データベース構築促進及び技術開発に関する報告書

英日キーワード変換機能をもつ データベース検索システムの開発

平成6年3月

財団法人 データベース振興センター 委 託 先 カ テ ナ 株 式 会 社



KEIRIN O

この事業は、競輪の補助金を受けて実施したものです。



		!
		e l
·		
		N.
		•

データベースは、わが国の情報化の進化上、重要な役割を果たすものと期待されている。今後、データベースの普及により、わが国において健全な高度情報化社会の形成が期待される。さらに海外に対して提供可能なデータベースの整備は、国際的な情報化への貢献および自由な情報流通の確保の観点からも必要である。しかしながら、現在わが国で流通しているデータベースの中でわが国独自のものは1/3にすぎないのが現状であり、わが国独自のデータベースの構築およびデータベース関連技術の研究開発を協力に促進し、データベースの拡充を図る必要がある。

このような要請に応えるため、(助データベース振興センターでは日本自転車振興会から機械工業振 興資金の交付を受けて、データベースの構築および技術開発について民間企業、団体等に対して委託 事業を実施している。委託事業の内容は、社会的、経済的、国際的に重要で、また地域および産業の 発展に寄与すると考えられているデータベースの構築とデータベース作成の効率化、流通の促進、利 用の円滑化・容易化などに関係したソフトウェア技術・ハードウェア技術である。

本事業の推進に当たって、当財団に学識経験者の方々で構成されるデータベース構築・技術開発促進委員会(委員長 山梨学院大学教授 蓼沼良一氏)を設置している。

この「英日キーワード変換機能をもつデータベース検索システムの開発」は平成5年度のデータベースの構築促進および技術開発促進事業として、当財団がカテナ株式会社に対して委託実施した課題の一つである。この成果が、データベースに興味をお持ちの方々や諸分野の皆様方のお役にたてば幸いである。

なお、平成5年度データベースの構築促進および技術開発促進事業で実施した課題は次の表のとおりである。

平成6年3月

財団法人 データベース振興センター

平成5年度 データベース構築・技術開発促進委託課題一覧

分 野		課 題 名	委 託 先
	1	CD-ROM によるテレビ視聴率データベースの構築	(料ビデオ・リサーチ
	2	輸入畜肉貨物の規格・重量等の検証用データベースの構築	五十嵐冷蔵㈱
	3	知的資源型データベースの調査研究	(株)ジャパンコミュニケーションズ
•			インスティテュート
社 会	4	ビジネス雑誌記事データベースの共同構築とその利用に関	経済文献研究会
,		する調査研究	
	5	新しい電子情報サービスに関する調査研究	㈱日本経済新聞社/㈱日経データ社
	6	研修用教材データベースのプロトタイプ作成	働大阪科学技術センター
	7	マイクロマシン技術情報データベースの構築調査	(財)マイクロマシンセンター
	8	異分野研究のための知的オリエンテーション・データベー	(株)けいはんな
		スシステムの構築	
	9	関西広域データベースセンター設立のための調査研究	関西データベース協議会
中小企業振興	10	地域活性化のための産・学交流支援データベースのプロト	東北インテリジェント・コスモス
地域活性化		タイプ作成	学術機構
	11	中小企業技術情報データベースの構築	(株)オーネット
	12	地域情報を対象にした分散協調型データベースシステムの	(株)エマーズ
		開発	
	13	電子デバイス情報の海外提供サービスに関する調査	電子デバイス情報サービス㈱
	14	英日キーワード変換機能をもつデータベース検索システム	カテナ(株)
海外		の開発	
	15	CD-ROM による 5 カ国対訳特許用語辞典及び関連諸制度	丸善㈱
		一覧の構築	
	16	人体形状画像データ合成のための技術開発	(組)人間生活工学研究センター
	17	OCRを利用したキーワード自動抽出に関する調査研究	(株)エレクトロニック・ライブラリー
	18	既存画像データのフォーマット変換システムのプロトタイ	(株)ジー・サーチ
		プ作成	
 技 術	19	安全研究における多重シソーラス・システム構築のための	㈱紀伊國屋書店
		基本安全用語データベースの開発	
	20	データベース検索サポートシステムのプロトタイプ作成	セントラル開発㈱情報図書館
			RUKIT
	21	グループワーク支援のための分散型トランザクション管理	㈱新世代システムセンター
		方式の調査研究	

目 次

1.	熴	安		1
	1.	1	目的	1
	1.	2	英日シソーラスの考え方	2
	1.	3	プロトタイプ・システムの機能範囲	3
	1.	4	稼働環境·····	5
2.	関連	シン	マテムの動向	6
	2.	1.	ディストリビュータが用意しているシステム	6
	2.	2	端末用ソフトウェア	2 0
	2.	3	機械翻訳ソフトウェア	3 2
3.	英日]シ:	ノーラス	3 7
	3.	1	アルファベット順リスト(シソーラス本体)	3 7
	3.	2	階層リスト	3 9
	3.	3	KWICリスト·····	3 9
	3.	4	シソーラスのアクセス	4 0
4.	プロ	1 1 2	タイプ・システム	4 8
	4.	1	システムの構成	4 8
	4.	2	入出力制御機能	4 9
	4.	3	検索キーワード作成機能体系	5 1
	4.	4	データベース検索支援機能	6 0
	4.	5	検索コマンド生成機能	8 0
	4.	6	通信制御機能	8 3
5.	実月	目シス	ステムに向けての課題	8 6
	5.	1	外的環境に関する課題・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	8 6
	5.	2	システム的な課題	8 7
	5.	3	内部シソーラスについての課題	9 0
	5.	4	プロトタイプ・システムの評価	9 0
6.	おオ	っりし	٠	9 2
	6.	1	データベース普及に関する提言	9 3
	6.	2	海外運用システム調査研究委員会	9 5



1. 概 要

1. 1 目 的

現在、わが国は、経済力を背景に、多くの分野においての国際協力が叫ばれている。そのため、多くの国々との平和的、文化的な交流が不可欠であり、正確で速やかな情報の交換が必要となっている。 国際間での情報伝達手段は、過去においても、電信、電話、衛星通信、等の様々な手段が利用されてきた。

本課題で着目する「オンライン・データベース・サービス」も情報伝達手段の一つであり、つぎのような利点から、現在、拡大傾向にある産業といえるだろう。

- ① 学術、医療、ビジネス、等の多くの分野の情報をユーザが検索しやすいように体系的に保管している。
- ② ユーザの検索しやすいような機能体系を整備している。
- ③ 知的所有権等の無形資産の価値が評価される時代になってきている。

しかし、反面、いろいろな、問題点もあがってきている。国内だけの問題点としては、

- ① コマンドが不統一である。
- ② 統一シソーラスが無い。
- ③ 自然言語を検索言語にして欲しい。
- ④ 検索手順が煩わしい。
- ⑤ オンライン・データベース・サービス産業の規模が小さい。

があり、他国からのデータベースの利用となると、国内的な問題にまして、つぎのような問題があげられる。

- ① 使用言語が違うため使いにくい。
- ② 国産データベースのシソーラスの分類方法が海外ユーザの利用法と整合しない。
- ③ 海外の端末は、日本語対応のものが少ない。

以上のように、海外からわが国のデータベースを利用する場合、問題点が多く、わが国のオンライン・データベース・サービス産業を育成する上での阻害要因となっている。

そこで、本課題では、英日キーワード変換機能を持ったデータベース検索システムを海外製のパー ソナル・コンピュータ上で作成し、国外のユーザの評価を受けることを目的としている。

そして国外の評価が、わが国のオンライン・データベース・サービス産業を普及・育成する上での、

問題提起となることが期待できる。

1.2 英日シソーラスの考え方

本課題では、既存の英日シソーラスを使用する。そのシソーラスのつぎの方針で作成されている。

- ① 日本語のシソーラス(日経シソーラスのビジネス分野)から作成されている。
 - 語の意味はシソーラス内での語の意味を分析して定めている。
- ② 英語ディスクリプタとして適切な用語を選択している。
 - 原則として実際使用している語であること。
 - 既存の英語シソーラス中のディスクリプタを優先する。
 - ・ 英国式綴り、複数形とする。
- ③ 日本語ディスクリプタと英語ディスクリプタを1対1で対応させている。
 - 英語ディスクリプタと同義語全体で重複をさける。

このような方針で、

- ① アルファベット順リスト
- ② 階層構造リスト
- ③ KWIC リスト

の3つの構造が用意されている。そして、それぞれ構造毎に用途が想定されている。

(1) アルファベット順リスト

シソーラスの基本構造で見出し語をキーにアルファベット順にソートしてある。見出し語には検索に使用できる優先語(preferred terms)と、優先語の同義語や類義語で検索には使用できない非優先語(non-preferred terms)とがある。

見出し語が優先語か非優先語かによって、中にある項目の種類が違ってくる。

- ① 見出し語が優先語のときの項目
 - 見出し語
 - スコーブノート 利用法や注釈
 - UF (Use For) 同義語、類義語になっている非優先語
 - NT 下位語
 - BT 上位語

- RT 関連語
- ② 見出し語が非優先語のときの項目
 - 見出し語
 - USE 検索に使用すべき優先語

(2) 階層構造

用語間の階層関係を表現したリストで、優先語だけを対象としている。階層はつぎのように表現される。

A

- A 1
- • A 1 1
- • A 1 2
- A 2
- • A 2 1

В

• B 1

「・」は、階層の深度を示している。「・」が多ければ多いほど下位概念の用語でより特定的な表現をする。階層を図示すると図1-1のようになる。

なお、同一階層の順番は、アルファベット順になっている。

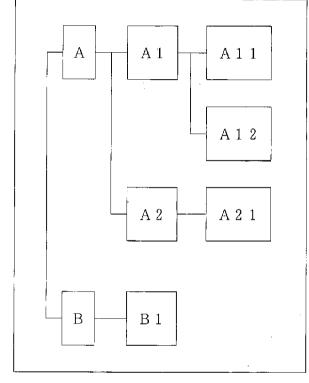


図1-1 階層構造

(3) **KWICリスト**

1つの単語から、その単語を構成要素にしているすべての用語を引くためのリストである。

優先語は、大文字で表現され、非優先語は小文字で表現されている。構成要素の単語は、アルファベット順に用語と一緒に並んでいる。

1. 3 プロトタイプ・システムの機能範囲

本プロトタイプ・システムは、海外から国産データベースを利用するにあたって機能を限定した、パイロット・システムの立場をとる。開発にあたって、つぎの3点を前提とする。

① 国産データベース・ディストリビュータは、現状のまま利用し、開発は、端末の操作性の向上

を目指す。

- ② 海外ユーザの使用言語の対象を英語に限定する。
- ③ 検索対象の国産データベースは、新聞記事情報(テキストデータ)のみを扱ったものにする。

従って、開発する部分は端末側のみで、端末に用意する機能と内容は、つぎのとおりである。

(1) 入出力制御機能

本プロトタイプ・システム操作全般を制御する。メニュー、ウィンドウ、ダイアログボックスなどの主に Windows 廻りの制御をする。

本プロトタイプ・システムでは、データベースを検索するときに自然語の検索質問文入力を許す関係上、入力した検索質問文から検索キーワードとして使用できる用語を切り出す機能が必要となる。 本機能は、検索質問文中に含まれている優先語及び非優先語を抽出する。

(2) 検索キーワード作成機能体系

データベースを検索する際に使用する、検索キーワードを作成するためのツール群でつぎのものが ある。

- ① 追加
- ② 修正
- ③ 削除
- ④ 全削除
- ⑤ アルファベット検索
- ⑥ 階層検索
- ⑦ 単語検索
- ⑧ 関連語検索
- ⑨ 詳細表示
- ① 日本語表示

(3) 二次検索支援機能

本機能は、検索キーワードのみの検索(一次検索)を行った結果として得た文献集合に別の検索条件を与えることによって該当件数を減らしていく操作の過程を容易にするためのものである。

