，利用者の要求に関連すると思われる情報まで含めて積極的に提供しようとする一連のサービス活動をいう場合もある。

このような状況から総合的に判断すると，利用者に少しでも多く役立つ情報を提供するためには，どのようにしたら良いかということが情報検索の究極の目的であるといえる。

この検索システムの設計には，システムの用途，範囲，蓄積や検索の処理方法などを決定しなければならないが，具体的には，

(1) 誰れが（システム側の人間）

(2) どんな種類の情報を（情報の選択）

(3) どの位の量（情報量）

(4) どんな方法で（蓄積の方法）

(5) どういう分析を行なって（情報の分析深度）

体系的に蓄積するかということであり，さらに検索のために

(6) どんな質問を（質問の様式）

(7) どんな方法で（検索処理の方法）

行なうかによって，システムが決まる。

システム設計上，蓄積される情報と検索される情報が，照合し易いようにすることも重要であり，蓄積と検索の情報を索引ずける場合の概念やレベルに食違いがないか十分検討を要する。とくに記事情報検索システムは，このようなシソーラス作成に関する索引言語の体系化や複合語・同義語・同音異語・異音同語の取り扱い方が検索システムの良し悪しを決めることになる。

この他，検索上の問題としては，質問者のレベルや質問の方法にもとずく質問の分析をどうするか，質問情報のキーワード（見出し語）をファイルの索引コード（標準的なキーワード）に変換するにはどのような方法をとるかなどがある。

8.2 検索システムの運用方法

情報検索システムは，その目的や環境に応じて，いろいろな検索形態が考えられるが，その中でも際立った例として前節でも述べたように蓄積されたデータ・ベースから，要求された情報だけを取り出す受動的なものと，システム側から積極的に利用者が必要と思われる関係情報を提供するものとでは，その検索の方法も異なってくる。

受動的検索は，利用者からの問い合わせのあった情報に対してのみそのキーワードをシソーラス・ファイルからとりだして，必要な索引コードを選択し，それに基ずきデータ・ファイルから該当する情報を抽出する方法である。

一方，能動的検索は，利用者1人1人の関心分野や要求情報に関するファイルを備えており，それを活用して新情報を入手するごとに，システムの利用者に垂流し的に配布しようとするもので，情報の選択提供 — SDI（Selective Dissemination of Information）— とよばれている。

また，一般的な情報検索システムにおける検索の処理方式について述べると，利用者（質問者）からの問い合わせ示指として，端末器や入力装置からシステムに対し，必要情報のキーワード 検索条件，出力の形式，エラーの場合の処理などの指定を行ない，さらに検索実行指令を与える。検索指令に従ってシステムは，まずインプットしたキーワードについて論理的な組合わせを行ない，その結果とシソーラス・ファイルとの突き合わせをする。つぎにシソーラス・ファイルから索引コード（標準的なキーワード）を選択して該当する索引コードをみつけ，それをキーにしてデータ・ファイルから必要な情報を求める。求めた情報は，システム内でキーワードと統合・編集したうえで，指定された出力形式により質問者に回答する。また質問のインプット・ミスや，未登録のキーワードなどがインプットされると，システムはいくつかの照合過程において常に異常をチェックし，その都度その状況を出力装置にアウトプットする。

この関係を図8.2.1に示す。

この他，キーワードに該当項目がなかったり，用語の使い方が不明な場合には，キーワードを分析・解釈して，利用者とシステムとの会話を通して利用者に標準的なキーワードをサポートできるようにシソーラス・ファイルを整備しておけば，より良い情報検索システムが設計できる。このようなサービスのできるシソーラス・ファイルを設計する場合の標準的なキーワードは同位語や同義語だけでなく，そのキーワードの上位のレベルのものや下位のレベルのもの，関連用語などについても含めたデータ・リンケージを考え，他のキーワードとの関係についても表現できるようにまとめておく必要がある。

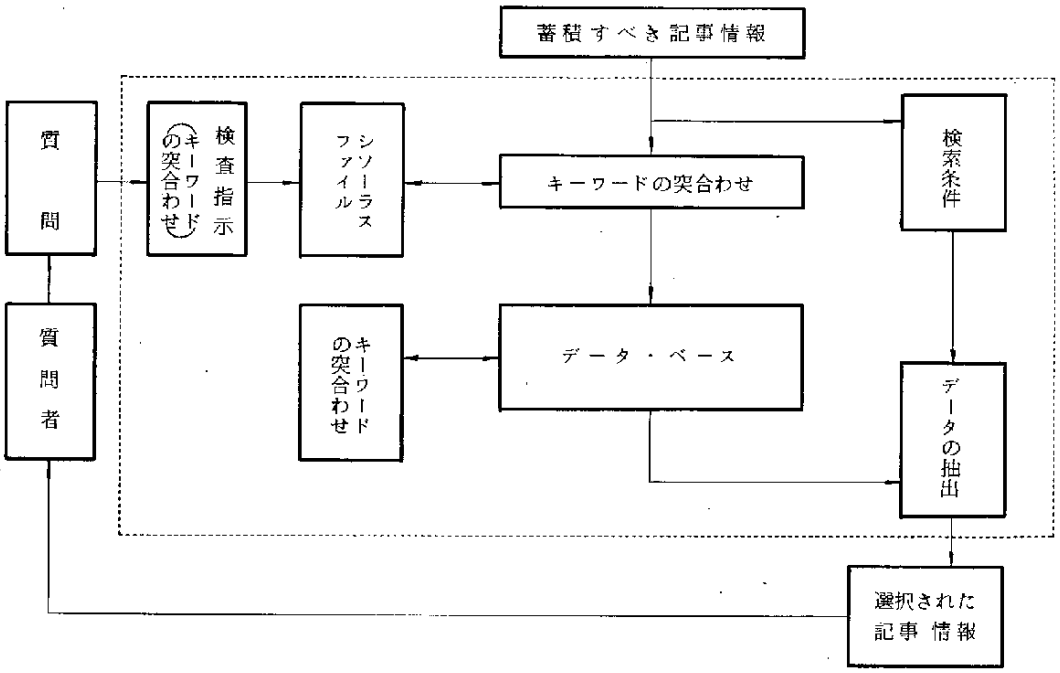


図 8.2.1　情報検索システムの形態

とくにオンラインを利用した情報検索システムにおいては，端末器を通した利用者とシステムの直接的な会話が不可欠であるので，利用者の不慣れな面を少しでも補えるように工夫や，代替用語の提供，用語の解説などの機能を付加することが必要である。

以上，本報告書では，これまでに記事情報検索のためのデータ・マネジメント・システムのあり方とその活用法について検討してきたが，これらの具体的なモデルとしては，コンピュータ関係の記事情報検索をとりあげ，第9章の事例研究で解説している。

事例研究：コンピュータ情報・検索システム（ＭＡＳＣＯＴ）

9. 事例研究：コンピュータ情報・検索システム（MASCOT）

(1) 研究開発への背景

コンピュータを使用した情報検索システムは，情報処理システムの重要な機能の一部分として早くから研究開発がなされてきた。その技法も，ＰＣＳの時代から今日にみるオンライン多重処理の可能な時代に至るまで，多くの方式が生みだされ，実用化されてきた。すなわち，分類コードをキーワードとした検索から，シソーラスの不要な連想記憶による自然語の検索に至るまで，学問の分野からも応用技術の分野からも，多角的に研究が進められている。

また，情報検索を利用面から見た場合には，経営情報システム（ＭＩＳ）を支援する重要な機能として，その果す役割は極めて大きく，意思決定のための情報をデータ・バンクから即時に，しかも必要な加工を施して効果的に提供することを可能にしている。

しかし，情報検索といっても，その取扱う情報の種類によって，また，利用目的によっていろいろな検索方法が考えられる。情報には数値情報，語句情報，文章情報，図形情報，音声情報，……，などの種類があるが，今日までのコンピュータによる検索システムは，主として数値データを取扱う統計情報検索システム，語句データを取扱う文献情報検索システムが，一般的に考えられてきた。情報の蓄積には，磁気テープ，磁気ディスクが主に使われ，シークエンシャル処理，ランダム・アクセス処理により情報の検索が行なわれてきた。

検索機能の必要性は当初，入出力処理ルーチンの一機能として要求され，それがコンパイラ，あるいは専用言語などの高水準言語の中で機能することが要求された。その後，単一目的だけではなく，多くの目的のために使用できるように，共用ファイルをもった検索システムの必要に迫られ，データ・ベースを基盤とする汎用データ・マネジメント・システム（ＧＤＢＭＳ）が作られるに至った。この段階に至って，検索システムは，データ・マネジメント・システムの一機能として取扱われるようになり，データの管理と密接な関連を持つようになった。

これに最近では，バッチ処理だけでなく，オンライン・リアルタイム処理，ＴＳＳ処理の機能が追加され，ＯＮＬＩＮＥ－ＤＭＳや，ＯＮＬＩＮＥ－ＦＭＳが主流を占るようになった。汎用データ・マネジメント・システムは，ホスト言語形式または専用言語形式により記述が容易で便利になっているばかりか，運用面からも多くの効果的な機能を持っているが，実際的な活用面から見ると，いまだ十分使いこなしているところは少なく，今後に多くの問題点をかかえている。

注(1) ＧＤＢＭＳ：Generalized Data Base Management System
(2) ＦＭＳ：File Management System
(3) ＧＤＢＭＳの種類，言語形式などについては４５－Ｓ００１「経営予測のためのデータ・マネージメント」を参照

われわれの周囲に存在する情報のうち，その８０％は不定形な情報であり，これらをできるだけ定形化，規範化して記録し，それを管理することの重要性が最近強く認識されてきている。情報検索の分野でも，文章情報，図形情報の取扱い方が重要視されつつある。人間が記憶したり，人間が管理していたこれら不定形な情報を機械の管理に置換えようとする試みが行なわれている。それらのいくつかの方法の中でも，マイクロフィルム記録，ホログラフィ記録などによる情報検索技術の開発はわれわれの注目するところである。しかし，この分野の多角的利用に対しては，いまだ，実験研究段階であり，実用化までにはまだ数年を要すると思われる。マイクロフィルムやホログラフィ記録によるデータを，コンピュータで管理する情報検索システムでは，データ群を一つのパターンとして，取扱うことができるため，インデックス・テーブルを完全にもてば，大容量のしかも不定形な，新聞記事や文献などの文章情報や，製図やカラー写真などの図形情報も適時検索することが可能になる。このような検索システムが可能になれば，コンピュータは，インデックス・テーブルの管理と不定形な大量情報の入出力管理において，重要な役割を果たすことになるであろう。

しかし，このような大容量記憶のための便利な情報管理システムが多くの利用分野で実施され利用されるまでの過渡的な段階として，不定形な情報のうち，とくに文章で表わされた情報，たとえば新聞記事，雑誌，文献などの記事情報が，現在使用されているコンピュータと周辺機器を使って，どのように管理でき，検索できるかを究明することも重要であろう。

最近では，ファクシミリ，マイクロデータ伝送の技術進歩もめざましく，これらを総合した記事情報サービスが注目されつつある。また，通信回線の民間への開放もあり，通信とコンピュータを直結した新しい情報提供サービスが可能になった。しかし，一概に情報提供サービスといっても，その目的や対象，方法などにより，いろいろなシステム構成が考えられ，それを支援する情報検索も目的とするシステムの状況に応じて，さまざまな形態をとることになる。

一般に，どのようなシステム開発にもいえることであるが，そのシステムの利用価値を計る尺度の規準をどのように置くかによって，そのシステムの価値評価は変ってくる。処理能力の向上，管理水準の向上，省力化，経済性，将来性などの面から，これらの全てを満足するシステムを作りあげることは，おそらく希であろう。処理能力の向上を図れば，通常，経済性はある程度犠牲にせざるを得なくなる。

そこで，検索システムを作るうえにも，その利用目的を限定して，どこのメリットを上げなければならないかを検討し，その重点部分を中心に総合的にシステムの価値評価を行なった上で，開発にとり組まなければならない。どのような種類・内容の情報が，どのような対象・範囲の利用者に，どのような方式，形式で提供され，それがどのような分野の目的に使用されるかによって，検索システムはそれぞれ特徴をもった「何々のための何々システム」といったような専用システムとなる。とくに，情報提供サービスを営利で行なう場合には，このような検討は非常に重要であり，目的に合った情報を適時に選択し検索して，見易い形式で安価に提供できるシステムにするための十分な調査研究をしなければならない。そのような利用サービスを行なうには，データ・バンクを常に有効な利用価値の高い情報で満しておかなければならず，そのためのメンテナンスが非常に重要にな

り，運用管理にはかなりの労力と時間を要することになる。また，検索された報告のサービス効果を高めるために，その表示にはタイプライタ，ラインプリンタ，ディスプレイなどを，緊急度に応じて十二分に駆使したシステム編成を考える必要がある。もちろん，バッチ処理サービスにするか，オンライン処理サービスにするかは，情報の緊急性，形態，量などにより決められることになろう。

このような背景のもとに，われわれはここで記事情報提供サービスのための記事情報検索システムとして，コンピュータ関連の記事情報をとりあげ，これらの大容量情報をカナ文字情報として整理し，各々の情報を限定した長さに要約して，コンピュータで効率よく管理するとともに，必要に応じて即時に出力できるようなシステムＭＡＳＣＯＴを想定し，実用化のための研究を行なうことにより，記事情報提供サービスのためのコンピュータシステムとしての管理技法，検索技法を実験的に考察することにした。

以後，コンピュータ記事情報・検索システムＭＡＳＣＯＴの概要について述べることにする。

(2) システムの概要

コンピュータ情報・検索システム（ＭＡＳＣＯＴ：ＭAnagement and retrieval System for COmputer's information）は，日々発生するコンピュータに関する政策，運営，技術開発，サービスなどの一連の記事情報を，コンピュータとその周辺機器を使用して効果的に管理し，それらの情報の中から利用者のニースに合ったものを選択し検索して，タイムリー報告するためのシステムである。すなわち，新聞，雑誌，文献，辞典などに記載されるコンピュータに関する，ハード面あるいはソフト面の文章情報を，カナ文字，英数字で６６文字以内に整理してデータを作り，それらのデータを使用してサービスの目的毎にファイルを編成し，さらにそれらを統括管理するデータ・ベースを確立し，データ・マネジメント・システムを造りあげている。ファイルは，インデックスド・シークエンシャルまたはランダム・アクセス・ファイルで編成している。これらのファイルで管理できるデータ件数は，およそ３５万である。検索は，会話形式で行ない，必要な情報を選択し，適時抽出できる簡単な専用言語を保有している。検索に必要なキーワードは，記事内容を最も良く表わす語句を数個，その文章の中から選び出すＫＷＯＣ方式をとっている。ＫＷＯＣ方式をとったのは，自然語をキーワードとして文頭から順次検索して行くＫＷＩＣによる方法では，異音同語，同音意語，関連語などの取扱いをコンピュータ側にまかせることになり，プログラムが複雑な割には処理効率が悪く，実用的でないと判断したためである。現在，研究を行なっているようなタイプのシステムでは，入力時に多少の労力がかかっても，キーワードの基本的な管理は人手で行なう方がよい。それは，キーワードとデータの結びつけをデータ収集段階で明確にしておくことにより，検索結果の利用効率を高めることになるからである。検索結果の報告は見易くするために出力項目の配列を考え簡潔に要領よくまとめるように配慮した。報告に緊急性を用しない場合は，情報の理解度を高めるため自動漢字編集システムを使用して漢字情報の出力も可能にした。

これら一連の処理を行なうコンピュータ・プログラムは，汎用データ・マネジメント・システム（ＲＡＰＩＤ：Retrieval And Production for Integrated Data）― 富

士通提供）をベースにしたコボル語によるプログラムとアセンブラ（ＦＡＳＰ）による独自のインバーテッド・ファイルをもった専用プログラムの２つの面から作成し，これらを比較検討してシステムの運用面での効率テストを行なった。

ＭＡＳＣＯＴは，ＦＡＣＯＭ２３０－６０ＴＳＳモニタＶの下で稼動するディスク・ベースのマネジメント・システムである。現段階においては，ＯＮＬＩＮＥ－ＤＢＭＳが富士通から提供されていないので，バッチ用のＤＢＭＳ－ＲＡＰＩＤを使用し，記事情報検索を行なうことにした。

ＭＡＳＣＯＴにおける記事情報サービスの主旨は，必要とするコンピュータ情報の所在を明らかにし，その概要を知らせることであり，また，情報源およびその詳細を知りたい向きには，出所，掲載頁などを提示して，その足掛りを教えることにある。ＭＡＳＣＯＴが将来自動検索サービス・システムとして巾広く利用されてゆくためには，当然マイクロシステムとの同期，サブシステム間の有機性が必要となろう。

本来，このような情報提供サービスのためのシステムを開発する場合には，利用面の制約条件をはっきりさせてからとりかかるべきであるが，われわれは，この際，利用価値，利用効果を全く無視してシステム設計に当らざるを得なかった。なぜなら，記事情報のオンライン検索とその提供サービスの実施例はいまだ無く，全く未開の分野であったからである。利用価値を高めるためには，やはり一定期間そのシステムを活用し，幾度か改良を行なって試行錯誤的に最も好ましい利用のパターンを見出して行くのが賢明であると思われる。このような観点から，本システムの研究開発は，あくまでコンピュータ・サイドに立って記事情報をどのように管理し，検索したら効果的か，記憶の効率化，管理の仕易さ，検索の迅速性といった３点からその技法を究明することにした。

ＭＡＳＣＯＴは，サブシステムとして①コンピュータ用語・検索システム（ＭＡＳＣＯＧ：ＭＡnagement and retrieval System for COmputer's Glossary），②コンピュータニュース・検索システム（ＭＡＳＣＯＡ：ＭＡnagement and ritrieval System for COmputer's Article），③コンピュータ文献・検索システム（ＭＡＳＣＯＬ：ＭＡnagement and retrieval System for COmputers Literature）から構成されている。

ＭＡＳＣＯＧは，コンピュータに関連する各用語とその意味を記録し，必要に応じて用語の意味を検索サービスするシステムであり，出典，記載頁，関連用語なども含めてタイプライタまたはディスプレイに表示し報告することができる。

ＭＡＳＣＯＡは，主要新聞あるいは専門分野の新聞に記載されたコンピュータに関連するニュースの要点を，カナ文字情報として記録管理しておくことにより，必要に応じて指定された条件（ＡＮＤ，ＯＲの機能を有し，その組合せ検索が可能）のニュースを会話形式で選択しながら検索し，タイプライタ，ラインプリンタ，漢字プリンタに印刷し報告するシステムである。

ＭＡＳＣＯＬは，内外の雑誌または文献の名称，発行年月，記載内容（題目），所属分野などを英字，カナ文字，記号などで整理し，管理することによって必要な分野の必要な内容が盛込れた雑誌や文献を適時検索し，サービスするシステムである。検索方式および出力様式は，ＭＡＳＣＯＡ

と同様である。これらのサブシステムについては後の節で詳しく説明する。

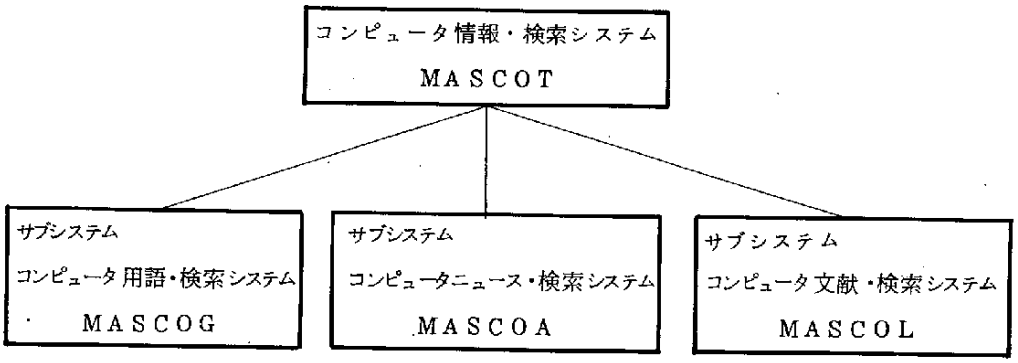


図9.0.1　MASCOTの構成

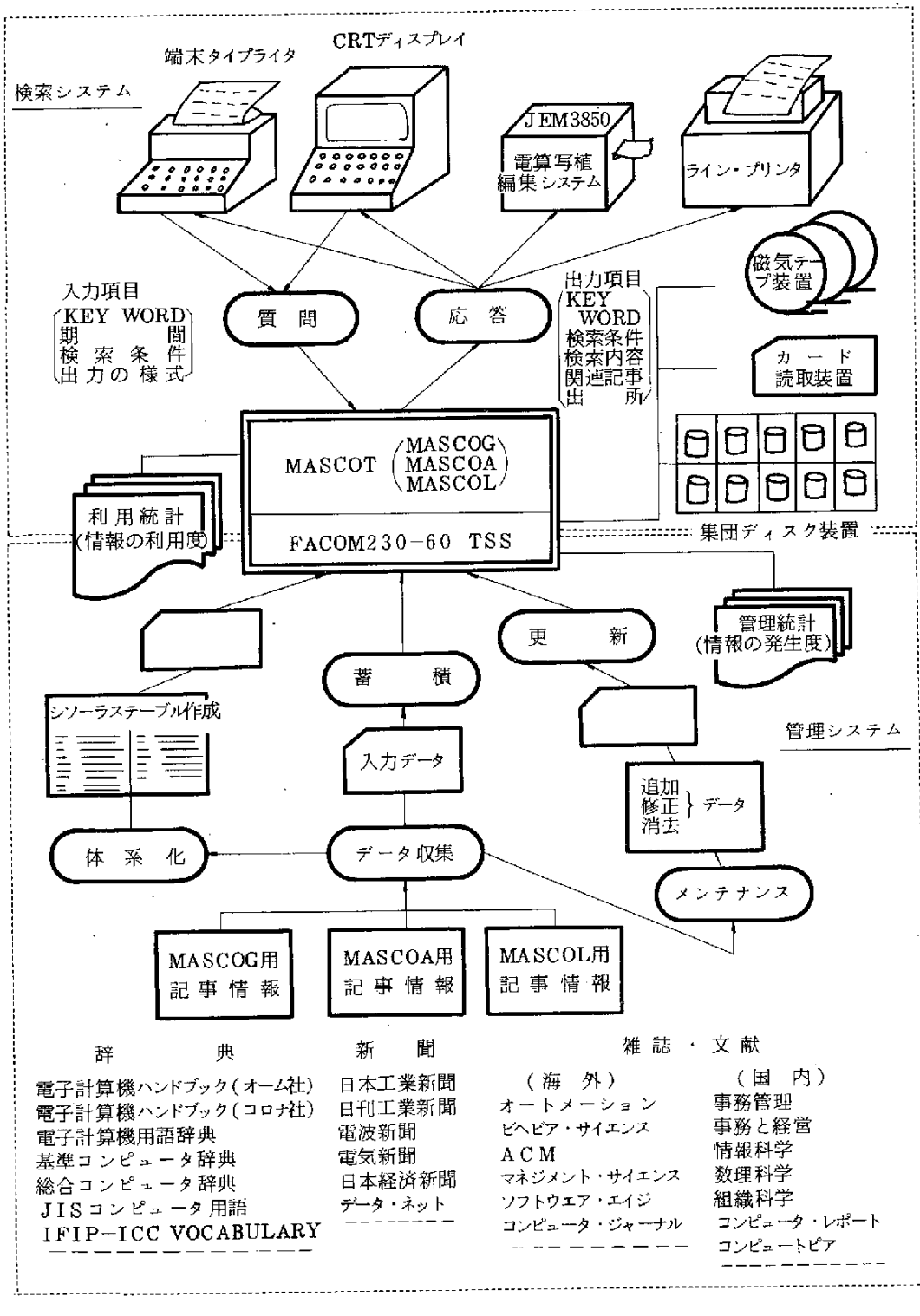


図9.0.2　コンピュータ情報・検索システム：MASCOT

9.1 サブシステム：コンピュータ用語・検索システム（ＭＡＳＣＯＧ）

(1) ＭＡＳＣＯＧの概要

コンピュータ用語・検索システム：ＭＡＳＣＯＧは，コンピュータに関連する用語とその情報を記録し，必要に応じて指定された用語の意味や出所を検索し，提供サービスするシステムである。コンピュータの分野で使われる用語は，

① ハードウエア設計上，製作上の用語
② プログラミング上の用語
③ ＥＤＰ運用管理上の用語
④ ユーザにおける利用分野の用語
⑤ 数学，経営工学などの周辺科学の用語
⑥ 電気，機械などの関連産業の用語

などがある。

ＭＡＳＣＯＧでは，これらの用語を整理し，ファイルに５０,０００語程度管理できるよう設計している。しかし，現在は実験段階でもあるので，使用データは２３,０００語に止め研究を行なっている。用語は８種類の用語辞典から引用しており，これら２３,０００語の中から同意語を整理してそのうちのおよそ１５,０００語を用いてデータ・ベースを作成している。用語は，辞典の用語をそのままキーワードとして使用している。用語間にシソーラスをもつことは，非常に大切なことではあるが，この分野の用語がいまだ定着していない現段階においては，整理が難しいので今回はそれを考えずにデータ・ベースを作成した。１件当りのデータは，用語，用語の意味，出所（辞典名，掲載頁），整理番号などの項目で構成しており，英数字・カナ文字で集団磁気ディスクＦＡＣＯＭ４７１Ｋに記録している。検索は，コンピュータ用語をキーワードとし，英数字・カナ文字を使用した語または句による自然語で行なう。用語の意味は，初心者用，専門家用の区別なく特定の８種類の辞典からその内容を整理し，６６文字以内に要約して入力データとしたもので，意味の難易やその内容が十分表現し切れないところがある。そこで，このような不足の情報を補うために，データの中に出所，掲載頁を盛込んでさらに詳しいデータが欲しい向きには，適当な辞典が選べるような足掛りを提供している。

ＭＡＳＣＯＧは，ＭＡＳＣＯＴモニタの下で働く検索用サブシステムであり，将来は他のサブシステム：ＭＡＳＣＯＡ，ＭＡＳＣＯＬと関連をもたせ，自由に組合せ検索ができるようにする予定である。データ・ベースは，汎用データ・マネジメント・システム：ＲＡＰＩＤを使用して運用している。検索した結果は，問合せ後数秒でタイプライタまたはＣＲＴディスプレイに報告できる。また，漢字文の出力が必要な場合には，電算写植編集システムを用意している。

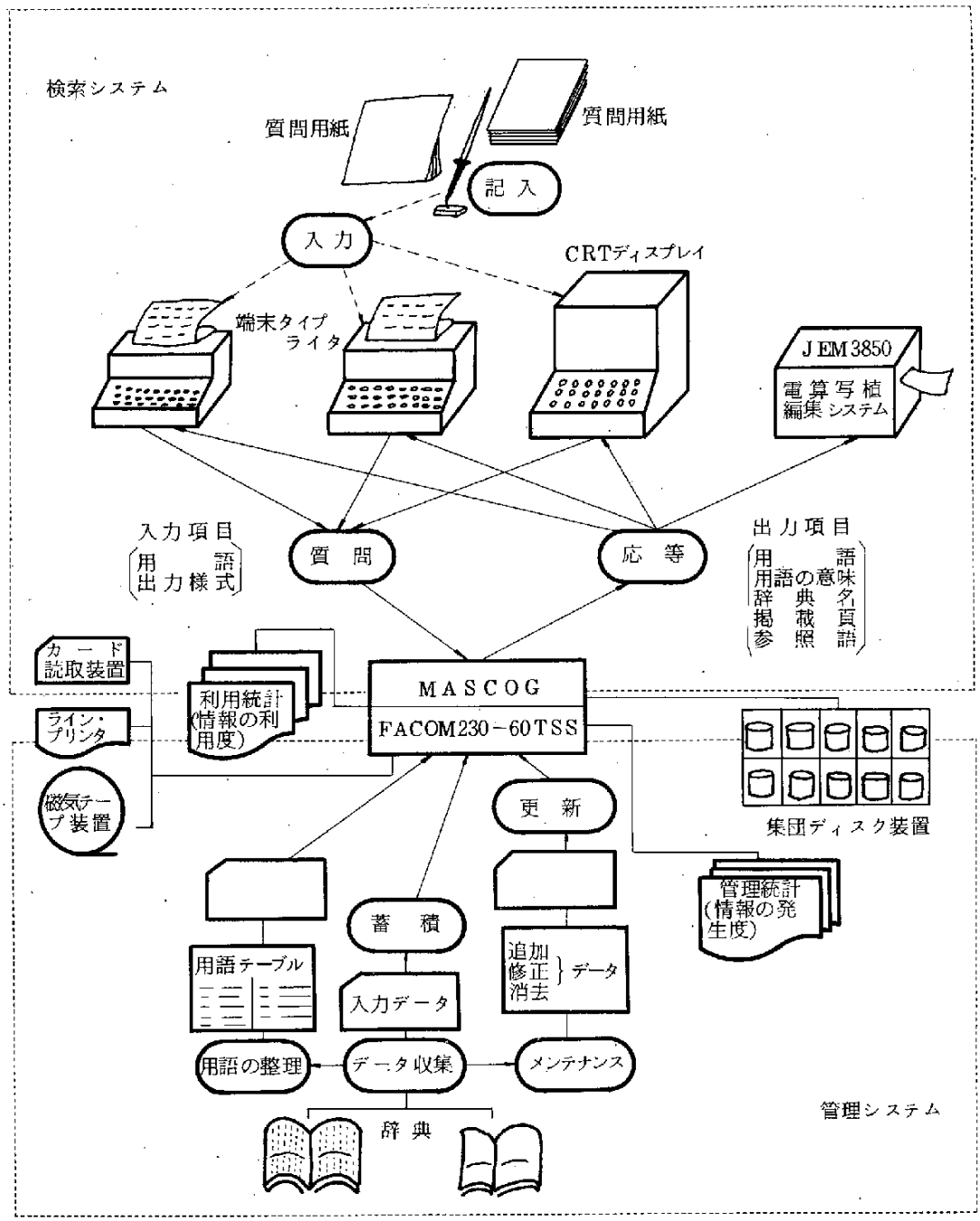


図9.1.1　コンピュータ用語・検索システム：MASCOG

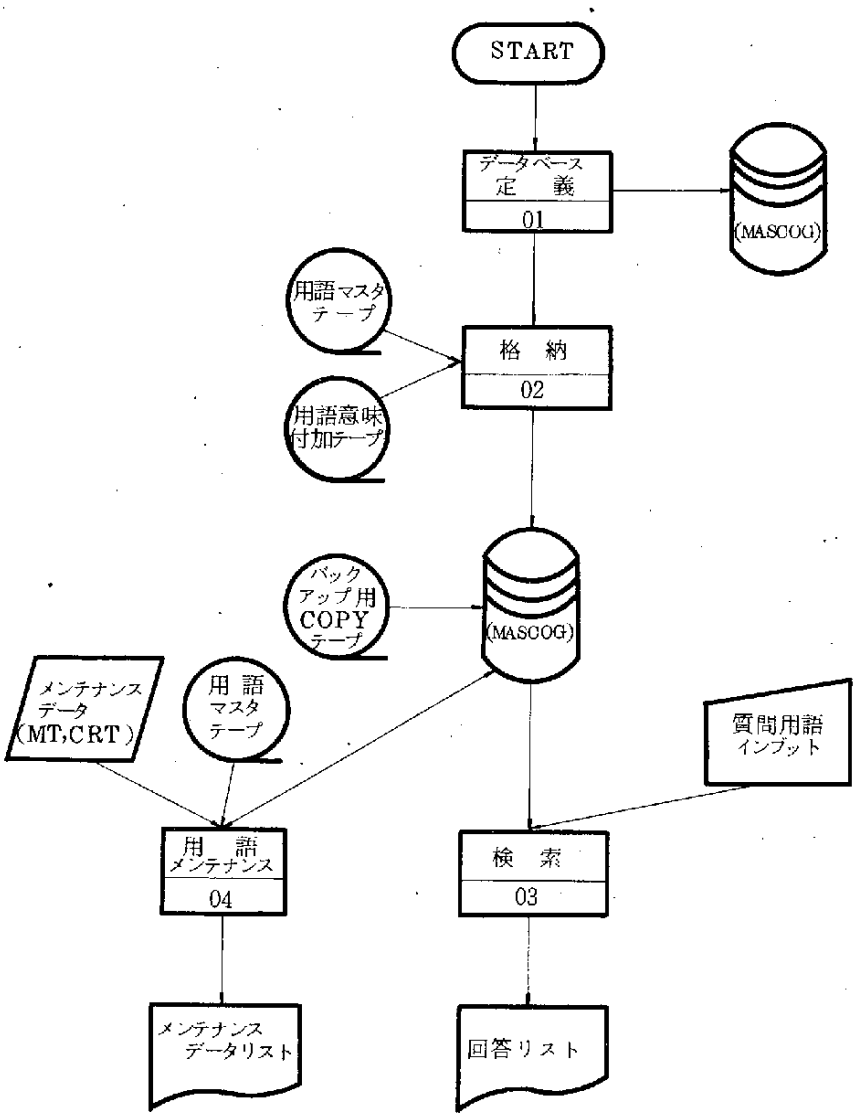


図9.1.2　　MASCOGにおけるシステム・フローチャート

(2)　データの収集方法

入力データとして，コンピュータ関係の辞典から，用語，用語の意味，辞典名，掲載頁などを収集した。利用面を考えると用語は，データ収集段階で初級者用とか技術者用とか，あるいは専門分野用とか区別して整理しておいた方が好ましい。データは，数種類の辞典から収集するため同語，同意語，関連語を整理し，同語は一つに，同意語・関連語は参照用語として一グループにまとめた。たとえば，「グラフィック・ディスプレィ」という用語には，同語として，グラフィックディスプレィ，グラフィック・ディスプレーなどをまとめ，同意語として，ディスプレィ，図形処理装置，ブラウン管ディスプレイ装置，さらに関連語として，ディスプレィ・システム，コンピュータ・グ

ラフィック，図形処理システム，グラフィック・データ・プロセッシング，図形処理プログラムなどをまとめた。これらをとりまとめる作業には，多くの労働と時間それに専門の知識を必要とした。

実験システムは，次の８種類の辞典から用語を引用し入力データとした。

① 図解・電子計算機用語辞典（日刊工業）
② 基準　コンピュータ用語辞典（日本事務能率協会）
③ 総合　コンピュータ辞典（日本経営出版会）
④ ＪＩＳ　情報処理用語（日本規格協会）
⑤ 電子計算機ハンドブック（オーム社）
⑥ 電子計算機ハンドブック（コロナ社）
⑦ 情報処理ハンドブック（光琳書院）
⑧ ＩＦＩＰ－ＩＣＣ　Vocabulary of Information Processing

整理されたデータは，コンピュータ・システムの入力データとしてカード（用語マスタ，用語意味データの２種類）に穿孔した。入力データ・フォーマットは次の通りである。

用語マスタ

整理番号	用　　語	出所１			出所２			出所３			
		辞典略号	掲載頁								
（５桁）	（３２桁）	(2)	(5)	(5)							

用語意味データ

整理番号	用　語　の　意　味	
（５桁）	（６６桁）	

(3) データ・リンケージの方法

当初，コンピュータ用語のシソーラスを作成するために体系化を企てたが，情報処理はいろいろな分野の集合であるため，一つの用語が何個所の分野とも関係をもつことが多く，特定のマスタに帰着させることはむずかしい。たとえば，「キャラクタ・ディスプレイ」という用語をとりあげた場合，ソフトウエアの分野，ハードウエアの分野，ユーザ利用の分野，プログラム言語の分野など，いろいろな分野のマスタと関連をもつため，取扱い易い木構造にデータ編成することができず，複雑なネットワーク構造をとらざるをえない。また，「キャラクタ・ディスプレイ」は，見方を変えればオンライン・サービスの端末機として使用され，オフラインのデータ・コレクタとしても使われる，さらに，ＭＩＳに不可欠な周辺装置でもあるなど，さまざまな分野との関連をもっており，さらにそれらの結びつきが強くなったり弱くなったり複雑な様相を呈している。現段階では，このような関係を全て満足するような複雑

なシソーラスを組みあげることはむずかしく，メンテナンスの問題と合せて考えるとその実用化は不可能に近いと思われる。以上のような状況から，この研究では無理に体系化は行なわず，特定の用語に対する参照用語として同意語，関連語を同一グループに集める程度に止め，シソーラスは強いて考えないことにした。データの整理は，収集した用語全てをＥＢＣＤＩＣコードで分類し，リストしたものを用いて行なった。整理した索引用語とその意味の一覧表の一部を次に掲載する。

ｵｰﾄﾏﾃｲﾌﾞｸ･ｺｰﾃﾞｨﾝｸﾞ	ﾌﾟﾛｸﾞﾗﾐﾝｸﾞﾆﾖｳｽﾙ ｼﾞｶﾝﾔ ﾛｳﾘﾖｸｵ ﾍﾗｼ ｱﾔﾏﾘｵ ﾅｸｽｺﾄｵ ﾓｸﾃｷﾆｼﾀ ﾌﾟﾛｸﾞﾗﾐﾝｸﾞ･ｼｽﾃﾑ
ｵｰﾄﾏﾃｲﾌﾞｸ･ｺﾝﾄﾛｰﾙ	ｼｽﾃﾑﾉ ｵｰﾄﾒｰｼｮﾝｶ ｼﾝｹﾝﾋﾉ ｾｯｹﾞｼﾀﾃﾞﾀﾃﾞﾅｸ ｺｳﾄｰﾅ ｶﾝﾘｺｳｶｵ ﾒｻﾞｽｺﾄ
ｵｰﾄﾏﾄﾝ	ﾆﾝｹﾞﾝﾉ ﾃﾉｳﾃｷﾄﾞｳｻｵ ｵｺﾅｲｴﾙ ｼﾞﾄﾞｳｷｶｲ
ｵｰﾊﾞｰﾌﾛｰ	ｱﾌﾚ: ｼｼｸｴﾝｻﾞﾝﾉ ﾅﾂｶｶﾞ ﾚｼｰｽﾀ ﾏﾀﾊ ﾙｲｻﾝｷﾉ ｱﾂｶｲｳﾙｶｽﾞﾉ ﾊﾝｲｶﾗ ﾊﾐﾀﾞｽｺﾄ
ｵｰﾊﾞｰﾚｲ	ﾄﾞｳｲﾂﾉ ｼﾞｮｺﾞﾉ ｺﾄﾅﾂﾀｽﾃｯﾌﾟﾃﾞ ﾅｲﾌﾞｷｵｸｿｳﾃﾉ ｵﾅｼﾞﾌﾟﾛｯｸｵ ｸﾘｶｴｼ ｾﾀｲﾙ ｼｽﾃﾑ
ｵｰﾌﾟﾝ･ｴﾝﾄﾞ	ｱﾀﾗｼｲｺｳｾｸｵ ﾂｲｶｼﾃﾓ ｼｽﾃﾑｾﾝﾀｲﾉ ｺｳｾｲｵ ﾍﾝｺｳｽﾙﾋﾂﾖｳﾉ ﾅｲｺﾄ
ｵｰﾌﾟﾝ･ｼｮｯﾌﾟ	ﾃﾞﾝｻﾝｷﾉ ﾌﾟﾛｸﾞﾗﾐﾝｸﾞﾔ ｷｶｲｿｳｻｵ ｹｲｻﾝｵﾖｳﾎﾞｳｽﾙ ﾌﾞｼｮﾉ ﾖｳｲﾝｶﾞ ﾐｽﾞｶﾗｽﾙｺﾄ
ｵｰﾌﾟﾝ･ﾙｰﾁﾝ	ｼｽﾃﾑｰﾁﾝﾆ ﾁｮｸｾﾂ ｸﾐｺﾏﾚﾙ ｻﾌﾞ･ﾙｰﾁﾝﾉｺﾄ
ｵｰﾌﾟﾝ･ﾙｰﾌﾟ	ｼﾞﾄﾞｳｾｲｷﾞｮｼｽﾃﾑﾃﾞ ｺﾝﾄﾛｰﾙ･ﾙｰﾌﾟｶﾞ ﾃﾞｷﾃｲﾅｲｺﾄ
ｵｰﾌﾟﾝ･ﾙｰﾌﾟｾｲｷﾞｮ	ｶｲﾙｰﾌﾟｾｲｷﾞｮ: ｼﾞﾄﾞｳｾｲｷﾞｮｼｽﾃﾑﾃﾞ ｺﾝﾄﾛｰﾙ･ﾙｰﾌﾟｶﾞ ﾄｼﾞﾃｲﾅｲ ｾｲｷﾞｮ
ｵｱｶｲﾛ	1ｺﾉ ﾆｭｳﾘﾖｸﾀﾝｼﾆ 1ｵｼﾒｽ ﾆｭｳﾘﾖｸｶﾞ ｱﾀｴﾗﾚﾀﾄｷﾆﾉﾐ ｼｭﾂﾘﾖｸﾀﾝｼﾆ 1ｶﾞ ｱﾗﾜﾚﾙ ｶｲﾛ
ｵｳﾄｳｼﾞｶﾝ	ｼｽﾃﾑｶﾞ ｱﾀｴﾗﾚﾀ ﾆｭｳﾘﾖｸﾆ ﾀｲｼﾃ ﾊﾝﾉｳｽﾙﾉﾆ ﾖｳｽﾙ ｼﾞｶﾝ
ｵｸﾘﾚｼﾞｽﾀ	ｼﾌﾄﾚｼﾞｽﾀ: 2ﾃﾝｿｳ ｸﾐｶﾗﾅﾙ ﾚｼﾞｽﾀﾉ 1ｼｭ
ｵｼﾎﾞﾀﾝﾀﾞｲﾔﾙﾃﾞﾝﾜｷ	ｶｲﾃﾝｼｷﾀﾞｲﾔﾙﾉ ｶﾜﾘﾆ ｽｳｼﾞﾆﾀｲｵｳｽﾙ ｵｼﾎﾞﾀﾝｵ ｵｼﾃ ｾﾝﾀｸ ｾﾂｿﾞｸｽﾙ ﾃﾞﾝﾜｷ
ｵﾌｼﾚｰﾄ･ｿｰﾄ	ﾏｰｼﾞ･ｿｰﾄﾉ 1ｼｭﾃﾞ ｲﾝﾌﾟｯﾄﾖｳﾉ ﾃｰﾌﾟｵ ﾉｿﾞｲﾀ Nﾎﾝﾉ ﾃｰﾌﾟﾃﾞ ﾏｰｼﾞ･ｿｰﾄｵ ｵｺﾅｳ
ｵﾌ･ｹﾞｰｼﾞ	ｲｯﾃｲﾉ ｷｶｸｶﾗ ﾊｽﾞﾚﾙｺﾄ
ｵﾌ･ﾗｲﾝ	ﾃﾞﾝｻﾝｿｼｷﾄｲｿｼｮﾆ ｼｮｳｻﾚﾃﾊｲﾙｶﾞ ｿﾚﾆﾖｿﾃ ﾃﾖｸｾﾂ ｾｲｷﾞﾖｻﾚﾃｲﾅｲ ｿｳﾁｵ ｿｳｼｮｳﾃｷﾆ ｻｽ
ｵﾝ･ﾗｲﾝ･ｼｽﾃﾑ	ﾃﾖｿﾂｾｲｷﾞﾖﾆ ﾀｲｽﾙ ﾋﾃﾖｿﾂｾｲｷﾞﾖﾉ ｼｮﾘﾎｳｼｷ
ｵﾌﾗｲﾝ	ｼﾞﾖｳﾎｳﾉ ﾃﾝｿｳｶﾃｲﾆｵｲﾃ ﾋﾄﾃﾞﾉ ｶｲﾆｭｳｵ ﾋﾂﾖｳﾄｽﾙ ｼﾞﾖｳﾀｲ
ｵﾌﾟｼﾞｪｸﾄ･ﾌﾟﾛｸﾞﾗﾑ	ﾌﾟﾛｾｯｻﾆﾖｯﾃ ﾎﾝﾔｸｻﾚﾀ ﾌﾟﾛｸﾞﾗﾑ
ｵﾌﾟｼﾞｪｸﾄ･ﾓｼﾞｭｰﾙ	ｿｰｽ･ｾｼﾞｺｰﾙｵ ｺﾝﾊﾟｲﾙｼﾃ ｷｶｲｺﾞﾆﾅｯﾀ ﾌﾟﾛｸﾞﾗﾑﾉｺﾄ
ｵﾌﾟｼﾞｪｸﾄ･ﾙｰﾁﾝ	ｵﾌﾟｼﾞｪｸﾄ･ﾌﾟﾛｸﾞﾗﾑ: ﾌﾟﾛｾｯｻﾆﾖｯﾃ ﾎﾝﾔｸｻﾚﾀ ﾌﾟﾛｸﾞﾗﾑ
ｵﾌﾟｺﾝ	OPCON: ﾀﾝｻｸﾎｳﾆﾖﾙ ｹｲｻﾝｾｲｷﾞﾖｵ ｵｺﾅｳﾀﾒﾉ ﾄｸｼｭﾅ ｾｲｷﾞﾖﾖｳ ｹｲｻﾝｷ
ｵﾌﾟｻｰﾁ	ｵﾍﾟﾚｰｼｮﾝｽﾞ･ﾘｻｰﾁﾉ ﾘﾔｸ
ｵﾍﾟﾗﾝﾄﾞ	ｹｲｻﾝｷﾉ ﾒｲﾚｲｺﾞﾉ ｺｳｾｲﾌﾞﾝﾉｺﾄ
ｵﾍﾟﾚｰｼｮﾝ･ｶﾞｲﾄﾞ･ｺﾝﾋﾟｭｰﾀ	ｾｲｷﾞﾖｼｽﾃﾑｾﾝﾀｲｵ ｶﾝｾﾝﾆ ｼﾞﾄﾞｳｶｾｽﾞ ﾆﾝｹﾞﾝﾆ ｿｳｻｵ ｵｺﾅｳｾﾙﾖｳﾅ ﾃﾞﾝｻﾝｷ
ｵﾍﾟﾚｰｼｮﾝ･ｺｰﾄﾞ	ｷｶｲｺﾞﾒｲﾚｲﾉ ｲﾖｸﾞﾃﾞ ｷﾎﾝﾃｷﾅ ｹｲｻﾝｷﾉ ｿｳｻｷﾉｳｵ ﾉﾍﾟﾃｲﾙ ｺｰﾄﾞ
ｵﾍﾟﾚｰｼｮﾝ･ﾐｽ	ﾃﾞｰﾀｼｮﾘｻｷﾞﾖｳﾆ ｵｲﾅ ﾆﾝｹﾞﾝｶﾞ ｵｶｼﾀ ﾔﾘｿｺﾅｲ
ｵﾍﾟﾚｰｼｮﾝｽﾞ･ﾘｻｰﾁ	ﾀｲｹｲﾉ ｳﾝｴｲﾆｶﾝｽﾙ ｾﾝﾀﾞｲﾆｶﾝｼ ｶｶﾞｸﾃｷｼｭﾎｳ ﾖｳｸﾞﾆﾖﾘ ｻｲﾃｷﾉ ｶｲｵ ﾃｲｷﾖｳｽﾙｺﾄ
ｵﾍﾟﾚｰﾀ	ｵﾍﾟﾗﾝﾄﾞﾉ ｴﾝｻﾞﾝﾅｲﾖｳｵ ｱﾗﾜｽ ｽｳｶﾞｸﾃｷｷｺﾞｳ: ﾒｲﾚｲﾉｲﾁﾌﾞ: ｹｲｻﾝｷﾉ ｿｳｻｲﾝ

オペ゛レーティング゛・システム	デ゛ンサンソシキオ ユウコウニ リヨウシヨウトヌル コントロール・システム
オリジ゛ン	フ゜ログ゜ラムヤ レコート゛・ア゛ロックノ ハシ゛メオシメス キオクソウチノ アト゛レス
オン・オフセイキ゛ヨ	シ゛ト゛ウセイキ゛ヨノ ホウシキノ イツシユ
オン・ライン	アル キキト アル キキカ゛ ラインテ゛ ケツゴ゛ウサレテイル シ゛ヨウタイ
オン・ライン・システム	ニユウリヨクテ゛ータカ゛ ハツセイケ゛ンカラ チヨクセツ ケイサンキニハイリ シユツリヨクテ゛ータモ チヨクセツ デ゛ンソウサレル システム
オンセイオウトウソウチ	シユツリヨクオ オンセイニヨツテ アタエル シユツリヨクソウチ
オンセイケツコ゛ウソウチ	T・S・Sテ゛ デ゛ンワカイセンオツカツテ タンマツキカラ ソウシ゛ユシンオヌル ハ゛アイニ ツカウ ヘンフクチヨウソウチノ イツシユ
オンセイシユウハ	シヨウヨウヒンヒツトシテ カイワニヒツヨウナ カチヨウシユウハスウナイノ アラユル シユウハスウ
オンライン	シ゛ヨウホウノ デンソウカテイニオイテ ヒトテ゛ノ カイニユウヌ ヒツヨウトシナイ シ゛ヨウタイ
カートリツシ゛	シ゛キテープ゛オ シユウノウシ シ゛キテープ゛ソウチニ ヨウイニ チヤクタ゛ツシ カツ ソノママノ ホリ゛ンモ ヨウイニシタ ヨウキ
カート゛	イツテイノ キソクニシタカ゛ツテ センコウオ オコナウ イツテイノ ケイシ゛ヨウスンホ゜ウオ ユウスル カミノ カート゛ノ コト
カート゛・イメーシ゛	カート゛ニハ゜ンチサレタ センコウイチオ ソノママ 2シンスウノ ヒ゛ツトトシテ キオクスル ホウホウ
カート゛・スタンカ	センコウカート゛オ トリアツカウ キキニオイテ ソノキノ ソウサオ シコウリヨウシタ カート゛オ ツミカサネ タクワエル ハ゛シヨ
カート゛・ハ゛ンチ	カート゛センコウキ; カート゛ニ モシ゛ スウシ゛ トクシユキコ゛ウオ センコウスルコトニヨリ シ゛ヨウホウオ キロクスル ソウチ
カート゛・ホツハ゜	カート゛オ キカイノナカニ オクリコムタメニ イレテオク トコロ
カート゛・リータ゛	カート゛ヨミトリキ; センコウカート゛オ ヨミトル デ゛ンシケイサンキノ ニユウリヨクソウチノ イツシユ
カート゛ケンコウキ	カート゛ニツイテ ケンコウオ オコナウ ソウチ
カート゛シヨウコ゛ウキ	カート゛オシヨウコ゛ウスル キノウオモツ ソウチ
カート゛セツケイ	カート゛ナイニ センコウキ ホンヤクインサツサセル フイールト゛ノ ハイチオ サタ゛メルコト
カート゛センコウキ	カート゛ニ モシ゛ スウシ゛ トクシユキコ゛ウオ センコウスルコトニヨリ シ゛ヨウホウオ キロクスル ソウチ
カート゛フ゛ンルイキ	センコウサレタ カート゛オ シ゛ト゛ウテキニ フ゛ンルイ セイリスル キカイ
カート゛ヘ゛ースノテ゛ンシケイサンキ	センコウカート゛オ シユタイトヌル デ゛ンサンキノ コト
カート゛ヨミトリキ	センコウカート゛オヨミトル デ゛ンサンキノ ニユウリヨクソウチノ イツシユ
カイキキヨクセン	Xニタイスル Yノ ヘイキンテキナ ケイコウオシメス キヨクセン Y=A+B*X+C*X**2
カイキケイスウ	サイショウ2シ゛ヨウホウテ゛ スイテイサレタ カイキキヨクセンノ ケイスウ
カイキフ゛ンセキ	アルインシノ テイト゛ マタハ リヨウノ タカ゛イニ チユウモクシテ リヨウト リヨウノ インカ゛カンケイオ フ゛ンセキヌル ホウホウ
カイキ゛ヨウ	インシ゛ノサイニ ツキ゛ニクルモシ゛オ ツキ゛ノキ゛ヨウニ インシ゛スルタメニ フ゜ラテンオ カイテンサセル ソウサ
カイケイキ	カクシユノ トウケイヒヨウヤ カイケイシヨルイ チヨウホ゜ノサクセイオ シ゛ト゛ウテキニオコナウ ソウチ
カイセン	ノホ゜リ クタ゛リノ リヨウホウコウノ ツウシンロニヨツテ コウセイサレル リヨウホウコウツウシンノ デ゛ンソウロ
カイセンコウカン	ハツシンカ゛ワノ カイセント チヤクシンカ゛ワノ カイセンオ チヨツケツシテ シ゛ヨウホウハ コウカンソウチオ スト゛ウリヌル

PROCEDURE-ORIENTED LANGUAGE	コウハンナ シュルイノ モンダ"イノ カイニ ショウサレ テツヅ"キナト"オ セイカクニ キシ"ュツシヤスイヨウニ ツクラレタ.ケ"ンコ"
PROCESS CONTROL SYSTEM	レンソ"ウテキナ オペ"レーションノ シ"ト"ウカオ モクテキトスル システム
PROCESS SUPERVISORY PROGRAM / 1800	IBM 1800 DACSニ タイシテ カイハツサレタ プ"ロセス セイキ"ヨヨウノ ソフトウエア
PROCESSING PROGRAMS	ショリ プ"ログ"ラム: シ"ヨウホウオ ショリシテ ケツカオ ウミダ"スコトオ オモナ キノウトスル プ"ログ"ラムノ ヨビ"ナ
PRODUCTIVE TIME	FAULTS アルイハ ERRORS ナシテ" サキ"ヨウノ ショリオ オコナウタメニ ツイヤサレル シ"カンノ コト
PROGRAM	AUTOMATIC DATA PROCESSINGニ オケル ショリコウテイオ セツメイ シタモノ
PROGRAM BODY	プ"ログ"ラム ホンタイ: プ"ログ"ラム プ"ブ"ンノ アトニ END ギ"ヨウオ ツケタモノ
PROGRAM CONTROL DATA	PL/1ノ プ"ログ"ラムノ ナカテ" ツカワレル テ"ータテ" プ"ログ"ラムノ シ"ッコウオ サユウスルモノ
PROGRAM CONTROL INSTRUCTION	プ"ログ"ラム セイキ"ヨ メイレイ: プ"ログ"ラムノ アセンブリノ シカタオ シメス メイレイ
PROGRAM DEVELOPMENT TIME	PROGRAMノ アヤマリオ ミツケル PROGRAM TESTINGノ アルイハ アクラシイ ソウサテシ"ュツオ シラヘ"ルタメノシ"カン
PROGRAM FLOWCHART	プ"ログ"ラム ナカ"レス": テ"ータカ" ショリサレル テシ"ュツオ ス"テ" ヒョウシ" シタモノ
PROGRAM LISTING	アセンブ"ラヤ コンパ"イルサレタトキノ ケ"ンシブ"ログ"ラムノ カクステートメントト モクテキプ"ログ"ラムノ ステートメントノ リスト
PROGRAM MODULE DICTIONARY	モシ"ュールオ ショリスル プ"ログ"ラムカ" ヨウキュウスル プ"ログ"ラム,モシ"ュールニ カンスル セイキ"ヨシ"ヨウホウノ アツマリ
PROGRAM PART	シ"ッコウブ"ン,FORMATブ"ン,DATAブ"ン オヨビ" ブ"ンカンスウ テイキ"ブ"ンノ アツマリ
PROGRAM SEGMENT	プ"ログ"ラムカ" ブ"ンカツ サレタモノ
PROGRAM SPECIFICATION	プ"ログ"ラムニヨッテ オコナワレル ショリキノウノ ロンリ コウソ"ウヤ ソノ ハンイナト"オ メイカクニ キテイスル キシ"ュツ
PROGRAM TESTING	テ"キアカ"ツタ プ"ログ"ラムカ" タタ"シク ハタラクカオ テストヨウノ テ"ータオ ツカッテ テストスルコト
PROGRAM UNIT	プ"ログ"ラム タンイ: シュブ"ログ"ラム マタハ フクブ"ログ"ラムオ イウ
PROGRAM-NAME	COBOLケ"ンシブ"ログ"ラムオ シキベ"ツスルタメノ コ"
PROGRAM-SENSITIVE FAULT	PROGRAM STEPS ヤ タイミング"ノ ス"レニ ヨッテノミ ショウス"ル FAULT (アヤマリ)
PROGRAMMED CHECK	プ"ログ"ラムノ 1ブ"ニ クミコマレタ ケンサノ テシ"ュツ
PROGRAMMING SYSTEMS	ケイサンキノ プ"ログ"ラム サクセイト ソウサオ ヨウイ カツ コウリツテキニ スルタメニ ツカワレル プ"ログ"ラムノ ソウショウ
PROLOGUE	シ"ヨマク: プ"ロツクノ カイシ"ニ オコナワレル ショリ
PROOF TOTAL	1セットニナッタ ケイサンテ" ソノツリアイオ ミルタメニ ケイサンノ 1ヨウソ タ"ケオ シュウケイシテ エラレタ コ"ウケイチオ イウ
PROPER STRING	キコ"ウノ キホンレツ: 'ト' オ ノゾ"ク キホンキコ"ウノ ニンイノ レツカ クウ
PROSPRO / 1800	IBM 1800 DACSニ タイシテ カイハツサレタ プ"ロセス セイキ"ヨヨウノ ソフトウエア
PROTECTION KEY	ホコ" キー: タスクニ フス"イシ タスクカ" シンコウシ"ヨウタイニ アルトキニ プ"ログ"ラムシ"ヨウタイコ"ニ アラワレルヒヨウシ"
PROVING	MACHINEニ アヤマリ" ナイカ トクシュナ TASKオ オコナウコトカ" デ"キルカ トイウコトオ シラヘ"ルコト
PROVING TIME	PROVING ニ ツイヤサレル シ"カン
PSEUDO-CODE	キ"シ" コート": キカイコ"テ"ハ ナク プ"ログ"ラミング"オ シヤスクスルタメニ ツカワレル ケ"ンコ"

PSEUDO-INSTRUCTION FORM	DUMMY INSTRUCTIONﾃﾞ ｶｶﾚﾀ DATAﾉ ITEMﾊ PSEUDO-INSTRUCTION-FORMﾃﾞ ｶｶﾚﾀﾄｲｳ
PSEUDO-OFF-LINE	ｷﾞｼﾞ･OFF-LINE;
PSEUDO-ON-LINE	ｷﾞｼﾞ･ON-LINE;
PSEUDO-VARIABLE	ｷﾞｼﾞ ﾍﾝｽｳ; ｸﾐｺﾐ ｶﾝｽｳﾒｲﾉ 1ｼｭﾃﾞ ｳｹｲﾚ ﾌｨｰﾙﾄﾞﾄｼﾃ ﾓﾁｲﾙｺﾄｶﾞ ﾃﾞｷﾙﾓﾉ
PUMP	PARAMETRIC OSCILLATOR ﾃﾞ ｱﾙ ｼｭｳﾊｽｳｵ ｱﾀｴﾙ ｺﾄ
PUNCH KNIFE	ﾊﾟﾝﾁ･ｱﾅｵ ｱｹﾙ ｷｶｲｿｳﾁﾉ ｱﾅｱｹ･ﾌﾟﾌﾟﾝｵ ｲｳ
PUNCH POSITION	ｾﾝｺｳｲﾁ; ｾﾝｺｳﾃｰﾌﾟ ﾏﾀﾊ ｶｰﾄﾞｼﾞﾖｳﾆ ｾﾝｺｳｻﾚﾙ ｲﾁ
PUNCHED CARD	ｾﾝｺｳｶｰﾄﾞ; ﾃﾞｰﾀｵ ﾋｮｳｼﾞｽﾙﾀﾒﾆ ｱﾙ ｷｿｸﾆ ｼﾀｶﾞｯﾃ ｱﾅﾉ ｱｹﾗﾚﾀ ｶｰﾄﾞ
PUNCHED TAPE	ｾﾝｺｳﾃｰﾌﾟ; ﾃﾞｰﾀｵ ﾋｮｳｹﾞﾝｽﾙﾀﾒﾆ ｱﾙ ｷｿｸﾆ ｼﾀｶﾞｯﾃ ｲｸﾂｶﾉ ｱﾅｵ ｱｹﾀ ﾃｰﾌﾟ
PUNCHING POSITION	ﾃﾞｰﾀ ｵ ｱﾗﾜｽ ｱﾅ ｶﾞ ﾊﾟﾝﾁ ｻﾚﾙ DATA CARRIER ﾉ ｲﾁ
PUNCHING STATION	PUNCHED CARD ﾃﾞ ﾊﾟﾝﾁ ｻﾚﾙ CARD TRACK ﾉ ﾌﾟﾌﾟﾝｵ ｲｳ
PUNCHING TRACK	2ﾂﾉ CARD TRACKｵ ﾓﾂ HACHINEﾉ 1ﾎﾟｳﾉ TRACKﾃﾞ PUNCHING STATIONｵ ｲｳ
PUNCTUATION CHARACTER	ｸｷﾞﾘﾓｼﾞ; , ; . ﾞ(ｲﾝﾖｳﾌ)) (ﾅﾄﾞｵ ｲｳ
PUSH-DOWN LIST	ﾃﾞｰﾀｺｳﾓｸﾉ ｼﾞｭﾝｼﾞﾖﾂﾞｹﾗﾚﾀ ｼｭｳｺﾞｳ
PUSH-DOWN STORE	ｱﾄﾆ ｶｷｺﾏﾚﾀ DATA ﾙﾄﾞ ﾊﾔｸ ﾖﾐﾀﾞｽﾖｳﾆ ｿｸﾗﾚﾀ ｷｵｸｿｳﾁ
PUSH-UP LIST	ﾃﾞｰﾀｺｳﾓｸﾉ ｼﾞｭﾝｼﾞﾖﾂﾞｹﾗﾚﾀ ｼｭｳｺﾞｳ
PUSHED-DOWN STACK	ｻｲｺﾞﾆ ﾊｲｯﾃｷﾀ ｿﾘｱﾃｶﾞ ｻｲｼｮﾆ ﾄﾘﾀﾞｻﾚﾙﾖｳｵ ｹｲｼｷｵ ﾓﾂ ｿﾘｱﾃﾉ ﾂﾐｶｻﾈ
PUT	ｻｷﾞﾖｳ ｸｲｷﾖﾘ ｼｭﾂﾘｮｸ ｸｲｷﾆ ｳﾂｽ ｿｳｻｵ ｵｺｳｾﾙ ﾏｸﾛﾒｲﾚｲ
@CB	ﾌﾟﾛｸﾞﾗﾏｶﾞ ﾃｲｷﾞｼﾀ ｷﾉｳﾉ ｼﾞｭﾝｼﾞ ｼｮｳﾉ ﾁｮｳｾｲｵ ｵｺﾅｳﾉﾆ ﾂｶｳﾚﾙ ｾｲｷﾞﾖ ﾌﾟﾛｯｸ
@ISAM	ｻｸｲﾝ ｼﾞｭﾝｼﾞ ﾃﾞｰﾀ･ｾｯﾄﾉ ﾚｺｰﾄﾞｵ ｼﾞｭﾝｼﾞﾆ ｶｲﾀﾘ ﾖﾝﾀﾞｰﾘｽﾙ ｱｸｾｽ ﾎｳｼｷ
@SAM	ｼﾞｭｳﾗｲﾉ ﾆｭｳｼｭﾂﾘｮｸ ｾｲｷﾞﾖｼｽﾃﾑﾃﾞ ﾓﾝﾄﾓ ﾋﾛｸ ﾂｶｳﾚﾃｲﾙ ｱｸｾｽ ﾎｳｼｷ
@QUALIFIED DATA-NAME	ｹﾂｺﾞｳｼ OFｶ INﾉ ｲｽﾞﾚｶﾄ ｿﾉ ｱﾄﾆ ﾂｲﾀ ｼｭｳｼｮｸﾒｲﾉ ﾋﾄｸﾐ
@QUALIFIED NAME	ｱﾙ ﾓｼﾞ ﾀﾄｴﾊﾞ ﾋﾟﾘｵﾄﾞﾃﾞ ｸｷﾞﾗﾚﾀ ﾌｸｽｳﾉ ﾅﾏｴｶﾗ ﾅﾌﾃｲﾙ ﾃﾞｰﾀ･ｾｯﾄﾒｲ
@QUANTISATION	ﾘｮｳｼｶ; ﾋﾄﾉ ﾄｼﾉﾖｳﾆ､ｱﾙ ﾘｮｳｼ(ｱﾙ ｴｳｳﾞﾝﾉ ｽｳﾃﾞ ｶﾅﾗｽﾞｼﾓ ﾋﾄｼｸﾅｸﾃﾓ ﾖｲﾉﾃﾞ ｱﾗﾜｽｺﾄ
@QUANTITY	ﾘｮｳ; ﾀﾝｼﾞｭﾝ ﾍﾝｽｳ､ﾊｲﾚﾂ､ﾅﾏｴ､ｽｲｿｳ ｵﾖﾋﾞ ﾃﾂﾂﾞｷﾉ ﾗｼｭﾙｲ
@QUARTER-SQUARES MULTIPLIER	X*Y=1/4*(X+Y)**2-1/4*(X-Y)**2) ｺｳﾄｳｼｷﾉ ｴﾝｻﾞﾝｵ ｵｺﾅｳ ANALOG MULTIPLIER
@QUARTZ DELAY LINE	DELAY LINEﾉ 1ｼｭﾃﾞ ﾁｴﾝｾﾝﾄ ｼﾃ ｽｲｼｮｳｵ ﾓﾁｲﾀ ﾓﾉ
@QUASI-INSTRUCTION FORM	DUMMY INSTRUCTIONﾃﾞ ｶｶﾚﾀ DATAﾉ ITEMﾊ @QUASI-INSTRUCTION-FORMﾃﾞ ｶｶﾚﾀ ﾄｲｳ
@QUEUED ACCESS METHOD	ﾌﾟﾛｸﾞﾗﾑﾄ ﾆｭｳｼｭﾂﾘｮｸｿｳｻﾉ ｱｲﾀﾞﾉ ﾃﾞｰﾀﾉ ﾃﾝｿｳｵ ｼﾄﾞｳﾃｷﾆ ﾄﾞｳﾃｷｮｳｻｾﾙ ｱｸｾｽﾎｳｼｷ
@QUEUED INDEXED SEQUENTIAL ACCESS METHOD	ｻｸｲﾝ ｼﾞｭﾝｼﾞ ﾃﾞｰﾀ･ｾｯﾄﾉ ﾚｺｰﾄﾞｵ ｼﾞｭﾝｼﾞﾆ ｶｲﾀﾘ ﾖﾝﾀﾞｰﾘｽﾙ ｱｸｾｽ ﾎｳｼｷ

(4) ファイリングの方法

〔データ・ベースの定義〕 データ・ベースは汎用データ・マネジメント・システム：ＲＡＰＩＤの定義ユーティリティＰＲＤＤＥＳＰのフォーマット指定により行なう。

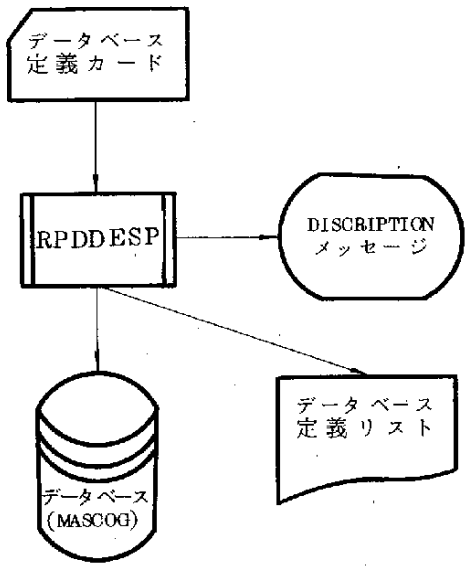


図9.1.3　ＭＡＳＣＯＧにおけるデータ・ベースの定義

〔データ・ベースの構造〕 ＭＡＳＣＯＧは，データの階層を持たない１レベルのレコードであるが，ＲＡＰＩＤの規定により，２レベル以上のレコード・タイプを持たなければならない。また，ソートされたマスタ・レコードをリング結合するなどのために，レベル１をダミー・レコードとしてレコードタイプ名ＤＵＭＭＹＫＥＹを設け，データ＂１９７１＂を格納した。レベル２には，コンピュータ用語に関する一連の情報を格納するマスタ・レコード（レコード・タイプ名：ＷＯＲＤＲＥＣ）を設け，レベルＩとレベル２の間をリファレンス・コード（アドレス情報）とリング（ＤＵＭＲＩＮＧ）で結合した。

ＤＵＭＲＩＮＧはＮＥＸＴ，ＰＲＩＯＲ，ＭＡＳＴＥＲの機能を持ち，正順，逆順，マスタと連結するリング構造である。

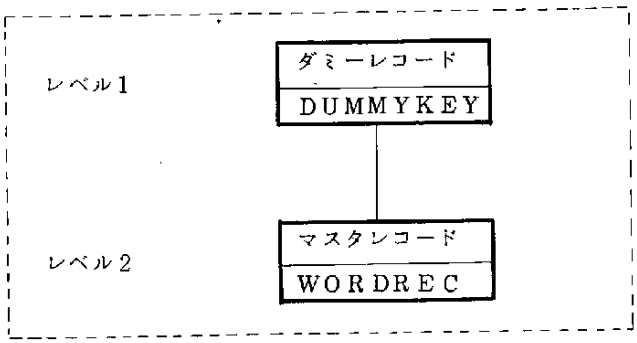


図9.1.4　　MASCOGにおけるデータベースの構造

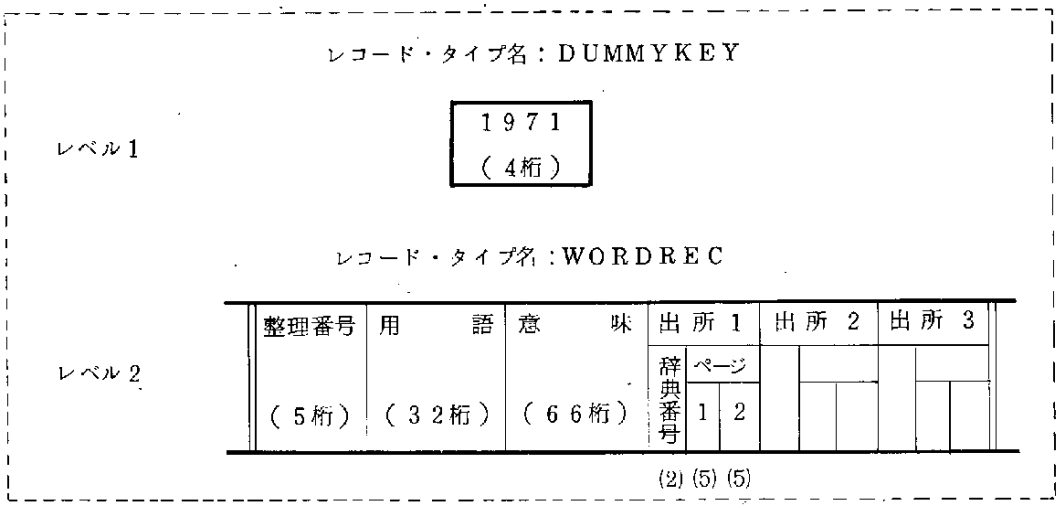


図9.1.5　　MASCOGにおけるファイルの構成

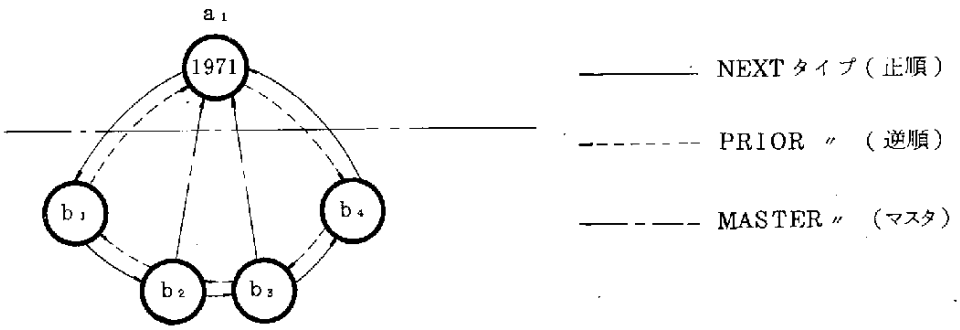


図9.1.6　　リング構造

〔データの格納〕　用語マスタと用語の意味データは，各々磁気テープに記録されており，用語マスタはEBCDICコード順，意味データは整理番号順に分類してある。データ1レコードのデータ

ベースへの格納は，各々の整理番号のマッチングにより行なわれる。格納は，まず，レコード・タイプのRTCSにリファレンス・コードを入れるため，PUT RANDOMによってDUMMYKEYを格納する。マッチしたデータはフィールドにセットし，キーワードとなる用語をランダマイズ・ルーチンを通して格納アドレスを決定すれば，データはリファレンスコードによる関連をもちながらデータ・ベースの指定されたWOROECに順次格納されてゆく。これらの一連の処理のうちに，各レコードはリングによって結合され，キーワードである用語はシークエンシャルにポインタで結ばれる。

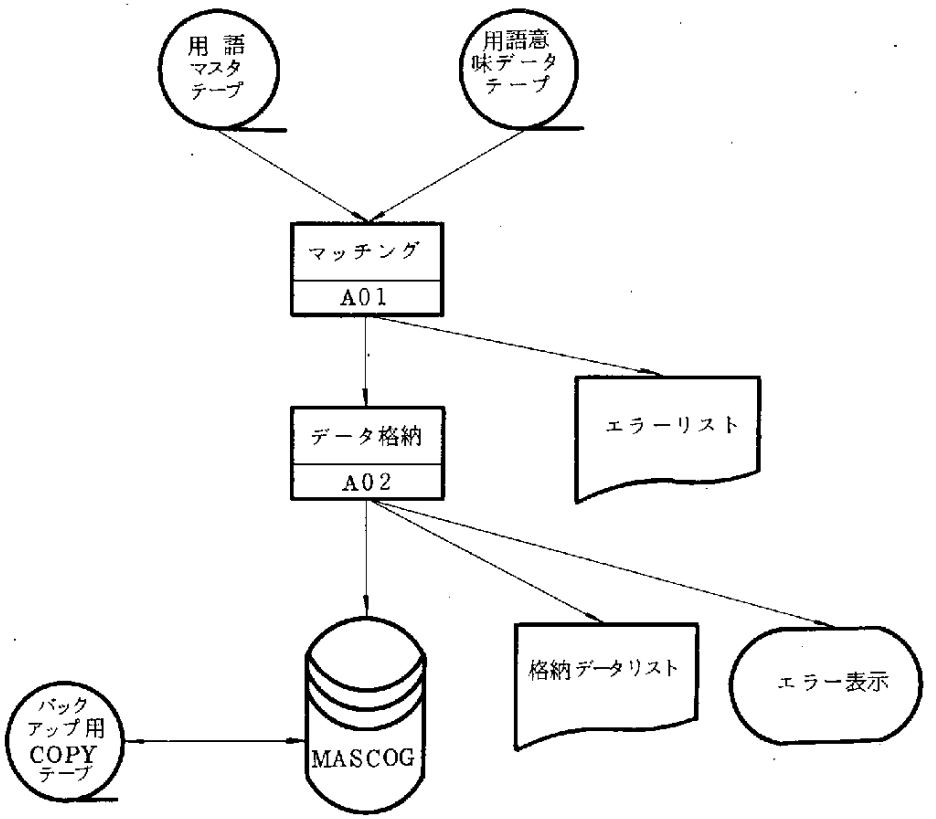


図9.1.7　データの格納

次にデータ・ベースへの格納におけるプログラムの一部を掲載する。

```
0080    320010/RAPID SECTION.
0081    320030/RAPID PAGE BUFFER 35.
0082    320040/RT DUMMYKEY.
0083    320050/01 DUM        PIC S9(4) COMP VALUE 1971.
0084    320060/RT WORDREC.
0085    320070/01 WD-R.
0086    320080/     02 R-CODE      PIC 9(5).
0087    320090/     02 R-WORD      PIC X(39).
0088    320100/     02 R-MEAN      PIC X(70).
0089    320110/     02 R-B1        PIC X(12).
0090    320120/     02 R-B2        PIC X(12).
0091    320130/     02 R-B3        PIC X(12).
0092    320140/     02 FILLER      PIC X(2) VALUE SPACE.
0093    400010 PROCEDURE DIVISION.
0094    400020 OPEN-RT.
0095    400030      OPEN INPUT KEEP-F MEAN-F.
0096    400040      OPEN OUTPUT LST-F.
0097    400050/     RPD-OPEN DATABASE UPDATE.
0098    400060 CLEAR-RT.
0099    400070      MOVE SPACES TO PRT.
0100    400080      WRITE PRT BEFORE ADVANCING C17.
0101    400081      ACCEPT TYPE-IN.
0102    400090 START-IN.
0103    400100/     PUT RANDOM DUMMYKEY.
0104    400110      IF SCB2 = '0175' GO TO MSG1.
0105    400120      IF SCB2 = '0171' GO TO MSG2.
0106    400130/     WHEN ERROR GO TO MSG3.
0107    400140 R-MT2.
0108    400150      READ MEAN-F AT END GO TO END-RT.
0109    400151      IF NO-2 NOT > TYPE-IN GO TO R-MT2.
0110    400160 R-MT1.
0111    400170      READ KEEP-F AT END GO TO ER-LST.
0112    400180      IF NO-1 NOT = NO-2 GO TO R-MT1.
```

```
FACOM 230-60     RAPID COBOL SOURCE LIST     -71.07.01-     (V-01/L-03)

CARDNO.  SEQNO. A  B

0113    400190      MOVE CODE TO R-CODE.
0114    400200      MOVE WD TO R-WORD.
0115    400210      MOVE MEANS TO R-MEAN.
0116    400220      MOVE H1 TO R-B1.
0117    400230      MOVE H2 TO R-B2.
0118    400240      MOVE H3 TO R-B3.
0119    400250/     PUT RANDOM WORDREC INT=90.
0120    400260      IF SCB2 = '0175' GO TO MSG4.
0121    400270      IF SCB2 = '0171' GO TO MSG5.
0122    400280/     WHEN ERROR GO TO MSG6.
0123    400290      MOVE DAA TO REF.
0124    400300      MOVE NO-1 TO CD.
0125    400301      MOVE WD TO WD-P.
0126    400310      MOVE BOOK5 TO HON.
0127    400320      MOVE BOOK2 TO HONS.
0128    400330      MOVE MEANS TO MN.
0129    400340      MOVE FM-1 TO P-FM.
0130    400350      WRITE PRT BEFORE ADVANCING C1.
0131    400360      MOVE FM-2 TO P-FM.
0132    400370      WRITE PRT BEFORE ADVANCING C1.
0133    400380      GO TO R-MT2.
```

(5) 検索の方法

検索は，語または句の自然語のキーワードで行なう。問合せおよび応答は端末タイプライタを用いるが，ＣＲＴディスプレィの使用も可能である。また，緊急を要さない場合は，漢字出力もできる。システムのレベル・アップとして，関連語検索やオンライン・サービスが当面の問題であり，用語の体系化とＭＡＳＣＯＧ専用のオペレーティングシステムを考えて行きたい。ＭＡＳＣＯＧにおけるキーワードとデータとの関係を次に示す。

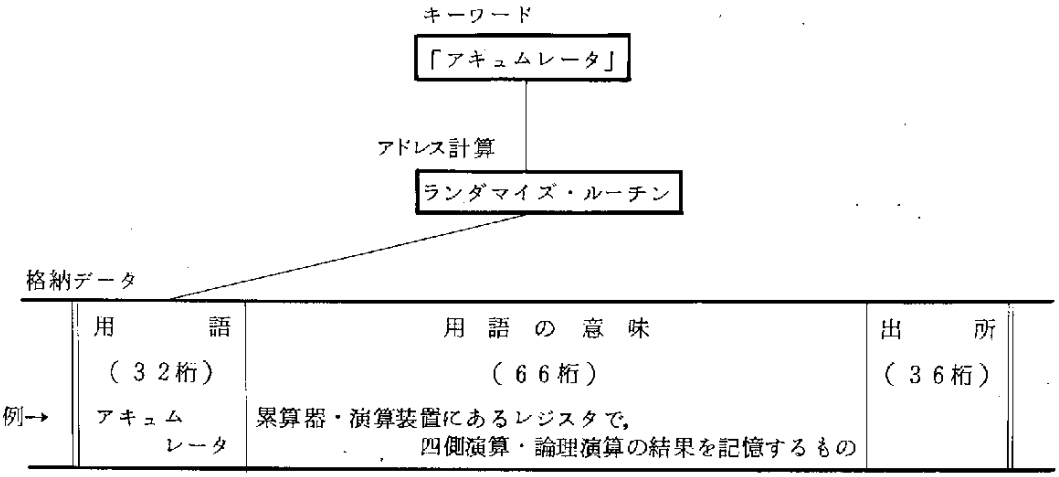


図9.1.8　MASCOGにおけるキーワードとデータとの関係

検索は，まず用語をキーボードから入力し，ランダマイズ・フィールドにセットする。ランダマイズ・ルーチンによってアドレス情報が得られると，そこに格納されているレコードがアウトプット・エリアにセットされ出力される。検索レコードが質問した用語と異なる場合に起るシノニムは，RAPIDが保有するユーティリティのランダマイズリストにその内容が提示される。検索用語のタイプイン・ミスおよび検索用語の未登録のメッセージは，その都度エラー表示される。ここで，MASCOGの検索プロセスを図9.1.9に示す。

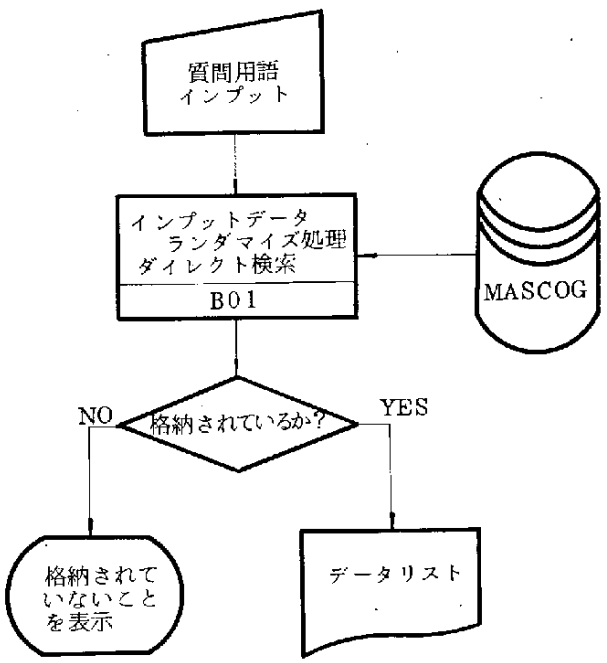


図9.1.9　MASCOGにおける検索プロセス

検索要求の手順と表示形式は次の通りである。

① READY表示……	ジュンビ・カンリョウ
② KEYWORD指定……	ケンサクヨウゴ：（３２文字以内で指定）
③ 出力TYPE指定……	タイプ：（出力形式の指定）
④ REQUEST命令……	RQ
⑤ ERROR表示……	ERROR：（その原因）

① ＭＡＳＣＯＧ用プログラムのローディングされると準備完了が表示される。

② 「ケンサクヨウゴ：」とシステムから問合せがきたら３２文字以内の英数字，カナ文字で検索用語を指定する。

③ タイプライタ，ＣＲＴディスプレィ，漢字出力などの指定を記号で行なう。

④ キーボード上のＲＥＱＵＥＳＴボタンを押し，検索の実行指示を行なう。

⑤ タイプイン・ミス，未定義用語などがリストされる。

検索結果は，図9.1.10のような出力様式で報告される。

① 端末タイプライタによる出力様式例

```
**********************************************************************************

ｹﾝｻｸ ﾖｳｺﾞ  : ｱﾄﾞﾚｽ

ﾖｳｺﾞ ﾉ ｲﾐ  : ﾊﾝﾁ ｼﾞﾖｳﾎｳｵ ﾃﾝｿｳ ｽﾙ ﾊﾟｲﾉ ﾃﾞｰﾀｺﾛ ﾏﾀﾊ ｲｷｻｷｵ ｱﾗﾜｽ ﾋﾖｳｼﾞﾉ ｺﾄ

ｼ ｭ ｯ ﾃ ﾝ  : JIS ｼﾞﾖｳﾎｳ ｼｮﾘ          (ﾆﾎﾝ ｷｶｸ ｷｮｳｶｲ)        P. 0801

            ﾂﾞｶｲ ﾃﾞﾝｼｹｲｻﾝｷ ﾖｳｺﾞ ｼﾞﾃﾝ (ﾆｯｶﾝ ｺｳｷﾞﾖｳ ｼﾝﾌﾞﾝｼｬ) P.   10

            ﾃﾞﾝｼｹｲｻﾝｷ ﾊﾝﾄﾞﾌﾞｯｸ      (ｵｰﾑ ｼｬ)               P. 10-8

**********************************************************************************
```

② 電算写植編集システムＪＥＭ３８５０による出力様式例

検索用語：線型計画法

用語の意味：互に関連し合っている活動の集を，最適にするための計算方法。
混合物を作る問題，人員配置の問題，輸送問題は好例。

出　　典：電子計算機用語辞典（日刊工業）　頁0138
電子計算機ハンドブック（コロナ社）　頁0089
総合コンピュータ辞典（ユニバック総研）　頁0154

図9.1.10　ＭＡＳＣＯＧにおける出力様式

検索におけるプログラムの一部を次に掲載する。

```
0090    333010/RAPID SECTION.
0091    333030/RAPID PAGE BUFFER 9.
0092    333040/RT WORDREC.
0093    333050/01 WD-R.
0094    333060/     02 R-CODE      PIC 9(5).
0095    333070/     02 R-WORD      PIC X(39).
0096    333080/     02 R-MEAN      PIC X(70).
0097    333090/     02 R-B1.
0098    333100/       03 B1        PIC X(2).
0099    333110/       03 P11       PIC X(5).
0100    333120/       03 P12       PIC X(5).
0101    333130/     02 R-B2.
0102    333140/       03 B2        PIC X(2).
0103    333150/       03 P21       PIC X(5).
0104    333160/       03 P22       PIC X(5).
0105    333170/     02 R-B3.
0106    333180/       03 B3        PIC X(2).
0107    333190/       03 P31       PIC X(5).
0108    333200/       03 P32       PIC X(5).
0109    333210/     02 FILLER      PIC X(2).
0110    400010 PROCEDURE DIVISION.
0111    400020 OPEN-RT.
0112    400030     OPEN OUTPUT LST-F.
```

```
FACOM 230-60     RAPID COBOL SOURCE LIST     -71.07.01-   (V=01/L=03)

CARDNO.  SEQNO. A  B
0113    400040/    RPD-OPEN DATABASE.
0114    400050     MOVE C17 TO CNTRL-CHA.
0115    400060     MOVE SPACES TO OUT-P.
0116    400070     WRITE PRT.
0117    400071     MOVE C5 TO CNTRL-CHA.
0118    400072     WRITE PRT.
0119    400080     MOVE C2 TO CNTRL-CHA.
0120    400090 IN-IR.
0121    400100     DISPLAY '*** ｸﾝｻｸ ﾖｳｺﾞ ｵ ﾀｲﾌﾟ ｲﾝ ｼﾃｸﾀﾞｻｲ (END ﾊ JOB-END)***'.
0122    400110     ACCEPT TYPE-IN.
0123    400120     IF TYPE-IN = 'END' GO TO END-RT.
0124    400130     MOVE TYPE-IN TO R-WORD.
0125    400140/    GET RANDOM WORDREC.
0126    400150     IF SCB2 = '0121' GO TO NO-REC.
0127    400160/    WHEN ERROR GO TO GET-ER.
0128    400170     IF R-WORD = TYPE-IN GO TO WRITE1.
0129    400171 CONT.
0130    400172/    CONTINUE.
0131    400173     IF SCB2 = '0121' GO TO NO-REC.
0132    400174     IF R-WORD NOT = TYPE-IN GO TO CONT.
```

(6) メンテナンスの方法

ファイル・メンテナンスには，データの追加処理，修正処理，消去処理がある。追加処理は，新造語や新生語のデータをつけ加えたり，未登録のデータを格納したりする場合に行なうものである。修正処理は，すでに格納されている用語，意味，出所などのデータの全部または一部を新しいデータと置換える変更や更新の作業である。消去処理は，格納されているデータが不要になった場合，あるいは間違って格納した場合に，そのデータを削除するために行なうものである。

〔追加処理〕メメンテナンス・データが，用語マスタにないことをチェックした後，データの格納（A01）の場合と同様，コマンドPUT　RANDOMを使って格納する。追加データが小量の場合は，キーボードから入力し格納する。