この操作には、理論演算で行うものと検索条件を与えるものとがある。

① 論理演算での操作

一次検索を行った結果として得られた複数の文献集合に対して、集合演算を行い新たな文献集合を 得るためのコマンドを作成する。

② 条件入力による操作

一次検索を行った結果として得られた文献集合の一つに対して、別の検索条件を入力して絞り込 みを行うためのコマンドを作成する。

(4) 検索コマンド生成機能

既に入力された一次検索(主にキーワード検索)と二次検索(論理演算、絞り込み検索等)の条件 を会話しているデータベースのコマンドに変換する。

(5) 通信制御機能

生成された検索コマンドを会話しているデータベースに送信するとともに応答を受信ファイルに格納する。送受信の手順は無手順で、送信条件(送信速度や漢字の対応等)はあらかじめ設定したもので行う。

(6) 回答データ編集機能

受信したデータベースの応答データを解読可能な形式に変換する。

1. 4 稼働環境

本課題で作成するプロトタイプ・システムは、国外のデータベース・ユーザを対象とした端末ソフトウェアである関係上、国際標準のハードウェアと基本 OS の構成であることが望ましい。

現在、世界のパーソナル・コンピュータを大別するとマッキントッシュ系と DOS/V 計の二系統がある。今回は、採用しているメーカーの多い DOS/V 系のマシンを採用する。

基本 OS は、DOS/V と MS-WINDOWS との組み合わせにして、オープン化を目指した稼働環境にする。

2. 関連システムの動向

国外のデータベースの利用を考えるとき、二つの点に留意しなければならない、一つは、簡単に検索することができること、もう一つは、言語の違いについてのサポートができることである。簡単に検索することができるという意味では、各ディストリビュータが独自の開発を行い、色々な形態がある。言語の違いについてのサポートは、ホスト・コンピュータ側で英日翻訳機能をもつものがあるが、検索する前にキーワードを変換する機能をもつものはない。現在、ホスト側、端末機側、その他のソフトウェアで、国外のデータベースを検索するときに有効だと思われるものの動向について、次に述べる。

2. 1 ディストリビュータが用意しているシステム

データベースの多くをみると、操作形態が徐々に変化してきている。会話型のメニューやメニュー 形式をもつ端末通信ソフトウェアが出てきたことは、確かに過去よりも使いやすくなっていると言え よう。

わが国のデータベースのサービスが開始されたのは、一番古いものでも昭和50年代の前半である。当時は、バッチ処理が主体でオンライン処理は、発展途上の時代だった。このような、コンピュータ事情の中で、構築されたデータベースは、学術情報、法令情報、特許情報等の文献データがデータ・ソースの主体となった。また、データベースには、文献や概要や抄録のみが格納されているリファレンス・データベースがほとんどで、文献の全体を格納しているファクト・データベースは稀であった。現在サービスされているデータベースの大部分は、従来方式の拡張形であり、検索の仕組みや、データの格納形態は、ほとんど変わりがない。どのように進歩してきたかと言うと、データ件数がふえてきたのは当然のこととして、統制語キーワードの辞書であるシソーラスの整備がある。シソーラスは、統制語キーワードを分野別に階層的に分類したものである。これに対する用語の追加、新たな分野の追加、再編成等の整備が行われている。他には、データを蓄積するときに付与するキーワードを始めとする検索項目の量が多くなってきたことが挙げられる。データに付与するキーワードは、シソーラスに掲載されている統制語キーワードだけでなく自由に付与する自然語キーワード(フリーターム)があり、この数が多ければ多いほど、ユーザがデータを検索するときのヒット率が高まる。

データベースの高度化には、前に述べたようにデータのメインテナンスとシソーラスのメインテナンスによりデータベース自体を成長させるアプローチだけでなくデータ検索の便宜を図るために操作方法を変更するアプローチがある。

操作方法の変更の根幹にあるものは、「コマンドからメニューへ」という動きにある。データ検索に熟達したユーザは、下手なメニューよりもコマンドの方が使いやすいと言うが、多くのユーザは、コマンド入力を嫌がる。

コマンド入力で操作するためには、データ検索の知識が必要であり、次に示す文献集合の場合の検 索の一例のように、自分の検索した結果がどのような状態にあるかを常に認識している必要がある。

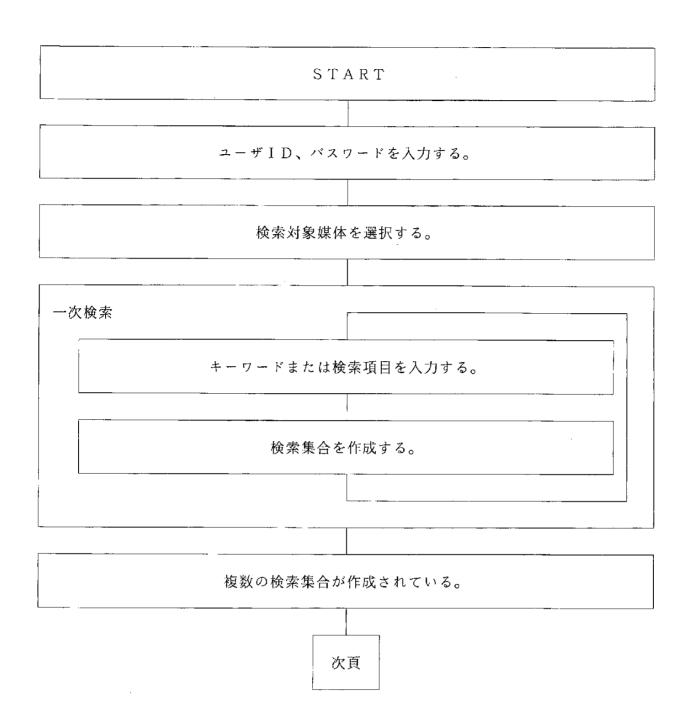


図 2-1 検索手続きの流れ(1/3)

検索集合の一本化 複数の検索集合に対するAND、OR、NOTのような論理演算を行う。 一件の検索対象の検索集合が作成されている。 二次検索 検索項目を入力する。 新たな検索対象の検索集合を作成する。 最終的な検索集合が作成されている。 次頁

前頁

図2-1 検索手続きの流れ(2/3)

表示、印刷などの出力処理

最終的な検索集合内のタイトルの一覧を表示する。

表示または印刷したい文献のタイトルを選択する。

選択した文献を表示または印刷する。

ログアウト

END

前頁

図2-1 検索手続きの流れ(3/3)

検索処理の過程は、以上のような形態が一般的である。従来のデータベース・サービスは、この処 理過程の全でがコマンド入力であった。現在は、メニュー選択方式が採用されていて、過去にコマン ド入力であったデータベースに関しては、コマンド入力方式に切り替えることもでき、熟達したデー タベース・ユーザは、この方法が使いやすい。

採用されているメニュー選択方式には、ホスト・コンピュータ側で制御する方式と端末側で制御する方式がある。

(1) ホスト・コンピュータ側で制御する方式

ユーザが端末機からコマンドを入力すると、その回答が画面に表示されたり、プリンタに出力されたりする。この操作の繰り返しがデータ検索である。

この方式は、ユーザが端末機を介してホスト・コンピュータと直接会話するという点において、コマンド入力方式と同じではあるが、コマンド入力は、ユーザの主体性における操作であり、メニュー選択方式は、ホスト・コンピュータがユーザを誘導していく操作である。例えば、新聞記事を検索する場合は、次のようになる。

① 検索対象媒体の選択

ア. コマンド入力の場合

>>MEDIA=AP

※「>>」は、コマンド入力プロンプト

図2-2 検索対象媒体の選択コマンド入力方式

イ、メニュー選択方式の場合

検索する新聞を選択して下さい。

- 1. A新聞(AP)
- 2. B新聞(BP)
- 3. C新聞(CP)
- 4. D新聞(DP)

BYE

 $\gg 1$

※「BYE」は、終了コマンド

図2-3 検索対象媒体の選択メニュー選択方式

- ② 検索項目の選択
 - ア. コマンド入力方式の場合 人力の必要なし。
 - イ、メニュー選択方式の場合

検索項目を選択して下さい。

- 1. 漢字キーワード
- 2. カナキーワード・
- 3. 掲載日付
- 4. 媒体名称
- 5. 掲載ページ

MAIN/BYE

>>1

漢字キーワードを入力して下さい。

>>

※ MAINは、メイン・メニューに戻る。

図2-4 検索項目の選択メニュー選択方式

- ③ 漢字キーワードの入力
 - ア. コマンド入力方式の場合

>>KW=漢字キーワード

図2-5 漢字キーワードの入力コマンド入力方式

イ.メニュー選択方式の場合

検索項目を選択して下さい。

- 1. 漢字キーワード
- 2. カナキーワード
- 3. 掲載日付
- 4. 媒体名称
- 5. 掲載ページ

MAIN/BYE

 $\gg 1$

漢字キーワードを入力して下さい。

>>漢字キーワード

検索集合/1 4件が作成されました。

内容を表示しますか? (Y/N)

>> N

検索項目を選択して下さい。

- 1. 漢字キーワード
- 2. カナキーワード
- 3. 掲載日付
- 4. 媒体名称
- 5. 掲載ページ

MAIN/BYE

>>

図2-6 漢字キーワードの入力メニュー選択方式

④ 検索集合「/1」(4件)のタイトル表示

ア、コマンド入力方式の場合

>>SHOW P1 /1

※ SHOWは、表示コマンド。P1はタイトル表示パラメータ。

図2-7 検索集合「/1」のタイトル表示コマンド入力方式

イ. メニュー選択方式の場合

検索集合 /1 8件が作成されました。

内容を表示しますか? (Y/N)

>> Y

表示形式を選択して下さい

- 1. タイトル
- 2. 概要
- 3. 内容全体

MAIN/FMENU/BYE

>>1

内容表示 検索集合:/1

- 1. 文献1 YY/MM/DD n文字
- 2. 文献 2 YY/MM/DD n文字
- 3. 文献3 YY/MM/DD n文字
- 4. 文献 4 YY/MM/DD n文字

MAIN/FMENU/BYE

>>

※ FMENUは、一つ前のメニュー。

図2-8 検索集合「/1」のタイトル表示メニュー選択方式

⑤ 検索集合「/1」(4件) 文献番号「1」の内容表示

ア. コマンド入力方式の場合

>>s HOW	P1 /1
内容表示 検	素集合:/1
1. 文献1	YY/MM/DD n文字
2. 文献 2	YY/MM/DD n文字
3. 文献3	YY/MM/DD n文字
4. 文献 4	Y Y / M M / D D n 文字
>>SHOW	P3 /1
内容表示。検	素集合 : / 1
1. 文献 1	YY/MM/DD n文字
	文 献 1 の 内 容
>>>	※ P3は、内容を表示するパラメータ。

図2-9 検索集合「/1」文献番号「1」の内容表示コマンド入力方式

イ. メニュー選択方式の場合

内容表示 検索集合:/1

1. 文献1 YY/MM/DD n文字 2. 文献 2 YY/MM/DD n文字 3. 文献 3 YY/MM/DD n文字 4. 文献 4 YY/MM/DD n文字 MAIN/FMENU/BYE $\gg 1$ 表示形式を選択して下さい 1. タイトル 2. 概要 3. 内容全体 MAIN/FMENU/BYE >>3 内容表示 検索集合:/1 1. 文献1 YY/MM/DD n文字 文献 1 の内容 表示形式を選択して下さい 1. タイトル 2. 概要 3. 内容全体 >> 1

図 2-10 検索集合「/1」文献番号「1」の内容表示メニュー選択方式

以上のように、コマンド入力方式は、データベース独自のコマンドやそのパラメータ、コード体系を覚えていれば最短距離の操作ができる。しかし、データ検索は、1ステップの操作毎に戻ってくる回答をユーザが見て判断することに神経を使うこともあって、コマンドは、できるだけ避けたメニュー選択方式を用いた方が広いユーザ層を見込めるし、データベースが、より身近になることが想定される。