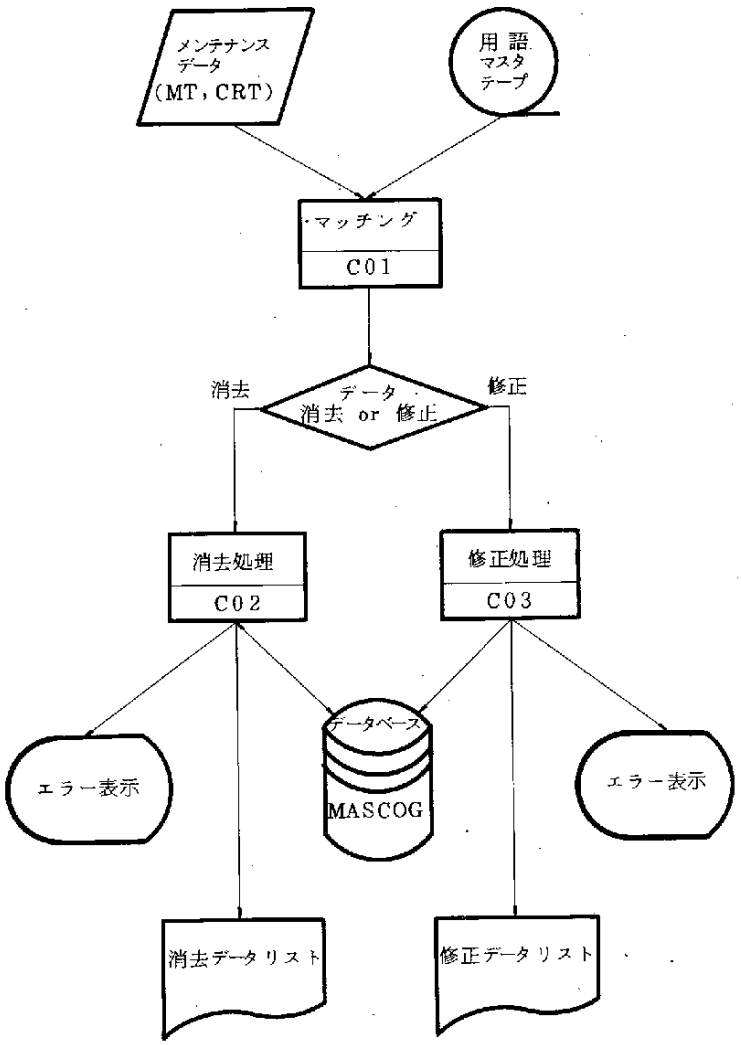


図9.1.11　MASCOGにおけるファイル・メンテナンスのプロセス

〔修正処理〕　用語の意味，出所（辞典略号，掲載頁）の修正を行なう。用語の修正は，追加または消去の方法で処理する。これは，用語そのものがキーワードであり，ランダマイズ・フィールドであるため制約上修正処理が行なえないためである。修正コマンド（ＭＯＤＩＦＹ）の実行は，ＧＥＴ　ＲＡＮＤＯＭで，修正すべき項目のリファレンス・コードをＲＴＣＳエリアにセットしてから行なう。その際，修正データは意味フィールドまたは出所フィールドに移されていなければならない。

修正の結果そのレコードは，旧レコードの格納されていたエリアに記録される。

修正レコードのアドレス情報，格納情報は，プリント・アウトしてその状況を確認する必要がある。

〔消去処理〕　消去処理は，削除すべきレコードのリファレンス・コードをＲＴＣＳエリアにセットし，消去コマンド（ＤＥＬＥＴＥ）により実行する。

消去結果，データ・ベース内の消去されたエリアはスペースになり，後に追加されるデータのための空領域となる。

消去したレコードの状況は，ＲＡＰＩＤユーティリティを利用し，プリント・アウトしてその結果を確認する必要がある。

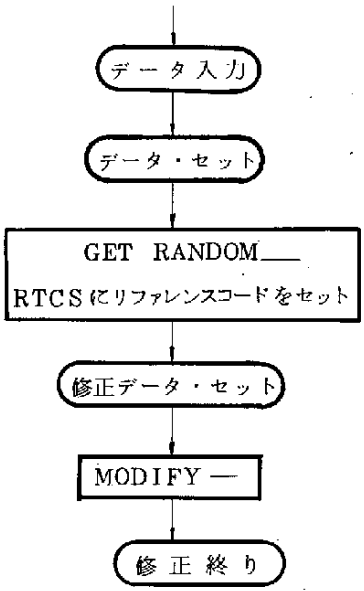


図9.1.12　　修正プロセス

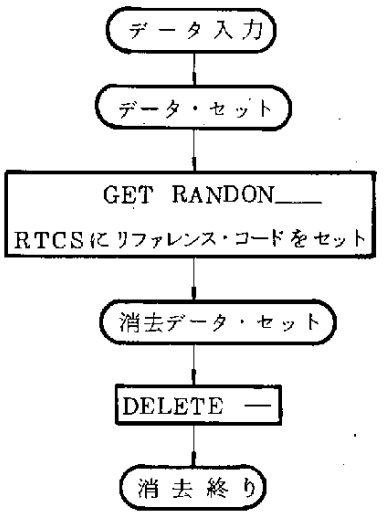


図9.1.13　　消去プロセス

9.2　サブシステム；コンピュータ・ニュース・検索システム（MASCOA）

9.2.1　汎用データ・マネジメント・システムRAPIDをベースにしたMASCOA

(1) システムの概要

コンピュータニュース・検索システム（MASCOA）は，コンピュータに関連する内外のニュース記事を英数字・カナ文字で1件当り66文字以内に要約し，データ・ベースで管理することによって，必要に応じて適時該当するニュースを検索し提供サービスするシステムである。

MASCOAで管理するコンピュータ・ニュースは，コンピュータ産業全般からみた場合およそ次の分野の記事情報である。

① 政策・経営
② 技術開発
③ 関連産業の動向
④ 生産・販売
⑤ サービス
⑥ 管理・運用
⑦ 利用動向

MASCOAは，MASCOTモニタの下で稼動するサブシステムである。データ・ベース作成には，汎用データ・マネジメント・システムRATPIDの保有する機能を活用しており，データはその管理によって集団磁気ディスクまたは磁気テープに格納している。MASCOAに収容できるデータはおよそ1,000,000件であり，そのうち利用度の高いあるいは新しいニュースについては，1次情報として200,000件を集団磁気ディスクに，比較的利用度の低いあるいは古くなったニュースについては2次情報として800,000件を磁気テープに，各々記録できるよう設計している。ファイル編成は，1次情報がランダム・アクセス・ファイル，2次情報がシークエンシャル・ファイルである。

これらのデータのアップ・デートは，別に用意するデータ発生統計，利用統計などの状況を検討しながら行なう。検索は，語または句からなる自然語をキーワードとして行なう。用語シソーラスは，十分利用者の要求を満すことはできないが，とりあえずこのような体系化ができるのではないかという程度のものを作成している。シソーラスの階層は，レベル0からレベル3までの4段階であり，これにレベル3の同意語を取扱うレベル3'という補足レベルを設けている。シソーラスをもつことは，マクロ的な面からもミクロ的な面からも検索でき，さらに，違った分野の違ったレベルのキーワードとも組合せて情報検索ができるなど巾広い利用が可能になることから便利である。どのニュースにも，記事情報であれば必らず，その文章の主題部分，目的部分，それを補足する部分が存在するが，MASCOAにおいてはこれらを，情報の発生源を示す事業体名，ニュースの目的を示す主要用語，ニュースの内容をより十分表現するための補足用語という形で文章の中の3つの部分からキーワードを抽出している。検索は，AND(*)機能とOR(+)機能をキーワード（8個まで使用可能）と自由に組合せて指定できる。

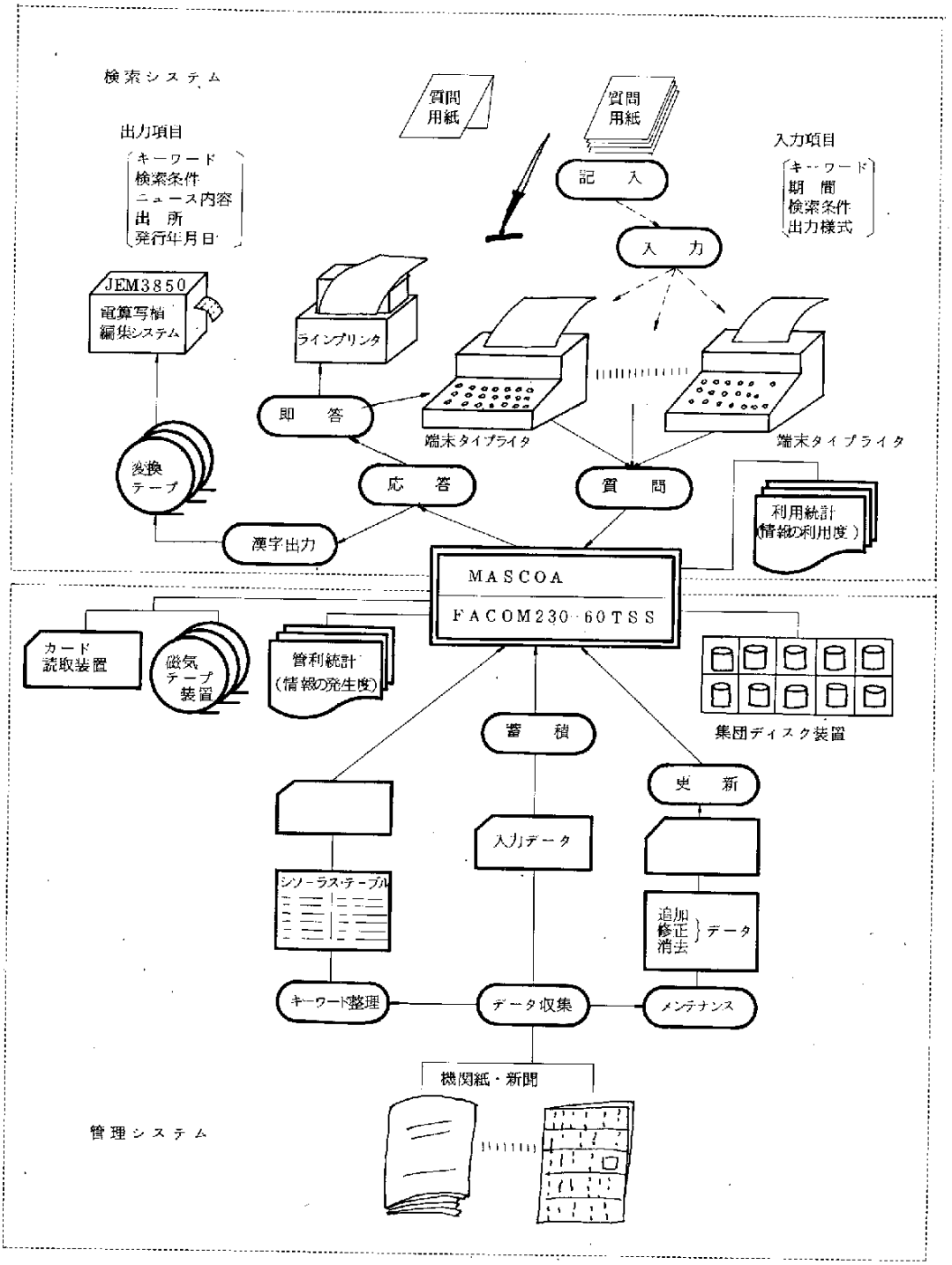


図 9.2.1　コンピュータニュース・検索システム：MASCOA

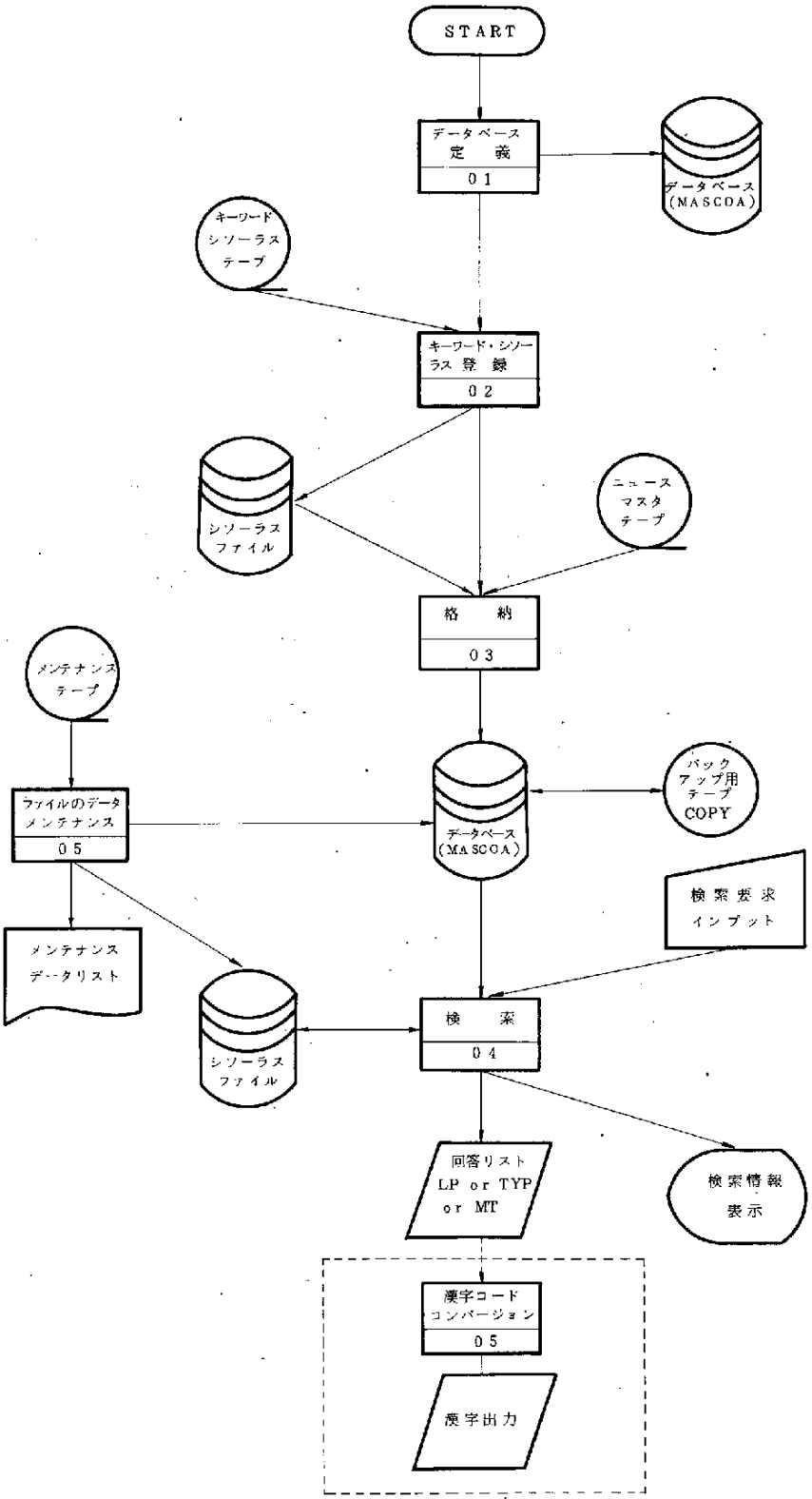


図9.2.2　MASCOAにおけるシステム・フローチャート

たとえば，４６年以後のＩＣの開発状況を知りたければ，（４６.＋４７）＊（ＩＣ）＊（カイハツ）と検索指定することにより，ニュースの件数や記事内容が即時にタイプライタまたはラインプリンタに報告できる。また，緊急を要しない場合はより見易くするために電算写植編集システムの利用も可能で，ＪＥＭ３８５０によるコンピュータ・ニュースの漢字出力ができる。この場合，ＭＡＳＣＯＡにおける出力情報はカタカナが中心であるので，これらのニュースをカナ漢字交換ルーチンを通し，文章中の用語をカナ文字コードから漢字コードに換えてテープにとり，そのデータをＪＥＭ３８５０のシステムに引き渡すという手続きが必要になる。

(2) データ収集の方法

コンピュータに関するニュース記事は，日本工業新聞，日刊工業新聞，電波新聞，電気新聞，日本経済新聞などの新聞，およびＤＡＴＡ　ＮＥＴ　ＮＥＷＳ，コンピュータ・メーカ発行のニュース誌などの機関紙から収集した。ニュース記事は各紙が同時に掲載したり，発行日を異にして掲載するなど重復するものが多々あるので，まずこれらを整理した。

それらの整理を終えたニュース記事は，事業体名，主要用語，補足用語などのキーワードを考慮しながら英数字，カナ文字で６６文字以内にまとめ資料表に転記した。出所，発行年月日，整理番号なども，このとき同時に記入した。

次に，転記したデータの中から３個以内でキーワードを抽出した。ＭＡＳＣＯＡにおけるキーワードは，いわゆるＫＷＯＣ方式をとり，語または句からなる自然語である。たとえば，「日立は，ＨＩＴＡＣ８４５０用磁気ディスク装置を開発した。」というニュースがある場合，事業体名として「日立」，主要用語として「磁気ディスク装置」，補足語として「開発」がキーワードとなる。キーワードは，事業体名，主要用語，補足用語，各々１個づつとは限らず，その数が２：１：０，１：０：１，０：３：０など，どのような割合でもかまわない。キーワードを抽出するにあたっては，そのキーワードがシソーラス・テーブルに登録されているか否かの照合が必要である。キーワードが登録されていない場合は，データを格納する以前に，それらのキーワードを一括して追加処理のためのメンテナンス・ルーチンを通し，テーブルに登録しておかなければならない。

データの収集段階では，コンピュータに関する巾広い知識をもった専門家が必要で，とくにデータの整理とキーワードの抽出においては，シソーラスをよりよいものに作りかえながらデータを作成してゆく高い技術と尨大な労力が必要である。

資料表に転記したデータは，入力データとして次のようなフォーマットでカードに記録する。

用語データ

整理番号	キーワード 1	キーワード 2	キーワード 3
（５桁）	（２０桁）	（２３桁）	（３２桁）

ニュース・データ

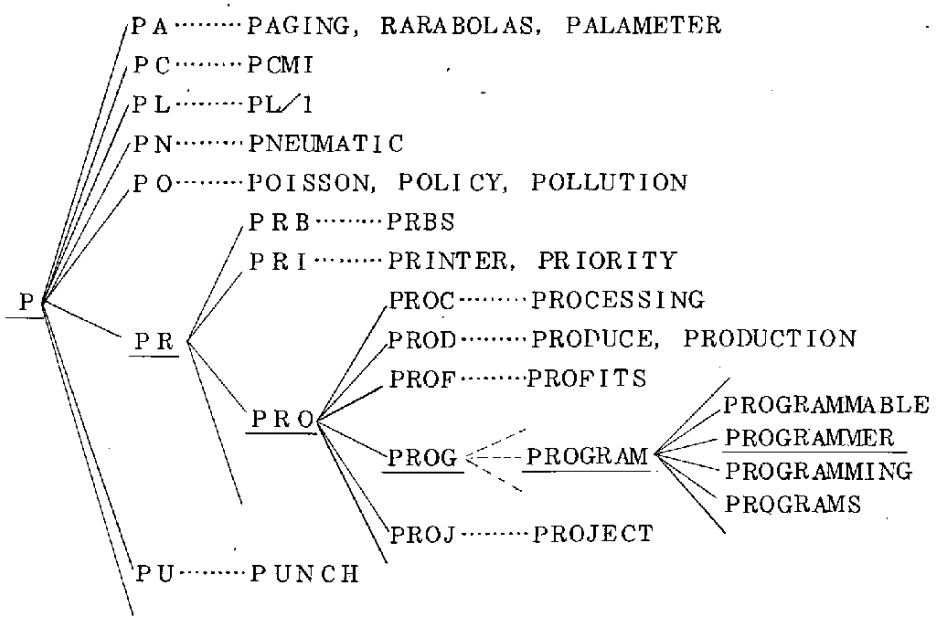
整理番号	出所コード	発行年月日	ニュース内容
（5桁）	（3桁）	（6桁）	（66桁）

(3) データ・リンケージの方法

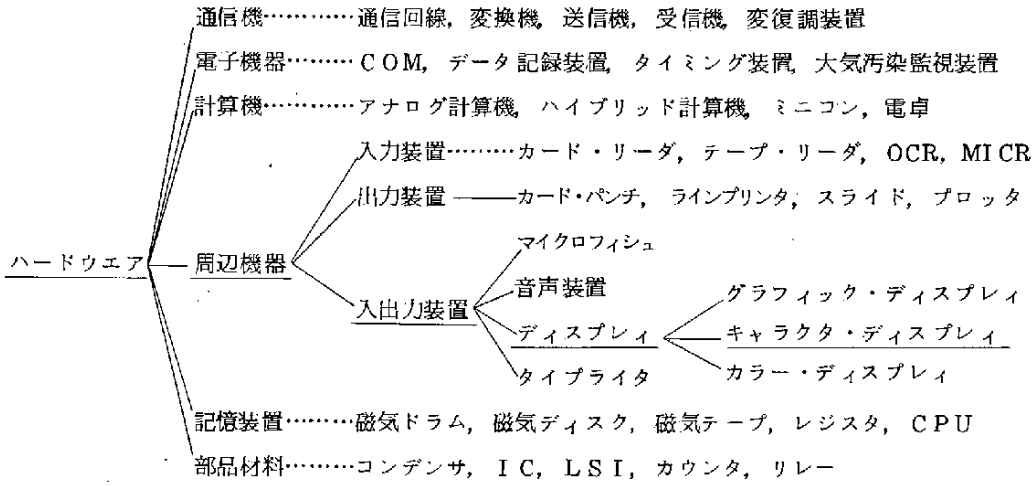
コンピュータ用語のデータ・リンケージは，現在のところ体系化され標準化されたものがない。コンピュータ分野の歴史は，いまだ浅く発展途上にあるため，このような成長段階におけるコンピュータ・ニュースの用語は，従来から引き続き使われてきた用語，新しく発生した用語，他の分野から導入された用語，省略して使われる用語，変形して使われる用語，不要となった用語などが入乱れて使われている。

このような状況においてコンピュータ用語の体系化を行なうことは非常にむずかしく，あえてシソーラスを造っても，リンケージの洩れや時間の経過にともない間違った関係を生ずる恐れがあり，十分利用に供するものにならないと思われる。シソーラスを常に満足のゆく状態にしておくためのメンテナンスは大変な作業で，コンピュータ・ニュースの分野のシソーラス造りのメリットが十分認識され，このために莫大な労力と費用と時間をかけることが許されなければ，この試みは達成できないであろう。

一般に用語シソーラスを造るには，語源あるいは語の構成から体系化を行なう場合と，活用面から体系化を行なう場合とがある。前者において，"PROGRAMMER"という用語をとりあげて，その周辺のシソーラスをみてみると次のようになっている。



また，活用面からみた場合のコンピュータ用語のシソーラスは，"キャラクタ・ディスプレイ"という用語をとりあげてみると，次のようになっている。



コンピュータ・ニュースで使われる用語のシソーラスは，これらの方法で体系化してゆくことができるが，この分野に存在する用語の全てをシソーラスで取扱うのではとりまとめ上，収拾がつかなくなる恐れがあるので，実用化するためには用語の選別または標準化を行なったうえで，これらの方法を利用する心構えが必要である。

そこで，MASCOAでは，利用面からはまだ十分満足ゆくものではないが，コンピュータ用語の際立った分野において，簡単なシソーラスを後者の活用面から把える方法で造ることにより，できるだけ実際の利用に近づけるための実験を試みた。

階層は，レベル0からレベル3までの4段階であり，レベル3の同意語を取扱うレベル3′を設けた。MASCOAでは，コンピュータ・ニュースの記事情報を，「何は」（主題部分），「何々を」（目的部分），「何々した」（動作部分）の3つの部分から把え，それぞれにレベル0で事業体名，主要用語，補足用語という総称をつけることにした。

レベル0におけるインデックス

事業体名
主要用語
補足用語

レベル1は，レベル0の大分類で表9.2.1に示すような，事業体では生産業，商業サービス業，計算機製造業，……，主要用語では標準化，大型プロジェクト，優遇措置，税制……といったキーワードである。なお，補足用語については活用面からの体系化が不可能であるので，最下位のレベル3とその関連のレベル3にのみキーワードをもつことにした。

レベル2は，将来レベル1とレベル3の間の中分類が必要になる可能性があるので設けたもので，

表9.2.1　レベル1におけるインデックス・テーブル

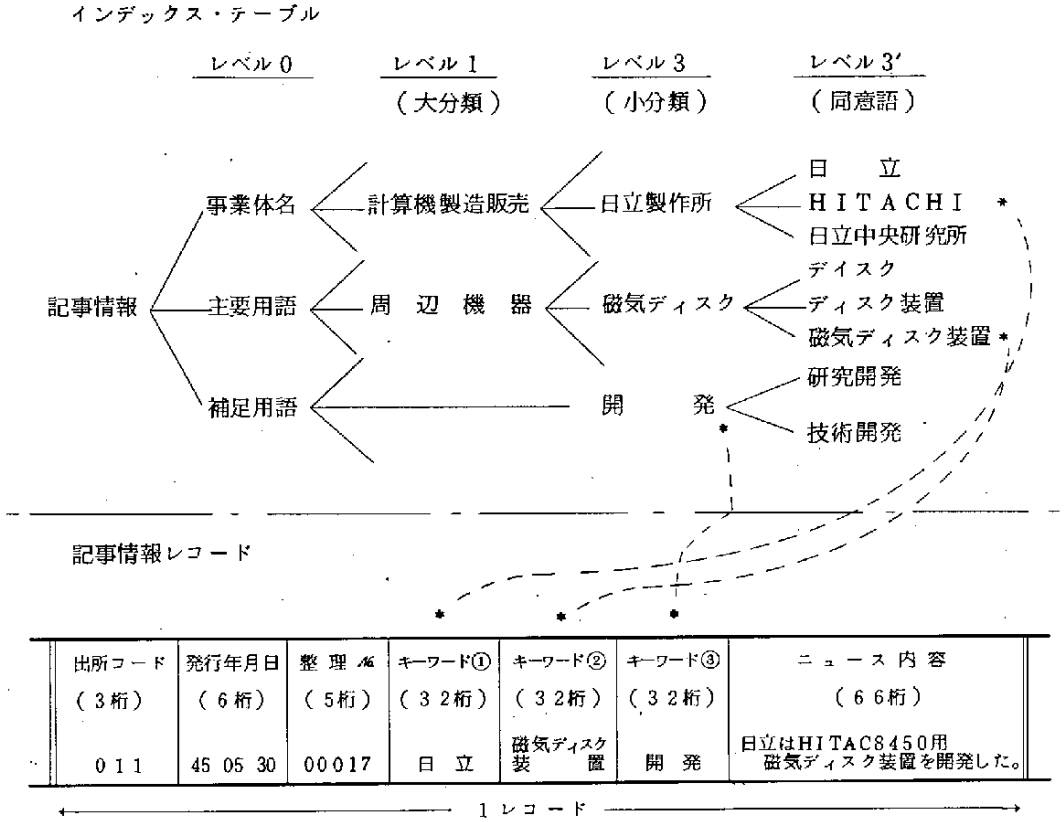
	（事　業　体）	31	オートメーション
01	生　産　業	32	通　信
02	商業・サービス業	33	システム開発
03	計算機製造販売	34	手法，処理方法
04	電子機器製造販売	35	汎用ソフトウエア
05	ソフトウエア会社・計算センター	36	特殊問題向プログラム
06	官　公　庁	37	言語プロセッサ
07	団　　体	38	管理プログラム
08	学　　校	39	周辺機器
09	病　　院	40	部品，材料
		41	記憶装置
	（主　要　用　語）	42	計　算　機
11	電子工業全般	43	電子機器
12	電子計算機全般	44	媒　　体
16	標　準　化	46	業務管理
17	大型プロジェクト	47	設立，合併
18	優遇措置，税制		
19	技術者教育	51	価　　格
20	情報処理の高度化	52	オンライン・リアルタイム
21	ソフトウエアの流通	53	処理サービス
22	情報産業の育成	54	展示会，講習会，講演会
23	特許，権利保護		
24	資金対策		（補　足　用　語）
		61	そ　の　他
30	制　　御		

現在のところ使用していない。ダミーレベルを設けることは，たとえばレベル1の計算機製造販売を，さらに電子計算機製造販売，ミニコン製造販売，電卓製造販売，会計機製造販売などと再分化したい場合に便利である。

レベル3は，用語をミクロ的に把えた場合，その用語の性質を表わす最小単位である。たとえば，日立製作所，磁気ディスク装置，パンチカードなどがこのレベルである。次に示す表はレベル3のキーワードの一部を事業体名，主要用語，補足用語に分けてリストしたものである。

事業体名		主要用語		補足用語	
03001012	ニホンCDC	3300118	ショウテイキショウホウシステム	6100061	キセイ
0300102	GE	3300119	シミンショウホウカンリシステム	6100062	キショユウテイケイ
03001021	GENERAL ELECTRIC	3300120	ショトウアナウスコウセイシステム	6100063	キショユウトウコウ
0300103	シャープ	3300131	ジテイメンショヨンシステム	6100064	キヨウカ
03001031	SHARP	3300151	ダイキボセンカンシテレメータシステム	6100065	キショユウ
03001032	ハヤカワデンキ	3300152	ターミナルシステム	61000651	サクセイキショユウ
0300141	タカチホコウエキ	3300161	ライキカイハツ	6100066	キショユウコウカン
03001411	タカチホ	3300181	データショウショウシステム	6100067	キヨウテイ
0300142	タイカーケイサンキ	3300182	データカンリシステム	6100068	キショユウショユウ
0300143	タケダリケン	3300183	テレメータシステム	6100069	キヨウリョク
0300151	チエルヘルシ	3300184	DIPS	61000691	キヨウチヨウ
0300161	DEC	3300185	データハンクセツケイ	6100070	キヨウトウケンキュウ
0300171	トウシバ	3300186	デンシカトケイシステム	61000701	キヨウトウカイハツ
03001711	トウキヨウシバウラデンキ	3300187	テストウエイキショウシショウホウショリシステム	6100071	キショウセイ
03001712	TOSHIBA	3300188	デンシショクシシステム	6100072	キヨウイクキカン
0300191	ニチデン	33001881	デンサンシャショクシステム	6100073	キヨウトウリョウ
03001911	ニホンデンキ	33001882	シャショクシステム	6100091	ケントウ
03001912	NEC	3300201	トウトウセツ	6100092	ケントウキュウ
0300192	ニホンミニコン	3300202	トツキヨショウホウケンサクシステム	6100093	ケイカク
0300193	ニホンMDS	3300203	トシコウツウシステム	61000931	シテキショウケイカク
0300194	ニホンビクター	3300204	TPDS	61000932	キホンケイカク
0300195	ニホンコロンビア	3300251	VERSION	6100094	ウンキ
0300196	ニッショウエレクトロニクス	3300252	バンクデンシトウヘンシュウシステム	6100095	ウンシュウ
0300197	ニホンシムキ	3300261	ヒデオシステム	6100096	ケイサイセイチョウ
0300198	ニホンケイサンキハンバイ	3300271	ヒデオ・デイスプレイシステム	6100097	ウラテイ
0300211	ハネウエル	3300272	フトウサンショウホウショリ	6100098	ウンキュウ
03002111	HONEYWELL	3300273	ファクシミリシステム	6100099	ウンキュウシャ
0300212	パロウス	3300274	プログラムカイハツ	6100100	ウンキュウショヨ
03002121	バローズ	3300281	マイクロフイルムシステム	6100101	ケイエイショウホウ
03002122	BURROUGHS	3300291	メデイカス	6100102	ケイエイ

レベル3'は，レベル3の同意語で異音同語，俗語，略語などである。たとえば，レベル3の磁気ディスクにおけるレベル3'にはMD，ディスク装置，磁気ディスク記憶装置などがある。MASCOAにおけるシソーラスとデータとの関係を図9.2.3に示す。



出所コード（3桁）	発行年月日（6桁）	整理No.（5桁）	キーワード①（32桁）	キーワード②（32桁）	キーワード③（32桁）	ニュース内容（66桁）
011	45 05 30	00017	日立	磁気ディスク装置	開発	日立はHITAC8450用磁気ディスク装置を開発した。

1レコード

図9.2.3　MASCOAにおけるシソーラスとデータとの関係

また，キーワードとニュース記事内容のリストの一部を次に掲載する。

スウチセイギョソウチ	フシ゛ツウハ FACOM-Rマタハ230-25オナイソ゛ウノスウチセイギ゛ョソウチFANUC250オカイハツ
スウリケイカクシステムMPS	ニチテ゛ンハ ハンヨウオオカ゛タスウリケイカクシステム MPSオ カイハツ
スウリケイカクヨウフ゛ログ゛ラム	ニホンUNIVACハ スウリケイカクヨウフ゛ログ゛ラム EMPS-カ゛ンマヨオ カイハツ
スコープ゛フ゛リンタ	イワツウハ CRTチ゛イスフ゛レイノカ゛ソ゛ウオインシヨスル スコープ゛フ゛リンタオ カイハツ
ススムコウキ゛ヨウ	ススムコウキ゛ヨウハ フ゛ラス゛マシ゛ュウコ゛ウハンノウオリヨウシタ ュウキコウフ゛ンシハクマクオサクセイ
ススムコウキ゛ヨウ	ススムコウキ゛ヨウハ フ゛ラス゛マオリヨウシタ ュウキコウフ゛ンシハクマクセイソ゛ウキ゛シ゛ュツオ カイハツ
スタンフオート゛ケンキュウシ゛ョ	スタンフオート゛ケンキュウシ゛ョハ NASAノエンジ゛ョチテ゛ シンセタ゛イノICトモイワレルVCオカイハツ
スタンレーチ゛ンキ	スタンレーチ゛ンキハ イリヨウキキニシンシュツシ コシ゛ンシ゛ョウホウケンサクシキヘ゛ツソウチオ カイハツ
スタンレーチ゛ンキ	スタンレーチ゛ンキハ シンクタンク スタンレー キ゛シ゛ュツケンキュウシ゛ョオ セツリツノ ケイカク
スタンレーチ゛ンキ	スタンレーチ゛ンキハ チ゛ンサンキヨウタンマツキカンレンフ゛ンヤヘ ホンカクテキニ シンシツ
スヘ゛ースレコータ゛	テイアツクハ シ゛ンコウエイセイセキサイヨウ スヘ゛ースレコータ゛ノ コクサンカニ セイコウ
スミトモシヨウシ゛	スミトモシヨウシ゛ョニチテ゛ンハ カコ゛シマシレクレーションセンターニCATVシセツオ セツチスル
スミトモシ゛ュウキ	スミトモシ゛ュウキハ メイアンヤ カタチオニンシキチ゛キル シ゛ンコウサノウロホ゛ツトノ カイハツオ スル
スミトモチ゛ンキ	スミトモチ゛ンキハ チ゛ンシキキオ リヨウサンカスルタメ ニホン・ニッシンチ゛ンキ・メイテ゛ントキヨウリヨク
スミトモチ゛ンキコウキ゛ヨウ	スミトモチ゛ンキコウキ゛ヨウ ハ アメリカウンュシヨウニ チ゛ータツウシンキ゛シ゛ュツノュシュツニ セイコウ
スミトモチ゛ンコウ	スミトモチ゛ンコウハ チ゛ンサンキヨウチ゛ータシュウシュウタンマツソウチオ ソレンニ ュシュツ
スミトモチ゛ンコウ	スミトモチ゛ンコウハ ハイセンノソリツト゛ステートカトセイエル フレキシフ゛ル・フ゛リントカイロオ カイハツ
スミトモチ゛ンコウ	スミトモチ゛ンコウハ ニッシンチ゛ンキトチ゛ータショリソウチ チ゛ータヒ゛ーチ゛セ゛ンメンテイケイ
スミトモチ゛ンコウ	スミトモチ゛ンコウハ フクオカシノ コウイキソウコ゛ウコウツウセイギ゛ョシステムオ カンセイ
スミトモチ゛ンコウ	スミトモチ゛ンコウハ ミリハト゛ウハカンノリヨウサンカノ ハ゜イロツトフ゛ランオ カンセイ
スミトモチ゛ンコウ	スミトモチ゛ンコウハ チ゛ンセンノサイテキチヨウリヨクオセトメル フ゛ログ゛ラムオ カイハツ
スミトモチ゛ンコウ	スミトモチ゛ンコウハ オフ゛トエレクトロニクスヨウ カ゛リュウムリンタンケツショウオ カイハツ
スミトモトクシュキンソ゛ク	スミトモトクシュキンソ゛クハ キオクソシチ゛イ コウタ゛ンセイコ゛ウキン ミスハ゜ンEL4オ カイハツ
スワセイコーシャ	スワセイコーシャ ハ コカ゛タ コウソクフ゛リンタ EP101ノ ショウヒンカニ セイコウ
セールス・エンシ゛ニアリング゛	ナコ゛マキコウハ セールス・エンシ゛ニアリング゛テ゛ ニッセイサンキ゛ョウト キ゛ヨウムテイケイ
セイカツシ゛ヨウホウセンター	マツシタチ゛ンコウ ハ ハイソウ キ゛ヨウム ノ コウリツカ ニ 'チュウフ゛ヨヤクセンター' オ カイセツ
セイキ゛ヨカイロ	フシ゛ツウハ ICセイキ゛ヨカイロオカセットニオサメタFUNUC10オ コウカイ
セイキ゛ヨソウチ	シンチ゛ンケ゛ンハ ROMホウシキニヨル モシ゛ュール セイキ゛ヨソウチ シーケンスフ゛ロセッサオカイハツ
セイキ゛ヨソウチ	オキ・フシ゛ツウハ フイルコ・フォート゛シャカラ チ゛ンケ゛ンセイキ゛ヨソウチノ キヨウリヨクオセトメラレル
セイキ゛ヨヨウコンヒ゛ータ	ニチテ゛ンハ セイキ゛ヨヨウコンヒ゛ュータノシンシリース゛NEAC3200オ ハツヒ゛ヨウ
セイキ゛ヨヨウコンヒ゛ュータ	シヤープ゛ハ セイキ゛ヨヨウコンヒ゛ュータノ シ゛シセセイサンニノリタ゛ス
セイキ゛ヨヨウソフトウエア	ヒタチハ セイキ゛ヨヨウソフトウエア ヘンテ゛ンシヨウケイサンヨウフ゛ログ゛ラムHISAPSオ カイハツ
セイキ゛ヨヨウテ゛ンサンキ	ヒタチハ ストレツチタイフ゛ノ セイキ゛ヨヨウテ゛ンサンキHIDIC500オ カイハツ

トスカルフ゛リンタBC1415P	トウシハ゛ハ ラインフ゛リンタシキ テ゛ンタク トスカルフ゛リンタBC1415Pオ カイハツ
トツキヨ	コウキ゛インハ トツキヨ・カ゛クシ゛ユツケンキユウノ リヨウシ゛ツタイノ チヨウサケツカオ トリマトメタ
トツキヨ	ヒタチハ トツキヨオ シヨウヒントシテコウカイスルコトオキメ トツキヨエイキ゛ヨウセイト゛オ ホツソク
トツキヨ	イテ゛ミラ・ムツミマイクロハ トツキヨコウホウノマイクロフイルムシステムオカイハツシ チカクハンハ゛イスル
トツキヨ	トウシハ゛ ハ カテイヨウVTRノ シ゛キキロクホウシキノ トツキヨオシハ゛テ゛ンカ゛ シンカ゛イトウツタエ
トツキヨ	アメリカRCAハ ワカ゛クニテ゛ンタクメーカニ RCAノトツキヨ シ゛ツシケイヤクオムスフ゛ヨウ ヨウキユウ
トツキヨ	フシ゛ツウハ ニホンハ゛ルスモータトノ トツキヨアラソイテ゛ライセンスケイヤクオムスフ゛コトテ゛ カイケツ
トツキヨケン	コウキ゛インハ オウカ゛タフ゛ロシ゛エクトノ ノウハウ・トツキヨケンシヨウオ ヒタチ・ニチテ゛ンニ キヨカ
トツキヨケン	アフ゛ライト゛テ゛ータ・リサーチシヤ ニ ソフトウエア テ゛ ハツノ トツキヨケン カ゛ アタエラレタ
トツキヨシ゛ツシケン	ニホンテ゛ンシケイサン ハ TSSニヨル カカ゛クケイサン サービ゛ス カイシ
トツキヨシ゛ヨウホウケンサクシステム	ニホントツキヨシ゛ヨウホウセンターハ トツキヨシ゛ヨウホウケンサクシステムノソフトウエアオ イチフ゛カンセイ
トツキヨシ゛ヨウホウケンサクシステム	ニホントツキヨシ゛ヨウホウセンターノ トツキヨシ゛ヨウホウケンサクシステムカ゛カンセイ コウカイシタ
トツキヨチヨウ	トツキヨチヨウハ マイトシ10マンケンノ セシ゛ヨウヒヨウトウロクノシンサニ テ゛ンタンキオ リヨウ
トツキヨチヨウ	トツキヨチヨウト ニホントツキヨキヨウカイハ ソフトウエアノネコ゛ニカンスル アンケートチヨウサオシタ
トツキヨチヨウ	トツキヨチヨウハ ニホントツキヨシ゛ヨウホウセンターオ 45ネンチユウニ ホツソク
トツキヨリヨウシハライ シ゛ユンヒ゛ キンセイト゛	ツウサンショウ ハ ライネンカラ トツキヨリヨウシハライ シ゛ユンヒ゛キンセイト゛オ ソウセツ スル
トツキヨリヨウシハライシ゛ユンヒ゛キンセイト゛	ツウサンショウ ハ テ゛ンシキ゛シ゛ユツ ニ トツキヨシハライシ゛ユンヒ゛キンセイト゛ オ ソウセツヨテイ
トツハ゛ンインサツ	タ゛イニホン・トツハ゛ンハ テ゛ンシフ゛ヒンキヨウカノタメ ノーテイツクエレクトロニクストテイケイコウシヨウ
トツハ゛ンインサツ	トウキヨウテ゛ンカ・トツハ゛ンハ シ゛キインサツカ゛イシヤ トウキヨウシ゛キインサツオ セツリツ
トヨタ	トヨタハ テキサス・インツルメンツト エレクトロニクスシ゛ト゛ウシヤフ゛ヒンノ キ゛シ゛ユツト゛ウニユウオス
トライアック	ミツヒ゛シハ カイテンシ゛テ゛モ セイキ゛ヨカノウナ サイリスタノトライアックオ カイハツ
トラツキング゛・スヘ゛クトラム・アナライサ゛	タケタ゛リケンハ シユウハスウソクテイヨウ トラツキング゛・スヘ゛クトラム・アナライサ゛オ カイハツ
トラックスケール	コウヨウテ゛ンシハ ケイリヨウシ゛ムシヨリオカンセ゛ンシ゛ト゛ウカシタ トラックスケールオ カイハツ
トランシ゛スタ	ヒタチ ハ テイオン シリコンハント゛ウタイ LTPトランシ゛スタノ コウキ゛ヨウカ ニ セイコウ
トランシ゛スタ	テ゛イハ゛イセスシヤ ハ カ゛ラスノトランシ゛スタト アタラシイハント゛ウタイオ カイハツ
トランシ゛スタ	カカ゛クキ゛シ゛ユツチヨウハ コウシユウハトランシ゛スタオトウシハ゛・レンソ゛クカコウソウチオヒタチニイタク
トランシ゛スタ	マツシタテ゛ンシハ オスタ゛ケテ゛ テ゛ンリユウオカエルPSSスイツチト トランシ゛スタオ カイハツ
トランシ゛スタ	テ゛ンキシケンシ゛ヨハ 10GHZテ゛アンテイスルDSMMOSトランシ゛スタオ カイハツ
トリマトメ	JTPDECハ フ゛ログ゛ラムリユウツウソクシンノタメ フ゛ログ゛ラムトウロクシユウオ トリマトメタ
トリマトメ	センフ゛シヨウハ カ゛ツコウニアル シチヨウキサ゛イノ セツヒ゛シ゛ヨウキヨウオ トリマトメタ
トリマトメ	テ゛ンシキカイコウキ゛ヨウカイハ ケ゛ンシ゛ヨウト シヨウライノテンホ゛ウオ トリマトメ ハツヒ゛ヨウシタ
トリマトメ	サンコウシン ハ シ゛ヨウホウサンキ゛ヨウ ハツテンノタメノ シチクニカンスル トウシンアンオ トリマトメタ
トリマトメ	JIPDEC ハ ヘ゛イコクニオケル シ゛ヨウホウシヨリサンキ゛ヨウノ シ゛ツタチヨウサオ トリマトメタ

ｿｳｻﾝ	ｿｳｻﾝﾄ ｷｶｲﾒｰｶﾊ ﾆﾎﾝｶｲﾖｳｷｷ ｷ*ｼ*ｭﾂｷﾖｳｶｲﾉ ｾﾂﾘﾂｵ ｹﾝﾄｳ
ｿｳｻﾝ	ｿｳｻﾝﾊ ｼ*ﾖｳﾎｳｼｮﾘｷ*ｼ*ｭﾂｼｬ ﾆﾝﾃｲｾｲﾄ* ﾀ*ｲ1ｶｲﾆﾝﾃｲｼｹﾝｵ ｵｺﾅｳ
ｿｳｻﾝ	ｿｳｻﾝｼｮｳﾊ ｷｶｲｺｳｷ*ﾖｳ ｼﾝｺｳﾘﾝｼ*ﾘﾁﾎｳﾉ ｴﾝﾁﾖｳﾊｵｺﾅﾜｽ*ｳﾁｷﾙﾎｳｼﾝ
ｿｳｻﾝ	ｿｳｻﾝﾊ ﾃ*ﾝｼｷｶｸｶｵ 7ｶ*ﾂｺ*ﾛﾏﾃ*ﾆ ﾎﾂｿｸ ｻｾﾙﾖﾃｲ
ｿｳｻﾝ	ｿｳｻﾝｼｮｳﾊ ｿﾌﾄｳｴｱｶｲﾊﾂﾆﾓ ﾎｼ*ﾖｷﾝｾｲﾄ*ｵ ﾃｷﾖｳｽﾙｺﾄｵ ｹﾝﾄｳﾁｭｳ
ｿｳｻﾝ	ｿｳｻﾝｼｮｳﾊ ﾆﾎﾝｻﾝｷ*ﾖｳｷ*ｼ*ｭﾂ ｼﾝｺｳｷﾖｳｶｲｵ ﾎﾂｿｸｻｾﾙ ﾖｳ ｽｽﾒﾃｲﾙ
ｿｳｻﾝ	ｿｳｻﾝｼｮｳﾊ ｺｳｺｳｷ*ﾖｳｾｲｻﾝﾆﾂｲﾃ ﾄｳｹｲﾌ*ﾝｾｷｹﾂｶｵ ﾊﾂﾋ*ﾖｳ
ｿｳｻﾝ	ｿｳｻﾝﾊ ﾊｲﾌ*ﾘｯﾄ*ICﾉ ｼ*ﾂﾀｲﾁﾖｳｻｵ ｶｲｼｼﾀ
ｿｳｻﾝ	ｿｳｻﾝﾊ 43ﾈﾝﾆｵｹﾙ ｼｭｳｾｷｶｲﾛﾉ ｾｲｻﾝｼ*ﾂｾｷｵ ﾏﾄﾒﾀ
ｿｳｻﾝ	ｿｳｻﾝﾊ ｼ*ﾖｳﾎｳｼｮﾘｼﾝｺｳｼ*ｷ*ﾖｳｷｷｮｳｶｲﾉ ｾﾂﾘﾂﾎﾂｷﾆﾝｶｲｽ ﾋﾗｲﾀ
ｿｳｻﾝ	ｿｳｻﾝﾊ 45ﾈﾝ9ｶ*ﾂﾆ ﾃ*ﾝｼ､ﾃ*ﾝｷｷｶｲｺｳｷ*ﾖｳ ｾﾂﾋ*ﾄｳｼｹｲｶｸﾁｮｳｻｵ ｼﾀ
ｿｳｻﾝ	ｿｳｻﾝﾊ ｼ*ﾖｳﾎｳｼｮﾘｷ*ｼ*ｭﾂｼｬ ｼｶｸﾆﾝﾃｲｾｲﾄ*ﾉ ｾﾝﾓﾝｲｲﾝｶｲｵ ﾎﾂｿｸｻｾ
ｿｳｻﾝ	ｿｳｻﾝﾊ ｺｳｷ*ﾖｳｳﾛﾎ*ﾂﾄｻﾝｷ*ﾖｳﾉ ｲｸｾｲｻｸﾆﾂｲﾃ ｹﾝﾄｳ
ｿｳｻﾝ	ｿｳｻﾝﾊ ﾃ*ﾝｻﾝｷｶｾ*ｲﾋｮｳｼ*ｭﾝｵ1/2ﾆｽﾙ ｺﾃｲｼｻﾝｾ*ｲｹｲｹﾞ*ﾝｿﾀｽ ｾｳｹﾙ
ｿｳｻﾝ	ｿｳｻﾝﾊ 46ﾈﾝﾄ* ﾃ*ﾝｻﾝｷｶﾝｹｲ ｼ*ｭｳﾃﾝｾｲｻｸｵ ﾄﾘﾏﾄﾒﾀ
ｿｳｻﾝ	ｿｳｻﾝﾊ 46ﾈﾝ ｷ*ﾖｳｼｭﾍﾟｿｺｳｺｳｷ*ﾖｳｾｲｻﾝﾐﾄｵｼｵ ﾊﾂﾋ*ﾖｳ
ｿｳｻﾝ	ｿｳｻﾝﾊ ﾆﾎﾝｻﾝｷ*ﾖｳｷ*ｼ*ｭﾂ ｼﾝｺｳｷﾖｳｶｲｵ ｾﾂﾘﾂｽﾙ ﾖﾃｲ
ｿｳｻﾝ	ｿｳｻﾝﾊ ｼ*ﾖｳﾎｳｻﾝｷ*ﾖｳﾘﾝｼ*ﾘﾁﾎｳｵ ｱﾗﾀﾒﾃ ｾｲﾃｲｽﾙ ﾎｳｼﾝ
ｿｳｻﾝ	ｿｳｻﾝﾊ ｶｸｷ*ﾆ 1969ﾈﾝ ｿｳｼｮｳﾊｸｼｮｵ ﾎｳｺｸ ﾊﾂﾋ*ﾖｳｼﾀ
ｿｳｻﾝ	ｿｳｻﾝﾊ ｺｸｻﾝ[Cｵｲｸｾｲｽﾙﾀﾒﾉ ｿｳｺ*ｳｾｲｻｸｵ ｳﾁﾀ*ｽﾎｳｼﾝ
ｿｳｻﾝ	ｿｳｻﾝｼｮｳｿｼｷｵ ｶｲｾｲｼ ﾃ*ﾝｼｷｷﾃ*ﾝｷｶ､ﾃ*ﾝｼｾｲｻｸｶｵ ｾﾂﾁ
ｿｳｻﾝ	ｿｳｻﾝｼｮｳﾊ ｼ*ﾖｳﾎｳｻﾝｷ*ﾖｳ ｲｸｾｲｷﾎﾝｼｻｸｵ ﾏﾄﾒﾀ
ｿｳｻﾝ	ｿｳｻﾝｼｮｳﾊ ｼ*ﾖｳﾎｳｻﾝｷ*ﾖｳ ｲｸｾｲｷﾎﾝｼｻｸｵ ﾏﾄﾒﾀ
ｿｳｻﾝ	ｿｳｻﾝﾊ ｼ*ﾖｳﾎｳｼｮﾘｷ*ｼ*ｭﾂｼｬ ﾆﾝﾃｲｼｹﾝ ｷｿｸｵｾｲﾃｲｼﾀ
ｿｳｻﾝ	ｿｳｻﾝﾊ 45ﾈﾝﾄ*ｼ*ﾖｳﾎｳｼｮﾘｼ*ﾂﾀｲﾁﾖｳｻｹﾂｶｵ ﾊﾂﾋ*ﾖｳ
ｿｳｻﾝ	ｿｳｻﾝﾊ ｻﾝｺｳｼﾝﾃ* ｼﾝｸﾀﾝｸ ｻﾝｷ*ﾖｳｾｲｻｸｶｶﾞ*ｸｾﾝﾀｰｺｳｿｳｵ ﾃｲｼ*
ｿｳｻﾝ	ｿｳｻﾝﾊ ｾｲｷ*ｮﾖｳ ﾐﾆｺﾝ､ﾒｰｶﾉ ｷｮｳﾄ*ｳｾｲｻﾝｶ*ｲｼｬﾉ ｾﾂﾘﾂｵ ｹﾝﾄｳ
ｿｳｻﾝ	ｿｳｻﾝﾊ ﾃ*ﾝｼｻﾝｷ*ﾖｳﾌ*ｵ ｾﾂﾁｽﾙ ｺｳｿｳｵ ｹﾝﾄｳｼﾀｲﾙ
ｿｳｻﾝ	ｿｳｻﾝﾊ ｼ*ﾖﾎｳｻﾝｷ*ﾖｳｲｸｾｲﾉﾀﾒ ｼ*ﾖｳﾎｳｼｮﾘｼｼﾝｺｳｼ*ｷ*ﾖｳﾀ*ﾝｵ ｾﾂﾘﾂ
ｿｳｻﾝ	ｿｳｻﾝﾊ ｼ*ﾖｳﾎｳｼｮﾘｼ*ﾂﾀｲﾁﾖｳｻﾉ ｹﾂｶｵ ﾏﾄﾒﾀ
ｿｳｻﾝ	ｿｳｻﾝﾊ ｷｶｲｺｳｷ*ﾖｳｼﾝｺｳｷﾎﾝｼｹｲｶｸｵ ﾍﾝｺｳｼ 45ﾈﾝﾄ*ｼ*ﾂｼｹｲｶｸｵ ｷﾒﾀ
ｿｳｻﾝ	ｿｳｻﾝﾊ ｼ*ﾖｳﾎｳｶｽｲｼﾝﾄ ｼｽﾃﾑｶｲﾊﾂ7､6ｵｸﾅﾄ*45ﾈﾝﾄ*ﾖｻﾝｵﾏﾄﾒﾀ
ｿｳｻﾝ	ｿｳｻﾝﾊ ｹｲｽｳｶ*ﾀﾃ*ﾝｻﾝｷﾖｳｼｭｳﾍﾝｿｳﾁﾉ ｼｭｳﾁｺｳｾｲｻﾝｵ ｼｼ*

(4) ファイリングの方法

〔データ・ベースの定義〕 データ・ベースの定義は，汎用データ・マネジメント・システム：ＲＡＰＩＤのＲＰＤＤＥＳＰに基づいて行なう。

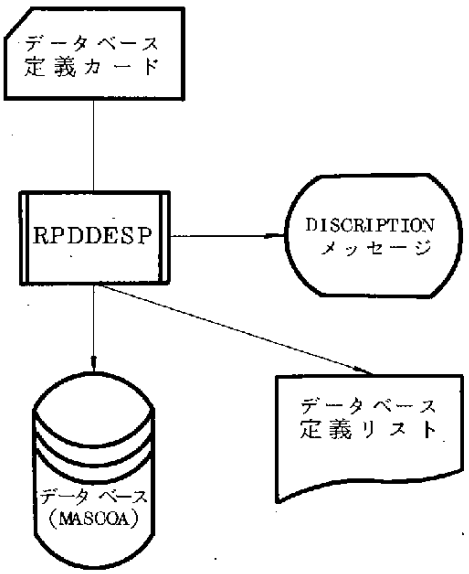


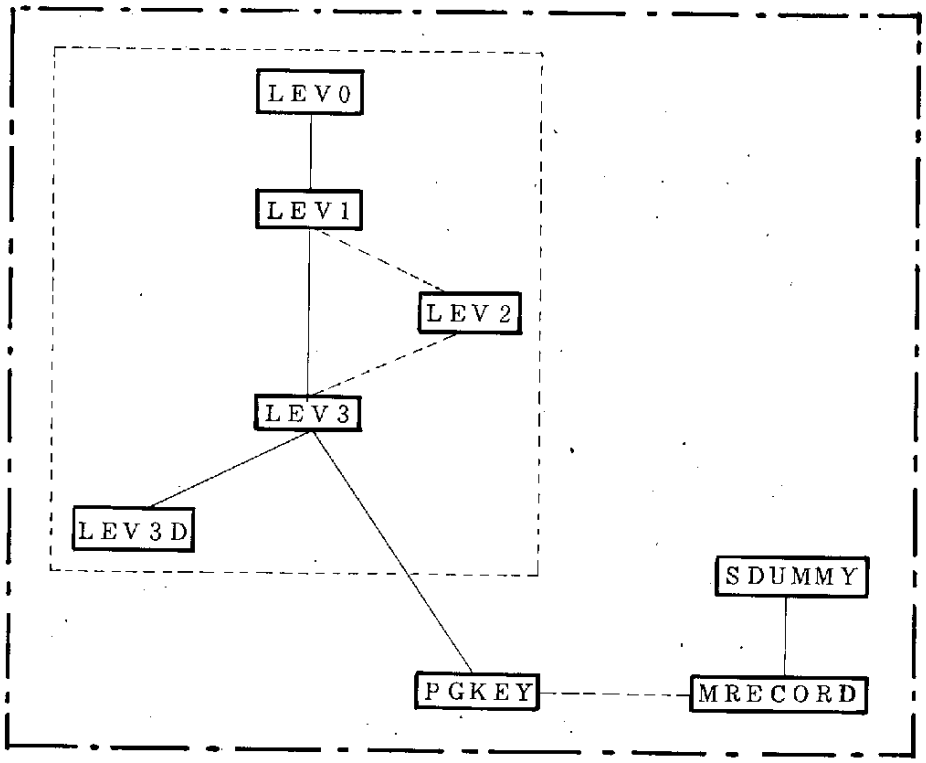
図9.2.4　ＭＡＳＣＯＡにおけるデータ・ベースの定義

〔データ・ベースの構造〕 ＭＡＳＣＯＡにおけるデータ・ベースの特徴は，４つの階層をもったキーワードをインデックス・テーブルに登録すること，キーワードとニュース記事内容を１対多の関係で連結していることなどである。

図9.2.5にみるように，キーワードのインデックス・テーブルでは，総称（ＬＥＶ０），大分類（ＬＥＶ１），小分類（ＬＥＶ３），小分類の同意語（ＬＥＶ３Ｄ），マスタ・レコードのアドレス情報を格納するレコード（ＰＧＫＥＹ）の各レコード・タイプを階層的にリングで結合している。レベル０からレベル２までの登録はアドレス指定により行ない，レベル３およびレベル３′はランダマイズ・ルーチンを通して登録している。

また，データ部分では，ニュース記事を発行年月日順に分類して格納するためのダミーレコード（ＳＤＵＭＭＹ），ニュース記事の内容を格納するためのマスタ・レコード（ＭＲＥＣＯＲＤ）をリングで結合している。

〔データの格納〕 ①キーワードの登録：レベル０からレベル３までのシソーラス・コード７桁とレベル３′の同意語判別コード１桁を分類し，コード順に登録して木構造のインデックス・テーブルを作成する。シソーラス・コードを使って登録することは，下位レベルのレコードのポインタを連続的に結合する場合に，上位レベルのマスタ・ポインタを利用できるという点で効果的である。登録されたキーワードは，シソーラス・テーブルのリストとしてプリンタに出力する。



レコード・タイプ名

ＬＥＶ０　：レベル０のキーワード（総称　）
ＬＥＶ１　：レベル１のキーワード（大分類）
ＬＥＶ２　：レベル２のキーワード（中分類）
ＬＥＶ３　：レベル３のキーワード（小分類）
ＬＥＶ３Ｄ：レベル３のキーワードの同意語
ＭＲＥＣＯＲＤ：マスタ・レコード（ニュース記事の内容）
ＳＤＵＭＭＹ：ＭＲＥＣＯＲＤを年度別に並べるためのダミーレコード
ＰＧＫＥＹ：ＭＲＥＣＯＲＤの格納アドレス情報

図9.2.5　　ＭＡＳＣＯＡにおけるデータ・ベースの構造

②ニュース記事内容の格納：データ格納の前処理として，ニュース記事のキーワードがインデックス・テーブルに登録されているかどうかをチェックしなければならない。ニュース記事にあるキーワードがインデックス・テーブルに登録されていない場合，そしてニュース記事に一つもキーワードがついていない場合は，その状況をエラーリストに出力し修正処理を施す。テーブル・サーチは，ランダム・アクセスで行なう。その順序は，まずレベル３を最初に検索し，引き続いてレベル３′，レベル１と検索してゆく。

データの格納は，発行年月日順に行なうため，最初に年度判定を行ないレコード・タイプＳＤＵＭＭＹから年度ポインタ（リファレンス・コード）をセットする。次に，マスタ・レコード

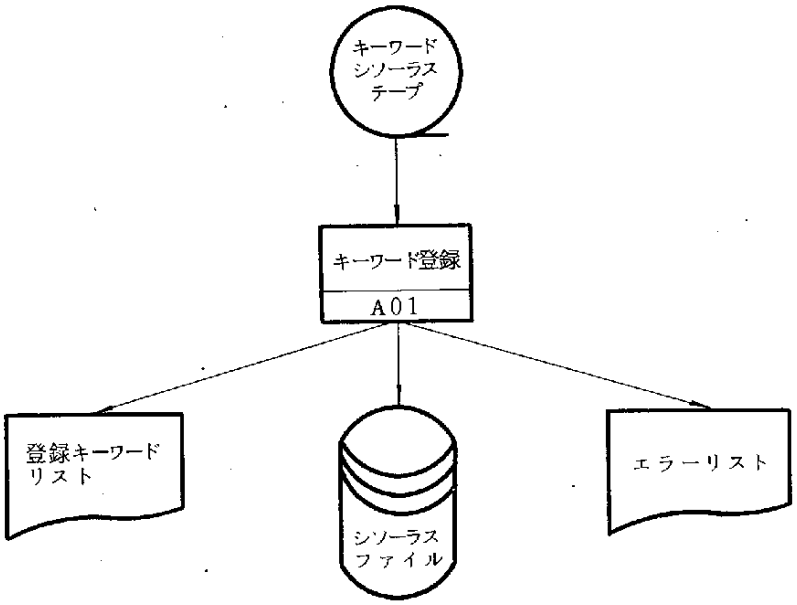


図9.2.6　用語インデックスの登録

(MRECORD)を格納エリアにセットしてニュース記事の内容を格納する。さらに，インデックス・テーブルとチェックして得られたレベル3のキーワードのポインタ(PGKEY)へ，マスタ・レコードのアドレス情報(リファレンス・コード)をセットしてリング結合を行なう。格納したデータは，リングしたキーワードとともにラインプリンタで出力する。

次にMASCOAにおけるデータの格納プログラムの一部を掲載する。

```
0123    008010/RAPID SECTION.
0124          /RAPID PAGE BUFFER 5.
0125    008030/RT  LEV0.
0126    008040/01  LZERO          PIC X(4).
0127    008050/RT  LEV1.
0128    008060/01  R1             PIC X(32).
0129    008070/RT  LEV3.
0130    008080/01  R3             PIC X(32).
0131    008090/RT  LEV3D.
0132    008100/01  R3D            PIC X(32).
0133    008110/RT  PGKEY.
0134    008120/01  KEYCD.
0135    008130/    02 RKEY        PIC 9(10).
0136    008140/    02 NENDO       PIC X(6).
0137    008150/RT  SDUMMY.
0138    008160/01  R-NEN          PIC X(4).
0139    008170/RT  MRECORD.
0140    008180/01  MREC.
0141    008190/    02 MR-N        PIC X(6).
0142    008200/    02 MR-S        PIC XX.
0143    008210/    02 MR-C        PIC X(7).
0144    008220/    02 FLD3        PIC XXX.
0145    008230/    02 MR-NEWS     PIC X(54).
0146    008240/    02 IR-COUNT    PIC 9(4).
0147    009010 PROCEDURE DIVISION.
0148    009020 OPEN-RT.
0149    009030     OPEN INPUT INPUT-F.
0150    009040     OPEN OUTPUT PRINT-F1.
0151    009041     OPEN OUTPUT PRINT-F2.
0152    009050/    RPD-OPEN DATABASE UPDATE.
0153    009060     MOVE SPACES TO PRT-FORM LST1 LST2 N12.
0154    009061     MOVE 0 TO TP-DAA.
0155    009062     MOVE SPACE TO R-NEN FLD3.
0156    009063     MOVE ZERO TO IR-COUNT.
0157    009070     WRITE PRT1 BEFORE ADVANCING C17.
0158    009080     WRITE PRT2 BEFORE ADVANCING C17.
0159    009090 ST1.
0160    009100     READ INPUT-F AT END GO TO END-RT.
0161    009110     IF NEN = R-NEN2 GO TO PS1.
0162    009120     IF NEN = 68 OR 69 OR 70 OR 71 OR 72 OR 73 GO TO GT1.
0163    009130     MOVE SPACES TO E-FM1.
0164    009131     MOVE '** NEN ERROR ** ' TO FLR3.
0165    009140     MOVE NGH TO E-NGH.
0166    009150     MOVE S-NO TO E-SNO.
0167    009160     MOVE CODE TO E-CODE.
0168    009170     MOVE KEYWD TO E-KEYWD.
```

```
CARDNO.  SEQNO. A  B
  0169   009180      MOVE E-FM1 TO LST2.
  0170   009190      WRITE PRT2 BEFORE ADVANCING C1.
  0171   009200      MOVE SPACES TO E-FM2.
  0172   009210      MOVE NEWS TO E-NEWS.
  0173   009220      MOVE E-FM2 TO LST2.
  0174   009230      WRITE PRT2 BEFORE ADVANCING C2.
  0175   009240      GO TO ST1.
  0176   010010 GT1.
  0177   010020      IF NEN = 68   MOVE A68 TO DAA.
  0178   010030      IF NEN = 69   MOVE A69 TO DAA.
  0179   010040      IF NEN = 70   MOVE A70 TO DAA.
  0180   010050      IF NEN = 71   MOVE A71 TO DAA.
  0181   010060      IF NEN = 72   MOVE A72 TO DAA.
  0182   010070      IF NEN = 73   MOVE A73 TO DAA.
  0183   010080 GT.
  0184   010090/     GET DIRECT.
  0185   010100      IF SCB2 = '0012'   GO TO DIS-DAA.
  0186   010110      IF SCB2 = '0019'   GO TO DIS-DAA.
  0187   010120/     WHEN ERROR GO TO ER1.
  0188   010121      MOVE R-NEN TO N12.
  0189   010130      IF NEN NOT = R-NEN2 GO TO D1.
  0190   010140 PS1.
  0191   010150      MOVE NGH  TO MR-N.
  0192   010160      MOVE S-NO TO MR-S.
  0193   010170      MOVE CODE TO MR-C.
  0194   010180      MOVE NEWS TO MR-NEWS.
  0195   010190/     PUT MRECORD  FROM  000011 TO  003985.
  0196   010200      IF SCB2 = '0175' GO TO NS1.
  0197   010210/     WHEN ERROR GO TO ER2.
  0198   010220      ADD 1 TO PUT-COUNT.
  0199   010230      MOVE ZERO TO SP-WK.
  0200   010240      MOVE 1    TO CN.
  0201   010250      MOVE ZERO TO SW.
  0202   011010      MOVE DAA TO MREC-DAA.
  0203   011020 NR. IF KEYS (CN) = SPACE GO TO SP-COUNT.
  0204   011030      MOVE KEYS (CN) TO R3.
```

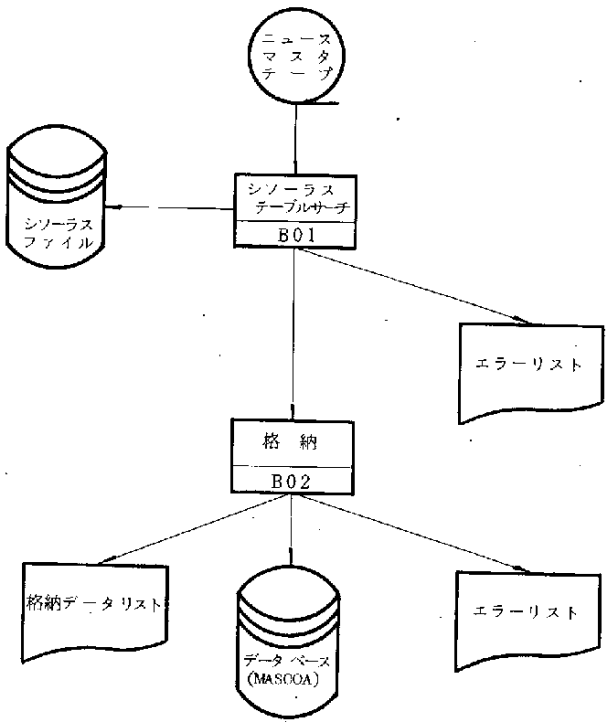


図9.2.7　データの格納

(5) 検索の方法

検索は，コンピュータ・ニュース記事の中で用いられる語または句からなる自然語のキーワードを使って，端末タイプライタから行なう。

検索結果は，その件数の多少に応じてラインプリンタまたは,端末タイプライタに出力する。また，見易い結果を得るために電算写植編集システムを利用した漢字出力も可能である。検索方法は，簡単な専用言語を用いキーワードと論理記号（ＡＮＤ，ＯＲ）とを組合せた会話形式で，必要な情報を探してゆくやり方である。たとえば，「日立および東芝における４５年度から４６年度の間のディスプレィ開発状況」を知りたい場合には，次のような検索指示と表示を行なう。

この例では，まず，ステップ１で検索情報の期間指定を行なっている。指定された期間内の情報は７８１件あるとの応答がある。

ステップ２では，ステップ１の範囲でディスプレィという情報が何件あるかを問合せている。応答は３２件である。

ステップ３では，ステップ２の条件のもとで開発が何件あるかを問合せている。応答は１４件である。

ステップ４，５では，それぞれステップ３の条件のもとで，日立，東芝の情報が何件あるかを問合合せている。応答は，それぞれ８件，５件である。

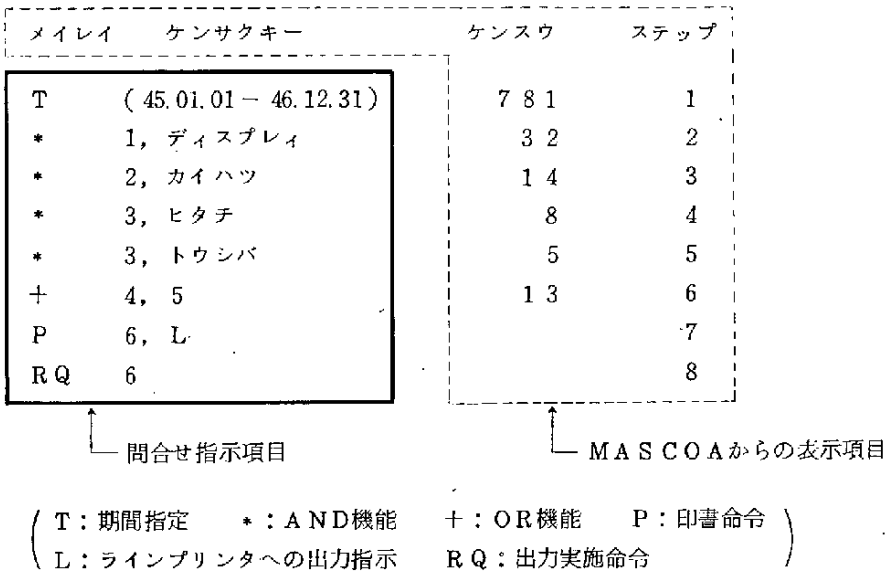


図 9.2.8　MASCOAにおける検索指示と表示例

ステップ6では，ステップ4とステップ5の情報を一つにまとめよという命令である。日立と東芝を合せた情報は，13件であるという応答がある。

ステップ7では，ステップ6の結果をラインプリンタに出力できるように準備せよという命令である。

ステップ8では，検索結果の出力を実行しなさいという命令である。

検索結果の出力様式は，図9.2.9に示す。

次に，MASCOAにおける検索プロセスを順を追って説明してゆく。図9.2.10にみるように，検索要求－1では，発行年月日の検索範囲を指定することにより，レコード・タイプSDUMMYの年度インデックスにおいて指定された期間のアドレス情報をリング・サーチして，該当するレコードの件数をカウントするとともに，その結果をタイプライタに表示する。

検索要求－2では，キーワードをタイプインすることにより，そのキーワードがインデックス・テーブルに登録されているかどうかをチェックする。キーワード・チェックの手順は，レベル3，レベル3′，レベル1の順序で行なう。キーワードが登録されていれば，命令キーに従ってキーコード(アドレス情報)検索を行ない，登録されていない場合は，エラーメッセージをプリント・アウトしエラーを修正して，改めて検索要求をやり直す。

インデックス・テーブルからのキーワードが検索されると，既に検索したデータがない場合はそのキーコードを一時バッファに記録し，ある場合はそのデータとAND(*)，OR(＋)の機能を使って論理統合を行ない，その結果を別のバッファへ記録する。

バッファは，第1バッファ，第2バッファ，第3バッファの3つを確保しておき，検索，統合処理のたびにバッファを移動する。古いバッファは，順次消去され，新しい結果と置換えられてゆく。

① 端末タイプライタによる出力様式例

ケンサクジョウケン	：(45. 01)*(IC)*(カイハツ)
キジナイヨウ	：スタンフォードケンキュウジョハ　NASAノエンジョデ　シンセダイノICトモイワレル　IVCヲ　カイハツシタ
シュッショ	：デンパシンブン
ハッコウネンガッピ	：45. 01. 20

② 電算写植編集システムJEM3850による出所様式例

☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆

検索条件　：	(43+44)＊(クリアリングセンター)
記事内容　：	科学技術庁は政府研究機関のクリアリングセンター構想を明らかにした。
発行年月日：	43年07月16日
出　　典　：	日本工業新聞

☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆

③ ラインプリンタによる出力様式例

ケンサクジョウケン：(44. 08)*(カイハツ)　　　テイキョウケンスウ：18

コウゾウケイカクケンキュウジョハ　ILCシャカラ　ソフトウエアカイハツヲ　ジュタクシタ(ニッカンコウギョウ, 44. 08. 02)

フジツウハFACOM－Rマタハ230－25ヲナイゾウスル　スウチセイギョソウチFANUC 250ヲ　カイハツシタ(デンパシンブル, 44. 0.8. 05)

フジツウハ　ガイブキオクソウチFACOM472Kディスクパックソウチヲ　カイハツシタ(デンパシンブン, 44. 08. 13)

トウシバキカイハ　トウシバトキョウリョクシテ　コンピュータ・ニュメリカル・コントロール・システムヲ　カイハツスル(ニホンコウギョウシンブン, 44. 08. 20)

図9.2.9　　MASCOAにおける出力様式

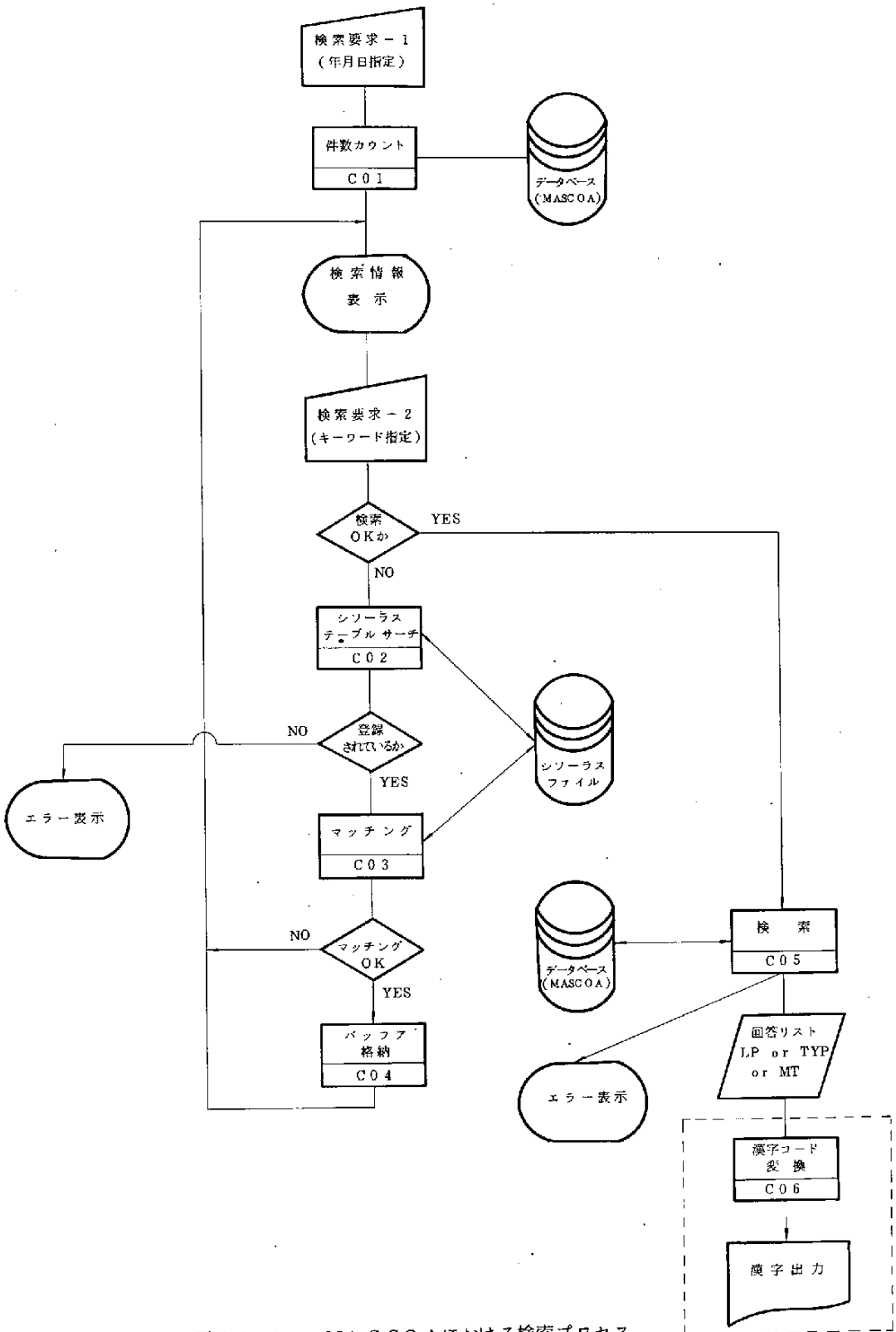


図 9.2.10　　MASCOAにおける検索プロセス

検索結果は，最終検索前に行なわれた３つの処理に対して，その内容が各々バッファに記録されているので，それぞれの指定を行なえばいずれのバッファの出力も可能である。

検索が終ってＰＲＩＮＴ命令がタイプされると，バッファ内にあるキーコード（アドレス情報）をサーチして，ダイレクトにデータベースから関係情報をとり出すことができる。バッファから取り出された情報は，一度プリント・エリアにセットしてから，ＰＲＩＮＴ実行命令とともにラインプリンタまたはタイプライタ，場合によっては磁気テープに出力することになる。磁気テープ出力は，カナ漢字コード・コンバートを施し，漢字混り文として別に報告する場合に使うためのものである。

次に，ＭＡＳＣＯＡにおける検索プログラムの一部を掲載する。

```
0237    011010/RAPID SECTION.
0238    011020/RAPID PAGE BUFFER 10.
0239    011030/RT  LEV0.
0240    011040/01  R0        PIC X(4).
0241    011050/RT  LEV1.
0242    011060/01  R1        PIC X(32).
0243    011070/RT  LEV2.
0244    011080/01  R2        PIC X(32).
0245    011090/RT  LEV3.
0246    011100/01  R3        PIC X(32).
0247    011110/RT  LEV3D.
0248    011120/01  R3D       PIC X(32).
0249    011130/RT  PGKEY.
0250    011140/01  ADCODE.
0251    011150/    02 RFCD  PIC 9(10).
0252    011160/    02 NCD   PIC 9(6).
0253    011210/RT  SDUMMY.
0254    011220/01  SKY       PIC 9(4).
0255    011230/RT  MRECORD.
0256    011240/01  MREC.
0257    012010/    02 R-NGH PIC 9(6).
0258    012080/    02 BANGO PIC XX.
0259    012090/    02 CD1   PIC XX.
0260    012100/    02 CD2   PIC XX.
0261    012110/    02 PG    PIC XXX.
0262    012111/    02 FILLER PIC XXX.
0263    012120/    02 NEWS  PIC X(54).
0264    012130/    02 IR-CT PIC 9(4).
0265    020010 PROCEDURE DIVISION.
0266    020020 OPEN-RT.
0267    020030/    RPD-OPEN DATABASE.
0268    020040     OPEN OUTPUT PRINT-F.
0269    020050     MOVE SPACES TO LSTFM CNS-OUT CONS1 OUT-A CONS3.
0270    020060     MOVE SPACES TO WK-C W-OUT.
0271    020070     MOVE ZERO TO B1 B2 B3 WKN C1 D-CNT IN-DAA U1 U2 U3 PCN.
0272    020080     MOVE ZERO TO OUT-COUNT.
0273    020090 FS. MOVE 1 TO K1.
0274    020100 F1. MOVE ZERO TO BF1 (K1) BF2 (K1) BF3 (K1).
0275    020110     ADD 1 TO K1.
0276    020120     IF K1 NOT > 1000 GO TO F1.
0277    020140     DISPLAY DP-1.
0278    020150     DISPLAY DP-0.
0279    020160 ACT.
0280    020170     ACCEPT CONS1.
```

```
FACOM 230-60      RAPID COBOL SOURCE LIST      -71.07.01-    (V-02/L-01)

CARDNO.   SEQNO. A   B
  0281    020180      IF INST = 'END '   GO TO TP-E.
  0282    020190      IF INST = 'NEXT'   GO TO TP-N.
  0283    020200      IF INST = 'IR-P'   GO TO TP-P.
  0284    020210      IF INST = 'IR-S'   GO TO TP-S.
  0285    020220      IF FLD12 IS NOT NUMERIC GO TO TYPE1.
  0286    020230      IF FLD12 > TP-NO        GO TO TYPE1.
  0287    020240      COMPUTE WKN = TP-NO - 2.
  0288    021010      IF WKN NOT < FLD1 GO TO TYPE2.
  0289    021020      IF FLD2 NOT = '*' AND NOT = '+' GO TO TYPE3.
  0290    021030 WD-GT.
  0291    021040      MOVE K-WD TO R3.
  0292    021050/     GET RANDOM LEV3.
  0293    021060      IF SCB2 = '0121' GO TO NG2.
  0294    021070/     WHEN ERROR GO TO NG1.
  0295    021080 UE.  IF K-WD = R3 GO TO OT2.
```

(6) メンテナンスの方法

ＭＡＳＣＯＡでは，めざましく発展を続けるコンピュータの分野における日々のニュース記事情報を取扱っていることからＭＡＳＣＯＴのどのサブシステムよりメンテナンスの重要性は大きく，頻繁にそれを行なわなければならないと思われる。ところが，これらの処理は実務上，データ管理者の立場からみると，そのサイクルを小刻にすることは非常に面倒であり，メンテナンス・データが発生するたびに更新していたのでは切がないので，週あるいは月単位の比較的長いサイクルで考えたいし，それらを一度に処理してしまいたいという希望がある。汎用データ・マネジメント・システムを使用した場合は，とくにこのメンテナンスに大きな時間を要するので，データ構造をやたらに複雑にしたり，データ量を増大にすることは，実務上避るべきであろう。

このような状況は，利用者側からみれば，常に最新のデータが欲しいのでアップ・デートを頻繁にしておいてもらいたいということであるし，データを管理する側からみればメンテナンスをできるだけ減らし，できることなら一括処理したいということになるのであろう。ＭＡＳＣＯＡでは，以上のようなことを十分考慮して，できるだけ情報提供サービスのニーズに応じられるよう，メンテナンス・データがある程度まとまった時点で一括して処理することにした。また，データの利用効率を上げるため個々のデータにつき利用統計をとり，古くなったデータと利用頻度の低いものについては管理限界を決めて，１次ファイル（磁気ディスク），から２次ファイル（磁気テープ）へ管理レベルを下げて保管することにした。ＭＡＳＣＯＡのメンテナンスは，これら１次情報，２次情報のニュース記事の内容と各階層のキーワードについて，追加，修正，消去の処理が自由に行なえるようにした。以下は，一連のメンテナンスのうちとくに１次情報に限って，そのメンテナンスの方法を述べて行くことにする。

〔キーワード・メンテナンス〕　インデックス・テーブルのキーワードの追加・修正・消去の処理を行なう。メンテナンス・データの入力は，カードまたは磁気テープから行なう。

キーワードの追加処理は，シソーラスのレベル番号の指定とマスタ・キーワード，ディテール・キーワード（レベル３’の場合はマスタ・キーワードのみ）を指示してリング結合を行ない登録す

る。処理は，データの格納の時と同様ＰＵＴコマンドによって実行する。

修正処理は，変更すべきキーワードが下位レベルにディテール・キーワード（レベル３の場合はマスタ・キー情報）とリングしているときは，いったん下位レベルのレコードのアドレス情報（リファレンス・コード）をバッファに導き，同時にリングから切離して処理する。修正が終ったならば，再度リング結合を行なって処理を完了させる。レベル３′の場合は，下位レベルがないのでダイレクトに修正できる。キーワードの修正は，シソーラスの変更につながることが多く，階層をいくつか同時に修正しなければならないことが多い。

簡単な例で修正処理を説明するために，いま１個のマスタ・キーの修正，つまりディテール・キーの代表語を変更する場合について考えてみると，このマスタ・キーのレコード・タイプがもつＲＴＣＳエリアにリファレンス・コードをセットし，修正データをワーク・エリアにセットしてＭＯＤＩＦＹコマンドを実行すれば，古いデータは新しいデータに置換えられる。たとえば，「ツウサン」というマスタ・キーを「ツウショウ・サンギョンショウ」と置換えたい場合には，まず，「ツウサン」の格納されているリファレンス・コードをＧＥＴ系コマンドでＲＴＣＳエリアにセットする。次に，修正用エリアに「ツウサンショウ・サンギョウショウ」をセットしてＭＯＤＩＦＹを実行すれば，修正は完了することになる。

レベル３，レベル３′のキーワードの修正は，これらがランダマイズ・フィールドであるため，実際の処理では修正すべきデータを消去処理で削除した後に新しいキーワードを追加するという形で登録し，リング結合する。

消去処理では，下位レベルにリングしているディテール・キーワードをどう処理するかという情報を，メンテナンス・データを入力する際に指示しておかなければならない。処理は，削除すべきキーワードのリファレンス・コードをＲＴＣＳエリアにセットして消去コマンドで実行する。

〔データ・ベース・メンテナンス〕　追加処理は，データの格納時の処理と同様に行なう。

修正処理は，記事内容の修正と抽出したキーワードの修正とに分けられる。記事内容を格納しているマスタ・レコード（ＭＲＥＣＯＲＤ）の修正は，格納されているアドレス情報をサーチして求め，新しいデータをセットして修正コマンドＭＯＤＩＦＹを実行すれば，元のアドレスに新しいデータが記録される。抽出キーワードの変更は，格納時にリングしたマスタ・キー情報を旧キーワードから切離したあと，新キーワードにリングすることで更新処理が行なわれる（ＳＥＴ　ＦＲＥＥコマンド，ＣＯＮＮＥＣＴコマンドを使用）

消去処理は，記事内容を格納しているマスタ・レコードの削除をまず行ない，ここで得られるアドレス情報と，入力データのキーワード情報からインデックス・テーブルにリングされているマスタ・キー情報（マスタ・レコードのアドレス情報，キーコード情報）を削除して処理は完了する。

図 9.2.11にＭＡＳ.ＣＯＡにおけるファイル・メンテナンスのプロセスを示す。

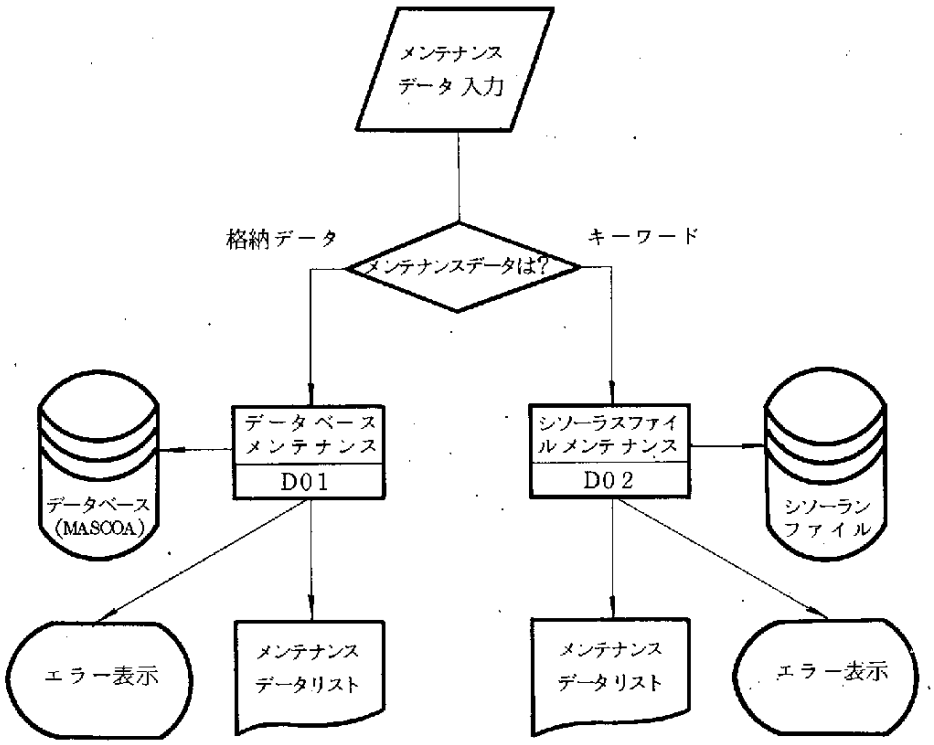


図 9.2.11　ＭＡＳＣＯＡにおけるファイル・メンテナンスのプロセス

9.2.2　専用データ・ベース・システム：ＭＡＳＣＯＡ

(1)　システムの概要

汎用データ・マネジメント・システム（ＧＤＢＭＳ）は，思想的に機能的に優れた特徴をもちながらも，ユーザ・サイドの業務がこれを活用できるまでに合理化され，体系化されていないため，ほとんど利用されていない状況にある。