(2) 端末側で制御する方式

ホスト・コンピュータ側でのメニュー制御は、従来からある会話型のコマンド入力方式の拡張形であるため、画面がスクロールしてしまう。端末側で制御するメニュー選択方式は、端末自体に通信機能、画面制御機能、コマンド生成機能を備えたソフトウェアを有償で添付するディストリビュータが登場し始めている。これを利用すると、コマンド入力が完全に無くなり、処理は画面単位になるためスクロールすることもなくなる。

回線接続後のデータ検索操作の画面例を次に示す。

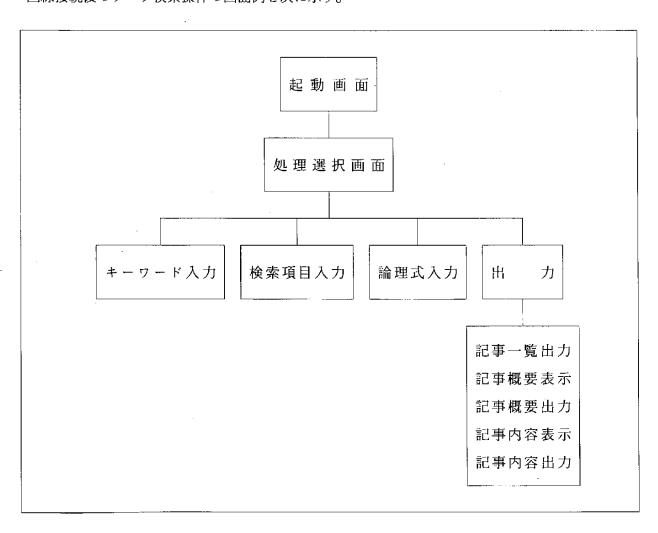


図2-11 回線接続後のデータ検索操作の画面例(1/7)

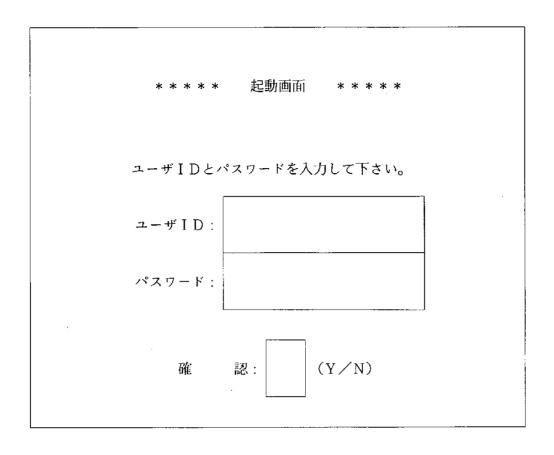


図2-11 回線接続後のデータ検索操作の画面例(2/7)

	処理選択画面 ****
検索集合一覧	
1. 検索集合1	2,500件
2. 検索集合 2	7 0 0 4
3. 検索集合3	1,2004
2面かえび松悪焦なた選	切してでナレーン
必要ならば検索集合を選	択して下さい=>
必要ならば検索集合を選 1. キーワード	択して下さい=> 9.終 了
1. キーワード	

図2-11 回線接続後のデータ検索操作の画面例(3/7)

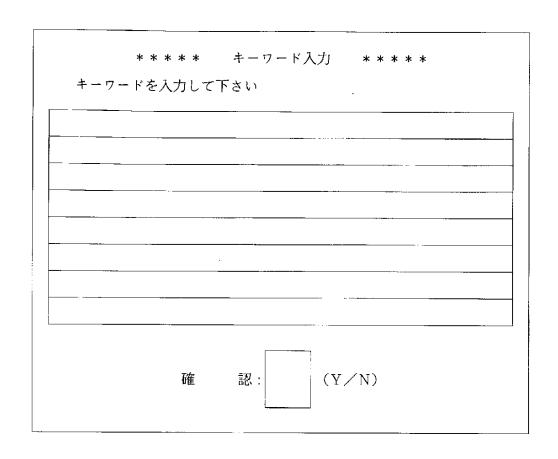


図2-11 回線接続後のデータ検索操作の画面例 (4/7)

* * * * * 検索項目入力 * * * * *	r
掲載日付 19 年 月 日 ~	
19 年 月 日	
掲載ページ 掲載媒体	
確 認: (Y/N)	

図2-11 回線接続後のデータ検索操作の画面例(5/7)

٠ ٨٠	+ 4+ 4	•	* * * *	1/100-11-47 (7)	ጥ ጥ ጥ ጥ
筷?	索集合一覧			•	
1.	検索集合1				2,500件
2.	検索集合 2				7 0 0 件
3.	検索集合 3	,			1,200件
論I	里式を入力し	,て ⁻	下さい。		
論耳 1	里式を入力し AND		· · · · · · · · · · · · · · · · · · ·	3	

図2-11 回線接続後のデータ検索操作の画面例(6/7)

		···
* >	**** 出 力 ****	
検索集合名		
記事一覧		
	·	
記事を選択して下さい	=>	
1. 記事一覧出力	4. 記事内容表示	
2. 記事概要表示	5. 記事内容出力	
3. 記事概要出力	処理を選択して下さい=>	

図2-11 回線接続後のデータ検索操作の画面例(7/7)

2. 2 端末ソフトウェア

データベース・サービスのメニュー化は、検索のノウハウに係る問題の解決までには至らないが、経験の浅いユーザでも簡単なデータ検索ができるようになった。このことは、ユーザ層を広げる意味でも有効なことである。しかしながら、本課題で取り上げているような「多言語間検索」についてのサポートは、成されていない。カテナ株式会社の端末用ソフトウェア「検索 Helper」は、本課題のように海外からわが国のデータベースを検索するのではなく、反対に、わが国からの英語圏のデータベースを検索することを意識して作成されたシステムである。このシステムを調査することにより、今後のデータベースの「多言語間検索」のアプローチの仕方、用意すべき機能、端末内部辞書のあり方等を考察する上での参考にすることができる。

「検索 Helper」は、本課題で作成したプロトタイプ・システムと同じように「データベースの操作性の向上」と「多言語間検索の対応するツール」を目指しているが、日本語を使うユーザが英語圏のデータベースを検索するという点で若干用意するツールの考え方が違ってくる。

まず、「検索 Helper」の次のような基本的な特徴を一つ一つ検証し、英語圏のユーザーが日本のデータベースを検索する際に、必要になりうる機能を考察した。

- ① MS-Windows のアプリケーション
- ② 単語切り出し機能
- ③ シソーラス照会機能
- ④ 関連語生成機能
- ⑤ 英日、日英キーワード返還機能
- ⑥ 検索機能体系
- ⑦ 英日翻訳機能
- ⑧ 通信機能

(1) MS-Windows のアプリケーション

低価格のパーソナル・コンピュータを端末機にする場合、単にデータベースの端末機としてでしか使用しないということは、四六時中データベースを利用していない限り、まずない。当然、別のアプリケーションを使用する要望が出てくる。アプリケーションをバージョン・アップしながら永続的に使用するためには、将来の基幹 OS を採用することが重要になってくる。その点、DOS/V と MS-Windows の組み合わせは、アップル社を除く大部分のメーカのパーソナル・コンピュータ上で稼働する。現時点での勢いを考えたら、採用するのが自然だと思われる。もちろんアップル社のマッキントッシュ・シリーズもその操作性を見たら捨て難いものがある。しかし、今回は、データの互換性やMS-DOS のテキスト・データの見やすさを考慮にいれて、DOS/V と MS-Windows の組み合わせを採用した。

MS-Windows は、ディスプレイ上に窓(ウィンドウ)を開いて、アプリケーションを走らせる OS である。MS-Windows そのものは、MS-DOS(DOS/V)上で動作している。MS-Windows を DOS / V 上で稼働し、その上でアプリケーションが稼働させるということは、単純に DOS/V 上で稼働 させるときと、次のような違いがある。

ア. 操作環境の向上

DOS/Vの環境では、処理の実行やプログラムの起動するためには、プログラムが存在するデバイスを意識しながらキーボードからコマンドを入力しなければならない。キーボード中心のこのような環境では、ファイル名やコマンド名などのスペルを覚えなければならず、特に初心者にとっては少々扱いにくい環境となっている。

それに対して、MS-Windows の環境では、プログラムの起動やその中での処理の実行は、マウスの操作が中心になる。コマンドの代わりに絵やメニューをクリックするだけで、処理が起動する。

また別のアプリケーションも操作が似ているので、一つのアプリケーションの操作を覚えれば、他のアプリケーションの操作も比較的覚えやすい。

イ、複数アプリケーションの同時実行

同時に一つの仕事しかできないことを「シングルタスク」という。DOS/V は、「シングルタスク」である。したがって、ある仕事を実行している最中に他の仕事をしなければならなくなったときには、最初の仕事を終了してからでなければ、他の仕事を実行することができない。

それに対して、複数の仕事を同時に行えることを「マルチタスク」という。MS-Windows では、 一つのウィンドウで一つのアプリケーションが実行でき、複数のウィンドウを開くことによって、複数のアプリケーションを同時に実行しているように見える。

MS-Windows は、厳密には、「マルチタスク」ではない。複数アプリケーションは同時に実行状態にあるが、動いているのは、1本だけある。他のアプリケーションは、一時停止の状態にある。止まっているアプリケーションを動かしたいときは、動かしたいアプリケーションのウィンドウをクリックすれば切り替えることができる。

また、AアプリケーションとBアプリケーション同時に実行状態にあったとして、Aアプリケーションで作成したデータをBアプリケーションに転送するような、別アプリケーション間でのデータの交換が可能になる。

(2) 単語切り出し機能

この機能は、検索質問文を自然言語の状態で入力したいという要望から生まれた機能である。現段階では、検索質問文として入力した日本語の文章から名詞を切り出すだけの機能である。例えば、次のような、検索質問文が入力されたとき、

[検索質問文]

パソコンのCPUの歴史について

切り出した単語は、次のようになる。

〔切り出し単語〕

パソコン

CPU

歷史

本来、このような機能は、日本語の文章を形態素に分解して、ある程度の意味解析をした上で、対象となるデータベータの統制キーワードに変換することが望ましい。突き詰めれば、文章の中には、内容の否定を表現するものもあれば、複数の言葉の関係を示すものもある。これから、文章の意味的要素から複数キーワード同士の関係を示す論理式を生成するまでを目標にしなければならない。このような解析は、現在の環境では、技術的にも非常に難しい。

英語圏ユーザが操作する環境での自然言語形式の検索質問文の入力は、英語の文章からキーワードを切り出すことになる。しかし、この発想は、わが国のユーザのものであり、英語圏のユーザは、コマンドの入力に慣れていたり、シソーラスの用語の表現形式が複数形だったり等の条件と前述したような文法的な解釈論にまで発展して行かなければいけないので、本課題では、この考え方を採用しなかった。

(3) シソーラス参照機能

日本語で範囲(指定した用語の漢字コードの範囲)を指定して、その範囲にある用語群を取り出す機能。通常のシソーラスは、階層構造に整理されているため、範囲を指定するケースは少ない。

	シソーラス参照
	参照するシソーラスの範囲を指定してください
FROM:	あ
TO :	Ŋ
	了 解 取 消
	シソーラス一覧
亜細亜	
圧力 安全	
鞍	
,	
	関連語 取消

図2-12 シソーラス参照機能操作例

本課題では、日英パイロット・シソーラスを利用して、範囲指定のほかに、階層上位からの検索 (上位概念語から下位概念語への検索)、単語による検索を用意した。特に、統制語キーワードを検索 する際には、メニューによる階層検索の機能があると便利である。

(4) 関連語生成機能

一つのキーワードから関連する用語群を取り出す機能。「検索 Helper」は、関連語同士は、同じグループのキーワードという考え方で整理しているファイルをアクセスしている。関連語は、普通、シソーラスに掲載してあり、検索のヒット率を上げるためのものである。

	シン	ノーラス一覧		
亜細 亜 圧力 安全 鞍				
	関 連 語	取	消	

図2-13 関連語生成機能操作例(1/2)

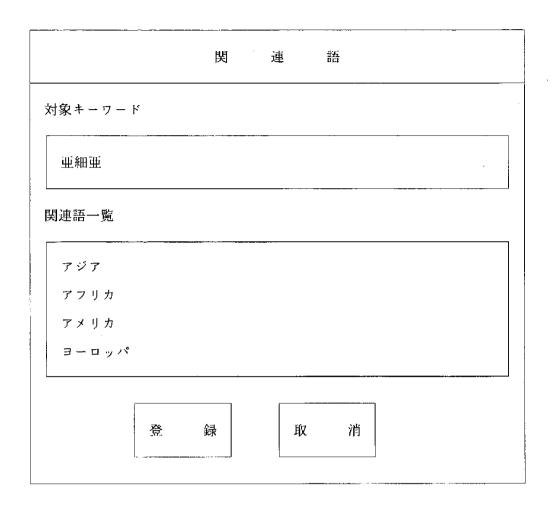


図2-13 関連語生成機能操作例(2/2)

本課題では、日英パイロット・シソーラス(端末内シソーラス)を用いて関連語を検索した。

(5) 英日、日英キーワード変換機能

英語の用語を日本語に変換したり、日本語の用語を英語に変換する機能で日本語、英語それぞれの 前方一致検索、後方一致検索もできる。

この機能は、ユーザの意志で使用されるものであるため、操作が多くなる部分がある。英語圏のユーザが日本のデータベースを検索するという前提が初めから分かっていれば、検索までの流れの中で、この機能を意識しなくても英日変換を自動的に行えばよい。

(6) 検索機能体系

データベースを検索は、前述した機能の組み合わせと検索コマンドの生成、送信、メッセージの受信、受信メッセージの編集といった機能から構成されている。通常は、次の手順で操作する。

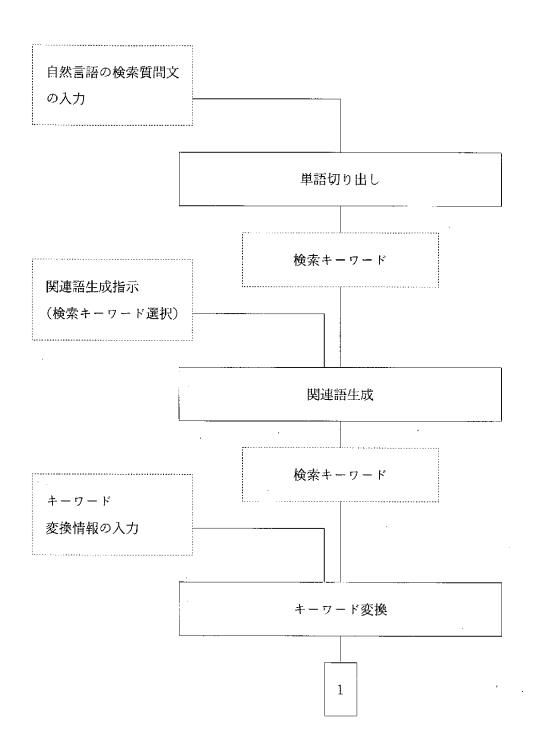


図2-14 検索機能体系の流れ(1/6)

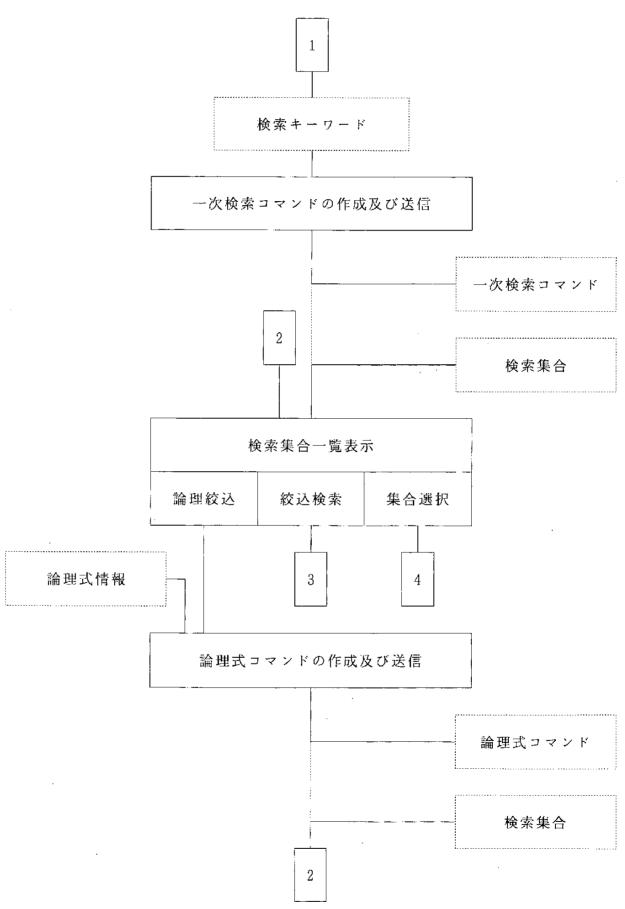


図2-14 検索機能体系の流れ(2/6)

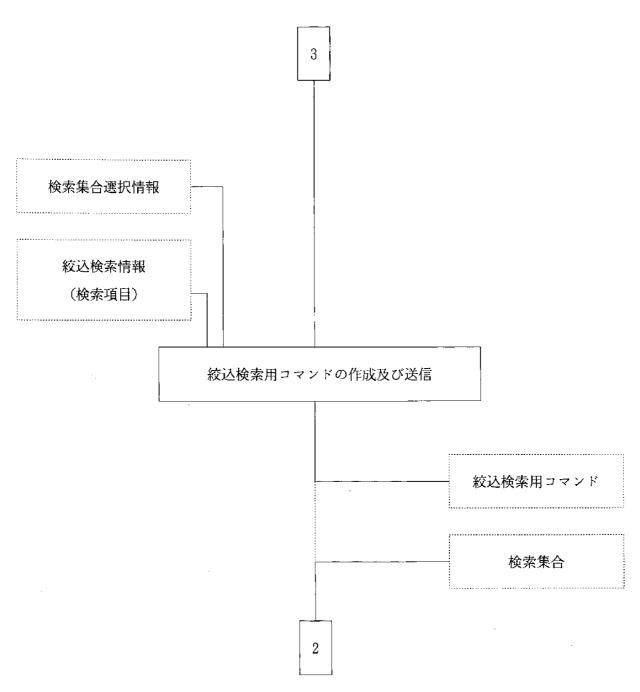


図 2-14 検索機能体系の流れ(3/6)

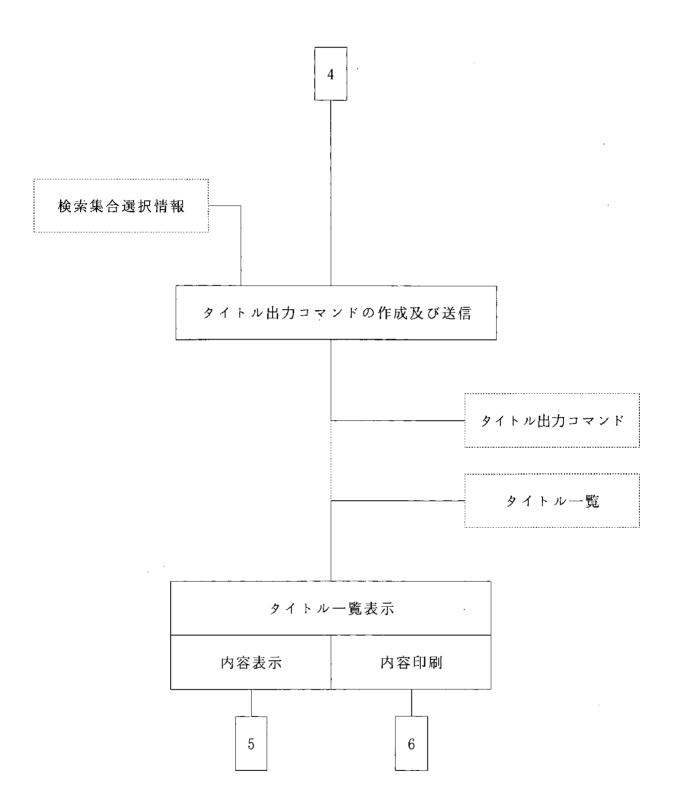


図2-14 検索機能体系の流れ(4/6)

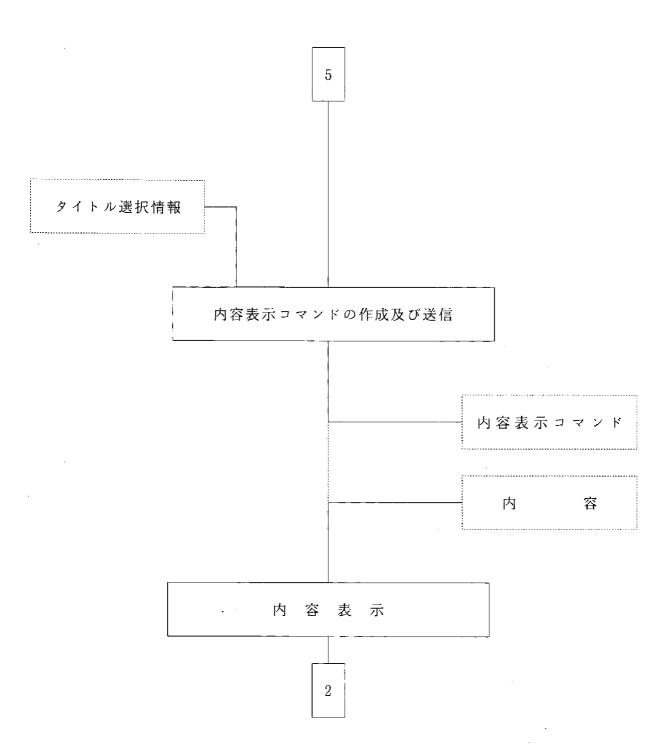


図2-14 検索機能体系の流れ(5/6)

-30 -

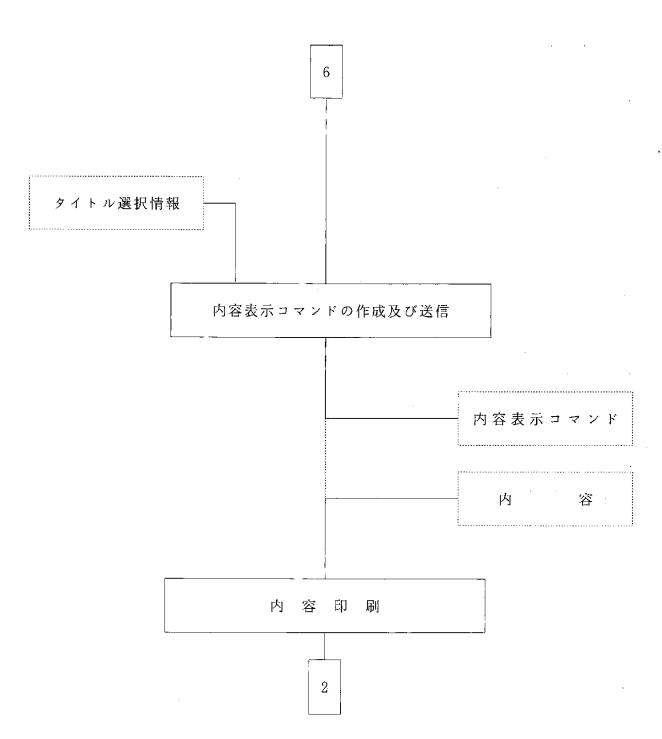


図2-14 検索機能体系の流れ(6/6)