また，多くの便利な機能を持っていることが実際の活用面ではかえって禍し，プログラミングにおいて特殊なコマンドを学ばねばならなかったり，管理・検索・レポート作成などの面で，いろいろな複雑な働きを盛沢山システム側に負わせるため，処理時間がべらぼうにかかってしまうという欠点がある。そこでＭＡＳＣＯＡでは，コンピュータニュース提供サービスという面から，このシステムに必要な最小機能を設定し，ＦＡＳＰ言語を用いてシステムを構築して記事情報提供サービスのための検索実験を行なうことにより，データの格納と検索の面から処理効率の向上を画ることにした。

ＭＡＳＣＯＡでは，キーワードを自然語でもち，これらのキーワードのシソーラス体系として階層構造を作っている。キーワードには，それぞれに対応した論理コードがつけられており，それらを管理するためにシソーラス・ファイルを設けた。また，データとキーワードを結びつけるた

め，インバーテッド・ファイルを作成した。インバーテッド・ファイルでは，論理コードとデータ（データ格納番地）が1対多の関係で連結されている。

論理コードは8桁からなり，頭の2桁がレベル1（大分類コード），次の2桁がレベル2（中分類コード），次の3桁がレベル3（小分類コード），最後の1桁がレベル3′でレベル3の関連データとなっている。

検索は，自然語でタイプライタから問合せることにより，シソーラス・テーブルで自然語のキーワードを論理コードに変換し，インバーテッド・ファイルから指定の論理コードを探しだせば，そこには求めるデータの格納番地が示されている。ＭＡＳＣＯＡでは，8つ以内のキーワードをANDまたはＯＲで結びつけた論理検索が可能である。条件に合ったデータは，格納番地をたどって引き出される。

このシステムでは，インデックス部にデータの所在を示すアドレス情報をもっている他，データ部においても，ページ情報部・レコード部などで個々のデータの所在の細目や，データ・リンケージのためのアドレス情報をもっており，変則的なインバーテッド・ファイルの構成となっている。

ファイル編成は，シソーラス・ファイル，インバーテッド・ファイルが共にシークエンシャル・ファイルである。データは，キーワード群をブロック単位で格納したパーテションド・ファイルである。検索は，キーワードを昇冪に整理したうえでダイレクト・アクセス処理を行なう。

このシステムの唯一の欠陥は，個々にデータのメンテナンスがしにくいことである。しかし，データ・ベースへの格納では通常のデータ処理におけるファイル作成と同じ位の時間ですむし，検索においては，ダイレクト検索の機能を失うことなく，即時的な処理が可能である。

なお，ＭＡＳＣＯＡの概要，データの形式，シソーラスの編成などについては，9.2.1節を参照されたい。また，ファイル構造については，４５－Ｓ００１「経営予測のためのデータ・マネジメント」の5.3節を参照されたい。

(2) シソーラス・ファイル

シソーラス・ファイルは，インバーテッド・ファイル作成および検索において，自然語のキーワードと論理コードの変換を行なうために必要である。ファイルは，ディスク・パックを使用し，順編成でシソーラス・ファイルを作成した。ディスク・パックに登録する前処理として，カードから磁気テープへキーワードを移し，それらのキーワードをＥＢＣＤＩＣコードのＡＳＥＮＤＩＮＧにソートする。このソートは，インバーテッド・ファイルを作成する際，自然語のキーワードから論理コードへ変換するために便利である。

また，検索を行なう際に，目的のキーワードをサーチする時間を短縮するために，シソーラス・ファイル作成時において相対ブロックに対して割当てられた情報を磁気テープに出力する。この磁気テープ記録をインデックス・テープ・ファイルと呼ぶ。

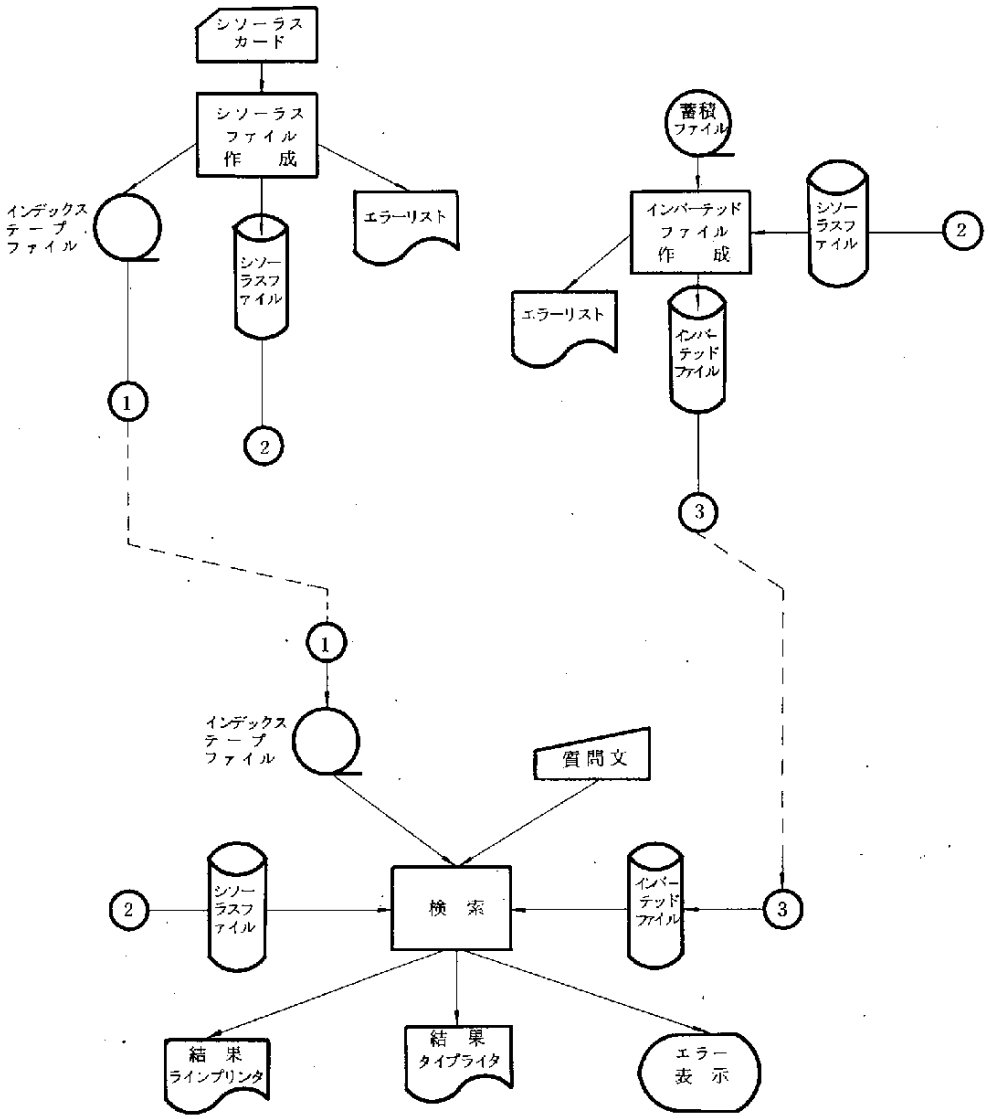


図9.2.12　専用システム：MASCOAのシステム・フローチャート

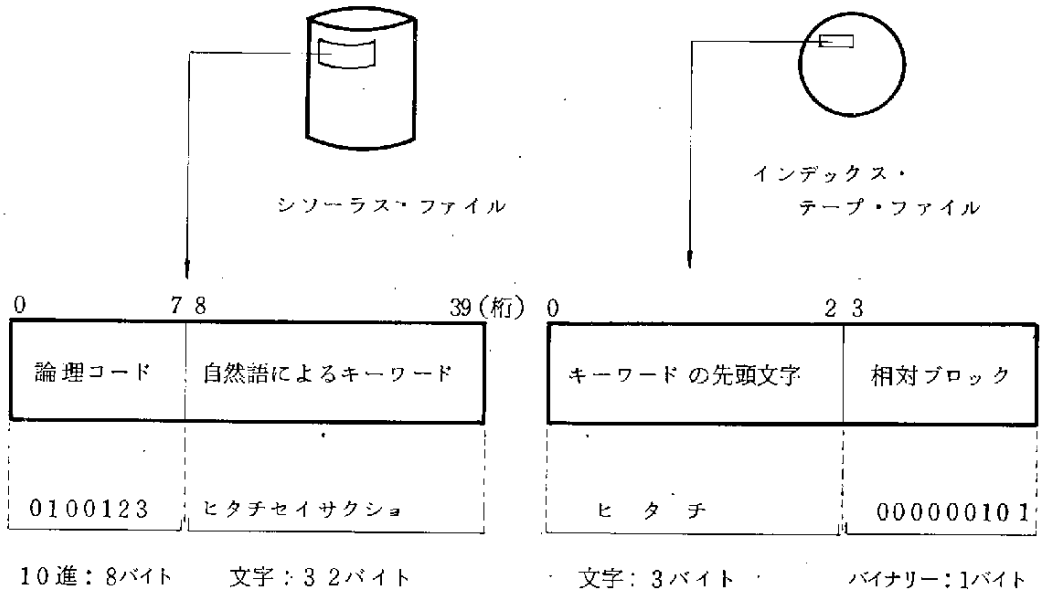


図9.2.13　シソーラス・ファイルとインデックス・テープ・ファイルの構成

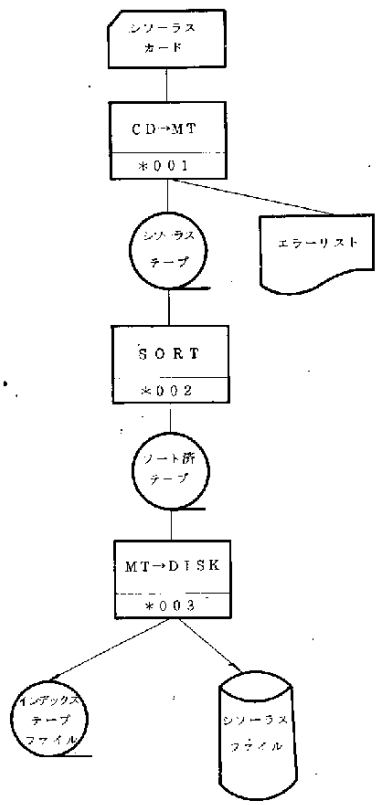
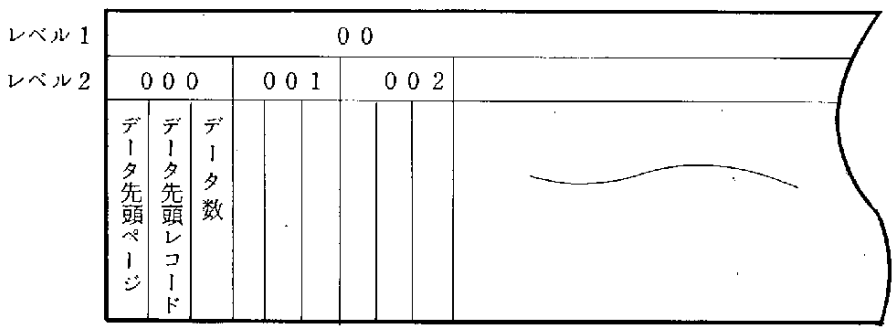


図9.2.14　シソーラス・ファイル作成プロセス

(3) インバーテッド・ファイル

インバーテッド・ファイルとは，インデックスのファィルのキーワードに対応して，いくつかの関連するデータが格納されているファイルが存在し，キーワードとデータがアドレス情報（直接アドレスまたは間接アドレス）で結ばれているファイル構造をいう。このシステムにおけるインデックス部は，マスタ・インデックスという形で存在し，1トラック：1ページとしたページ構造において，レベル1のコードの全てにこれらのページを割付ける。各ページには，そのコードに属するデータが格納されているページ情報（1ページ：1アドレスとした相対ページ・アドレス）を記録している。ここでいうレベル1とは，論理コードの上位2桁のことである。マスタ・インデックスには，データ部のページ情報を記録しているが，その他にページ内におけるデータの位置とそのレベルに関するデータ数も同時に記録している。

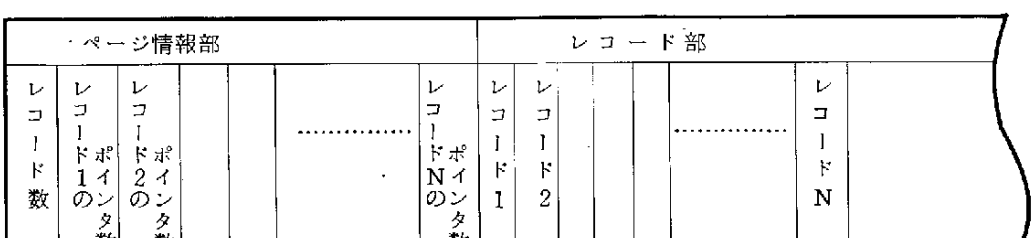
マスタ・インデックス部の構成



データ先頭アドレスとは，ページ内におけるレベル3のデータの最初の位置のことである。

データ部は，ページ情報部とレコード部から成り，ページ情報部の先頭にはレコード数が記録されている。このレコード数には，1ページ分のデータ・レコード数が記録されている。これは，データ部にページ情報部とレコード部が同一ページ内に存在するため，レコード部のスタート番地を計算するうえで必要なためである。ポインタの数を示すのは，1データに最大3個以内のポインタがつけられているためである。

データ部の構成



レコード部の構成

レコード部 （レコードN）													
	ページ・アドレス	レコードの位置	データの内容	ポインタ1						ポインタ2						
				レベル1コード	リングページ	リングレコード	レベル3コード	リングページ	リングレコード	レベル1コード	リングページ	リングレコード	レベル3コード	リングページ	リングレコード	.....

また，レコード部の各データの先頭にあるページとレコードは，検索終了後データを抜出すために使われる。

レベル1・コードおよびレベル3コードは，検索時において検索テーブルに格納し，次の条件がANDの場合に参照する。

リング・ページは，同じ属性のデータが次のページにある場合を考えて付加えてある。

リング・レコードは，同じ属性のデータがどこにあるかを示すもので，次のデータの位置を示している。このシステムでは，ファイルの構成図にも示す通り，データの内容が格納されている領域にもアドレス情報が存在するという変則的なインバーテッド・ファイルを構成している。図9.2.15にインバーテッド・ファイル作成プロセスを示す。

(4) 検索の方法

最初に検索時におけるインデックスの参照について述べると，キーワードがレベル1の場合は，マスタ・インデックスをレベル1の論理コードでサーチしてページ先頭アドレスおよびデータの先頭番地を知る。マスタ・インデックスよりページ・アドレスを得たならば，次にデータ部分を参照する。データの参照は，タイプインしたキーワードの論理コードが検索する論理コードの中で一番小さいとき，あるいは前回タイプインしたキーワードの論理コードが，現在タイプインしたキーワードの論理コードよりも小さいときにのみ行なう。

また，キーワードがレベル3の場合は，まず，論理コードのレベル1の部分でマスタ・インデックスをサーチする。次に，マスタ・インデックスよりページ・アドレスを得てデータ部を参照する。すなわち，データ部のフォーマットよりレコード数とポインタ数を知り，計算によりデータの位置を求めることになる。ポインタの抽出は，レベル3コードから行なう。キーワードがレベル3であるにもかかわらず，まずレベル1のコードでサーチするのは，レベル3の情報がレベル1の中に細目として存在するためである。

検索で指示するキーワードは，8個以下の複数個であるので，MASCOAではこれらの検索キーの個数だけテーブルをサーチしなければならない。サーチは，マスタ・インデックスにおいて論理コードの小さい順に行なわれる。これは，論理コードの小さい順にデータ部のポインタがつけられているためである。

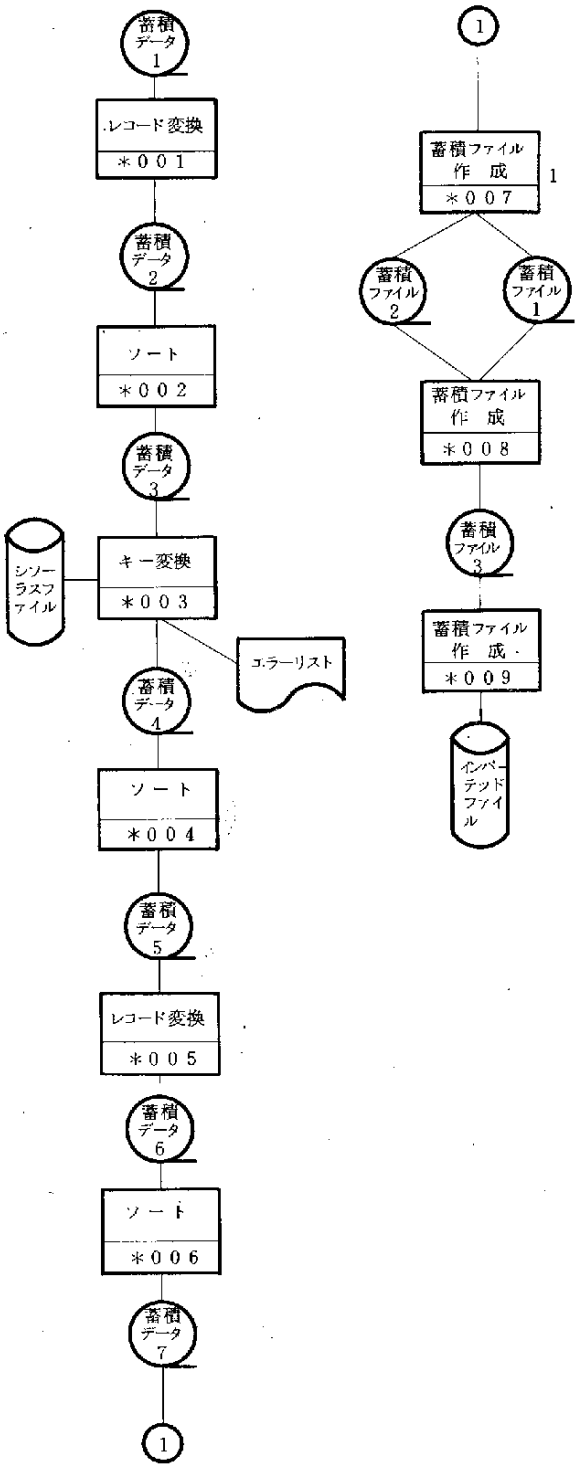


図9.2.15　インバーテッド・ファイル作成プロセス

次に，検索方法について述べると，最初にキーワードをタイプインする場合は，該当するページ・アドレスよりポインタを抽出する。キーワードがレベル1ならば，そのページ内のポインタは全て該当する。キーワードがレベル3ならば，まず，マスタ・インデックスのレコードのレベル1の部分からページ・アドレスを知り，データ部のレベル3のポインタのレコード数をサーチする。該当したポインタは，検索テーブルに格納し，チェック・ビットをONにする。チェック・ビットONとは，検索時に該当したデータのポインタを検索テーブルに格納するが，後に該当したことを知るためにコード化テーブルのポインタに印をつけておく。これがチェック・ビットONである。検索が全て終った段階で，チェック・ビットONになっているものを集めれば，それが検索する情報である。

2回以降のタイプインについては，レベル3のキーワード検索について述べる。

AND検索を行なう場合は，キーワードに対応する論理コード群の中からレベル1の一番小さいものを取出し，前回行なった検索の論理コードとの間で大小を比較する。

前回検索の論理コード ＜ 今回検索の論理コードの場合は，検索テーブルのレベル3の部分を今回タイプインしたキーワードでサーチし，合致しておれば前回のキーワードの中に今回のキーワードがあったことになり，ANDの条件が成立したことになる。このとき行なうサーチの回数は，タイプインのキーワードの数である。該当コードに対しては，チェック・ビットをONにする。

前回検索の論理コード ＞ 今回検索の論理コードの場合は，今回タイプインしたキーワードの論理コードにおけるレベル1のページ・アドレスを得て，そのページ内の全てのポインタを検索テーブルに一時記録し，前回のコード化テーブルより取り出して，検索テーブルの中から前回のコードと合致したものを探しだす。合致したものについては，コード化テーブルのチェック・ビットをONにする。

OR検索を行なう場合は，該当ページをコードし，検索テーブルに追加する。ポインタの抽出は，論理コードのレベル1のマスタインデックスよりページ・アドレスを得て，そのページ内のレベル3のデータとの突合せを行ない，一致していれば検索テーブルにポインタを追加する。このときチェック・ビットをONにする。

9.3 サブシステム：コンピュータ文献・検索システム（ＭＡＳＣＯＬ）

(1) ＭＡＳＣＯＬの概要

コンピュータ文献・検索システム（ＭＡＳＣＯＬ）は，国内または海外のコンピュータに関する雑誌や文献の記事をデータ・ベースに蓄積し，管理することにより，必要とする文献の名称，発行日，著者名，掲載頁，題名，分野を適時提供サービスする情報検索システムである。ＭＡＳＣＯＬで管理する雑誌・文献の範囲は，おおむね次の７項目に大別できる。

① ハードウエアと製造技術
② ソフトウエアと製作技術
③ 周辺科学と応用技術
④ システム開発
⑤ 情報処理部門の運営と管理
⑥ 政策および経営
⑦ 要員教育

ＭＡＳＣＯＬでは，これらの雑誌・文献を分野別分類コードで整理したり，その内容を最も良く表わす語または句を題名の中から１個または２個抽出してキーワードを決定している。入力データは，これら分類コードや用語のキーワードに雑誌・文献コード，発行年月，掲載頁，題名，著者名を組合せて１レコードとしている。データは，固定長の項目で構成している。システムは，汎用データ・マネジメント・システムＲＡＰＩＤを使用し，データ・ベースを作成して，集団磁気ディスクＦＡＣＯＭ４７１ＫでＭＡＳＣＯＬデータを管理している。ＭＡＳＣＯＬ用のファイルには，およそ１００,０００件のデータが収容できる。ファイル編成は，ＲＡＰＩＤがその機能として保有するランダム・アクセス・ファイルである。

キーワードのシソーラスは，現在のところもっていない。その理由は，ＭＡＳＣＯＧと同様，用語の体系化が難しいためである。これを補うものとして，ＭＡＳＣＯＬには分野別分類コードが用意されているが，この分野は重点分野という意味で，コンピュータに関連する際立った分野を３２とりあげ，雑誌・文献に最も合った分野のコードを分類コードとしてこの中から採用している。

検索のためのキーワードは，期日，分類コード，用語などを合せて６個以内であれば，いくつでも指定できる。キーワードを６個と制限したのは，ＭＡＳＣＯＬにおける検索では３～４個のキーワード指定が多いこと，ページ・バッファをＣＰＵメモリ上にできるだけ大きくとり，レスポンス・タイムを短縮することなどによるためである。検索機能としては，ＭＡＳＣＯＡと同様ＡＮＤ（＊）とＯＲ（＋）をもち，これとキーワードとを組合せて，自由に利用者が指定できる。たとえば，（４４＋４５＋４６）＊（システム開発）＊（ＭＩＳ）＊（機械工業）は，「機械工業におけるＭＩＳのシステム開発」について掲載したもので，４４年～４６年に発行された雑誌・文献から検索するという意味である。

データの管理，検索，および報告は，全て英数字，カナ文字を用いて行なう。端末機は，タイプライタ，ラインプリンタ，また緊急を要さない場合は，必要に応じて電算写植編集システムを利用

することができる。ＭＡＳＣＯＬは，ＭＡＳＣＯＴモニタの下で稼動する検索システムであるが，現在は他のサブシステムとの間の連繋をもっていない。コンピュータ用語のシソーラスができ，専用のオンライン・サービスのシステムができた時点では，当然これらのサブシステム間の有機性は必要になろう。

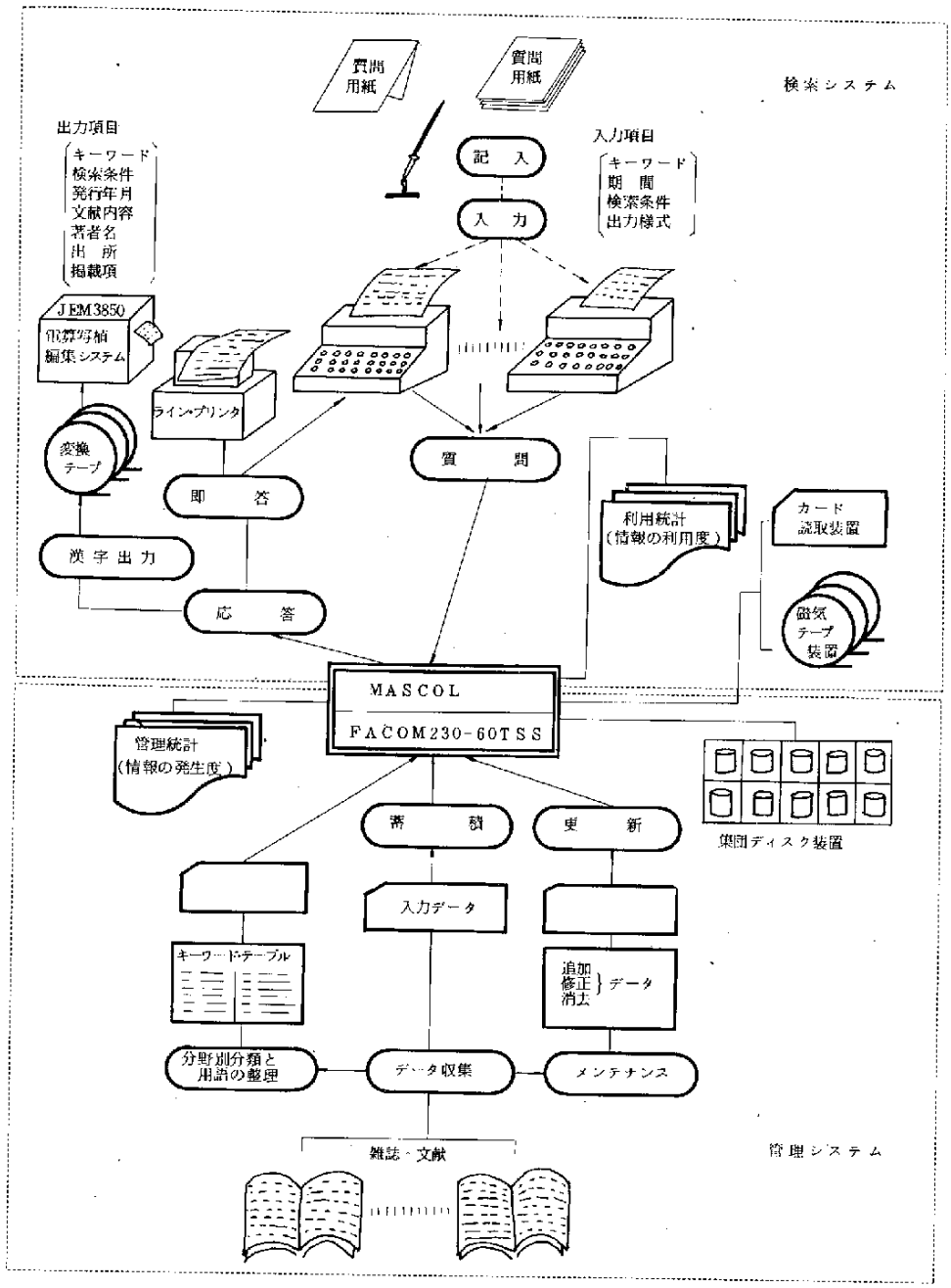


図9.3.1　コンピュータ文献・検索システム：ＭＡＳＣＯＬ

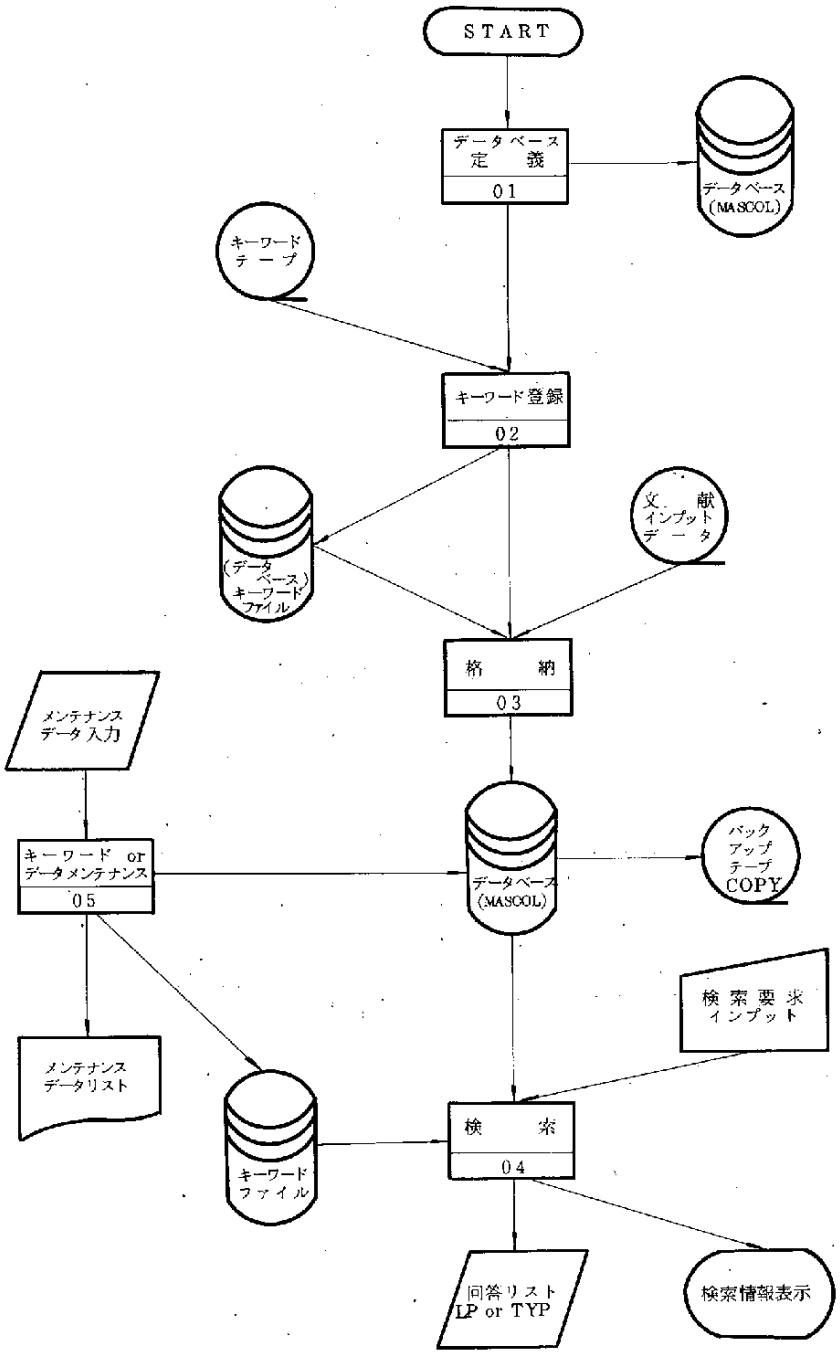


図9.3.2　　MASCOLにおけるシステム・フローチャート

(2) データの収集方法

現在，発行されている雑誌・文献から国内約１００種類，海外約４０種類をとりあげ，その中からコンピュータに関する情報を選定して入力データのための資料とした。国内資料は，事務管理，事務と経営，データ通信，情報科学，組織科学，数理科学，コンピュータ・レポート，コンピュートピア，日経エレクトロニクスなどの雑誌がおもなものであり，コンピュータ重要語索引集（情報科学研究所）などから引用した。海外資料は，Automation，Behavioral Science，Business Automation，Communications of The ACM，Datamation，Management science，Software Ageなどの雑誌を中心に海外雑誌・文献情報（日本能率協会）から引用した。

これらの資料を基に，まず文献の名称，掲載頁，発行年月，題名（タイトル）を資料表に整理した。データの転記には，コンピュータに入力可能な英数字，カナ文字を用いた。次に，転記された資料表の個別のデータにつき，その表題を最も良く表わす分野を分野別分類コード表から選び出し，その番号を資料表に記入した。さらに，その表題の中からキーワードとなるべき重要語を，語または句の形で抽出し，資料表に記入した。用語については，インデックス・テーブルに登録されているかどうかをチェックし，同音異語，異音同語，関連語などを整理してテーブルを修正した。整理の済んだ入力データは，用語データ，文献データ，継続データの３種類のカードに納めた。継続データは，題名が６６文字以内に納まらない場合に使用するもので，継続番号を記入して２枚まで追加することができる。これらデータのカード・フォーマットは，次の通りである。

用語データ

文献コード	掲載頁	用語 (1)	用語 (2)	
3桁	4桁	３２桁	３２桁	

文献データ

文献コード	掲載頁	継続番号	発行年月	分類コード	題名（タイトル）
3桁	4桁	1桁	4桁	2桁	６６桁

継続データ

文献コード	掲載頁	継続番号	題名の続き	
3桁	4桁	1桁	６６桁	

(3) データ・リンケージの方法

データ・リンケージは，ＭＡＳＣＯＧ同様考えていない。データ・リンケージを造りあげる方法として，一つは語源から用語の構成を体系化してシソーラスを造るやり方であり，もう一つは活用面から分類整理し体系化してシソーラスを造るやり方である。前者は，国語問題を研究している機関で早くから着手されていながら，ＭＡＳＣＯＴでこれらの技術を応用できる段階には至っていない。また，後者では，情報処理用語のＪＩＳはすでに造られたものの用語の数は少なく，コンピュータの分野で使われている用語のほとんどがまだ標準化されていない，などコンピュータ用語の体系化はむずかしい状態にある。このような状況のなかで，ＭＡＳＣＯＬにおいて，このシステムを利用してもらう人達に満足してもらえるコンピュータ用語のシソーラスを作りあげることは，不可能に近いものと判断したため，あえてこのために労働を費すことをしなかった。いまだ，コンピュータの分野は日進月歩その進展が他の分野に比べはるかに速く，新技術，新システムが続々と誕生している。そこで生れる新生語，新造語，改良語，俗語，関連語，消滅語は数限りなく散在し，これらを含めてコンピュータ用語のシソーラスを管理してゆくことは，とうてい困難と思われる。

そこでＭＡＳＣＯＬでは，データ・リンケージと同様，コンピュータ文献のうち比較的利用者の多いマクロ的な位置にある用語を整理し，重点用語として，それらに関する情報を提供サービスすることも重要であると思われるので，重点分野別分類コードを設定し，このコードを指定することにより，一般の検索用語よりも簡単に早く検索手続きを済ませ，報告を得ることができるような機

表9.3.1　ＭＡＳＣＯＬ用分野別分類コード

01	政　策	18	ロボット
02	経　営	19	制　御
03	情報化社会	20	オンライン
04	人間と組織	21	データ通信
05	創造性	22	コンピュータ
06	意思決定	23	ミニコン
07	未来問題	24	ファクシミリ
08	流　通	25	マイクロフィルム
09	自動化，省力化	26	図形処理
10	情報処理	27	情報検索
11	システム	28	シンクタンク
12	公　害	29	ＯＳ
13	標準化	30	プログラミング
14	教　育	31	プログラム言語
15	手法・方式	32	データの管理と帳表作成
16	周辺科学		
17	予　測	50	その他

能をシステムにもたせた。これらの重点語の主なものは，経営，情報化社会，未来問題，人間と組織，システム，情報処理，データ通信，コンピュータ，プログラミングなど３２語である。

ＭＡＳＣＯＬでは，さらに的確な情報を検索するために，特定の用語と組合せてキーワードを指定することにより，より利用目的に合致した雑誌や文献を探し出すことができる。なお，用語のインデックス・テーブルは数字がアセンディング順，英字がアルファベット順，カナ文字は５０音順に整理した。

(4) ファイリングの方法

〔データ・ベースの定義〕 データ・ベースの定義は，汎用データ・マネジメント・システム：ＲＡＰＩＤユーティリティＲＰＤＤＥＳＰの指定に基でいて行なう。

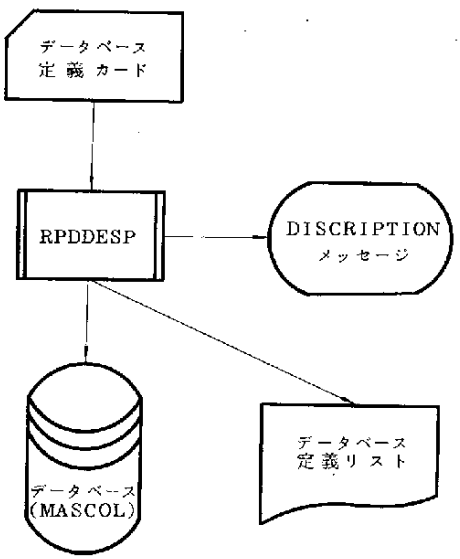
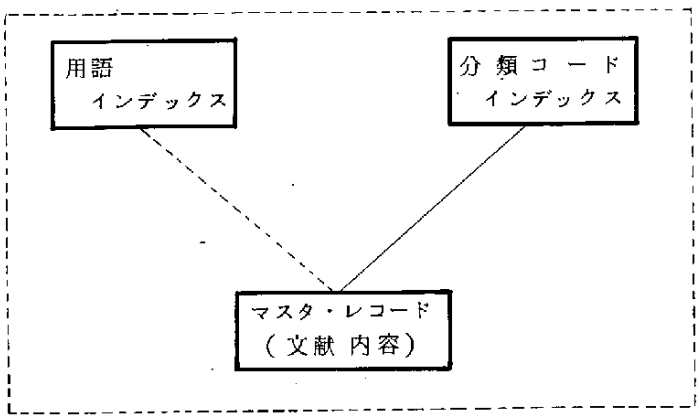


図9.3.3　ＭＡＳＣＯＬにおけるデータ・ベースの定義

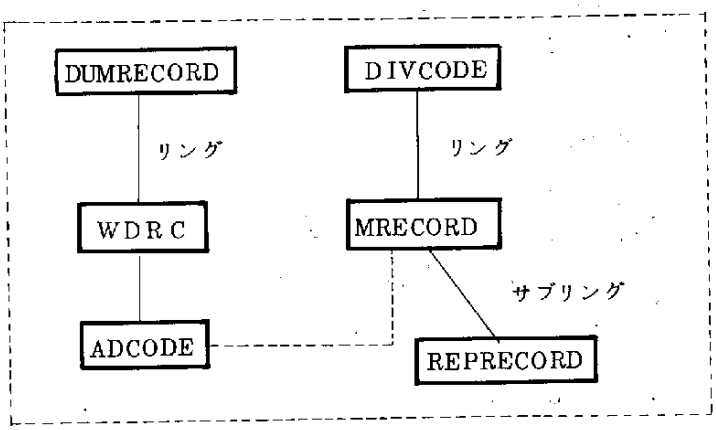
〔データ・ベースの構造〕 キーワードは，分類コードと用語から構成しており，分類コードと，文献内容を記録するマスタ・レコードとは，リングで結合している。用語インデックスは，ランダマイズ・レコードとして登録している。用語とマスタ・レコードとの関係は，用語を登録する際に，マスタ・レコードのアドレス情報を得ることによって結合している。

ここで，これらの関係をさらに細かに述べると図9.3.4の右側の関係にみるように，分類コードをページ指定するためのレコード・タイプ（ＤＩＶＣＯＤＥ），文献内容を格納するためのレコード・タイプ（ＭＲＥＣＯＤＥ），文献内容を追加し継続するためのレコード・タイプ（ＲＥＰＲＥＣＯＲＤ）を，各々階層的にリングで結合している。一方，図左側をみると，用語インデックスはあらかじめＥＢＣＤＩＣコードでソートされているものを使用するので，これをランダマイズ・レコードと関係づけをためのレコード・タイプ（ＤＵＭＲＥＣＯＲＤ），用語インデックスをライダ



勝 9. 3. 4　　ＭＡＳＣＯＬにおけるデータ・ベースの基本構成

マイズ・ルーチンを通し格納するためのレコード・タイプ（ＷＤＲＣ），用語と文献内容とを関係づけるため文献内容が格納されているマスタ・レコードのアドレス情報を得るレコード・タイプ（ＡＤＣＯＤＥ）を，各々階層的にリングで結合している。



レコード・タイプ名

DUMRECORD；キーワード・ソート用ダミー・レコード
ＷＤＲＣ　　；用語
ＡＤＣＯＤＥ　；ＭＲＥＣＯＲＤの格納アドレス情報
ＤＩＶＣＯＤＥ　；分類コード
ＭＲＥＣＯＲＤ　；文献内容
ＲＥＰＲＥＣＯＲＤ；継続レコード

図 9. 3. 5　　データ・ベースの構造

〔データの格納〕　①用語インデックスの登録：登録する用語インデックス・テープは，全キーワード３２桁がＡＳＥＮＤＩＮＧでソートされているものを使用する。ＭＡＳＣＯＬにおける用語

インデックスは，ＭＡＳＣＯＡ同様ランダマイズ・レコードとして格納するため，ダイレクト検索ができる。また，用語があらかじめソートされていることにより，ランダマイズ・ルーチンのアルゴリズムで生ずるアドレス指定でも，それらの用語はデータ・ベースのページに関連的に格納されるので，バッファへのアクセス回数を減し，検索を速める意味で効果的である。ランダマイズにおけるシノニムは，ＲＡＰＩＤシステムのランダマイズ・リストを利用して適当な処置を行なう。ＲＡＰＩＤのユーティリティにより，登録した用語やエラー情報はプリンタへリストする。

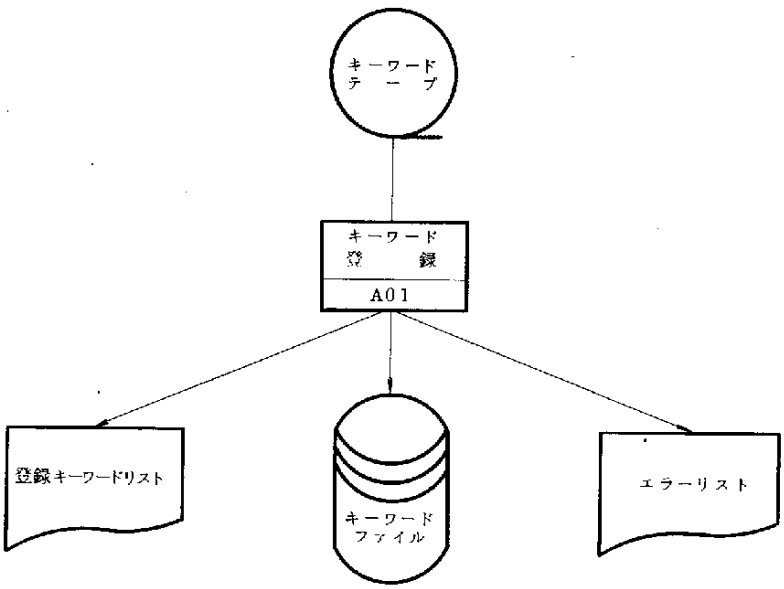


図 9.3.6　　用語インデックスの登録

②文献内容の格納：格納データと用語のインデックス・テーブルとのチェックは，格納データ１文献につき１～２個の用語（キーワード）をもっており，それらの用語がインデックス・テーブルに登録されているかどうかを調べるものである。キーワードが登録されていないレコードについては格納処理は行なわず，データ・エラーとしてリストする。文献内容を記録するマスタ・レコード（ＭＲＥＣＯＲＤ）は，分類コード別にリングで結合するとともに，文献内容を補足する継続レコードとサブリングをつくっている。マスタ・レコードを格納したアドレス情報（リファレンス・コード）は，用語インデックスと関連づけるためＡＤＣＯＤＥへ登録する。格納したデータは，キーワードとともにＲＡＰＩＤユーティリティを使用してリストし，その状況を確認する。

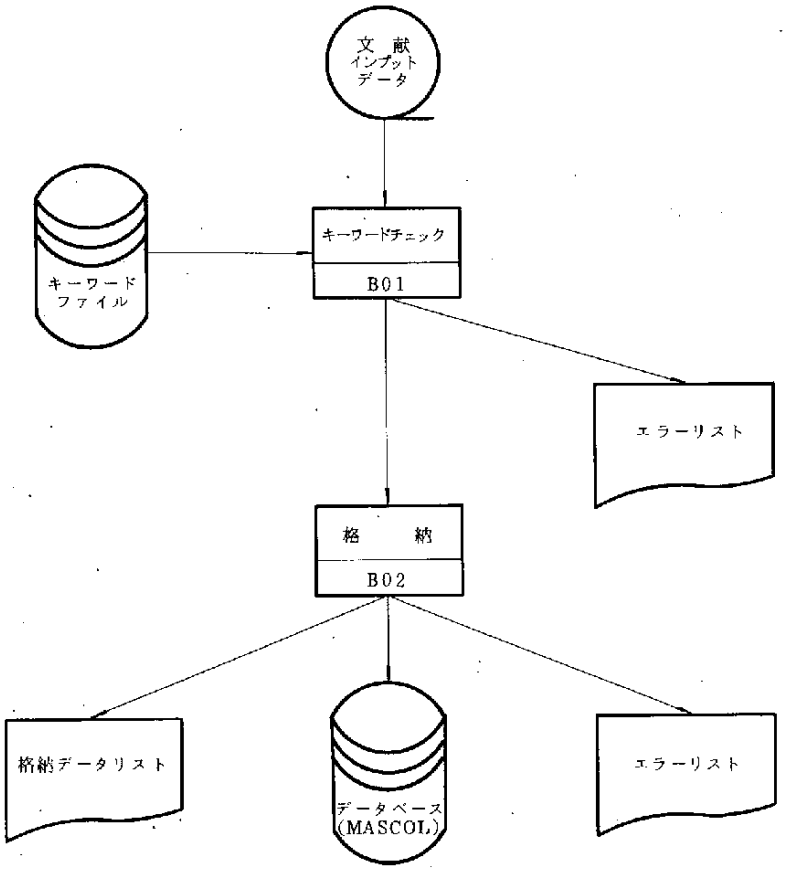


図9.3.7　　データの格納

(5)　検索の方法

検索は語または句からなる用語，または分類コードをキーワードとして行なう。問合せは端末タイプライタから行ない，応答はその件数の多少によりラインプリンタまたは端末タイプライタに行なう。緊急を要しない場合は，電算写植編集システムにより漢字出力もできる。検索方式は，簡単な専用言語を用いた会話方式で行なう。キーワードとデータとの関係は，図9.3.8の通りである。

問合せにより，キーワードがインデックス・テーブルに登録されている場合には，検索条件に従ってその命令を実行し，登録されていない場合には，インプット・エラーとしてその旨をタイプライタ表示する。用語は，ランダマイズ・レコードとして格納されているので，検索はダイレクト・サーチできる。分類コードの検索は分類表を参照し，インデックス・アドレス情報を求めて行なう。インデックス・アドレス情報とは，ＤＩＶＣＯＤＥに格納されている分類コードのリファレンス・コードであり，リング・マスタのポインタである。

キーワードが登録されているものについては，リング検索を行ない検索件数の表示を得る。検索結果の報告を受けたい場合は，出力の様式と実行命令を指示すれば，検索条件，出力件数，文献内容などがリストされる。すなわち，検索実行の命令を指示すると，ＰＲＩＯＲオプションの機能に

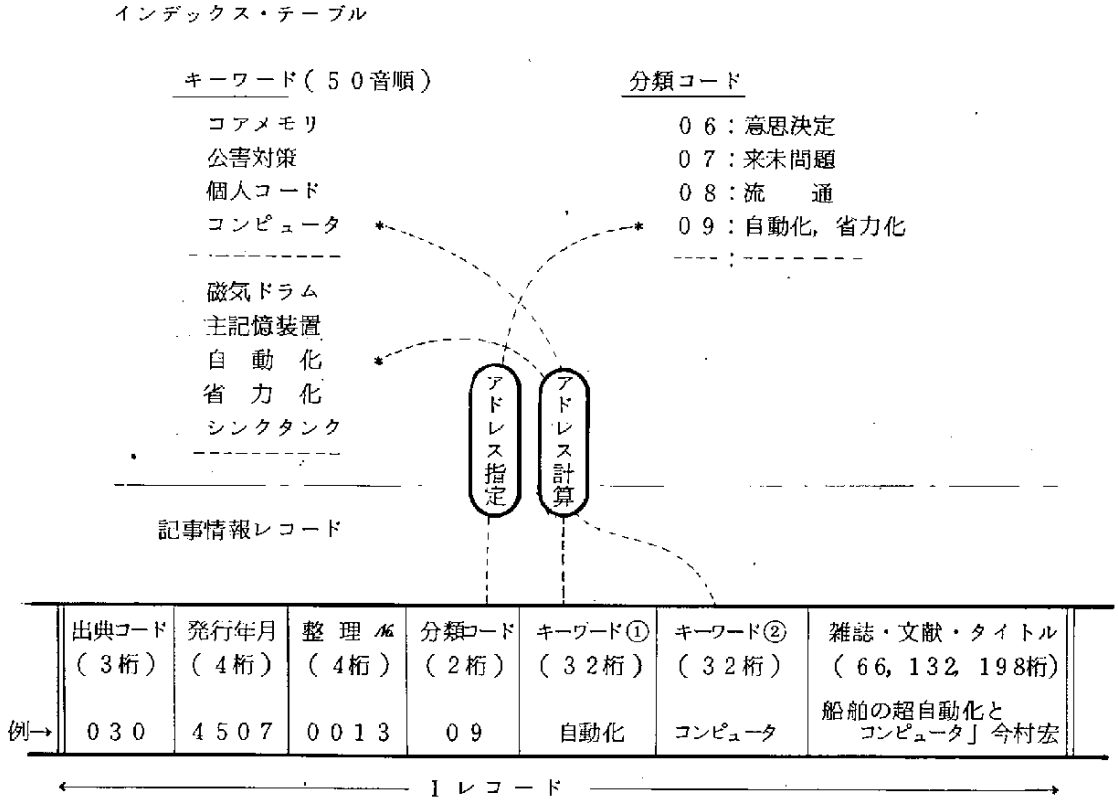


図9.3.8　　ＭＡＳＣＯＬにおけるキーワードとデータとの関係

より逆順のリング検索が行なわれ，アドレス情報としてのページやレコード番号などのリファレンス・コードを探し出し，アウトプット・エリアに文献データが導かれて出力される。この一連の検索プログラムは，図9.3.9に示す通りである。

検索指示および表示は図9.3.１０のような方法で行なう。

図9.3.１０の例では，機械工業におけるＭＩＳのシステム開発について掲載されている記事を４４年～４６年に発行された雑誌文献の中から検索する場合の指示とその表示方法である。

まず，ステップ１で４４年～４６年の文献情報を問えば，５２５件あるという応答がある。

ステップ２では，ステップ１の条件のもとにシステム開発が何件あるかという問に対し，５４件という応答である。

ステップ４では，ステップ３の条件のもとで機械工業の情報が何件あるかという問に対し，２件という応答である。

ステップ５では，ステップ４の結果の報告を受けたいので，その情報をタイプライタに出力できるようにせよという指示である。

ステップ６は，検索の結果の出力を実行せよという命令である。

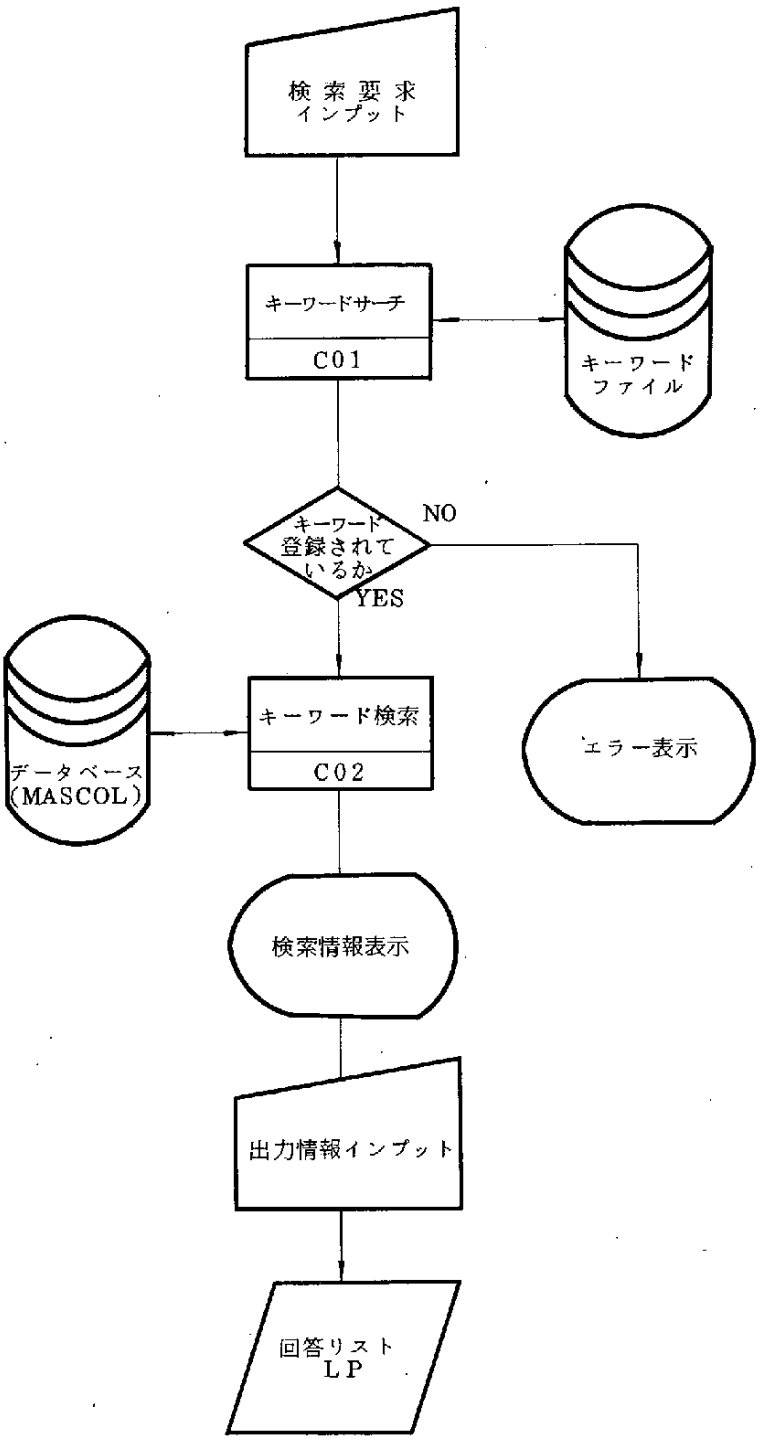


図9.3.9　　MASCOLにおける検索プロセス

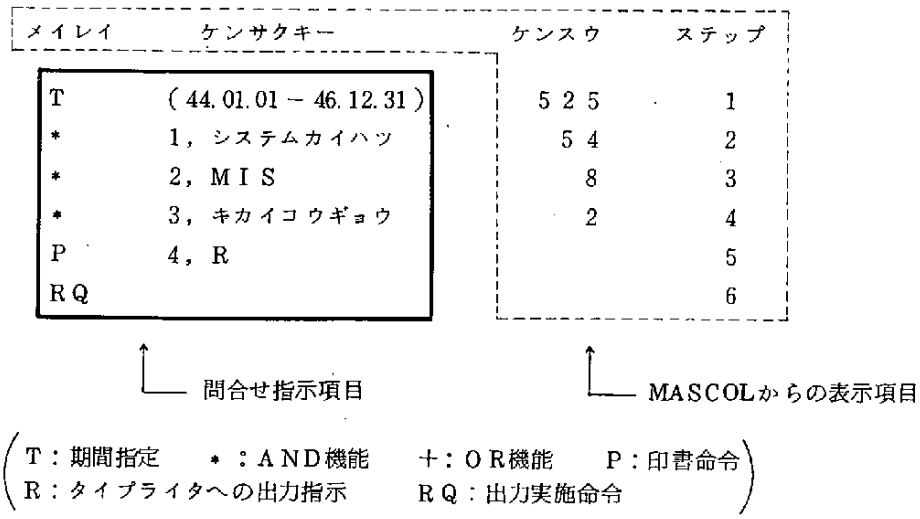


図9.3.10　MASCOLにおける検索指示と表示例

MASCOLにおける検索は，だいたい以上のようなステップを踏んで行なうが，検索結果の報告書の出力様式については，図9.3.11のような3つのタイプの形式で表示できる。

(6) メンテナンスの方法

MASCOLにおけるメンテナンスは，大別して，キーワードのメンテナンスとデータ・ベースのメンテナンスに分かれる。

〔キーワード・メンテナンス〕　キーワードのメンテナンスは，消去処理および追加処理によって行なう。修正処理は，キーワード格納フィールドがランダマイズ・フィールドであるため，PAPIDの規定上この機能が使えないことから消去または追加処理で代行する。

消去処理は，消去キーワードをレコード・タイプエリアにセットし，リファレンス・コードをサーチしたうえで，レコード・タイプのもつRTCSエリアにそのリファレンス・コードをセットして，消去コマンド（DELETE）で実行する。

追加処理は，キーワード登録ルーチンと同様にして行なうが，同一のキーワード登録を避るため，あらかじめキーワード・チェックをしてから処理を行なう。メンテナンスしたキーワードはプリンタへ出力し，その状況を確認する。

〔データ・ベース・メンテナンス〕　マスタ・データのメンテナンスには，消去処理，修正処理，追加処理がある。メンテナンス処理を行なうには，メンテナンス・データのアドレス情報をインプットしなければならない。アドレス情報は，データ格納ルーチンで出力したリファレンス・コードであり，メンテナンスは，これらのコードを参照し，キーワードにリングされているキーコードと関連づけて処理を行なう。

消去処理には，格納ルーチンでリストしたリファレンス・コードとキーワードの情報がインプット・データとして必要である。消去処理は，リファレンス・コードをマスター・データのレコー

① 端末タイプライタによる出力様式例

```
ケンサクジョウケン：（４５＋４６）＊（ミライモンダイ）＊（ソウゾウセイ）
タイトル　　：ミライガクニオケル　ソウゾウテキコントン ── ヒカクミライガク
　　ノタチバカラ」サイトウセイイチロウ
シュッテン　　：ジョウホウカンリ　　P. 0022
ハッコウネンガッピ：45. 07
```

② 電算写植編集システムＪＥＭ３８５０による出力様式例

```
検索条件：（公害）＊（コンピュータ）
題　　名：公害と戦う・コンピュータは公害追放の武器たり得るか」臼井健治
雑誌文献名：コンピュートピア　　頁. 0025
発行年月　：４５年９月
```

③ ラインプリンタによる出力様式例

ケンサクジョウケン：（45. 09）＊（オンライン）　テイキョウケンスウ　０１６

セイユジョノＭＩＳＴ　オンライン・コンピュータ」セキマナブ　（ＣＯＭＰＵＴＥＲ ＲＥＰＯＲＴ，Ｐ００２１，45. 09）

セイサンノ　オンラインシステム ── シンエイ　キミツセイテツジョノ　ゼンボウ」ノサカヤスオ（ケイソウ，Ｐ．００３９，45. 09）

オンラインシステムノ　セッケイト　ジッサイ(6)」ノグチアキオ（ビジネスコミュニケーション，Ｐ.００８１，45. 09）

オンラインシステムノ　パフォーマンスソクテイト　ソノジツレイ」アイザワタカシ，ハギモトタクゾウ（ビジネスコミュニケーション，Ｐ.0068，45. 09）

図9.3.11　ＭＡＳＣＯＬにおける出力様式

ドタイプにおけるＲＴＣＳエリアにセットし，消去コマンド（ＤＥＬＥＴＥ）によって行なう。この際に，下位レベルにリングされた題名の継続レコード（ＲＥＰＣＯＲＤ）をも同時に消去する。

また，キーワードとリングしているＡＤＣＯＤＥは，キーワード検索およびリング検索を行なって該当するものを探し出し消去する。

修正処理は，リファレンス・コードをＲＴＣＳエリアにセットし，データ・フィールドのエリアに修正すべきデータをセットして修正コマンド（ＭＯＤＩＦＹ）によって実行すれば，同一アドレスに新しい内容が記録される。

追加処理は，データの格納と同様の入力フォーマットで追加データを作成し，マスタ・データおよびキーワードをＰＵＴ　ＲＡＮＤＯＭのコマンドを用いて格納する。入力装置は，データ量に応じて，タイプライタ，カード読取装置などを使い分ける必要がある。

メンテナンス・データの処理状況は，ラインプリンタへ出力し確認する。ＭＡＳＣＯＬにおけるファイル・メンテナンスのプロセスを図9.3.12に示す。

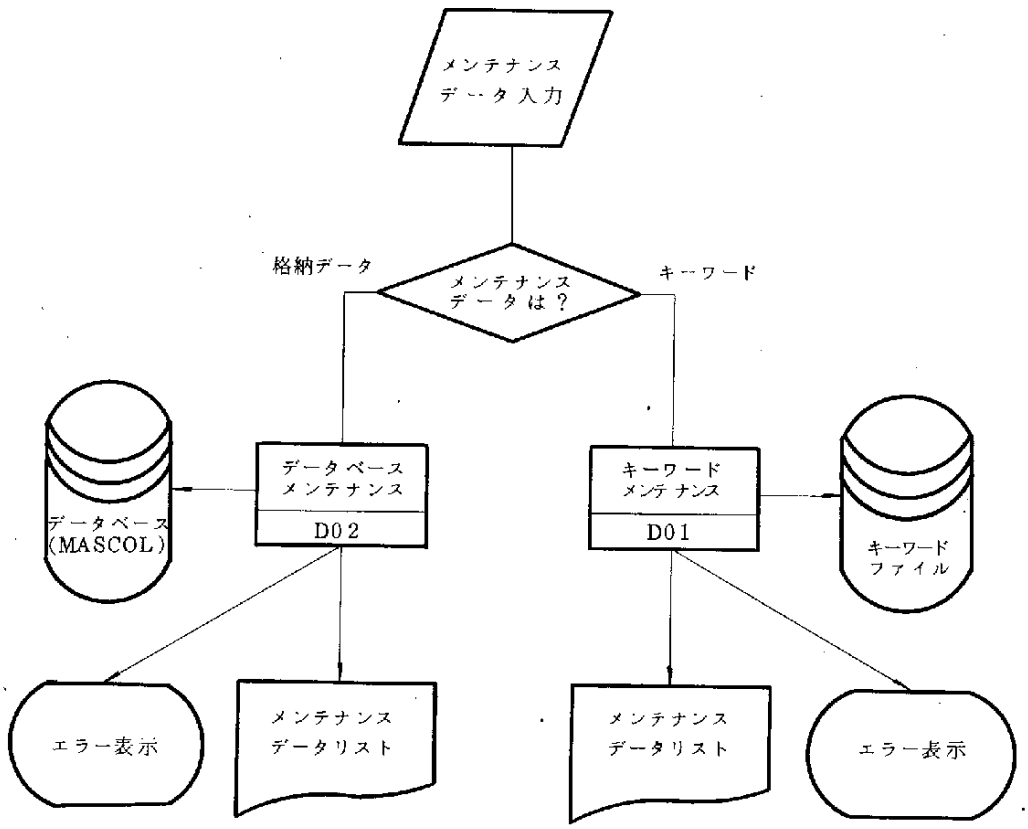


図9.3.12　ＭＡＳＣＯＬにおけるファイル・メンテナンスのプロセス

汎用データ・マネジメント・システム活用上の問題点

10. 汎用データ・マネジメント・システム活用上の問題点

メーカやソフトウェア会社から提供されている汎用データ・マネジメント・システム（ＧＤＢＭＳ）は，データの管理機能，検索機能，加工機能，レポート作成機能など豊富なユーティリティの機能をもち，しかもプログラミングを使易くするために特殊な言語をもっているといわれており，コンピュータ・ユーザにとってはシステム作成上このうえない便利な道具と思われている。

ＧＤＢＭＳが発表されて久しいが，しかし，これらのシステムをユーザ・サイドで真に有効に使いこなしている例はほとんどないといってもよい。現在，脚光をあび各界で注目されているＮＡＳＡの管理システム，ＮＨＫの番組総括システム，ＭＥＤＬＡＲＳなどの文献検索システム，鉄道の旅客サービス・システム，銀行の預金管理システム・信販のクレジット・サービス・システムなどは全て専用システムであり，独自のシステムでデータ・ベースを創りあげている。

これだけ便利なＧＤＢＭＳがありながらこれらが使われていない理由として，ＧＤＢＭＳのデータ・ベースに格納されるまでの前段階の用意が，ユーザにとってあまりにも大きな負担になりすぎることである。データ・ベースの概念として多目的のファイルを有機的に結合し共通的に管理しなければならないが，このためのデータの構成・ファイル編成が重要なキー・ポイントである。設計上の不備が格納に莫大な時間を要し，メンテナンスに大きな労力をかける原因となっている。また，データの格納からレポート作成処理に至るまで，機能面で優れたものを持ちながら十分な検討を怠ると，実際の活用面ではシステムを運用する段階で全く使いものにならないということにもなりかねる。

一方，プログラミングを容易にするため，データ・ベースの定義・格納・検索・メンテナンスなどの機能において特殊な言語をもっているが，これらの言語の働きを理解するために結構時間がかかる。どちらかといえば，これらのコマンドは専門家向きの言語といえるであろう。

この実験の結果判ったことは，ＧＤＢＭＳの全ての機能を業務にフルに発揮させようと思うと，現実には処理時間と労力の点で使いものにならなくなる恐れがあるということである。検索が主体であれば格納やメンテナンスの面で機能を落す必要があるし，レポート作成が主体であれば検索やメンテナンの面で機能を落す必要があるなどＧＤＢＭＳの機能を十分理解したうえで使い分けしなければならない。検索時間を速めるためには，複雑なデータ構造はできるだけ避け，階層を減らしポインタを少なくしなければならない。また格納を速めるためには検索と同様，データの構造はできるだけ簡単なものがよく，ソートリングは少なくした方がよい。

このようにＧＤＢＭＳは，必らずしも活用面からみて実際的に使い易いシステムとはなっていない。しかし，システム担当者がＧＤＢＭＳの機能を十分使いこなせる技術力をもち，企業のシステムが体系化されたうえで業務がデータ・ベースを作成できる状態に管理されていれば，ＧＤＢＭＳの利用効果は大きいことが疑いない。データ・ベースの概念は，企業の業務管理において無駄なデータを省き

総合的・システム的な方向を確立する目標として注目すべきであるが，同時にＧＤＢＭＳを企業のシステムの中で使いこなしてゆくという努力もまた必要であろう。

カタカナ漢字変換システム

11. カタカナ漢字変換システム

11.1 カタカナ漢字変換システムの動向

最近のコンピュータの普及・発展ぶりには，目をみはるものがあるが，それとともに，われわれの取り扱う情報の種類や量もまた飛躍的に増大している。

従来の汎用コンピュータは，カナ文字，英字，特殊符号しか取り扱うことができなかったため，コンピュータによる情報処理は，おのずから計算事務を中心とする数字情報に限定されてきた。しかし，情報処理分野における文字情報の重要性が認識されるにつれて，にわかに本来の漢字を中心とした文字情報への要請が高まってきた。

従来からの処理ではコンピュータで文字情報を扱う場合，漢字をカナに変換することによってかろうじて切り抜けてきた。しかし最近，コンピュータによる漢字処理への要請が高まってきており，漢字処理システムの開発がハードウェアの面でもさかんに行なわれ，4～5種類の漢字処理機器がすでに開発されている。

コンピュータで漢字を扱うとなると，それに付随していろいろな問題がでてくる。まず，漢字の数が欧米にくらべてべらぼうに多いということがあげられる。

一般の処理における文字の種類としては，最低5,000字位のものが必要であろうと思われる。そのためにインプット／アウトプットへの負担は大きくなる。とくに，アウトプットは，機械的に処理するので問題はないが，インプットはどうしても人的な要素が100%をしめるので問題である。コンピュータで漢字を処理する場合に問題となるのは，インプットにその大きな比重がかけられているという点である。

漢字処理のためのインプットデータの作成は，通常，漢字テレタイプによってインプットをする。従来からの処理として，すでに漢字をローマ字，カナ文字に変換して，コンピュータのインプット・ファイルとしてとってある場合の漢字出力は大変である。しかし，これもカナ文字を中心にしたデータをコンピュータで自動的に漢字に変換できるとすれば，時間，コスト，システム，などの面で漢字処理への移行は，可能である。

このような背景から，最近になって，コンピュータによる漢字処理システムの活用を容易にする方法の一つとして，「カナやローマ字」の入力をいかに「漢字かな混り文」に変換するかということが問題点となっている。

11.1.1 カタカナ漢字変換システムの概略

カタカナ文字を漢字に変換するには，次のような2通りの場合が考えられる。

① すでにカナ文字で漢字かな混り文をインプットしてある場合の変換

② 漢字かな混り文をカナ文字に直してインプットする場合の変換

この両者の変換処理は，いくつかの点で異なる。その大きな相異は，漢字をカナ文字に直した場合

におこる。

① 同音異議語

② ワカチ書き

などの問題によるものである。

「同音異義語」は，カタカナで「コウテイ」と書いた場合

工程，公定，行程，考訂，公邸，更訂，肯定，皇帝，校定，校訂，校庭，航程，高低，高弟……

などこれに対応する漢字が沢山あるので，カナを漢字になおす場合にこのような区分をどう処理するかということである。

また，「ワカチ書き」とは，たとえば

ヒルルスバンニコイ

とカナで書かれている場合

昼　留守　晩に来い

なのか

昼　留守番　に来い

なのかわからない。このためインプットの段階で分かち書きをしなければならない。

つまり

ヒル　ルス　バン　ニ　コイ

ヒル　ルスバン　ニ　コイ

というような語の区切りが必要になる。

他にいろいろな変換上の問題点はあろうが大きくわけては，この2つのポイントをコンピュータのなかでいかに処理するかということが大きな問題である。通常，コンピュータでカナ文字を漢字混り文に変換する場合，このような条件を区分するためある種のファンクション処理が必要である。

そのため，漢字かな混り文をカナ文字になおす場合は，変換のシステムにあったファンクションをあらかじめ挿入することができるので問題はないが，すでにカナ文字化されているデータについては，ファンクション処理がされていないので変換は大変である。

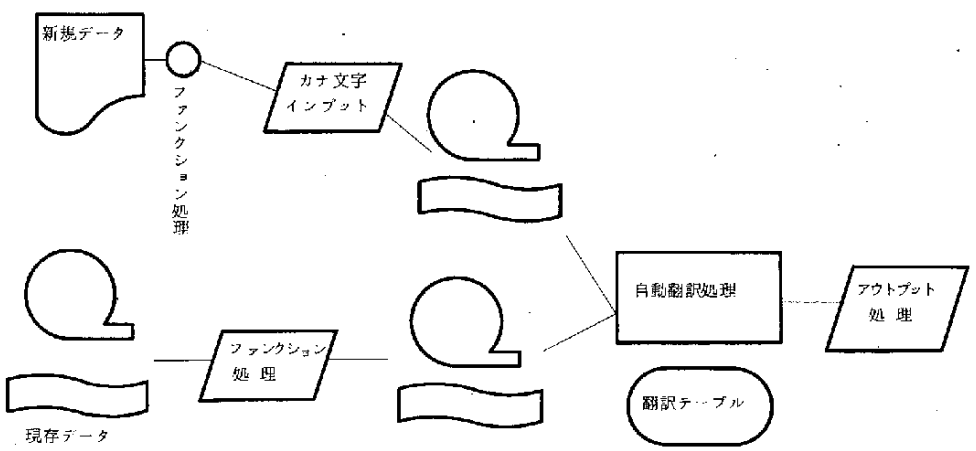


図１１.１.１　変換処理の典型的な例

1 1.1.2　入力処理および自動翻訳

カナ文字を漢字かな混り文に変換するための入力として，ファンクション処理が必要なことはすでにのべた。このファンクション処理は，翻訳の方式によりほぼ決まる。よって，ここでは一般的にどのような翻訳の方式があるかを述べ以後のシステムの参考にしたい。

翻訳方法は大きくわけて

① 文字単位

② 語単位

にわかれる。

1 1.1.3　文字単位の翻訳

(1) 音・訓を個別に区分する方法

漢字1字，1字のヨミガナを見出しとするテーブル（辞書）を用いて，データのカナ文字を変換する方法である。つまりヨミガナ単位にテーブルとつき合せを行ないデータを漢字1字分ずつ変換していく方式である。

この方法では，単位切りがどうなるかわからない。しかし「シンデンズジドウカイセキソウチ」（心電図自動解析装置）などでは，どのように単位切りされても困らない。この方法の，もっとも有利な点は「ヨミガナは変わらない」という点である。単語には，新しい言葉が生じるが，漢字には，このようなことはない。この方式では，同音異義語の処理がむずかしいという欠陥がある。入力方法の例をあげると，「この方法で索引する」という場合には，

※コノ・ホウ・ホウ※デ・サク・イン※スル／というような表現方法をとる。

(2) 音訓の両方から区分する方法

漢字1字1字のヨミガナを音と訓の両方から，インプットし，音訓の両方のテーブルより一致する語を選ぶ方法である。

たとえば「ホウ／カタ」という語に対しては下記のようなテーブルより検索する。

音：報，方，法，砲，芳 ……

訓：片，形，方，肩，型 ……

この方法で索引すると，音訓両方より一致したコードを選択できるので同音異義語などの問題はなく変換率は非常によい。ただしこの方法は，インプットの時の手間がかかる。

入力の表現形式の1例をあげると「この方法で」は，

※　コノ・ホウ／カタ・ホウ／ノリ※デ ……

となる。

(3) ツクリやヘンで区分する方法

この方法は，漢字1字，1字をヘンとツクリに区分するやり方である。たとえば「法」の入力表現方法は

・サンズイ／サル・

となる。

この方法はすでに漢字かな混りをカナ文字に変換してあるデータに関しては適用できないであろう。また，インプットのパンチ量は多くなるし，漢字をどのようなヘンとつくりにわけるかの判断がむずかしい。いずれにしてもあまりよい方法とはいえない。

1 1.1.4 単語単位による翻訳

(1) 分野区分，使用度数による方法

この方法は，語単位と違って単語単位に変換を行なう場合に用いられる。この方法では，変換テーブルが大きくなることと同音異義語の区分がつきにくくなるという欠点がある。そのため，分野区分，使用度数などによる何らかの優先順位で漢字を選択する必要がある。この方法ではユーザのインプット処理の手間や，またコンピュータによる変換のさいの完全翻訳のための訂正が大変である。入力方法の1例をあげると「この方法で検索する」は

「コノ・ホウホウ△デ・ケンサク△スル」と表示する。翻訳後の出力様式の例として

「・トウキ△ハ・カキ△ニ・キ△ヲツケヨウ」は

「〔冬季・冬期〕は〔火気，火器，夏季，花器〕に〔気，希，帰，貴，期……〕をつけよう。」という表現になる。

(2) 品詞判別方法

単語単位の変換であるが単語のうしろにくる品詞で，ある程度同音異義語を区分しようとする方法である。たとえば，

＜カトウ＞サンは

加藤

下等

過当

などから語尾の"サン"により〔加藤〕が選ばれる。このように語尾から人名や普通名詞・形容動詞の区分をする方法である。

この方法は単語テーブルの他に品詞区分，文法辞書などが必要となるので，テーブルが大きくなり，処理時間もかかる。

1 1.1.5 翻訳システムの基本

カナ文字をコンピュータにより漢字かな混り文に翻訳するためのシステムは，コンピュータが有する記憶装置や周辺機器によって変ってくる。以下その概略について述べる。

(1) インプット処理

Ⓐ すでにカナ文字データとして保管されているものについては，翻訳テーブルの方式によって，ファンクション処理をする。ファンクション処理は記票を人手で行ない，あとはコンピュータによって処理する。

もし，ディスプレイがあれば，ディスプレイで処理が便利である。

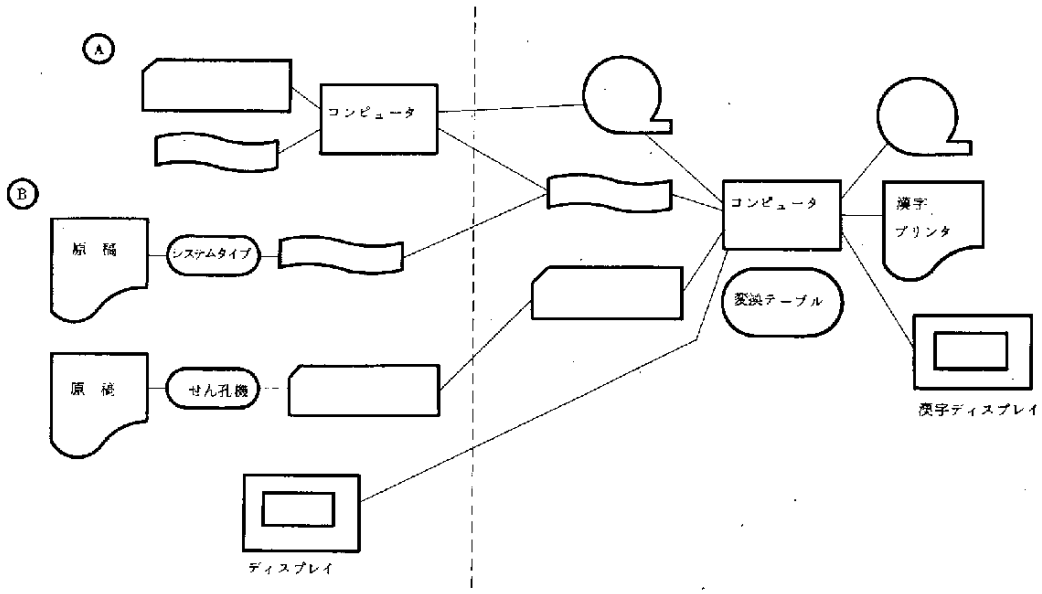


図１１.１.２　カナ漢字変換処理の概略

Ⓑ　新しくインプットする原稿に関しては，紙テープ，カード，などを使用するが，ディスプレイインプットも可能である。ただし，ⒶⒷともディスプレイの活用には，コンピュータを専有してしまわないように処理をマルチやＴＳＳで行なうか，または効率的なミニコンを使用しないとコスト高になるおそれもある。

(2)　プロセス（翻訳）処理

翻訳のためのテーブル作成方法については，前項に述べてきたのでここではインプットデータを，どのように翻訳するか，つまりテーブルより漢字を検索する場合にどのような検索方法をとるかを述べる。

Ⓐ　シークェンシャル編成処理

磁気テープベースの検索と同じ要領でインプットデータの検索を頭から順番に行なう。この場合の翻訳テーブルには，ランダムにファイルされている場合と，ある種のシークェンシャルにファイルされている場合の２通りがある。

いずれの場合も翻訳テーブルが大きくなると時間がかかる。

Ⓑ　ダイレクト編成処理

インプットデータを処理するための翻訳テーブルの位置が，そのインプットデータのもつ固有のＫＥＹにより決定されるような翻訳テーブルの編成をいう。

この方式はデータと翻訳テーブルの間にある種の条件や計算式がある場合で，検索時間は速いがテーブルの更新は大変である。

Ⓒ ツリーの応用

先にダイレクト編成による方法をのべたが，この方法はさらに，検索する翻訳テーブルの項目の間に階層関係を見つけ出し，その項目ごとに翻訳テーブルにファイルする方法である。

1	2	3
Z_5	Y_1	
		X_3
		X_4
	Y_2	
		X_2
		X_3
		X_4

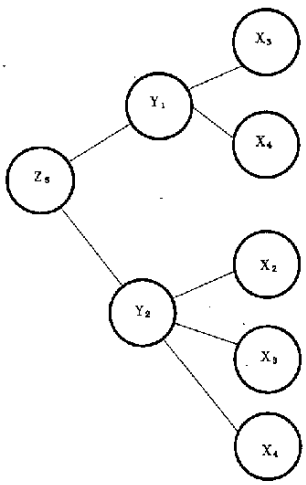


図１１.１.３　階層によるデータ編成

この方式では，組み合わせが多くなりテーブルそのものを多くもたなければならない反面検索時間は短縮される。その他いろいろな編成が考えられるが，ここでは基本的なものを紹介するに止める。

(3) アウトプット処理

アウトプットおよび完全データ作成については，２通りの方法が考えられる。

それは，

・漢字プリンタ

・漢字ディスプレイ

の活用である。しかし，このアウトプットデータにしても前述したように一度で完全な翻訳が行なわれることは少ない。１次アウトプット（翻訳）を行ない，何回か校正したのちに最終（完成）アウトプットまたは，最終完全ファイルを完成させるのが一般的な方法であろう。

この校正は，先のアウトプット機器（方法）により異なる。

Ⓐ 漢字プリンタによる場合

図１１.１.４に示すように翻訳ずみのアウトプットとインプット原稿とを校正し（赤字を入れ），赤字をパンチカードか紙テープに穿孔する。文の行または，節ごとには整理番号をつけ，コンピュータによる訂正をやりやすくする。訂正は，

① 部分訂正（字または単語ごとになおす）

② 全体訂正（整理番号がついている行，節すべてをなおす）

③ 挿　入

④ 削　除

などを組み合せて行なう。

コンピュータでは，訂正データと一次翻訳ファイルについている整理番号によって，マッチングし，内容を訂正する。

Ⓑ 漢字ディスプレイを使用する場合

コンピュータの補助記憶にある翻訳すみデータをディスプレイを使ってインプット原稿通りに校正するやり方である。この方法は，一番最適のように思えるが，コンピュータをかなり専有するのでコスト的には問題がある。

Ⓒ 漢字プリンタ，漢字ディスプレイを併用する場合

この方法は一次校正を漢字プリンタのアウトプット用紙で行ない，赤字部分の訂正をディスプレイで行う方法である。Ⓑの方法と異なって，すべてのデータをディスプレイ上でみる必要がなく，赤字のみを訂正すればよいので時間的にも短縮される。訂正の量があまりない場合は有効な方法である。

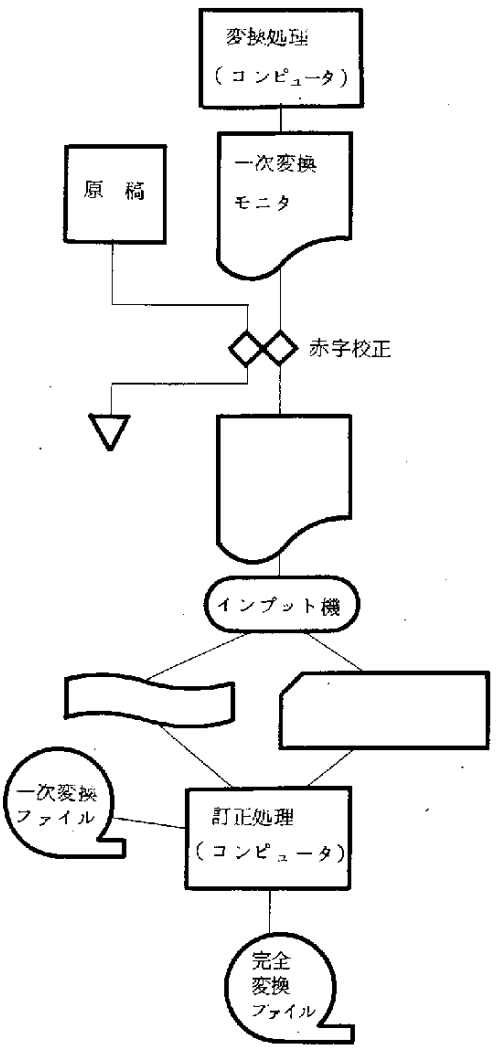


図１１.１.４　漢字プリンタを使用した訂正処理

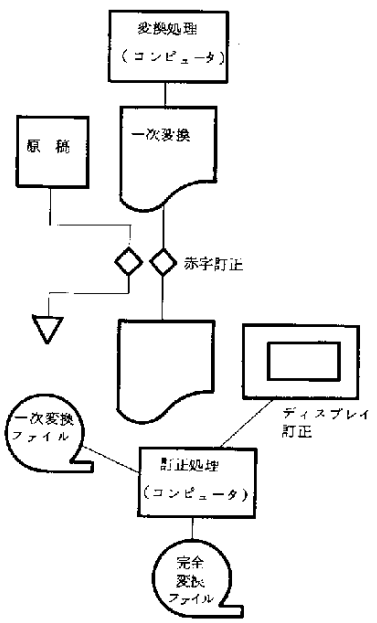


図１１.１.５　漢字プリンタ・漢字ディスプレイを併用した訂正処理

11.1.6　カタカナ漢字変換システムの開発状況

カナ文字入力に対する漢字かな混り文へのソフトウェア開発は，今迄あまりおこなわれていなかった。

しかし，漢字処理機器の開発によって，最近，漢字インプットの一方法としてにわかに注目をあびてきた。そのためどの程度のものがあるのか，はっきりした状況はわからない。現在，発表されているものとしては，次に示すものが有名である。

(1)　国立国語研究所

文字単位の変換テーブルをもち，音・訓を別々にして，翻訳する方式である。つまり，漢字1字1字のヨミガナを見出しとするテーブル（辞書）を用いて，データのカナ文字とコレート（辞書引き）をして行く方法である。（ローマ字文はカナ文字に変換した上で処理をする）ヨミガナ単位にテーブル・マッチング（つき合わせ）を行なって，データを漢字1字分ずつ変換して行く，このヨミガナ方式の最も有利な点は，「ヨミガナが変わらない」という事実であるといわれる。単語はつぎつぎに新語が生ずるが，漢字はまったく新しいよみがなが生ずることはない。この方法では，テーブルの容量も小さくてすみ，更新も非常にらくにできる。

(2)　日本ソフトウエア株式会社

単語単位の変換テーブルをもち，品詞判別方法をとっている。つまり単語の後にくる品詞である程度の同音異義語を区分している。

このシステムは，ローマ字を漢字かな混り文に翻訳するシステムでローマ字→カナ文字→漢字かな混り文と翻訳していく。辞書としては，単語辞書，文法辞書をもっており，それぞれ，自立語，付属語のあつまりをテーブルにもっている。

(3)　沖電気株式会社

情報処理振興事業の一環として委託された「漢字かな混り文変換プログラム」である。

その仕様の概略は，次のようなものである。

Ⓐ　入　　力

①　入力データはかた仮名かローマ字とし，媒体は紙テープまたは，磁気テープとする。

②　入力は，本文ブロック，ダミーブロックなどから構成される。本文ブロックは，一続きの意味内容をもつ有効情報の集まりとし始め，終り，本文，識別詞等から構成する。本文はいくつかの段落から成り，段落は文節から構成する。ダミー・ブロックはコメントなど無効情報の集まりとする。

③　識別詞

各種の識別詞を用意し，略語宣言，消去，段落の識別，原表記のままの出力，確認の繰返し，その他が容易に行なえる。

Ⓑ　処理プログラム

処理プログラムの機能は大略，次のような機能をもつものである。

①入力変換機能

②翻訳，変換機能

③出力編集機能

④辞書作成，更新機能

⑤その他

(4) その他

その他，NHK，電算機メーカなどで研究または実施しているといわれている。とくに，インプットがカナ文字情報ではないが，２１ビットで構成される速タイプというインプットマシンがある。

この速タイプは，裁判所などで使用されているが，このインプットマシンを使って紙テープのインプットデータを作成し，その２１ビット構成のデータをコンピュータにより漢字かな混り文に変換するシステムが日立と日電で開発されている。

これなども，ある特定の符号を漢字かな混り文に変換するシステムであり，カナ文字を漢字かな混り文に変換するシステムの例として興味深いものである。

１１.２ ＭＡＳＣＯＴにおけるカタカナ漢字変換システム

１１.２.１ システムの目的

ＭＡＳＣＯＴは図１１.２.１のような基本システムのもとに構成される記事情報検索システムである。すなわち，このシステムは情報源として新聞，文献，雑誌(国内，国外)，辞典，その他から，コンピュータ，周辺機器，電子部品，基本技術，応用技術，政府の動向，メーカの動向，ユーザの動向，文献紹介，用語の意味などの情報を提供するシステムである。

インプットデータは，ＫＥＹおよび漢字情報をカナ文字に変換してカードに穿孔しており，アウトプットは，タイプライタ，文字用キャラクタ・ディスプレイ，ラインプリンタ，ＭＴなどに行なっている。最近，コンピュータによる漢字処理の問題がクローズアップされてきており，記事情報をすべてカナ文字で出力するのは，ユーザに対して，親切な処理とはいえないので，研究課題としてＭＡＳＣＯＴのシステムでは漢字処理を加えその出力を可能にした。