2. 3 機械翻訳ソフトウェア

「多国間検索」を行うとき、機械翻訳は、不可欠である。データベースでの利用法は、次の3通りの方法がある。

- ① ホスト側に翻訳機能を持っている形態。
- ② 端末側に翻訳機能を持っている形態。
- ③ 外部委託、ディストリビュータの翻訳サービス等の完全に分離した形態。

③の場合は、翻訳の機械化を完全に無視している形態なので、ここでは、言及しない。①の場合は、ホスト側の負担が大きく費用もかかる。②の形態が実現するもっとも、低コストになる。

(1) 翻訳システムの機能

機械翻訳自体の精度については、現段階において、完璧を求めるのは、難しい。英仏変換のように 文字や文法が比較的似ているものは、文法解析が進歩すればかなりの精度が期待できる。しかし、ポルトガル語圏の人が英語の文書を読むときは、敢えて、翻訳しないという現実がある。日本語のよう に文法が特殊で、文字コードが2バイトである言語は、1バイトで表現できる文字を使用している言語とは、かなり扱いが違ってくる。

端末機で翻訳するということは、言い換えるとパーソナル・コンピュータ用の機械翻訳ソフトウェアが必要ということになる。

現段階においては、英日翻訳システムのみが流通し始めたばかりであり、その翻訳精度もユーザが満足するには至っていない。せいぜい人間が翻訳するときの補助的な役割をするといったところである。

機械翻訳システムは、大型、汎用コンピュータからワークステーション、パーソナル・コンピュータとダウンサイジングしてきた。機械翻訳システムは、速い処理速度、大きなRAMの容量、大きなディスクの容量が要求されているので、実際に、パーソナル・コンピュータでの稼働に耐えられるようになったのは、平成に入ってマッキントッシュやウィンドウズが普及してきてからである。

一般的な、構成は、次のとおりである。

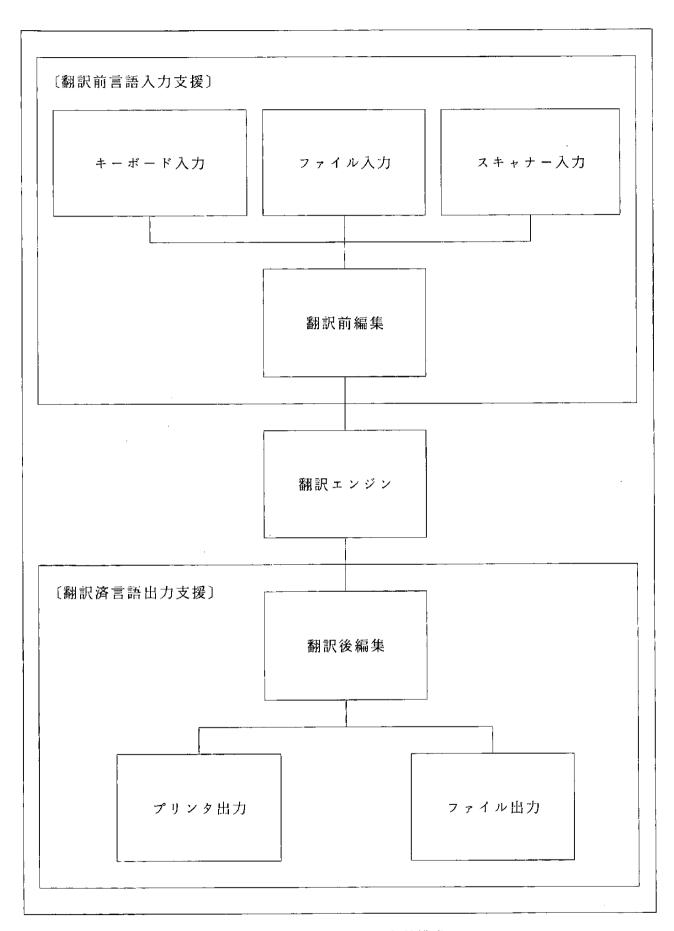


図2-15 翻訳システムの一般的構成

構文トランスファー方式の翻訳エンジンでの処理は、次のようになる。

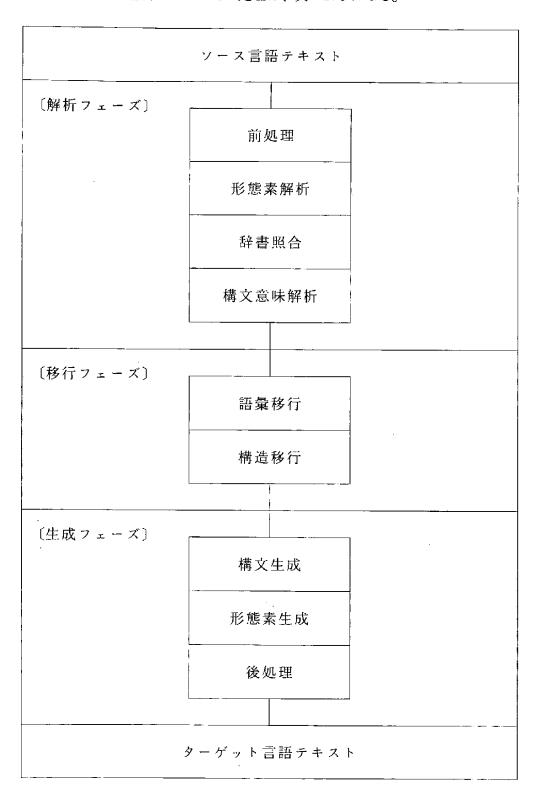


図2-16 構文トランスファー方式の翻訳エンジンでの処理

(2) 翻訳システムとウィンドウズ

前述のように翻訳システムには、イメージデータから入力するマルチメディア的〜側面と翻訳前後の編集機能のようなデストップ・パブリッシング的〜側面がある。このような特徴は、マッキントッシュやウィンドウズで稼働しやすいソフトウェアであると言うことができる。

また、データベース検索システムも MS-Windows で稼働するので検索結果のデータを翻訳システムへ実行中に移行することができる。

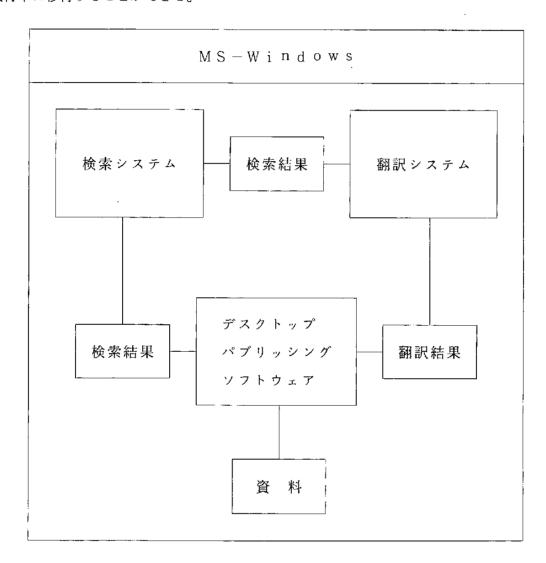


図2-17 Windowsモデル

このように、Windows ソフトウェアを組み合わせることで、色々なバリエーションが考えられる。

(3) 本課題と翻訳システム

本課題で作成するシステムは、英語圏のユーザがわが国のデータベースを検索する関係上、日本語の検索結果を英語に翻訳する必要がある。しかしながら、パーソナル・コンピュータで稼働する日英翻訳システムの事例は少ないし、翻訳制度もまだまだである。

英日翻訳システムよりも日英翻訳システムの方が日本語の文法を解析しなければならない分、技術的な難しさをもつことも原因の一つである。

したがって、今回の開発では、日英翻訳システムの将来的改善を期待して、いつでも追加できる状態にしておく。

3. 英日シソーラス

本課題で採用する日英パイロット・シソーラスは、国産データベースの特定分野のシソーラスに英語を付与した形態の端末内シソーラスである。元のシソーラスは、「日経シソーラス」のビジネス分野が中心になっている。この国産シソーラスの用語に英語が一件ずつ割りつけている。

この、シソーラスは、3つの用途を想定して、作成されている。

- ① 付加情報(日本語、ローマ字、関連語、同義語、類語)を検索したい。
- ② 階層順に検索したい。
- ③ 単語で検索したい。

これらを実現するために、次の3種類のファイルを用意している。

- ① アルファベット順リスト (シソーラス本体)
- ② 階層リスト
- ③ KWICリスト

内容の詳細は次に述べる。

3. 1 アルファベット順リスト(シソーラス本体)

シソーラスの用語を英訳したものを、アルファベット順に整理したもので、一つの用語に対して、ローマ字、漢字、スコープノート (用語の使用法、注釈)、上位語、下位語、関連語、類語、同義語が掲載されている。

- 見出し語(優先語) - 階層コード 日本語ディスクリプタ ACCOUNTING (C3.3.1) — 企業会計(KIGYO-KAIKEI) NT ALLOWANCES 引当金(HIKIATE-KIN) ANNUAL REPORTS アニュアルレポート (ANNUAL-REPORT) ASSETS 資産 (SHISAN) ASSETS REVALUATION 資産再評価 (SHISAN-SAIHYOUKA) AUDITING 企業監査(KIGYO-KANSA) BREAKVEN POINT 下位語 損益分岐点(SON'EKI-BUNKI-TEN) **DERECIATION** 減価償却(GENKA-SHOKYAKU) DISCLOSURE ディスクロージャー (DISCLOSURE) EQUITY CAPITAL 自己資本(JIKO-SHIHON) FINANCIAL STATEMENTS 財務諸表(ZAIMU-SHOHYO) LIABILITIES 負債 (FUSAI) BT FINANCIAL MANAGEMENT 上位語 財務管理(ZAIMU-KANRI) CORPORATE ADVERTISINS — 見出し語(非優先語) USE INSTITUTIONAL ADVERTISING 企業 PR (KIGYO-PR) INSTITUTIONAL ADVERTISING 企業 PR(KIGYO-PR) UF CORPORATE ADVERTISING

図3-1 アルファベット順リスト例

記号の意味は、次のとおりである。

見出し語が優先語の場合

SN: スコープノート (scope notes)。このシソーラスにおける用語の特定の意味や用語の使い方の説明。

UF: (use for)。同義語や類義語になっている非優先語。UFとして示した非優先語からこの見出し語へのUSE 参照があることを示している。

NT:下位語 (narrower)。

BT:上位語 (Broader terms)。 RT:関連語 (related terms)。

・見出し語が非優先語の場合

USE:検索に使用すべき優先語を示す。

3. 2 階層リスト

シソーラスの用語を英訳したものの、上位、下位の階層関係を示したリストで、階層が深くなるほど用語の前の「・」が多くなっている。また、同一階層は、アルファベット順に並べられている。

C3.3	FINANCIAL MANAGEMENT	財務管理(ZAIMU-KANRI)
C3.3.1	• ACCOUNTING	企業会計(KIGYO-KAIKEI)
C3.3.1.1	· · ALLOWANCES	引当金(HIKIATE-KIN)
C3.3.1.1.1	· · · ŁOAN LOSS ALLOWANCES	貸倒引当金
		(KASIDAORE-HIKIATE-KIN)
C3.3.1.1.2	• • • RETIREMENT ALLOWANCE RESERVES	退職引当金
		(TAISYOKŲ-HIKIATE-KIN)
C3.3.1.2	• • ANNUAL REPORTS	アニュアルレポート
		(ANNUAL-REPORT)
C3.3.1.3	· · ASSETS	資産(SHISAN)
C3.3.1.3.1	· · · CREDITS	債権(SAIKEN)
C3.3.1.3.1.	l · · · · BAD DEBTS	不良債権(FURYO-SAIKEN)
C3.3.1.3.2	· · · FINANCIAL ASSETS	金融資産(KIN'YU-SHISAN)
C3.3.1.3.3	· · · HIDDEN ASSETS	含み資産(FUKUMI-SHISAN)
C3.3.1.3.4	· · · INVENTORIES	棚卸し資産
		(TANAOROSHI-SHISAN)

図3-2 階層リスト例

3. 3 KWICUXF

用語を単語別に分類しアルファベット順に並べたものである。用語が単語 3 語から構成されていた としたら、リスト上、三回登場することになる。

				
	PRODUCT	ADAPTABILITY		製品融通(SEIHIN-YUZU)
	FUND	ADJUSTMENT		資金融通(SHIKIN-YUZU)
	SHIPMENT	ADJUSTMENT		出荷調整(SHUKKA-CHOSEI)
	MARKETS FOR	ADULTS	us	e MARKETS FOR THE MATURED PERSON
		ADVANTEST	(CO.)	アドバンテスト (ADVANTEST)
		ADVERTISING	•	商品 PR(SHOUHIN-PR)
		ADVERTISING	CAMPAIGNS	セールスキャンペーン
				(SALES-CAMPAIGN)
		ADVERTISING	EXPENDITURE	広告費(KOUKOKU-HI)
	CORPORATE	ADVERTISING	us	e INSTITUIONAL ADVERTISING
	EXAGGERATED	ADVERTISING		誇大広告(KODAI-KOUKOKU)
	INSTITUTIONAL	ADVERTISING		企業 PR(KIGYO-PR)
	CORPORATE LEGAL	AFFAIRS		企業法務 (KIGYO-HOUMU)
-1				

図3-3 KWICリスト例

3. 4 シソーラスのアクセス

3種類のシソーラスの利用法は、つぎのとおりである。

(1) アルファベット順検索

用語の範囲を指定して、その範囲の用語をすべて取り出す。ユーザは、FROM, TO の 2 つの入力 欄に用語を入力する。英和辞典に例えると FROM に入力する用語が A に近い方、TO に入力する用語が Z に近い方であり、英日シソーラスに集録している用語の内 FROM と TO の間にある用語をすべて取り出す。内部コードの大小関係は、次のようになる。

FROM入力欄の用語 ≤ 検索用語 ≤ TO入力欄の用語

(2) 階層順検索

上位概念語から下位概念語へ検索する。例えば、次のような、シソーラスの階層関係の場合、

Α

- A 1
- • A 1 1
- • A 1 1 1
- • A 1 1 2
- • A 1 1 3
- • A 1 1 4
- • A 1 2
- • A 1 3
- A 2
- • A 2 1
- • A 2 2
- • A 2 3
- A 3

В

- B 1
- • B 1 1
- • B 1 2
- • B 1 3
- • B 1 4
- B 2
- B 3
- B 4
- • B 4 1
- • B 4 1 1
- • B 4 1 2

階層を図式化すると、

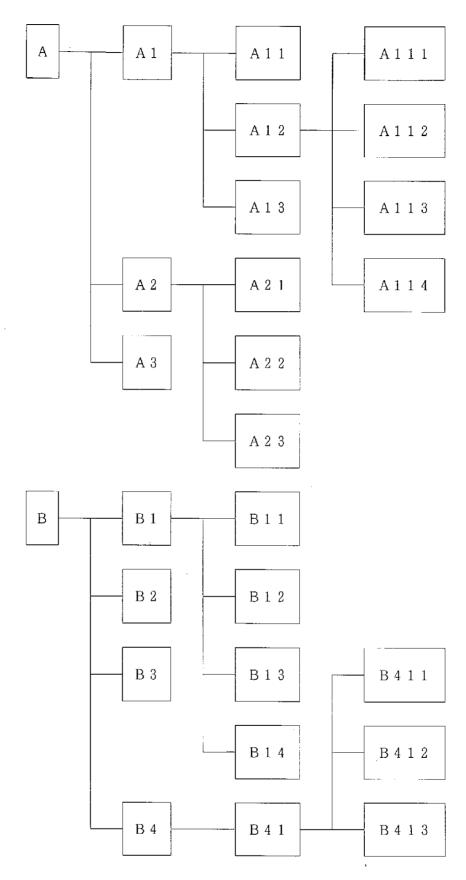


図3-4 階層の図式化

階層順検索は、上位概念語から下位概念語へ検索するか、下位概念語から上位概念語へ検索するという操作になる。検索の遷移図は、次のようになる。

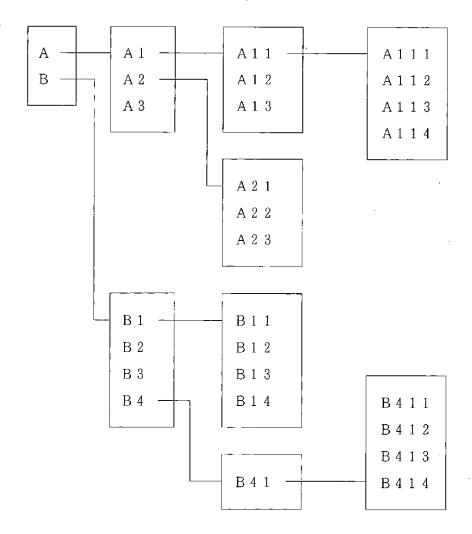


図3-5 階層検索の遷移図

上図の四角に囲まれた部分が一度に表示される単位である。この単位を1グループとしてデータを整理すると、システム設計上のデータの管理が容易になる。内部的には、次に示すインデックス・ファイルを作成する。1用語1レコードでグループNOでソートされている。用語アドレスは、アルファベット順に整理されたファイルのアドレスを指す。

用語アドレス	下位グループ 先頭アドレス	上位グループ ・ 先頭アドレス	自グループ NO
	•	•	•

図3-6 インデックス・ファイルの構造

(3) 単語による検索

単語を入力して、その単語を含む用語をすべて取り出す。英日シソーラスの KWIC リストを利用する。KWIC リストは、同一の単語を含む用語を集めているので、キーとなる単語で用語を検索するためのものである。入力した単語から用語を検索する方法は次のとおりである。

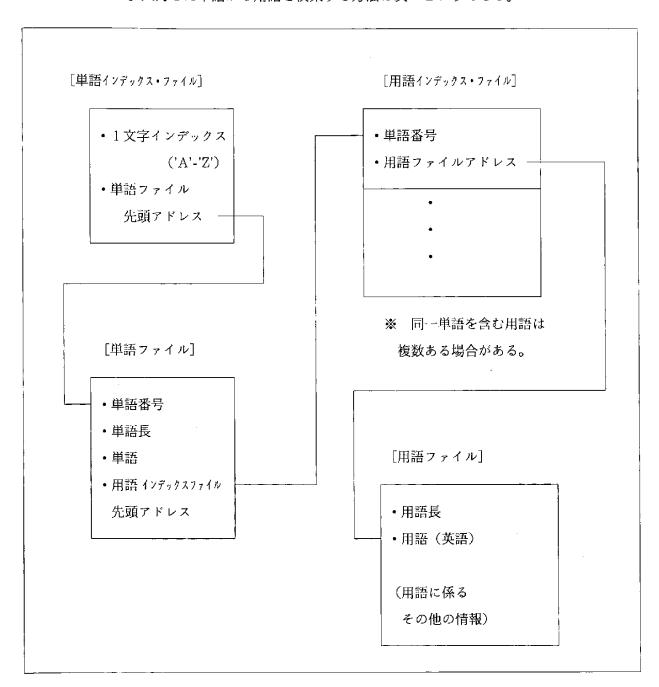


図3-7 単語による検索の構造

(4) 関連語検索

選択した用語の関連語をあれば取り出す。アルファベット順リストには、英文用語毎の日本語、ローマ字、上位語、下位語、関連語、類語、同義語が収録されている。関連語は、各用語に近い概念の用語で元の用語と併用することによって、検索率を高めることができる。すなわち、検索キーワードを補完する役割を持っている。選択された用語から関連語を検索する方法は次のとおりである。

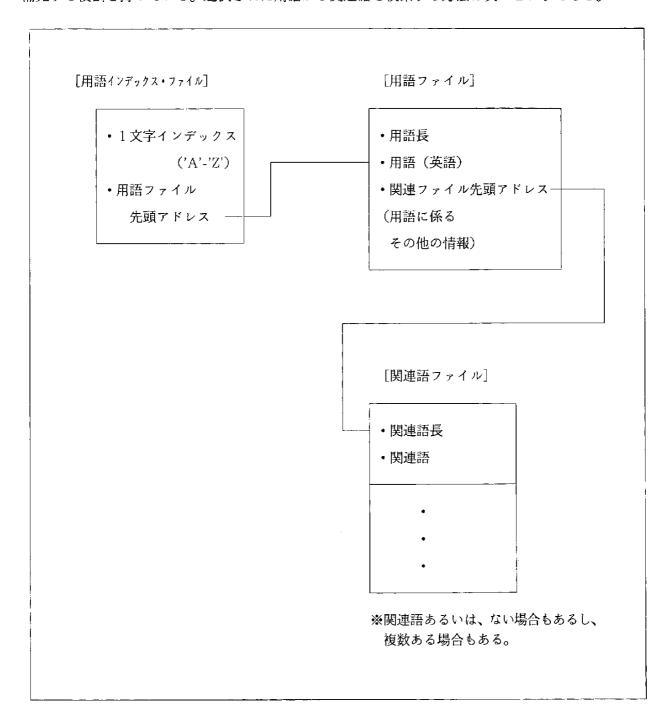


図3-8 関連語検索の構造

(5) 詳細表示

選択した用語のローマ字、漢字、スコープノート(用語の使用法、注釈)、上位語、下位語、関連語、類語、同義語を表示する。選択された用語から詳細情報を検索する方法は次のとおりである。