このＭＡＳＣＯＴにおける漢字処理のシステムのうち本節では次の①，③を中心に述べる。

① インプットデータの作成

これは，すでにカナ文でインプットした漢字データを，翻訳テーブルで自動翻訳し，漢字かな混り文として完成させて MASCOT漢字情報検索のマスタとする。

② プロセス処理

検索要求に合致するよう，キーワードを論理的に組合せて該当する記事情報をコンピュータで検索し，磁気テープに出力する。

③ アウトプット処理

検索し，磁気テープで出力した漢字かな混りの記事情報を，漢字プリンタ（5,０００字の文字を所有）の所定の用紙に印刷する。

④ データの提供

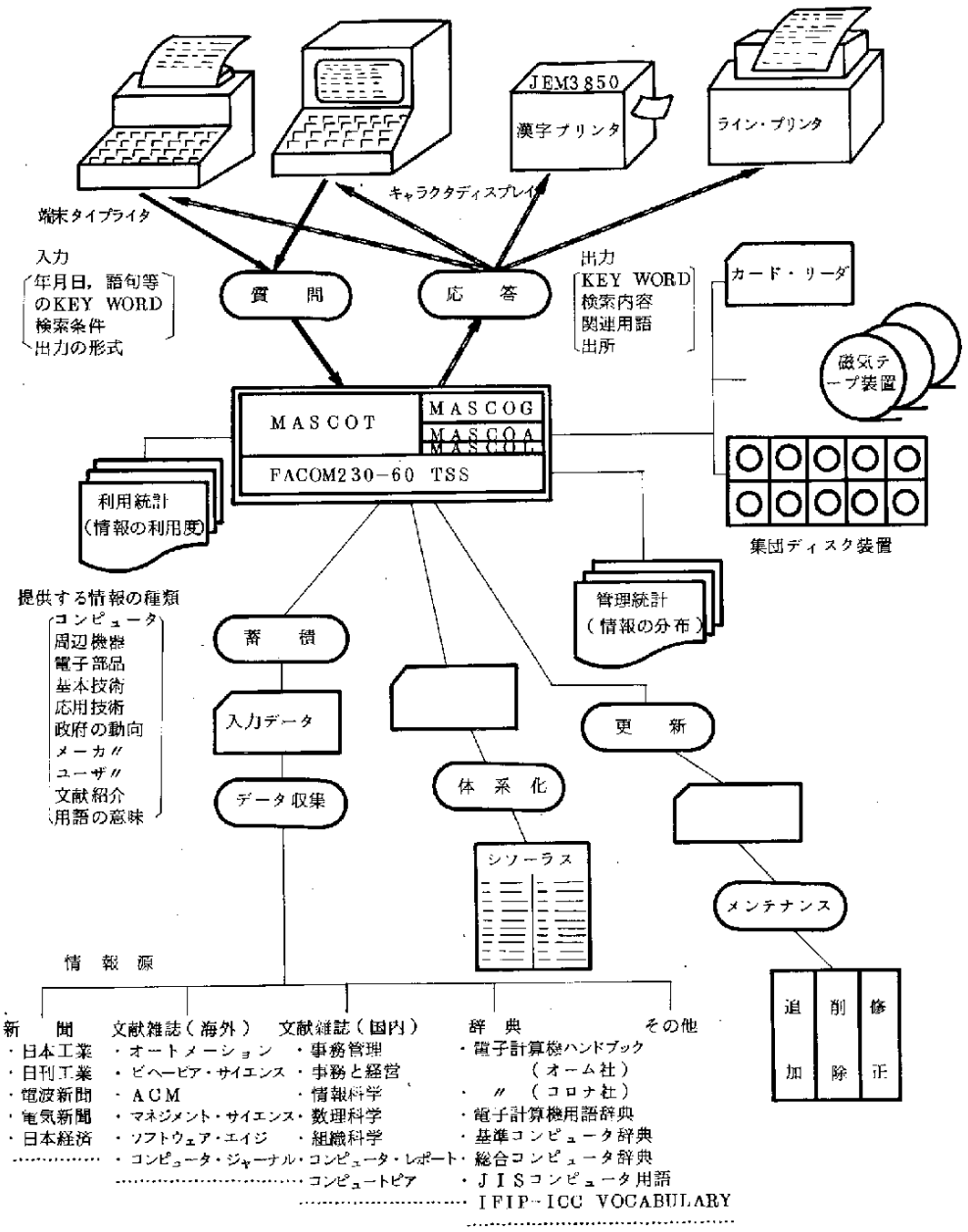


図１１.２.１　コンピュータ情報・管理システム≪ＭＡＳＣＯＴ≫

所定の用紙に印刷された報告書を，要求者に提供する。また，コンピュータに収録されているデータを，定期的にメンテナンスするため自動編集用電算写植システムを活用して，漢字かな混り日本語文としての記事情報をリストする。

11.2.2 処理の方法

このシステムにおける処理は，すでにカナ文字で，ファイルされているデータを漢字かな混り文に変換することである。記事情報のファイルには，変換するためのファンクション類は，一切入っていない。このため，原ファイルにファンクションを加える作業が必要となるが，このシステムではできるだけ，ファンクション類を少なくする方法を検討した。

以下その概略を述べる。

〔1〕 入力方法

カナ文字文を漢字かな混り文へ変換するためにおこる最大の問題は，データのセグメンテーションを如何に有効におこなうかということである。この問題を解決するために今迄種々の方法が考えられ実験されてきた。このシステムでは，それらの方法の中から一番単純でわかりやすい方法をとりあげ検討することにした。このシステムでは，基本的には，単語単位にカナ文字を変換する方法をとっている。

(1) 完全分離方式（A方式）

この方法は漢字（外来語も含めて）とひら仮名の部分を完全に分離して入力するやり方である。漢字の変換は，次に記すB方式と共通で単語の単位でセグメントして入力する。この形式の1例を示すと

・A　B　・C　D　・E・F・G　H

漢字　漢字　漢字　漢字　漢字

先頭に・があったら・または△（SPACEも含めたSpecial charactor）がくるまでは漢字とみなして漢字テーブルを検索する。

ex1. 原文：通産　電子工業課　IBMの価格分離について，検討中

入力：・ツウサン　・デンシ・コーギョウ・カ　IBM　ノ　・カカク・ブンリ　ニツイテ　・ケントウ・チュウ

ex2. 原文：需要予測の　手法と　その運用に関する　セミナー

入力：・ジュヨウ・ヨソク　ノ　・シュホウ　ト　ソノ・ウンヨウ　ニ・カンスル　・セミナー

規　則　漢字変換するものは，必ずその単語の前後の形が次のどちらかの形式をとらなければならない。

1°　・×××・

2°　・×××S

(2) 不完全分離方式（B方式）

考え方としては，A方式に少々手を加えたものである。この方法ではA方式のように原文を漢

文部分とひら仮名部分にわける必要がない。この形式の一例をあげると

・ A ・ B ・ C ・・D ・　　・ E ・・F ・・G ・ H

漢字　　漢字　漢字　　漢字　漢字　漢字

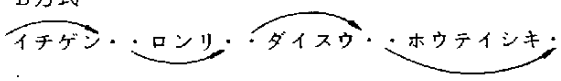
B方式では，・から・までが1語に該当し，・は次のセグメントには作用しない。

ex1. 一元論理代数方程式

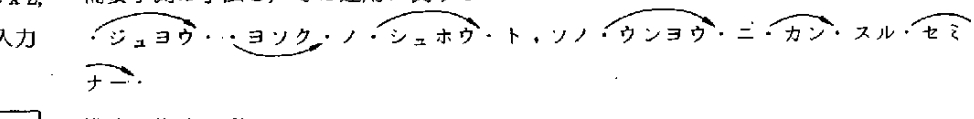
A方式

・イチゲン・ロンリ・ダイスウ・ホウテイシキ

B方式



ex2,　需要予測の手法と，その運用に関するセミナー

入力　 

規則　漢字変換する形

1°　・×××・

2°　・×××S

〔2〕漢字テーブル

翻訳テーブルの作成については，2通りの考え方がある。

① あらゆる分野に通用するテーブルを所有する。

② 専門分野ごとのテーブルと比較的利用度の高い一般用語のテーブルを所有する。

前者は，同音異語など用語の整理が大へんであるとともに，テーブルが非常に大きくなり，コンピュータの容量や処理時間も増大する。

後者は，テーブルを2種類サーチをする必要があるが分野ごとに処理するのできめのこまかい処理が可能である。

本システムでは，後者を採用している。また，翻訳の形式は，前述したとおり単語単位にもつことを基本としている。

漢字テーブルの作成の方法についてもいろいろ考えられるのでここで代表的な例をあげて検討してみる。

(1) 発生語登録方式

この方法では，テーブル中に発生する単語をすべて所有する。漢字テーブルの検索は1回ですませる。たとえばロンリダイスウホウテイシキ→論理代数方程式

(2) 語群分離方式

基本単語だけテーブルに登録しておけばよいので(1)に比べてテーブルは小さくすむ。ただこの方法は入力のフォーマットの制約（区切）を間違えると変換できなくなる。

テーブル例

「論理代数方程式」について作れば

ロンリ　　論理

ダイスウ　　代数

ホウテイシキ　　方程式

となり，入力方法は，ロンリ／ダイスウ／ホウテイシキとなる。

このようなセグメンテーションを行なえば，変換結果は「論理代数方程式」と出力できる。また，入力方法として

ロンリ／ダイスウホウテイシキ

ロンリダイスウ／ホウテイシキ

などのセグメンテーションも行なえるが，このようないろいろな組合せを考えるとテーブルが非常に大きくなってしまうという欠点がある。

しかし，翻訳するときどんな区切りになっていても翻訳できるようなテーブルが利用者にとっては好ましい状態であるといえる。このような考え方から「論理代数方程式」のテーブルは，次のような形でもつことができる。

ロンリ	論理
ダイスウ	代数
ホウテイシキ	方程式
ダイスウホウテイシキ	代数方程式
ロンリダイスウホウテイシキ	論理代数方程式

しかし，これだけの変換用語を用意しても，テーブルに全てのセグメンテーションのケースが盛込まれていない。たとえば，

ロンリダイスウ／ホウテイシキ→論理代数／方程式

の出力は

ロンリダイスウ方程式

となり，完全翻訳ができない。ここでも，インプットの区切りの方法がいかに難しいかがわかる。

このような不完全翻訳における問題を或る程度救う方法として，最長マッチ方式とかチェーン方式などがあげられるが，これらも問題点を含んでいるので前記の方式でおこなって段階をおって修正・進展させていく方法がよいように思われる。

〔3〕 外来語（カナ表示）の扱いについて

このシステムでは当初，外来語も漢字とみなし，テーブルを作成する予定であった。しかし，このシステムの中心となるコンピュータの用語は，外来語のしめる割合が非常に多く，ある部分では外来語が，漢字より多くなるので，外来語をテーブルとして持たず入力形式のまま出力する方法をとる必要がおこってきた。そこで次の2つの方法を考えてみた。漢字と外来語（カナ文字）をファンクションで区分するため，

(1) 漢字と外来語（カナ文字）のファンクションを別に考える。

(2) 漢字と外来語（カナ文字）の区分は，漢字テーブルより該当するものは漢字で，該当しないものは外国語あるいは未登録語としてカナ文字で出力する。

上記の詳細について例をあげて説明すると(1)の表示形式は次のとおりである。

$t_1 \cdots\cdots t_i / *u_1 \cdots\cdots u_j * / v_1 \cdots\cdots v_k / \# w_1 \cdots\cdots w_e \# / x_1 \cdots\cdots x_m$

ひら仮名　漢字　ひら仮名　カタカナ文字　ひら仮名

たとえば，「日本ユニバック２３日からジャーナリストを対象にコンピュータセミナーを開催する。」というニュースは記事情報ファイルでは

「ニホンユニバック２３ニチカラジャーナリストヲタイショウニコンピュータセミナーヲカイサイスル」

であるが，変換するための区切り入りファイルでは

「＊ニホン＊＃ユニバック＃２３＊ニチ＊カラ＃ジャーナリスト＃ヲ＊タイショウ＊ニ＃コンピュータ・セミナー＃ヲ＊カイサイ＃スル」

となり，漢字とカナ文字をファンクションで区分する。つまりファンクション＊の部分は漢字で，＃の部分はカナ文字でアウトプットする。

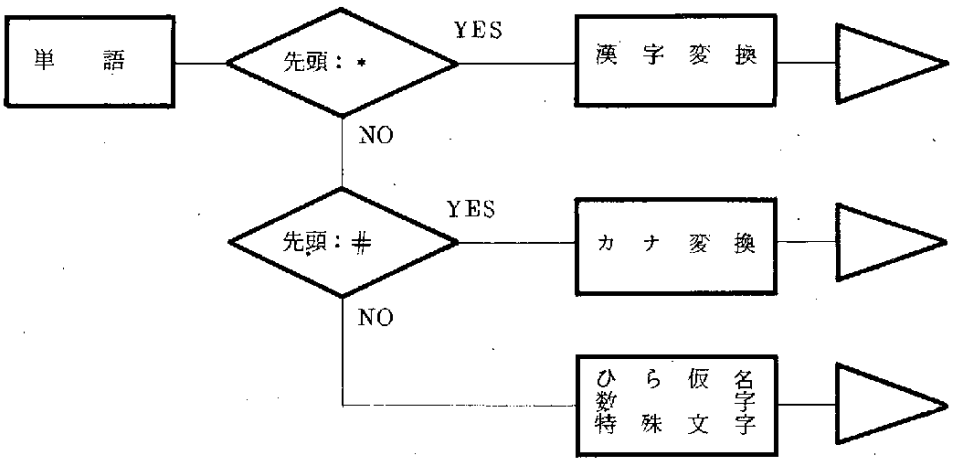


図１１.2.1　(1)　の　形　式

この方法では，漢字とカナ文字が別々に処理されるので時間的には，(2)に比べて遅い。

また，(2)の表示形式は

$t_i \cdots\cdots t_1 / * u_1 \quad u_j * / v_1 \cdots\cdots v_k / * w_1 \cdots\cdots w_e * / t_1 \cdots\cdots x_m$

(1)で使用した例でいえば次のようにセグメンテーションを行なう。

「＊ニホン＊＊ユニバック＊２３＊ニチ＊カラ＊ジャーナリスト＊ヲ＊タイショウ＊ニ＊コンピュータ・セミナー＊ヲ＊カイサイ＊スル」

(2)では漢字とカナ文字を同ファンクションで共有するために，ファンクションを使いわける必要がな

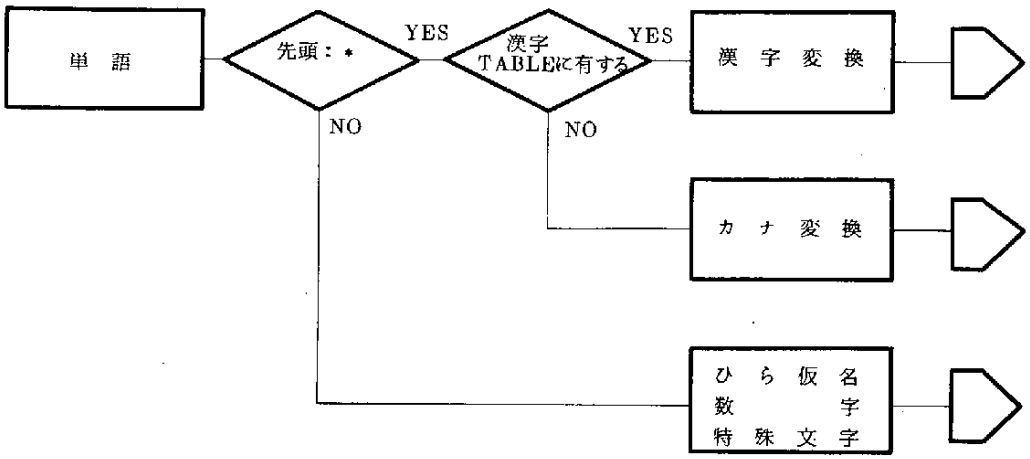


図１１.2.2　(2)　の　形　式

(2)は(1)に比べて，漢字のみならずひら仮名，数字，特殊文字など，すべて，漢字テーブル上で処理されるので処理時間がかかる。

〔４〕翻訳テーブルの決定

これまでにカナ文字に対する，漢字かな混り文変換システムの基本的な検討をいろいろしてきたが，ここで本システムの翻訳テーブルを次のように決定した。

(1)　単語分離について

文章中にあらわれた漢字をどのように分離したら最良であるかということは，非常に難かしい問題である。いろいろ検討した結果ここでは粗雑な方法であるが，単語の構成要素として「接頭語」「基本単語」「接尾語」という考えを導入し，文章中にあらわれた漢字の分離を進めて行くことにした。

1)　接頭語：たとえば「非」（ひ），「新」（しん），「再」（さい），「不」（ふ），「無」（む），「前」（せん）のように

①　この文字の後に漢字が２字以上続く形式のもの。

②　この文字を外しても単語として意味が通じるもの。

このような語頭の文字を接頭語と定義する。この場合ＭＡＳＣＯＴでは次のように表現する。

新機種→新／機種

新規　→新規

なお，下段は普通の単語であるが取扱上の相異を示すために掲載した。

2)　接尾語：たとえば「法」（ほう），「度」（ど），「面」（めん），「的」（てき），「率」（りつ），「値」（ち），「論」（ろん），「用」（よう），「数」（すう）のように

①　最後尾にこの文字がくるもの

②　この文字を外しても，単語として意味が通じるもの。

このような単語の末尾にくる文字を接尾語と呼ぶ。ＭＡＳＣＯＴで取扱う主なものをあげると，次のようなものがある。

一致法 → 一致／法
方法 → 方法

安定度 → 安定／度
精度 → 精度

位相面 → 位相／面
全面 → 全面

機械的 → 機械／的
目的 → 目的

誤差率 → 誤差／率
確率 → 確率

固有値 → 固有／値
価値 → 価値

構造論 → 構造／論
結論 → 結論

伝票用 → 伝票／用
費用 → 費用

２進数 → ２進／数
代数 → 代数

接頭語，接尾語の追加は適宜おこなう。但し同音異義語のあるものは除く（同音意義語の多いものは，そのものをつけて基本単語とする）。たとえば，

～せい	～か	～ひょう	～けい
～制	～化	～票	～系
～製	～課	～表	～形
～性	～科		

このシステムで今回行なった接尾語のわけ方はまちまちであるので，除々に是正し，その字のもつ性質を考えて分類していきたい。

(2) 漢字テーブルの構成

漢字テーブルに登録する単語としては，次の６種類の構成になっている。

漢字テーブル
1. 接頭語Ⅰ
2. 接頭語Ⅱ＋基本単語
3. 基本単語
4. 基本単語＋接尾語Ⅱ
5. 接尾語Ⅰ
6. 外来語

ここで接尾語Ⅰとは，(1)であげたものである。また，接尾語Ⅱとは，未整理のもの及び同音異義語の多いものである。

(3) 漢字テーブル・ファイルの構成

漢字テーブルをコンピュータで処理する際の，ファイルの方法は大きな問題であるのでここで前述の入力処理の方法にもとづいて，漢字テーブルのファイル構成を述べる。

① 同音異義語の扱い

ファイル上にＫＥＹをつけ，その後に同音異義語の漢字コードをつけた。たとえば，

	キョウリョク	*	協力・強力	LF

というテーブルがあった場合，入力データ

「・ハンバイ・ニ・キョウリョク・ナ・ソシキ・ヲ」

に対して出力データは

「販売に（協力・強力）な組織を」

となる。このように同音異義語の出力に対しては［＊］ＫＥＹを読むことにより（― ・ ―）という形で表示する。

このような処理方法は，ことごとく初歩的な思考であるが，ＴＥＳＴをおこなう上では非常に楽な方法である。一応，以上のような基本的な構想のもとにシステムを完成させるが，カナ文字を漢字かな混り文になおすときのすべての条件を満足させるには，まだまだ多くの対策が必要である。

前述したように，このシステムの変換は全分野にわたるものでなく，コンピュータという特定の分野における用語変換であることに注意していただきたい。

11.2.3 システムの構成

前述したような各種の処理方法を検討した結果，このシステムの概要を次のように決めた。

〔1〕漢字テーブル作成

(1) 漢字テーブル作成の手順

まず，カナ文字から漢字かな混り文への変換を行なうために変換テーブルを作る作業がある。

この作業を行なうため，現在，カナ文字でファイルされている磁気テープをコンピュータにかけ頻度順にそれぞれの単語（カナ文字）をプリントアウトした。

この頻度順にアウトプットしたカナ文字単語データに対して，同音異義語の区分，単語，ダブリのチェック，カナ文字単語の漢字への変換，などを行ない，変換テーブルの原稿を作成した。この原稿を漢字テレタイプにより紙テープ（２バイト／１語）に穿孔しコン

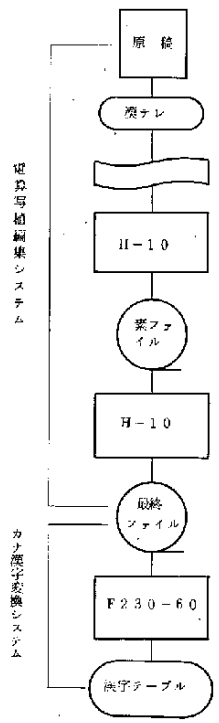


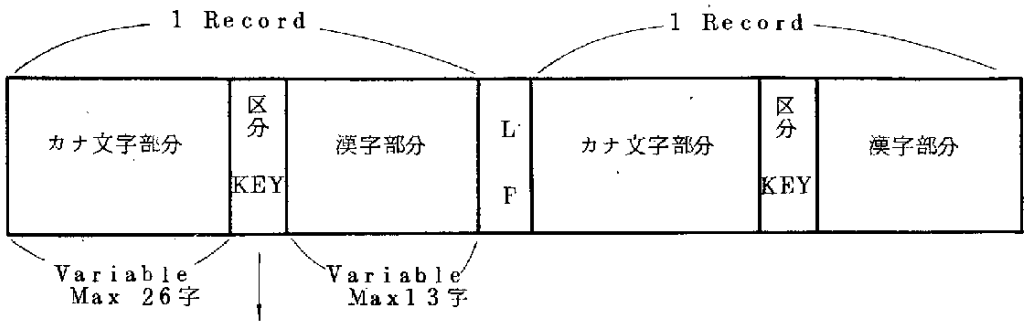
図11.2.3 カナ文字漢字変換テーブル作成プロセス

ピュータを通して，磁気テープに記録し素ファイルを作成した。その後，素ファイルに訂正処理を施し，最終ファイルを完成させた。

(2) I／Oフォーマット

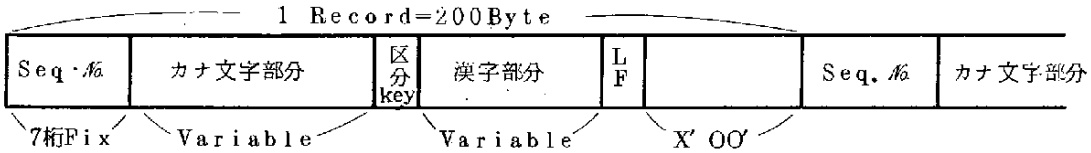
漢字テーブル作成のためのI／Oは，下記のとおりである。

① 紙テープフォーマット



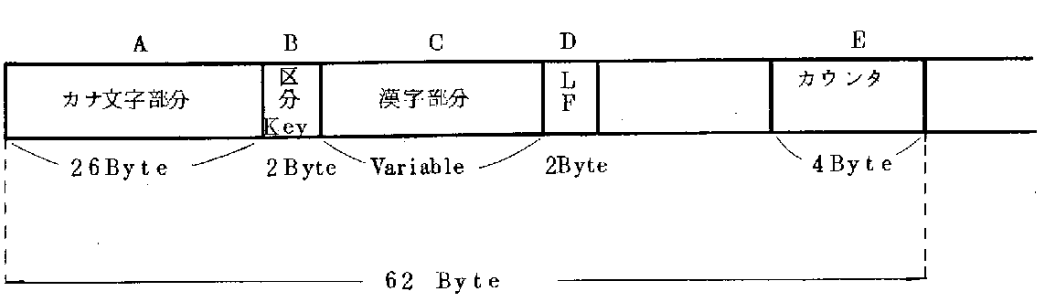
コードは，漢字テレタイプコードを取扱うために1字／2バイトであらわす。1レコードの項目は，カナ文字部分，漢字部分よりなり，各々の項目は，変長（Variable）である。

② 素ファイルの構成



素ファイルの1レコードの構成はSeq.№　カナ文字部分，漢字部分よりなる。Seq.№は訂正のときに必要な番号でレコードの番地をしめす。Seq.№は漢字変換のさいには必要がない。

③ 最終ファイルの構成



最終ファイルの1レコードの構成は，カナ文字部分，ＫＥＹ区分，漢字部分，ＬＦ区分，カウンタなどの項目よりなる。

A項目：検索ＫＥＹで固定長である。（ＥＢＣＤＩＣコード順），空白部分はSpace（X'40'）でうめる。

B項目：固定長で { □（漢テレの全角のSpace）X'1310' ——→ 単数対応
　　　　　{ ☆ X'2510' ——————————→ 複数対応

C項目：変長で$\frac{L}{F}$（X'1F2F'）がくるまでのデータをカナ文字に対応させる。Max13字（26Byte）

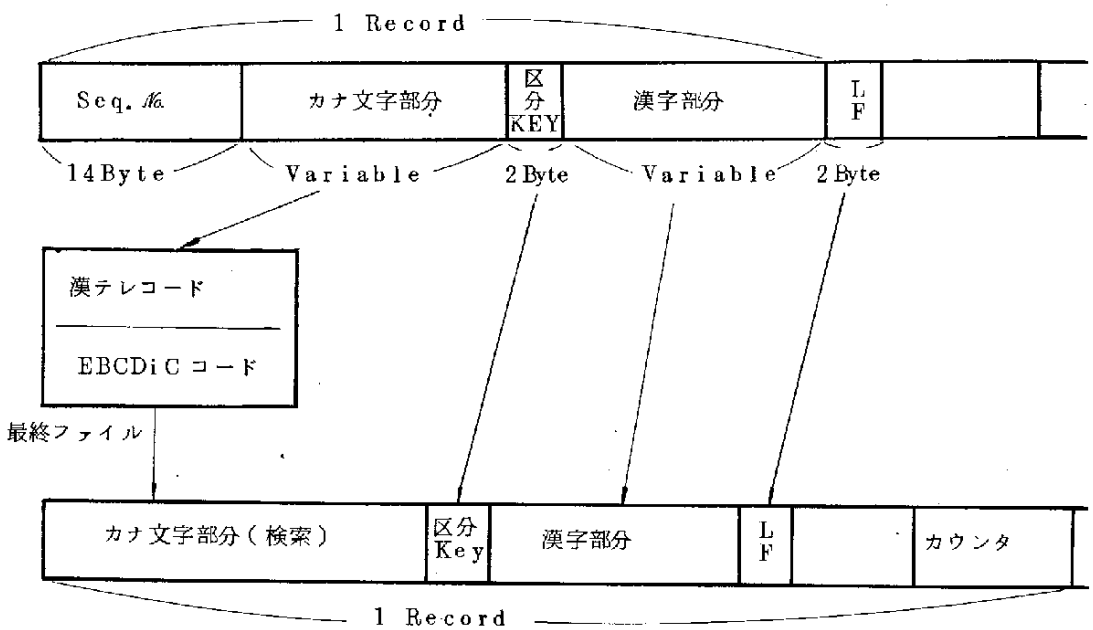
D項目：固定長で区切り

E項目：固定長で頻度数のカウンタ（頻度統計に使用）

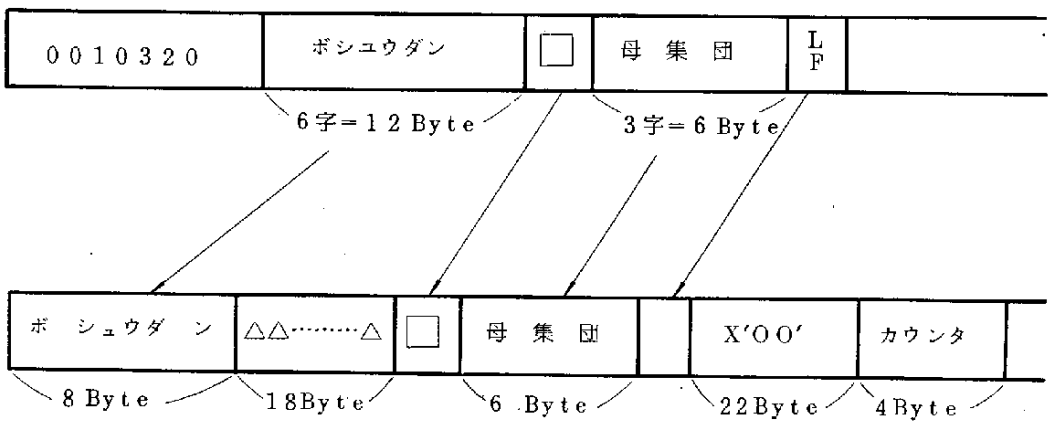
なお，ここでいうFixおよびVariable は項目の桁数であって位置ではない。

(3) 素ファイルから漢字マスタ（最終ファイル）への処理

素ファイルから最終マスタを作りあげる処理は以下のとおりである。



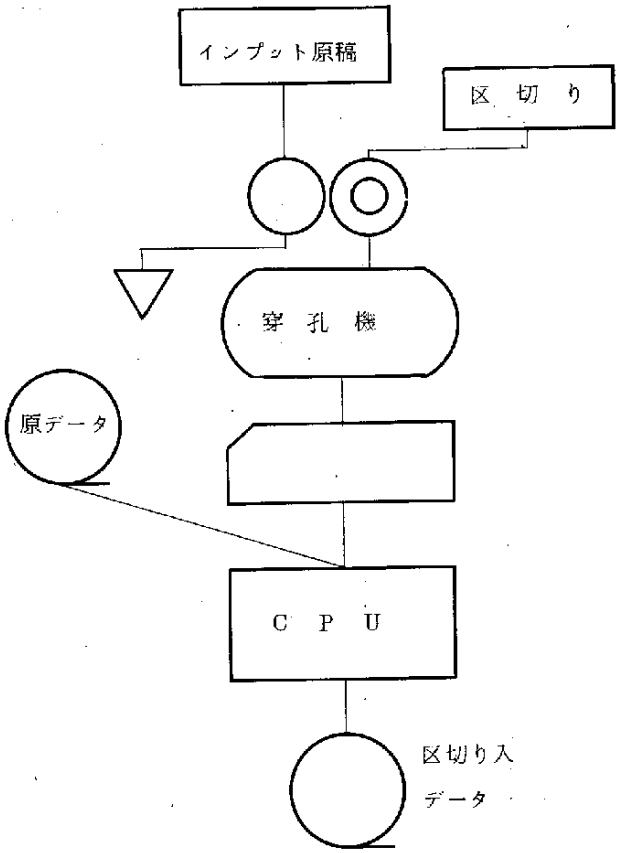
例



なお，漢テレコードの半濁音，濁音は，｛ダ：タ゛　プ：フ゜｝のような形で取扱う。

〔2〕 入力データの整理

漢字かな混り文に変換する原データとしての記事情報に，変換システムが必要とするファンクション（単語の区切）を挿入し，変換可能なデータとして整理するのがここでの作業である。



(1) 入力形式（区切り方）

入力に対するデータは＊または＃をファンクションとして下記のような条件でインプットする。

$\underbrace{U_1\ U_i}_{①}\ *\ \underbrace{V_1\ \cdots\ V_j}_{②}\ *\ \underbrace{W_1 \cdots W_k}_{①}$

$\#\underbrace{X_1 \cdots\ X_e}_{③}\#\quad \underbrace{Y_1\ \cdots\cdots Y_m}_{①}$

$*\ \underbrace{Z_1\ \cdots\ Z_n}_{②}\ *$

① ひら仮名・英数字・特殊記写変換

② 漢字変換

③ 外来語（カナ文字）変換

漢字および外来語に対応するものは，必ず次のような形であらわされる。

1. ＊ $x_1 x_2$ …… x_i ＊（漢字変換）
2. ＃ $y_1 y_2$ …… y_j ＃（外来語変換）
3. ＊ $x_1 x_2$ …… x_i　（漢字変換）
4. ＃ $y_1 y_2$ …… y_j　（外来語変換）

（注） 1と2の形式と3と4の形式の違いは，1．2．の形式の最後の印が単語の終結を意味し，3.4.の形式の最後の␣が終結を意味すると同時に空白をも意味していることである。これらの相異を例に示せば次のようになる。

＊ヨソク＊スル　→ 予測する

＊ヨソク＊␣スル
＊ヨソク␣スル　→ 予測␣する

＃アウトプット＃スル　→　アウトプットする

＃アウトプット␣スル
＃アウトプット＃␣スル
} →アウトプット␣する

また「一元論理代数方程式」は次のようなセグメンテーションを行なう。

＊イチゲン＊＊ロンリ＊＊ダイスウ＊＊ホウテイシキ＊

これを

＊イチゲン＊ロンリ＊ダイスウ＊ホウテイシキ＊

とすると

「一元ろんり代数ほうていしき」

と変換される。

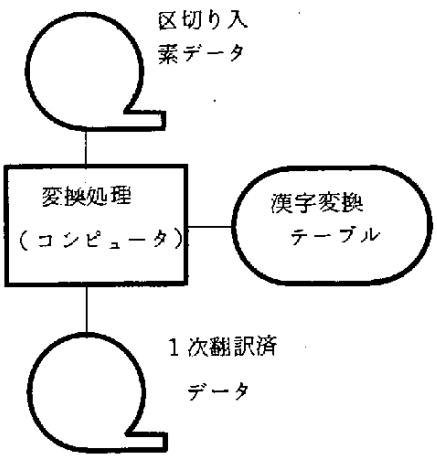
もし漢字テーブルに＊ロンリ＊がないときは

①は一元ロンリ代数方程式のようにカタカナでアウトプットする。

上記のようにインプットデータのセグメンテーション（区切り）は，完全翻訳を指向するための非常に重要な問題である。

〔3〕 データ変換

変換するために必要なファンクションの挿入が完了した記事情報のデータ・ファイル（区切入り素データ）は，コンピュータにより，単語単位で漢字かな混り文に変換される。



(1) データの変換

変換方法は，以下のような方法で行なう。まず文章の1区切（ファンクションから次のファンクションまで）が読みこまれる。その1区切の頭にあるファンクションが「＊」と「＃」，では処理が異なる。

「＊」は，漢字処理のために漢字テーブルサーチを行なって，コード変換するための印である。

もし，テーブルに該当する単語がなければ，単語はカナ文字に変換される。

「＃」は，外来語であるカナ文字および，英数字，特殊記号のコードをテーブルサーチして変換処理するための印である。

「＃」，「＊」以外のものは，ひら仮名コードをサーチして変換する。このように，カナ文字漢字変換とは，カナ文字（ＥＢＣＤＩＣコード）を漢字コードに変換することを意味するものである。

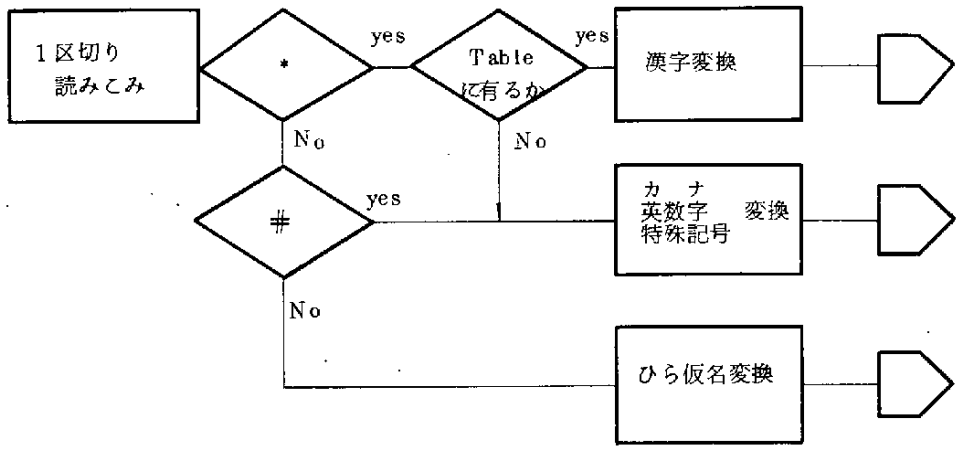
例をあげると

ex1. ＊タカチホ＊　＃Ｌ２０００＃ヲ　＊ハッピョウ＊　＃ハード＃ト……

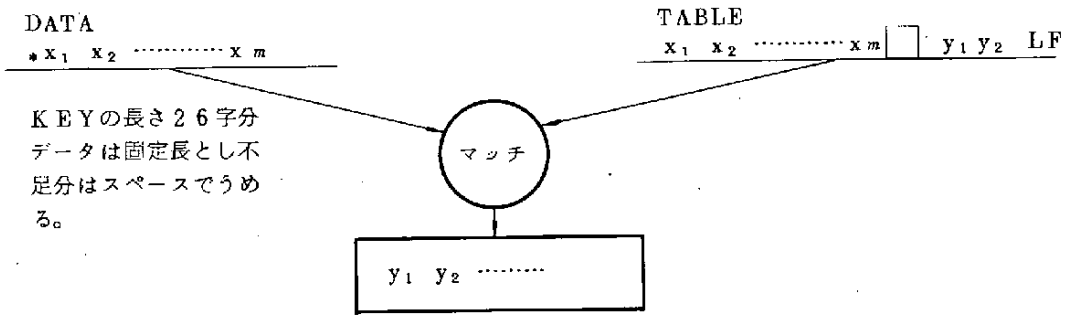
→　高千穂　Ｌ２０００を　発表　ハードと……

ex2. ＊ツウサン＊　＊コウギョウ＊＊カ＊デ＊カカク＊＊ブンリ＊……

→ 通産　工業（科・化・課）で　価格分離……



(2) 漢字テーブルの検索方法は，データの区切り部分と，テーブルのカナ文字部分をコンピュータ内部でマッチングさせ，一致したときに，テーブルの漢字部分を取り出しファイルする。



KEYが複数対応をとる場合（同音異義語）には（　，　）でくくって出力する。完全翻訳を行なう場合は後の作業として（　，　）から適当な語を選択し，その部分の訂正を行なう。

〔4〕 入力データの訂正

前述したように単語単位のカナ文字→漢字かな混り文への変換は，100%コンピュータで行なうのは無理である。そこで1次変換データを訂正し，完全な変換済データを作る必要がある。そのための訂正をいかに簡単にするかはシステム運用上重大な問題である。以下その訂正方法について述べる。

(1) 訂正方法の概略

漢字変換の終った，1次変換ファイルに対して

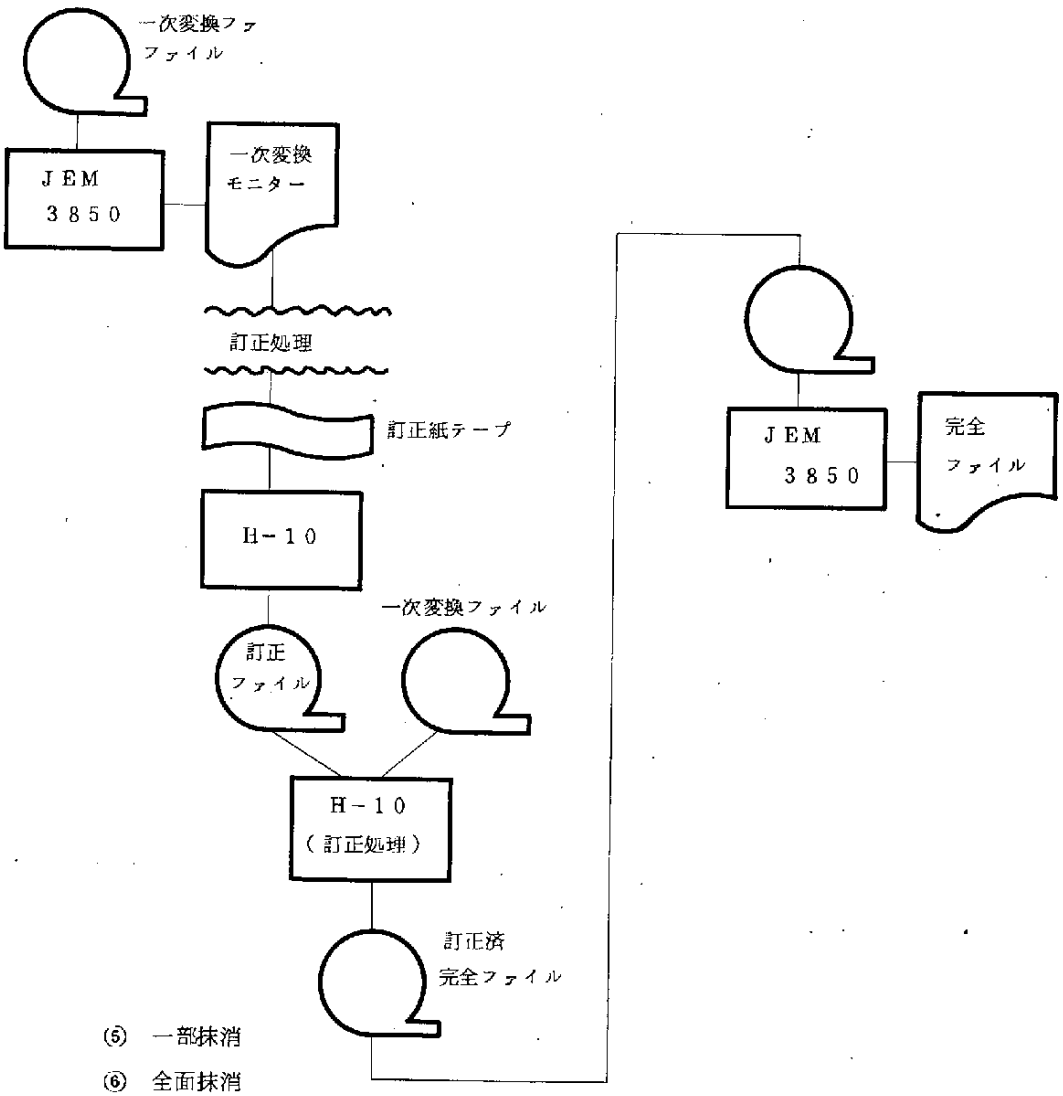
① 一部訂正（単語単位）

② 全面訂正（文章単位）

③ 選択（同音異義語のあるとき）

④ 新規挿入

図11.2.4　データの訂正処理



⑤　一部抹消

⑥　全面抹消

の操作を行なう。

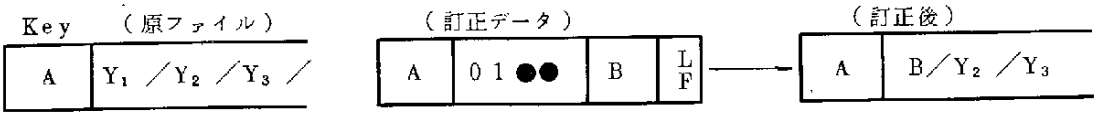
(2)　一部訂正の方法

文章１行中の一部の単語を訂正する場合で，その〔ＫＥＹ〕とその位置〔X_1，X_2〕（文の先頭からいくつ目にあるかを指示して訂正する。

ＫＥＹ	X_1 X_2	● ●	差し換え分	ＬＦ	

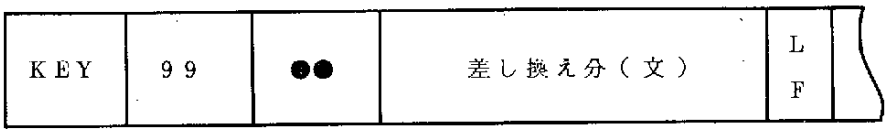
〔● ●〕部分は，無視〔さし換え分〕は，訂正するための正しい単語を入れる。

たとえば



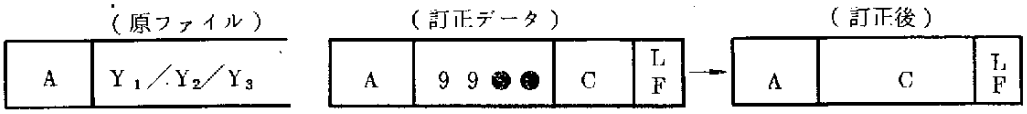
この例では，訂正指示〔０１●●〕により原ファイルの１番目のデータX_1 が抹消され，新しいデータBが挿入される。

(3) 全面訂正の方法



文章１行分のデータを全面訂正する場合に，記号〔９９〕を使用する。〔● ●〕は無視。〔差し換え分〕には，文章１行中全部のデータが入る。

例をあげると

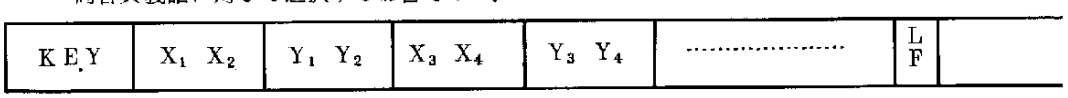


ただし，C＝{ X_1／X_2／……}

訂正指示〔９９●● C〕により，原ファイルの１行〔Y_1，Y_2，Y_3 ……〕は，〔C＝X_1／X_2／X_3 ……〕に全面訂正される。

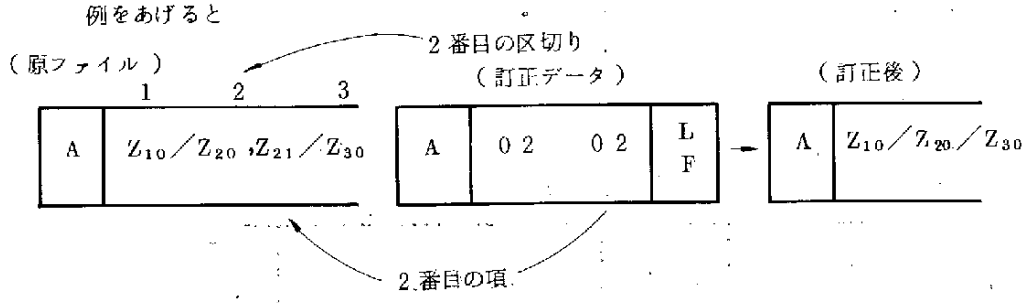
(4) 選択の方法

同音異義語に対して選択する場合である。



〔X_i X_{i+1}〕と〔Y_i Y_{i+1}〕は対になっている。〔$X_i X_{i+1}$〕は文の先頭から幾つ目かを指示する のであり，〔Y_i Y_{i+1}〕は異義語のうち何番目かを指示するためのものである。

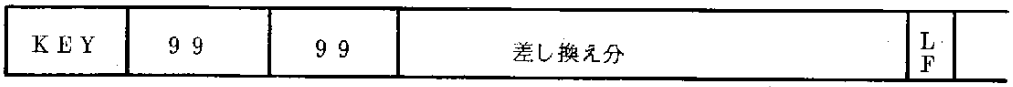
例をあげると



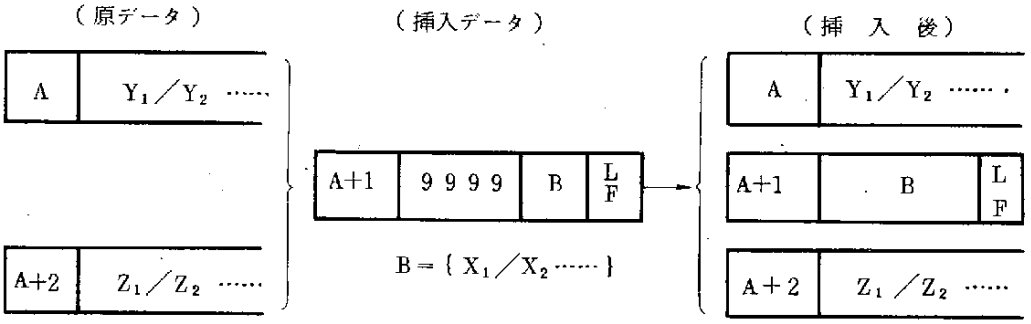
訂正指示〔０２，０２〕は，原ファイルの２番目のデータ〔Z_{20}・ Z_{21}〕 のうち，さらに２番目を削除せよという指示である。この指示によりZ_{21}が削除される。

(5) 新規挿入の方法

新規挿入の場合は〔９９，９９〕および〔ＫＥＹ＝Ａ〕を入力することにより，新規挿入分に〔Ａ＋１〕の内容が追加される。キーコードは，あらかじめ用意された補助コードに加算されるか，また，新規挿入された文章中の行以降のＫＥＹに自動的に＋１ずつ加算されをかいずれかの方法がとられる。



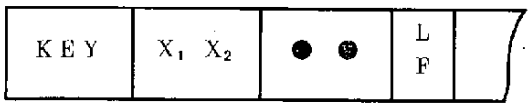
例をあげると



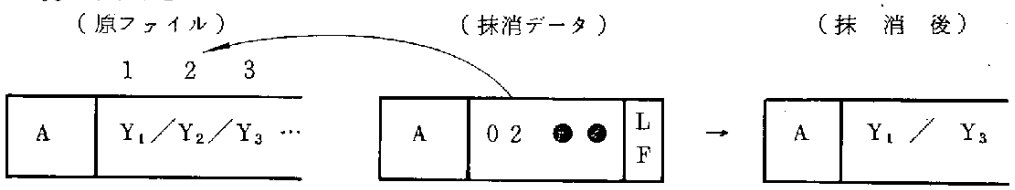
訂正指示〔９９，９９〕によりＡ行とＡ＋２行の間にＡ＋１行の内容が挿入される。

(6) 一部抹消の方法

文章中１行の単語を部分的に削除する場合で，その削除すべき行中の〔ＫＥＹ〕および先頭の単語から幾くつ目かの指示〔X_1 X_2 〕をすることにより，一部抹消を行なう。



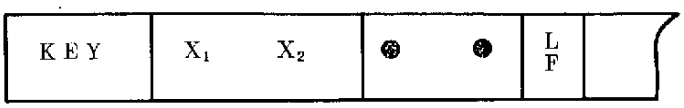
例をあげると



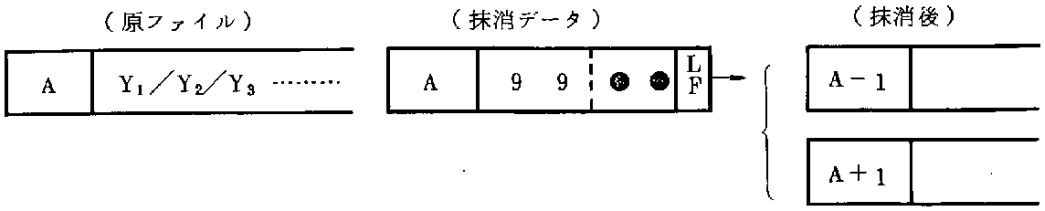
訂正指示〔０２●●〕（ただし次に指示されるデータを指示してはならない）によって，原ファイルの２番目のデータ Y_2 が抹消される。

(7) 全面抹消の方法

文章中１行をすべて抹消する場合で〔ＫＥＹ〕および〔９９〕を指示することにより全面抹消を行なう。



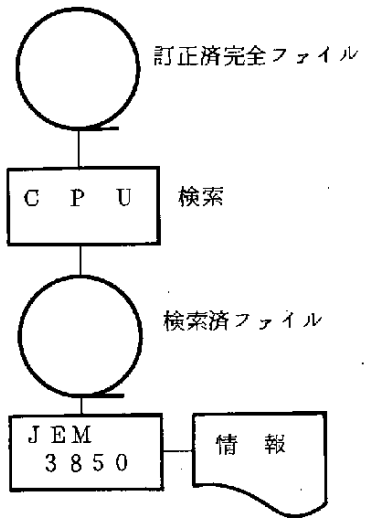
例をあげると

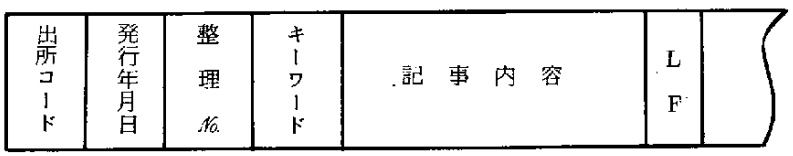


抹消処理によりA行のデータは，全部抹消される。

〔5〕 データ検索およびデータのアウトプット

完全に訂正された記事情報は，漢字かな混り文のマスタとしてファイルされる。マスタ・ファイルに収録された完全翻訳の記事情報は，漢字プリンタにより漢字かな混りの日本語文として各種のフォームにアウトプットされ提供される。漢字かな混り文のマスタテープ・ファイル・フォーマットは，下記のとおりである。





〔6〕 ハードウエアの概略

このシステムに使用されるハードウェアーは，

① 漢字テレタイプ

（5,000字インプット可能）

② HITAC－10汎用コンピュータ

・16KW（32KB）

・MT（3台）

・PR（131KW）

・PTR

・CR

③ ＪＥＭ３８５０高速漢字処理システム

・８ＫＷ

・ＭＴ（１台）

・ＰＴＲ

・モニタ・プリンタ

・ハードプリンタ　｛所有文字種5,０００字｝

④ ＯＵＫ－９３００

・３２ＫＢ

・ＭＴ（８台）

・ＰＲ

・ＣＲ

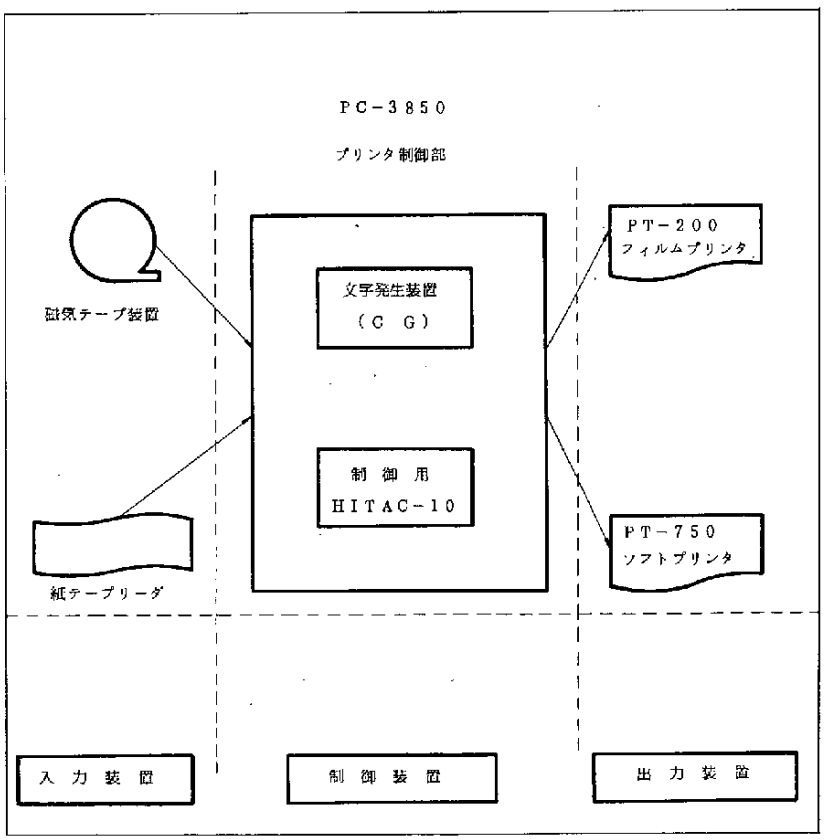
⑤ ＦＡＣＯＭ２３０－６０

・磁気ディスク４７１Ｋ

・ＭＴ（６台）

〔7〕漢字処理システムの概略

ＪＥＭ－３８５０電算写植編集システム構成は下記のとおりである。



入力装置

・コンピュータ……各種コンピュータに接続可能

・磁気テープ装置…ＭＴ－１００Ｄ９トラック８００ＢＰＩ

・紙テープリーダ…ＰＲ－３５０スルーレイト５００例／秒，インクリメント１２０列／秒

ＰＣ－３８５０プリンタ制御部

・文字種……５Ｋ，１０Ｋ，１５Ｋ，２０Ｋ

・外　字……1.　２５Ｋ字／ユニット，×１ユニット交換可能，他に５０字／カセット×５交換可能

・文字サイズ……５ポイントから２４ポイント（６級～３２級）

・文字の幅……８／８～１／８ cm

・送り量……字間・行間とも0.5 mm インクリメント

・組体裁……縦・横混植可能

・変形プリント……長・平体（任意率）各３種および上つき・下つき

・プリント速度……１５０字／秒・９０００字／分

・文字書体……モトヤ書体

出力装置

〔ＰＴ－２００フィルムプリンタ〕

・プリント方式……ラインプリント

・カメラ部……露光後にマガジンに収容・銀塩感材

・感光材料……印画紙・フィルム１００，１５０巾　１００メートルロール

・プリントサイズ……３０パイカ

〔ＰＴ－７５０ソフトプリンタ〕

・プリント方式……ラインプリント

・現像方式……安定化処理方式

・感光材料……スタビライズドペーパＢ４．Ａ４．Ｂ５　巾１００ｍロール

・ページカット……２００～４００ mm 以内

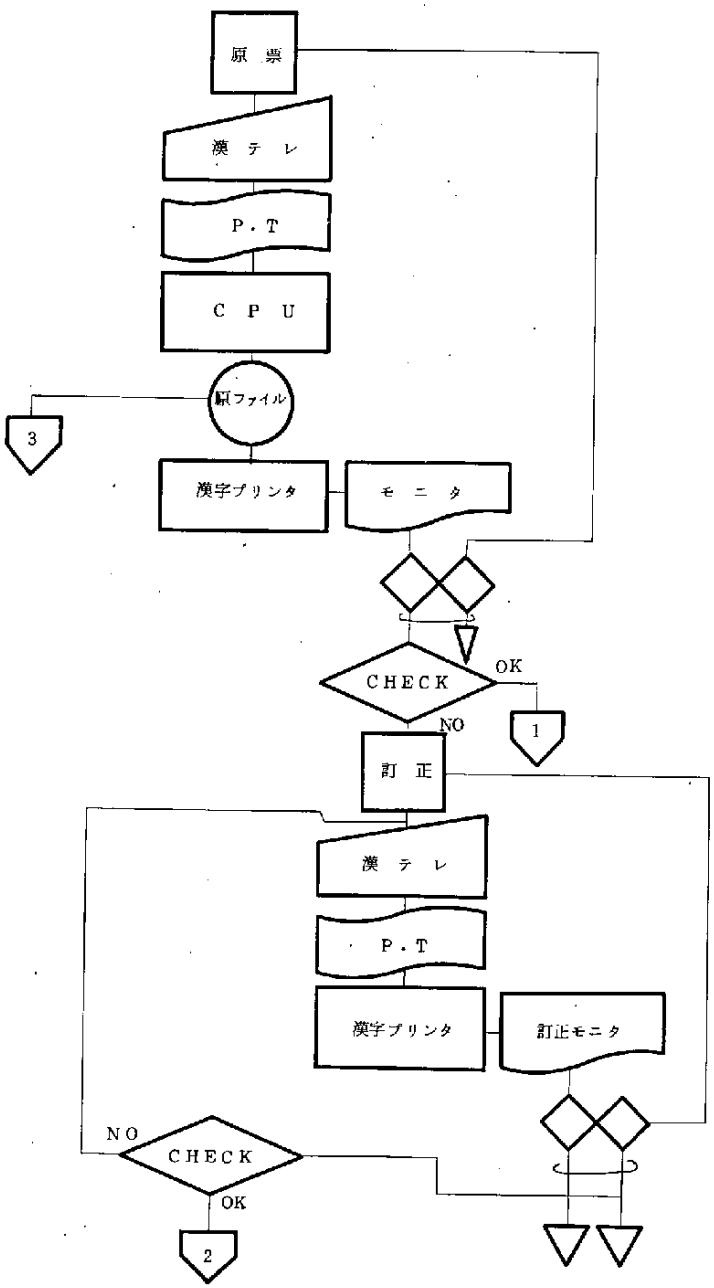
・ビジネスフォーム…フォーム合成プリント可能

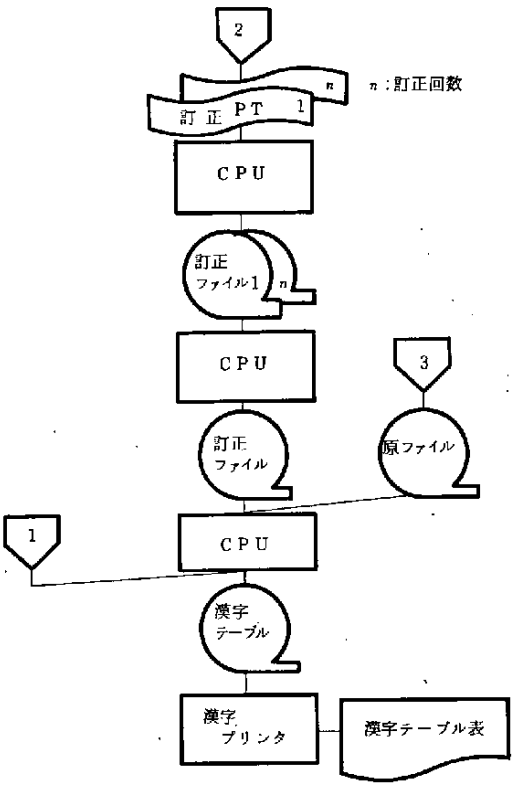
11.2.4 システム・フローチャート

(1) カナ→漢字混り　変換FLOW

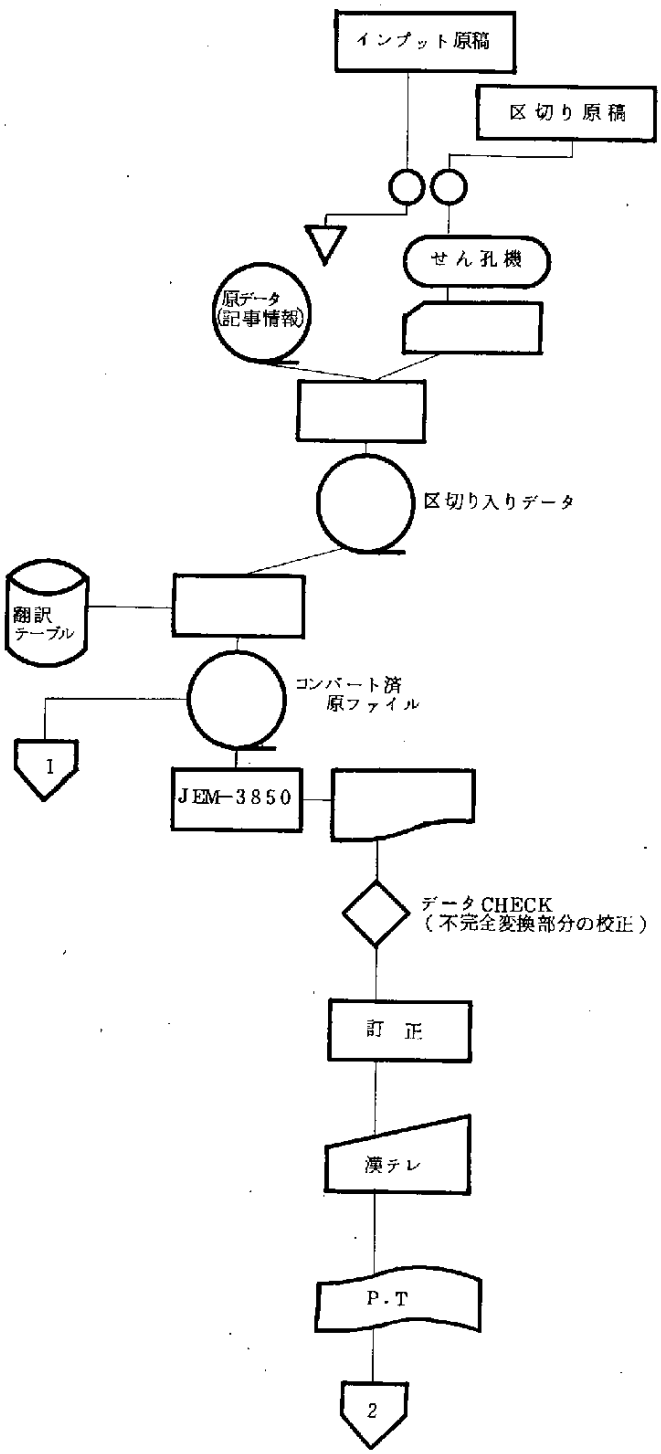
補助記憶としては，磁気テープを用いているが，ドラム，ディスクに置きかえても同じ処理となる。

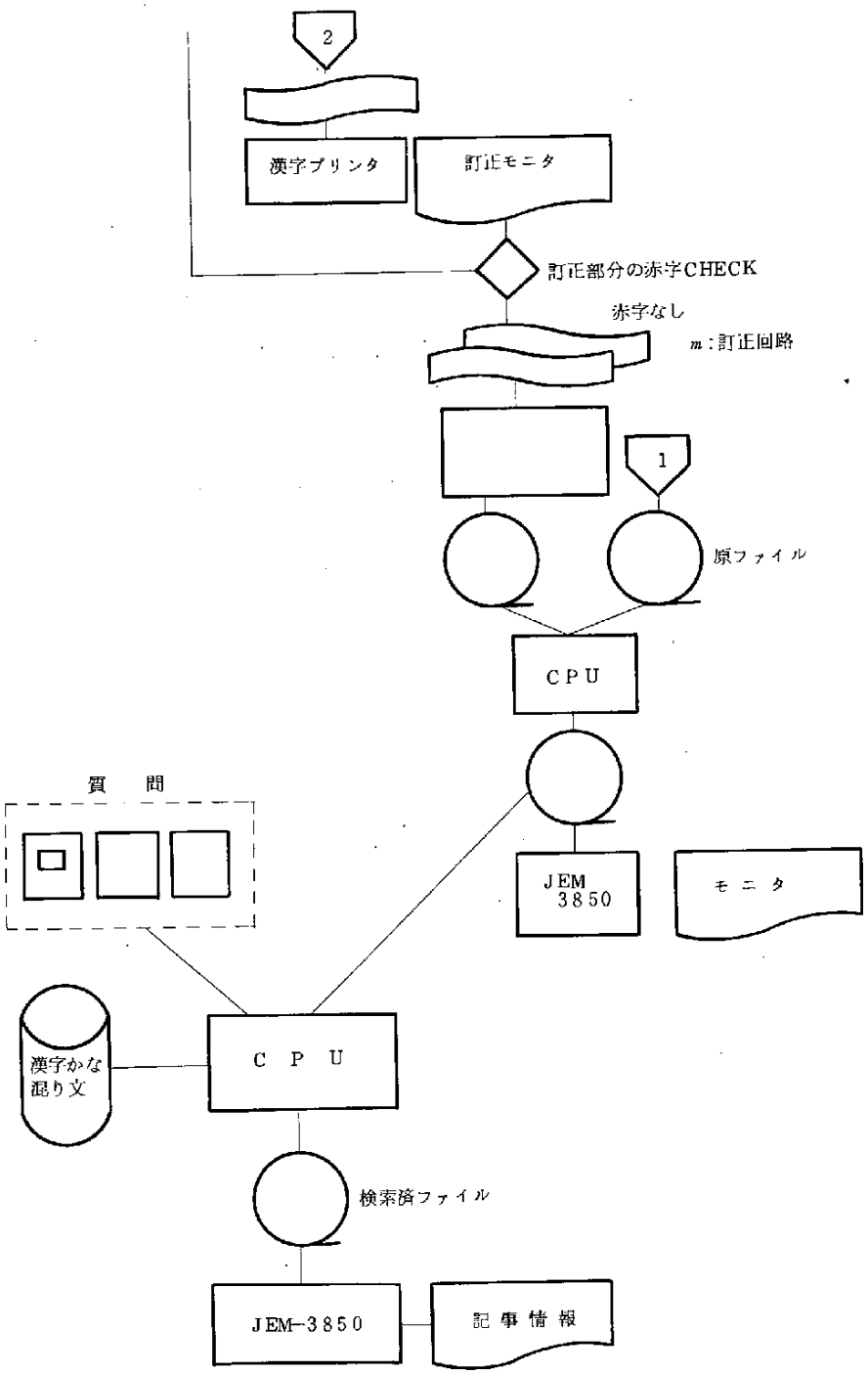
(1) 漢字テーブル作成




2
訂正 PT 1 n
n：訂正回数
CPU
訂正ファイル 1 n
CPU
3
訂正ファイル
原ファイル
1
CPU
漢字テーブル
漢字プリンタ
漢字テーブル表

(2) 漢字かな混り文完全データ作成




2
漢字プリンタ
訂正モニタ
訂正部分の赤字CHECK
赤字なし
m：訂正回路
1
原ファイル
CPU
質　問
JEM
3850
モニタ
漢字かな
混り文
C P U
検索済ファイル
JEM-3850
記事情報

11.2.5 変換実例

これまで述べてきた仕様に基づき事例研究を進めてきたが，記事情報の漢字かな混り文変換とその1次出力は，図11.2.5に示すとおりである。

図11.2.5　漢字かな混り日本語文の出力例

```
1)  70/01/06(01)D.00   ユニバックソウケン　図形認識などの　研究に　着手
2)  70/01/06(02)D.00   日本計算機　イバラキ工場お　増設
3)  70/01/06(03)D.00   日本計算機販売　札幌，静本に　支店お　新設
4)  70/01/06(04)D.00   日本エコノミストセンター　22日　▶企業における　システム適応法の　セミナー▶　開催
5)  70/01/06(05)D.00   日本アドレソグラフ，マルチグラフは　卓上電子複写機　A-M500お　発売
6)  70/01/06(06)D.00   アメリカ　マグニファイル社　ミニ，コン用　磁気ディスク　発表
7)  70/01/06(07)D.00   アメリカ　ボシロム社　グラフイック；レコーダー　発売
8)  70/01/06(09)D.00   インターシル社　6カゲツナイに　日本に　MOS，ICで　合弁会社　設立　予定
9)  70/01/06(10)D.00   アメリカ　ノベーション社　タイプ用　ボウオン装置　発売
10) 70/01/06(11)D.00   中小企業チヨウ　（情報・乗法）（化・算）対策お　45年の　重点サク㊀　とりあげる
11) 70/01/06(12)D.00   ミヤザキ電算センター　受託ギヨウムのゾウダイで　OUK1050お　増設
12) 70/01/06(13)D.00   広島市　行政ギヨウムの　機械化に　NEAC2200-200，300お　設置した
13) 70/01/07(01)D.00   タカチホ　BIT社の　ミニ，コン　販売　開始
14) 70/01/07(02)D.00   サハラ市ヤクショ　事務合理化に　USAC1500お　導入
15) 70/01/07(03)D.00   スエーデン性　マンモスタンカーに　最大の　電算機　設置
16) 70/01/07(04)D.00   アメリカ　ベリフエラルズ社　高速アクセスメモリー装置お　発表
17) 70/01/07(05)D.00   アメリカ　コンピユータ，メカ社　テイレンな　パイ，ダイレクショナル，テーブリーダ　発表
18) 70/01/07(06)D.00   イセキ納期　HITAC8210　導入し　全国的　流通体制　確率へ
19) 70/01/07(07)D.00   ヒタチゾウセン，川崎重工は　4月おメト　に　電算機による　人事管理へ
20) 70/01/07(08)D.00   東京信用金庫提界の　シンキン共同事務センター　設立　グタイカへ
21) 70/01/07(09)D.00   ソウサンのIC産業お　援助するため　▶IC　信頼性センター▶お　設置する　方針
22) 70/01/07(10)D.00   アメリカ　マツキンゼー社　今年　3月　日本支社お　設立
23) 70/01/08(01)D.00   IBM　大幅な　人換カイカク
24) 70/01/08(02)D.00   JSC（日本システムクリエーション）　真ソフトウエアガイシヤ　真発足
25) 70/01/08(03)D.00   アメリカ　プラクチカル社　ミニ，プリンターお　開発
26) 70/01/08(04)D.00   アメリカ　リツカーシヤ　コンピユータ（使用・仕様）の　ギヤンブル装置　開発
27) 70/01/08(05)D.00   東北金属　タイオウ輸出　強化のため　西ドイツに　販売ガイシヤお　設立
28) 70/01/08(06)D.00   アメリカ　ベル研究　あたらしい　デイスプレー材料お　開発した
29) 70/01/08(07)D.00   海外（電器・電機）通信（協力・強力）（回・会）　インド，セイロン，パキスタンに　シセツダンお　派遣
30) 70/01/09(01)D.00   トウシバ　GEの　TSS技術の　導入で　ツウサンへ　申請
31) 70/01/09(02)D.00   ミツビシ▶キンヨウ（回・会）▶が　（情報・乗法）産業への　進出お　本格検討
32) 70/01/09(03)D.00   デンデン　フジツウFACOM，JIP，IBMに　回線サービス　認可のミコミ
33) 70/01/09(04)D.00   デンデン　3分つから　カクカイトップの　データ通信教室お　開催
34) 70/01/09(05)D.00   ケイダンレンで　（情報・乗法）処理ゼイセイ　新設は　ふりとの　みかたでる
35) 70/01/09(06)D.00   インターテック社　SDCの　ソフトウエア　輸入で　よいみとうし
36) 70/01/09(07)D.00   農主リツチ対策本部　ケンナイに　コウジウチおき防止た　りコーナどお　しヨウニン
37) 70/01/09(08)D.00   航空ジエイタイ　静脈（情報・乗法）の　電話線送実験に　（成功・製副）
38) 70/01/09(09)D.00   ナオエツ電子（工業・興業）が　ちかく　シリコン生産　お　開始
39) 70/01/09(10)D.00   アメリカショウムショウ　シドニーで　▶データ通信，グラフイック，データ，システムズ▶（展・店）
40) 70/01/09(11)D.00   ダイイチギンコウ　総合オンラインお　めざし　FACOM230-60お　導入
41) 70/01/09(12)D.00   アメリカ　デジタル，エクイツプメント社　ミニコン　PDP11　2機種お　発表
42) 70/01/09(13)D.00   西ドイツIBMは　国内　電算機価格お　2から　4%ひきあげと　発表
43) 70/01/09(14)D.00   トウシバ　アメリカ　リツトン，メデイカル，プログツシヤト　イヨウ電子の　技術提携
44) 70/01/09(15)D.00   中部電力　▶デンリヨクリヨウキンケイサんの▶キカイかに　OCRお▶サイヨう
45) 70/01/10(01)D.00   古河サンスイカイも　（情報・乗法）サービスガイシヤの　設立お　検討
46) 70/01/10(02)D.00   ▶ソフトウエアシンコウ協議会▶お　8社が　結成へ
47) 70/01/10(03)D.00   経済調査研究会　研究局発財団の　設立お　提言
48) 70/01/10(04)D.00   ユウセイショウ　データ送信で　法改正要項お　2月（中・注・柱）　まとめる
49) 70/01/10(05)D.00   長野県庁で　電算機の始動　10月に　本格化
50) 70/01/10(06)D.00   日本エコノミストセンター　2月　▶電子計算機　シユウヘンキキ▶の　セミナーお　開催
51) 70/01/10(07)D.00   ガツケン　配送ギヨウムの合理化お　テツテイするため　FACOM230-25お　増設
52) 70/01/10(08)D.00   ビジネスコンサルタント　8月　社内研修用に　IBM1130お　導入
```

図について説明を加えると

～～～～は，漢字テーブルにないためカタカナ出力された単語である。

ソウケン → （総研）

ギョウム → （業務）

――― は，セグメンテーション（区切り）が悪かったために起きたデータ・エラーである。

═══ は，該当する用語がテーブルにあるにもかかわらずセグメンテーションがまずかったためにカタカナ出力となったデータ・エラーである。

10) のサクニは漢字ファンクションキー「＊」の位置を間違って「＊サク＊ニ」を「＊サクニ＊」としたためのエラーである。正しくは「策に」と変換する。

[　] は，テーブルが完全に作られていないために起きたエラーで，誤変換されたものである。これは同音異義語の取扱いが不十分なためによるもので複数対応にしなければならず，テーブルの修正が必要である。

シン：新 → 真

カクリツ：確立 → 確率

デンキ：電気 →（電器，電機）

△ は，漢テレパンチェラーのまま，テーブルに入っていたものが出力されたもので，当然訂正されなければならない。

以上のようなカタカナ漢字変換から変換率を調べてみると，５００件のデータについて次のような結果が得られた。

データ件数	５００件	
漢字変換必要数	2,７４３件	
正変換数	2,０２９件	（７４.０％）
単数対応	1,８５２件	（６７.５％）
複数対応	１７７件	（　６.５％）
誤変換数	１１２件	（　４.１％）
単数対応	７３件	（　２.７％）
複数対応	３９件	（　１.４％）
漢字テーブルにないカタカナ出力	６０２件	（２１.９％）

複数対応とは，カッコ（，）に囲まれた，いわゆる同音異義語のある漢字コードであり，２次以後の修正作業によって正しい漢字に直す。

１１.３　漢字処理システムの今後の方向

コンピュータで漢字かな混り文を処理しようとする場合の一例を述べてきたが，さらに本格的にコンピュータで漢字かな混り文を処理しようとするときは，以下のような点を研究課題にする必要がある。

11.3.1 漢字インプット方式

今回の実験はＭＡＳＣＯＴのサブシステムとしてカナ文字のデータを漢字かな混り文に変換する方法を進めてきたが，始めから漢字データを入力できる場合には漢字テレタイプでインプットした方が処理効率がよいと思われる。

実際，現状の漢字テレタイプでは，１分間に７０～９０字のスピードで穿孔することができる。これは，カナ文字換算すると１４０～１８０字／分にあたり，通常のタイプライタが１分間に２００字前後であることから，現状でも漢字穿孔は決して不利ではない。また漢字インプットもいろいろ開発されてきており１００～１５０字／分位のスピードをもつものの開発も時間の問題であろう。

さらに，情報量を比較してみると，漢字かな混り文とカナ文字文では，同じ情報量に対し記憶場所を占める割合がだいたい１：２であり，コンピュータを使用した場合には，メモリ，その他の点で現在のシステムとほとんど同じ位の能率で処理できるものと思われる。

費用についていえば，カナ文字が１字３０銭～３５銭，これに対して漢字は８０銭前後であるから先ほどの１：２の比率を適用すると６０銭：８０銭となる。

漢字をカナ文字に変換する手間，校正時の読みずらさ，アウトプットの判別の不明瞭さを考えれば，コンピュータによる漢字処理がシステムにおいて決して高価でなく，時間的にも不利ではないことが判る。

なお，システムの構成上どうしても漢字かな混り文とカナ文字の両方をファイル上に所有したい場合には，漢字かな混り文でインプットし，このデータをカナ文字変換したほうが費用的にも，時間的にも効果的であると思われる。この方法をとれば，入力時に，日本語をカナにわざわざ変換する手間が必要でなくなり，（実際のシステム運用上では，この作業が一番大変である）校正など人手のかかるところが大幅に削減される。

11.3.2 マイクロフィルムとの結合

コンピュータで漢字情報を扱う場合問題になるのは，その情報の価値であろう。実際，コンピュータによる情報処理は，インプット，大量ファイルの蓄積，検索，アウトプット処理など相当に高価なものとなろう。そこでマイクロフィルムの活用がシステム運用上クローズアップされる。

しかし，すべてがマイクロフィルム＝ＣＯＭシステムに適しているものではない。マイクロフィルムは，

1) 磁気テープや磁気ディスクでは，記録された情報を眼でみることは，できない。また，図や写真，設計図をコンピュータで直接扱えない，などの欠点を補う。
2) 情報を検索したり，また一度取り出した情報を元に戻したりするのに特別の機械が必要でないので，検索コストがかからない。
3) 情報記録のたゆのフィルム費用は低廉である。

などの特性を生かしたものに利用したほうがよい。

また，「特許情報サービス」「技術情報サービス」などの文献検索システムは，

(1) 検索インデックスの多様性

(2) ファイリングの複雑性

(3) 検索の迅速性，頻度性

(4) アウトプット方法の多様性

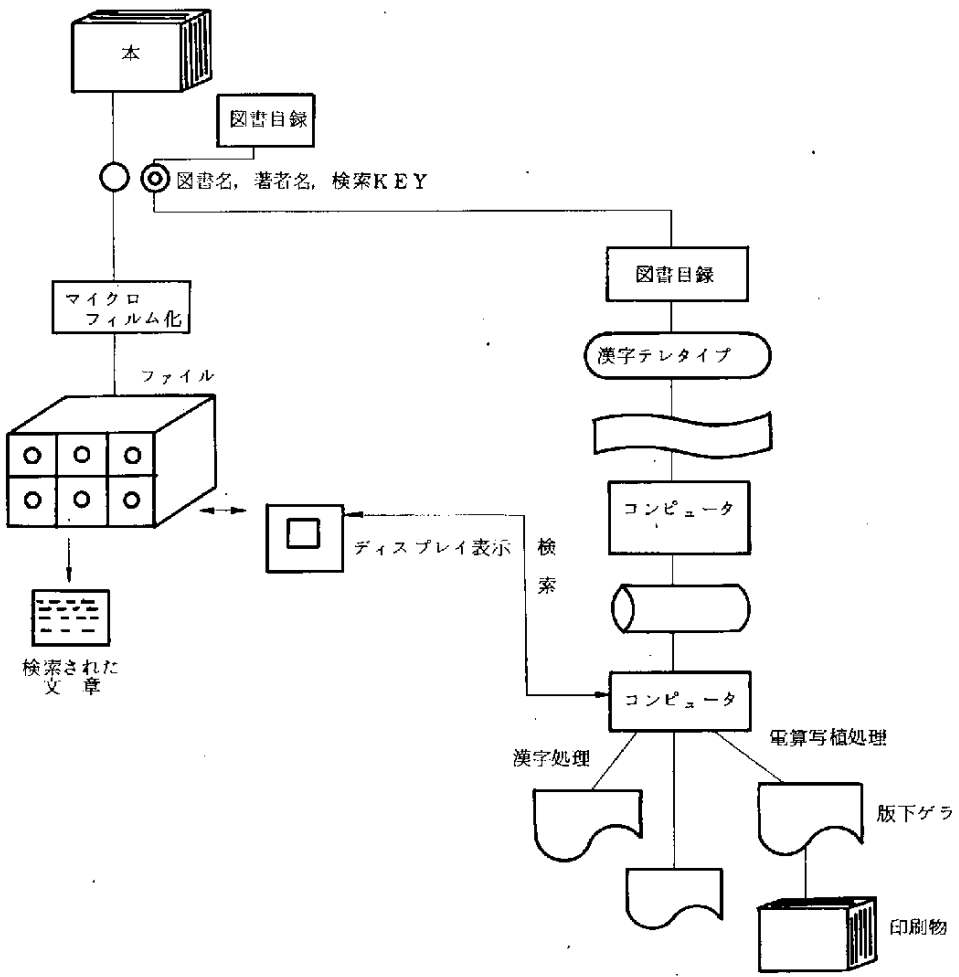
(5) 追加，改定の処理

(6) 出版物へ結びつき

などの条件からマイクロフィルムとコンピュータとの連繋をもたせて処理する方法が最も有効であると思われる。

さらに，「図書データ」のようなものでも，図書そのものは，マイクロフィルム化し，マイクロフィルムの特性を生かすし，図書目録は，コンピュータにインプットし，管理面でその特性を生かすことができる。

図１１.3.1　マイクロフィルムを使用した図書管理



いずれにしても，今後の情報処理において大量のデータを，安価に，迅速に管理するためには，コンピュータとマイクロフィルムを有効に結びつけたシステムの開発が必要であると思われる。

コンピュータによる漢字かな混り文の処理は，いわば「データ・バンク」における処理の一部であろうと思われる。「データ・バンク」といわれるもののアウトプットは，必ずそのデータを使用した印刷物の刊行という処理に結びつく。印刷物を作成する場合，コンピュータでアウトプットした漢字かな混り情報を，さらに活字をひろって版下を作成していたのでは意味がない。そこでコンピュータと結びついた電算写植編集システムが必要になる。

この電算写植編集システムを利用するためには，あらかじめコンピュータによって自動編集処理をする必要がある。

自動編集処理は

○ 入力処理

ファンクション処理

スペース・コード処理

キャラクタピッチ付加

○ 組版処理

ページ割付処理

ノンブル柱処理

○ 棒組調整処理

ジャスティフィケーション・禁則・分離禁止・行頭・行末・中央揃え・ダブ・均等配分

ルビ・数式・合成・ハイフォネーション

○ 図版処理

図版・写真・見出しなどの割付

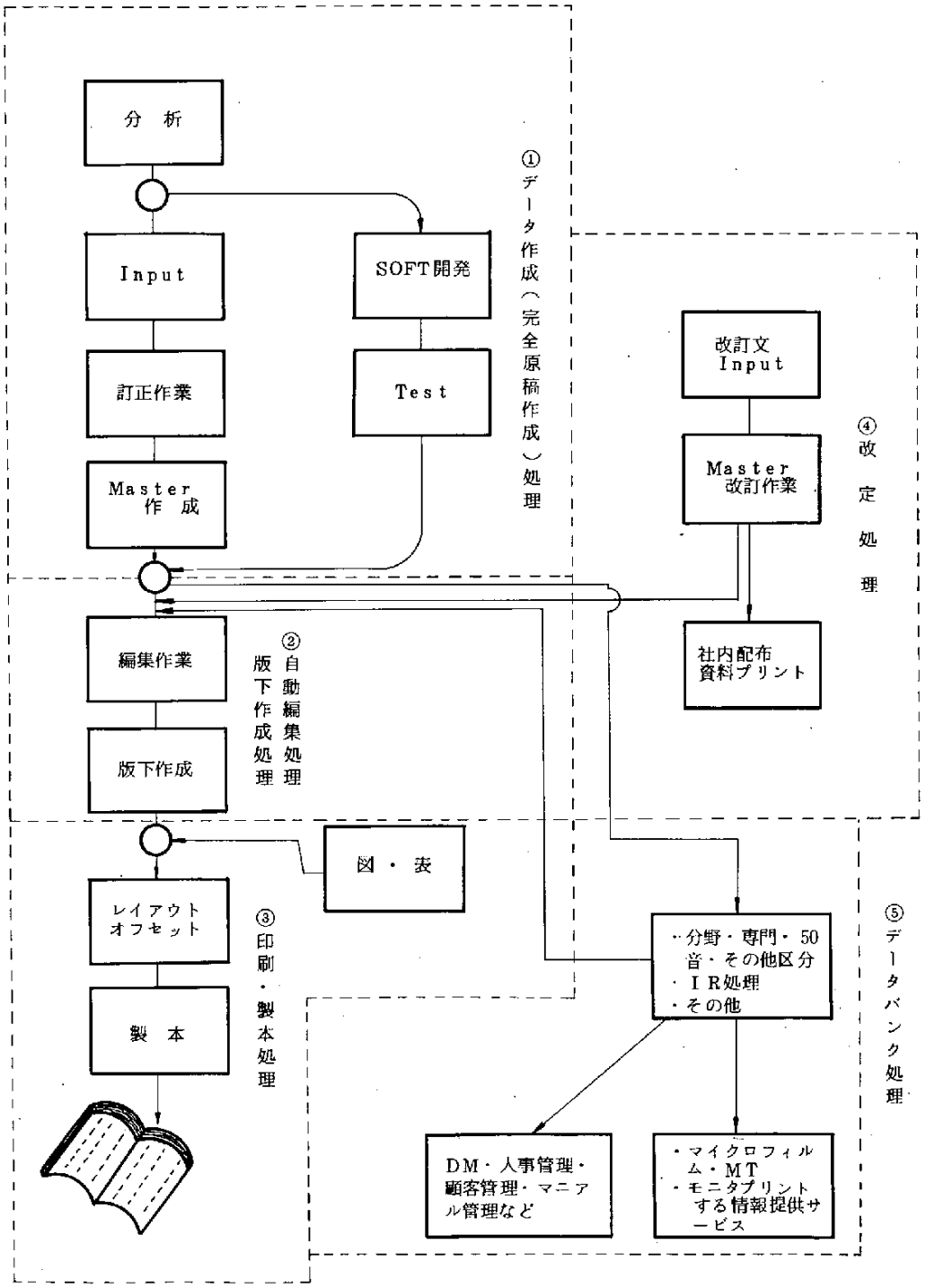
スペース処理

などからなり，ソフトウェアとしてプログラム化したものである。

このカナ文字→漢字かな混り文変換システムで使用された，ＪＥＭ－３８５０は，これらの機能を一部に含んでいる。

電算写植編集システムは，コンピュータより磁気テープでアウトプットしたデータを高速で印刷物の版下として編集するものである。このように「データ・バンク」の情報処理は，さらにコンピュータを媒体として広がって行き，漢字かな混り文への処理がより広い分野に活用されて行くことであろう。

図１１.３.２　自動編集処理システムの概要



昭和47年3月発行

発行所　財団法人　日本情報処理開発センター
東京都港区芝公園3丁目5番8号
機械振興会館内
TEL（434）8211（代表）

印刷所　三協印刷株式会社
東京都渋谷区渋谷3－11－11
TEL（407）7316

46－S006