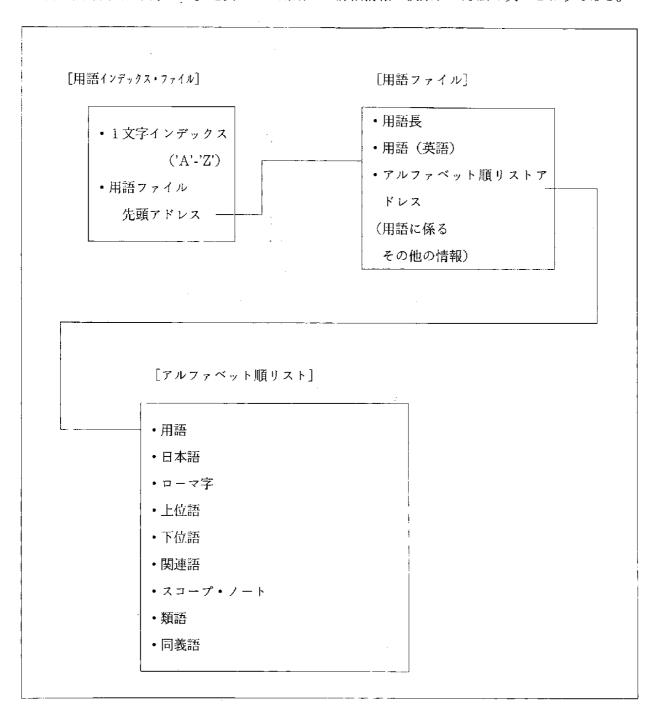


図3-9 詳細表示の構造

(6) 日本語表示

検索候補のキーワードを日本語に翻訳して表示する。検索候補のキーワードから日本語を検索する 方法は次のとおりである。

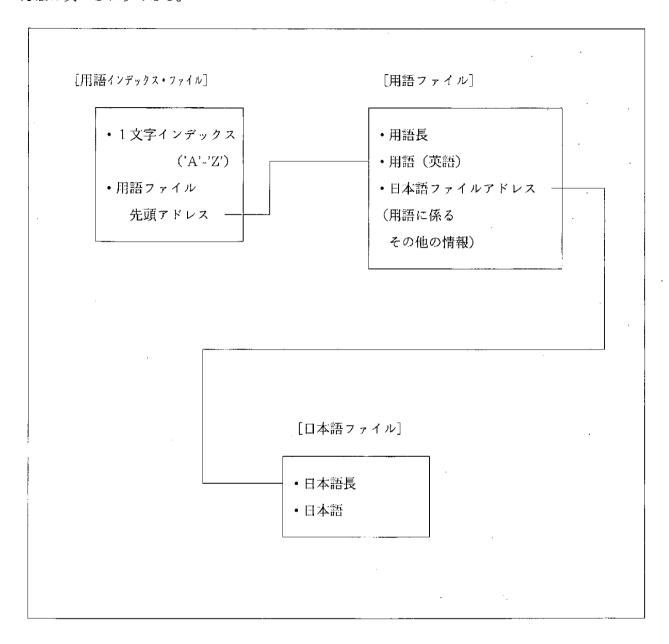


図3-10 日本語表示の構造

4. プロトタイプ・システム

本プロトタイプ・システムは、キーワードの入力支援、データベースの検索支援、データベース検索にともなう通信制御を中心としたシステムである。

4. 1 システムの構成

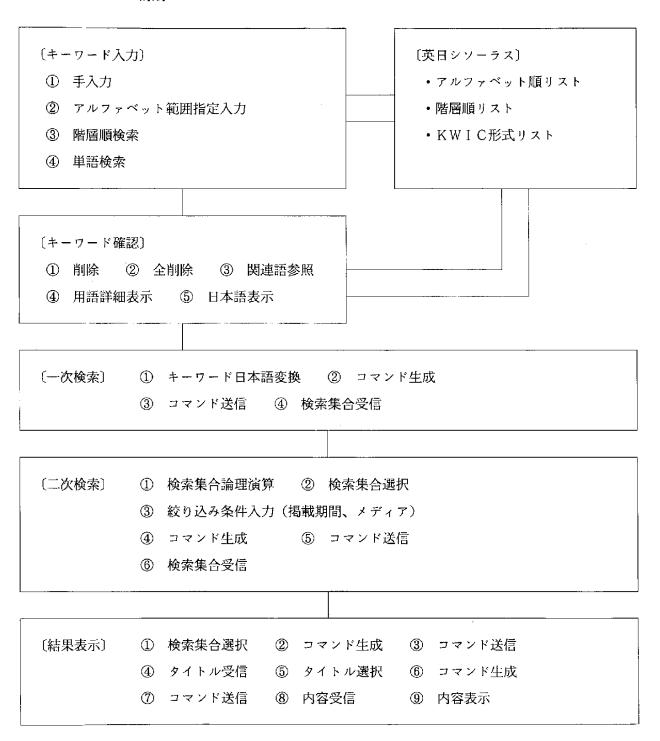


図 4-1 システムの構成

4. 2 入出力制御機能

本プロトタイプ・システムは、MS-Windows V 3. 1を採用することによって操作性の向上を図った。Windows は、ディスプレイに窓を開いて、アプリケーションを実行する OS である。Windows 自身は、MS-DOS 上で稼働している。従来の、MS-DOS の環境では、処理の実行やプログラムを起動するためにユーザは、プログラムの存在するデバイスを意識しながらキーボードからコマンドを入力しなければならない。キーボード入力が中心になっていた環境では、コマンドの名前などを覚えなければならず、初心者にとっては特に使いにくい環境になっている。それに大して、Windows の環境では、処理の実行やプログラムの起動は、マウスを使った操作が中心になる。アプリケーションの実行中でも特定の名称を入力する以外は、すべてマウスを使ったオペレーションになる。コマンドを入力するかわりにアイコン、メニュー、ダイアログボックスのボタンをクリックすることでほとんどのことができる。

(1) 起動

アイコンをクリックする。

(2) 基本ウィンドウ

システムが起動するとつぎのウィンドウが表示される。

Keyword Search Option Quit		
KEYWORD		

図 4-2 基本ウィンド

このウィンドウの二段目に「Keyword」、「Search」、「Option」、「Quit 」と表示されているが、この列がメニュー・バーと呼ばれている。操作は、ここをクリックすることによって始まる。「KEYWORD」と表示されている下の矩形の部分は、リストボックスといい、ここでは、一次検索を行うときのキーワードを表示する。

(3) $\times = = = -$

メニュー・バーの項目をクリックしたとき、それぞれのポップアップ・メニューを表示する。

① Keyword

一次検索に使用するキーワードを作成するための支援体系でつぎのようなメニューを表示する。

Keyword		
new		キーワードを入力し、リストボックスに登録する
update		キーワード・リストボックスのキーワードを修正する
delete		キーワード・リストボックスのキーワードを削除する
all clear		キーワード・リストボックスのキーワードをすべて削除する
alphabetical		シソーラスをアルファベット順に検索しリストボックスに登録する
hierarchies		シソーラスを階層順に検索しリストボックスに登録する
word		シソーラスを単語を検索しリストボックスに登録する
related		リストボックスのキーワードの関連語をシソーラスから検索する
details	<u></u>	リストボックスのキーワードの詳細情報を表示する
japanese	 	リストボックスのキーワードを日本語表示する

図4-3 キーワード・メニュー

② Search

電話回線の接続/切断、データベース検索の一連の手続きを行う。メニューは、つぎのとおりである。

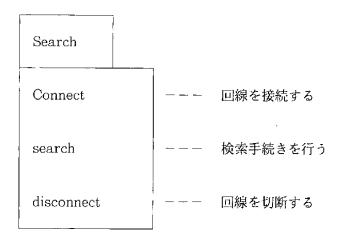


図4-4 サーチ・メニュー

③ Option

データベースの選択や条件設定を行う。メニューは、つぎのとおりである。

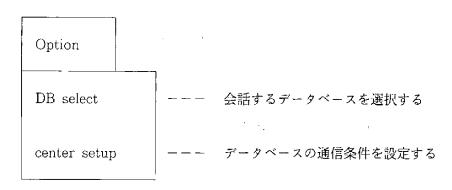


図4-5 オプション・メニュー

4. 3 検索キーワード作成機能体系

英語の検索キーワードを入力して日本のデータベースを検索するために、本プロトタイプ・システムでは、英日シソーラス・パイロットモデルを使用している。このシソーラスは、前にも述べたとおり、つぎの性質をもっている。

- ① 日本のデータベースを基本にしている。
- ② 英語と日本語は、1対1である。
- ③ アルファベット順に整理してあり、関連語等の情報を備えている。
- ④ 上位概念から下位概念までを階層的に整理している。

⑤ 単語で検索できるように整理している。

検索キーワード作成機能体系は、以上の性質をもつ内部シソーラスを利用して、1次検索に利用するキーワードの候補を作成するためのものである。

(1) 新規入力 (new)

キーワードをユーザ手入力する機能である。この機能は、シソーラスを利用しないで、ユーザに自由に入力をさせるものである。検索するときに、ここで入力した用語がシソーラスに存在すれば、対応する日本語が検索キーワードになる。Keywordメニューで new を選択すると、次のダイアログボックスが表示される。

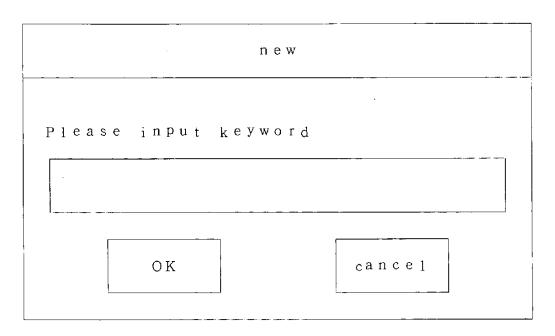


図4-6 新規入力ダイアログボックス

ここで、キーワードを入力して「OK」ボタンをクリックすれば、基本ウィンドウにキーワードが 追加される。

(2) 修正 (update)

基本ウィンドウの中の一つのキーワードを修正する。

基本ウィンドウのリストボックスに表示されているキーワードを一つクリックし、Keywordメニューで update を選択すると、次のダイアログボックスが表示される。

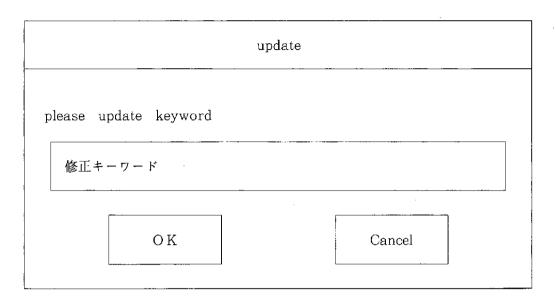


図4-7 修正ダイアログボックス

このダイアログボックスで表示されているキーワードを修正して「OK」ボタンをクリックすると 該当する基本ウィンドウのキーワードが置き換わる。

(3) キーワード削除

基本ウィンドウに表示されているキーワードを選択して削除する。

基本ウィンドウのリストボックスに表示されているキーワードを一つクリックし、Keywordメニューで delete を選択すると、次のダイアログボックスが表示される。

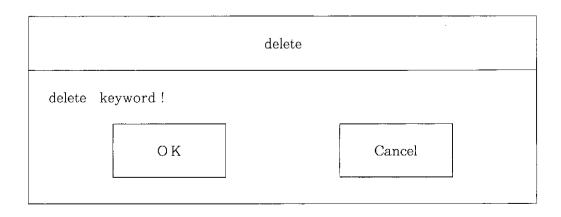


図4-8 キーワード削除ダイアログボックス

このダイアログボックスで「OK」ボタンをクリックすると該当する基本ウィンドウのキーワードが削除される。

(4) 全キーワード削除 (All Clear)

基本ウィンドウのキーワードを全て削除する。

Keywordメニューで All clear を選択すると、次のダイアログボックスが表示される。

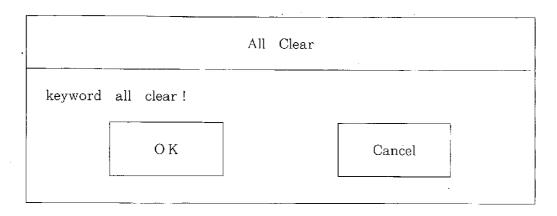


図4-9 全キーワード削除ダイアログボックス

このダイアログボックスで「OK」ボタンをクリックすると基本ウィンドウのキーワードが全て削除される。

(5) アルファベット順範囲指定によるシソーラス照会 (alphabetical)

指定した範囲のシソーラスを照会する機能。シソーラスの検索範囲は、入力したふたつの文字列の 文字コードの間にあるものである。

Keyword メニューで alphabetical を選択すると、次のダイアログボックスが表示さまれる。

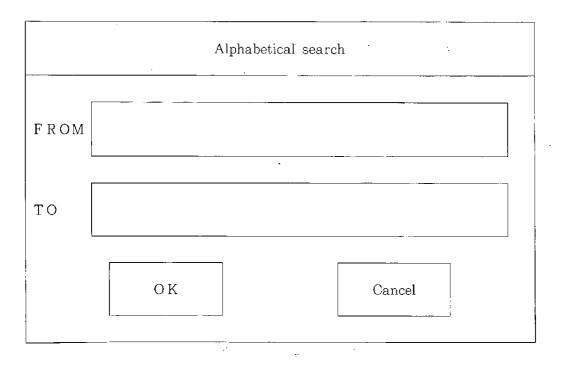


図4-10 アルファベットサーチ・ダイアログボックス

このダイアログボックスの「FROM」と「TO」に用語を入力するとキーワード選択ダイアログボックスが表示される。

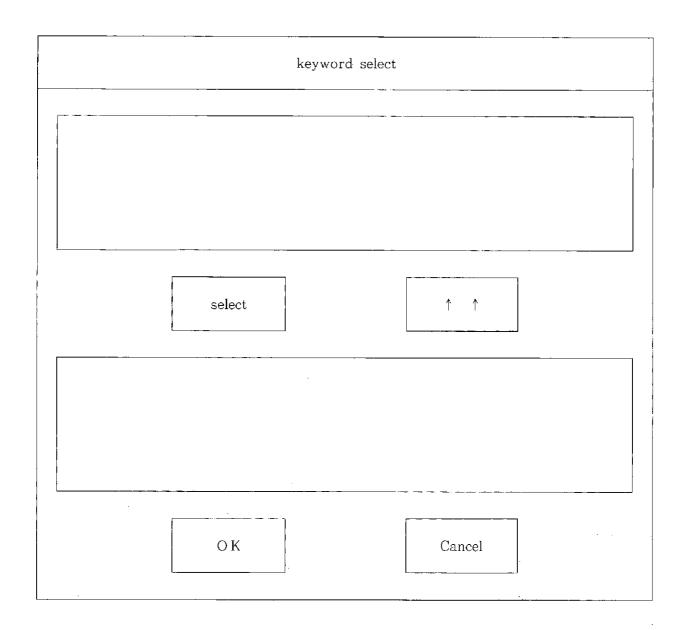


図4-11 アルファベット検索用キーワードセレクト・ダイアログボックス

このダイアログボックスの上のリストボックスに選択した範囲の用語が表示される。この中から検索キーワードとして採用する用語を決める。採用したい用語をクリックした後、「select」ボタンをクリックすると、その用語が下のリストボックスに移る。この操作は、採用したい用語が複数あっても可能である。また、下のリストボックスから、上のリストボックスへ戻したいときは、戻したい用語をクリックして「 $\uparrow \uparrow$ 」ボタンをクリックする。採用したい用語が確定したら「OK」をクリックすることによって、基本ウィンドウに追加される。

(6) シソーラスの階層順参照 (hierarchies)

階層順にシソーラスを照会したいときにこの処理を行う。

Keyword メニューで hierarchies を選択すると、次のダイアログボックスが表示される。

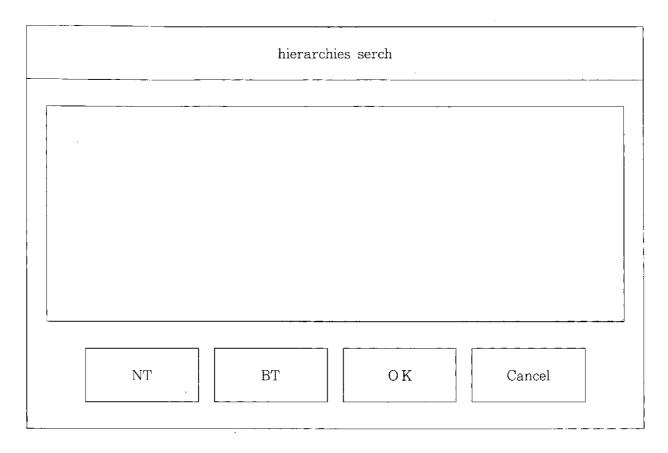


図4-12 階層サーチ・ダイアログボックス

はじめリストボックスの中には、最上位概念語の一覧が表示される。「NT」ボタンは、選択した用語の1階層下位の用語の一覧を表示する機能である。「BT」ボタンは、現在表示されている用語グループの1階層上位の用語を含んだ用語グループの一覧を表示する検索キーワードとして採用したい用語を見つけたら、その用語をクリップして「OK」ボタンをクリックすると基本ウィンドウにその用語が追加される。

(7) 単語入力によるシソーラス照会(word)

入力した単語を含んでいる用語をシソーラスの中から選択する。

Keywordメニューで word を選択すると、次のダイアログボックスが表示される。

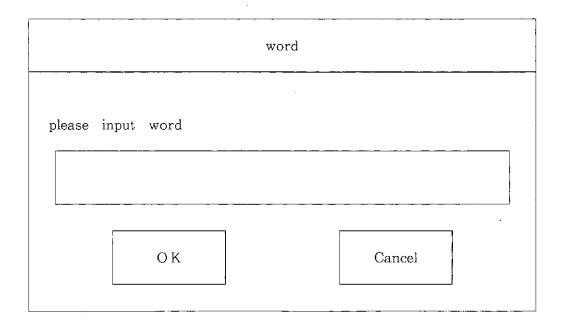


図4-13 単語入力ダイアログボックス

このダイアログボックスに単語を入力し「OK」ボタンをクリックするとキーワード選択ダイアログボックスが表示される。

	keyword select
select	↑ ↑
ОК	Cancel

図 4-14 単語用キーワードセレクト・ダイアログボックス

このダイアログボックスの上のリストボックスに選択した範囲の用語が表示される。この中から検索キーワードとして採用する用語を決める。採用したい用語をクリックした後、「select」ボタンをクリックすると、その用語が下のリストボックスに移る。この操作は、採用したい用語が複数あっても可能である。また、下のリストボックスから、上のリストボックスへ戻したいときは、戻したい用語をクリックして「↑↑」ボタンをクリックする。採用したい用語が確定したら「OK」をクリックすることによって、基本ウィンドウに追加される。

(8) 関連語検索(related)

基本ウィンドウに表示されているキーワードを選択してその関連語をシソーラスから検索する。 基本ウィンドウのリストボックスに表示されているキーワードを一つクリックし、Keyword メニューで related を選択すると、キーワード選択ダイアログボックスが表示される。

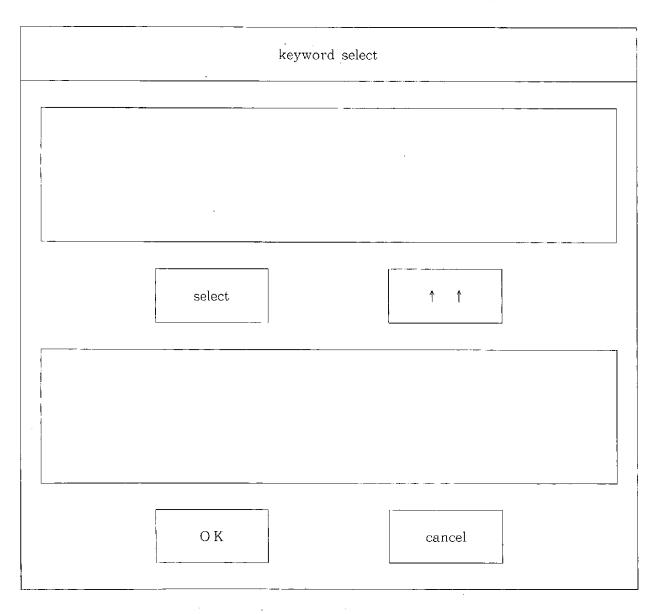


図4-15 関連語セレクト・ダイアログボックス

このダイアログボックスの上のリストボックスに選択した範囲の用語が表示される。この中から検索キーワードとして採用する用語を決める。採用したい用語をクリックした後、「select」ボタンをクリックすると、その用語が下のリストボックスに移る。この操作は、採用したい用語が複数あっても可能である。また、下のリストボックスから、上のリストボックスへ戻したいときは、戻したい用語をクリックして「 $\uparrow \uparrow$ 」ボタンをクリックする。採用したい用語が確定したら「OK」をクリックすることによって、基本ウィンドウに追加される。

(9) 詳細表示 (details)

基本ウィンドウに表示されているキーワードを選択してそのキーワードの内容をシソーラスから検索する。

基本ウィンドウのリストボックスに表示されているキーワードを一つクリックし、Keyword メニューで details を選択すると、次のダイアログボックスが表示される。

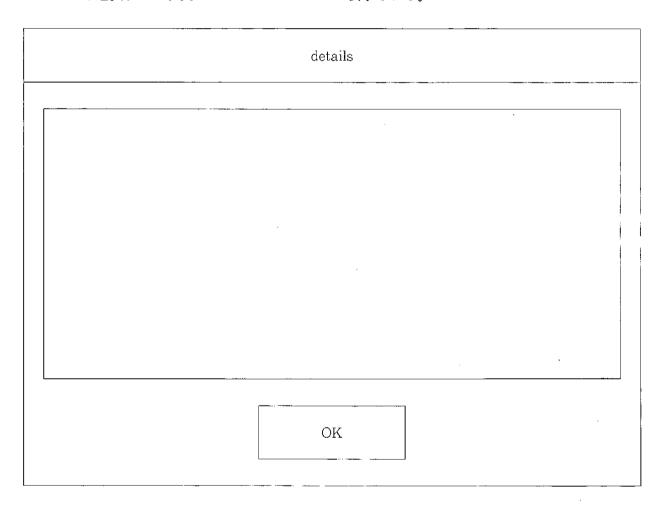


図4-16 詳細表示ダイアログボックス

(10) 日本語表示 (japanese)

基本ウィンドウに表示されているキーワードの日本語訳をシソーラスから検索する。 Keyword メニューで japanese を選択すると、次のダイアログボックスが表示される。

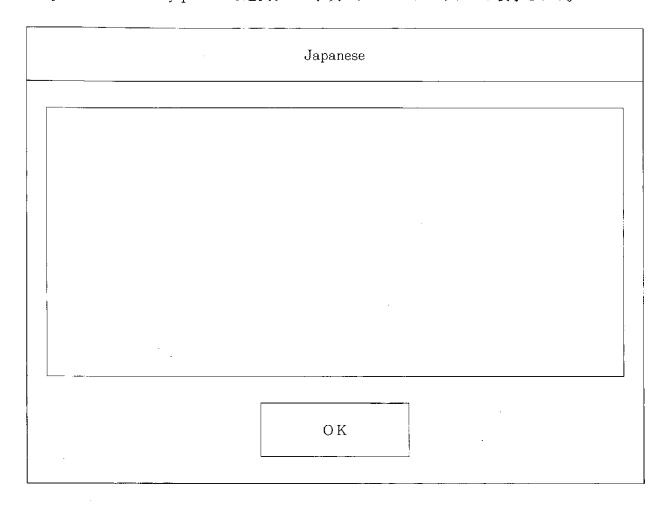


図4-17 日本語表示ダイアログボックス

データベースの一次検索は、このとき、表示される内容で検索される。

4. 4 データベース検索支援機能

データ検索は、ユーザの必要な情報をいかに効率的に取得することができるかが、ポイントになってくる。本プロトタイプ・システムで扱う新聞記事情報は、タイトルや抄録のみのリファレンス・データベースではなく、全文を収録したファクト・データベースである。正確に言えば、写真、絵、図、表を除いたファクト・データベースと言えよう。従って収録件数が多いので、絞り込むための条件の与え方が通常の通信ソフトを使用した検索でも難しい。

しかし、本プロトタイプ・システムの目的は、可能な限り、難解な操作を避けることにあるので、 使用するコマンドは、必要最小限にした。しかも、このコマンドは、手で入力するわけではなく、ダ イアログボックスでの操作を行っていくうちに、自動的に生成される。

想定した検索手順は、つぎのとおりである。

- ① 検索キーワードの作成検索キーワード作成機能体系で検索するキーワードを作成する。
- ② 検索キーワードの和訳 検索キーワード作成機能体系で作成したキーワードでシソーラスを検索して和訳する。シソー ラスに存在しなかったときは、そのまま検索キーワードとして採用する。
- ③ 検索キーワード個別の検索和訳されたキーワードで一つずつ検索する。このとき、キーワードの個数分の検索集合がセンター側で作成される。
- ④ 検索集合に対する論理演算 既に存在する検索集合に対してAND、OR、NOT等の論理演算を行う。このとき、新しい 検索集合がセンター側で作成される。
- ⑤ 特定検索集合に対する条件検索 ユーザがさらに絞り込みたい特定の検索集合に対して、再検査を行う。検索条件は、新聞記事 情報の場合、記事が収録されている日付と、記事が収録されている媒体で概ねの検索結果は、得 られる。
- ⑥ 特定検索集合のタイトル一覧表示 ユーザが指定した検索集合に含まれるタイトルの一覧表示をセンターに要求する。
- ⑦ 特定検索集合のタイトル選択による内容表示 タイトル一覧から内容の見たいものを選択し、内容を表示する。

以上の操作をウィンドウズ化することによって、ユーザ・インターフェースの向上を図る。検索の 流れを図示すると次のようになる。

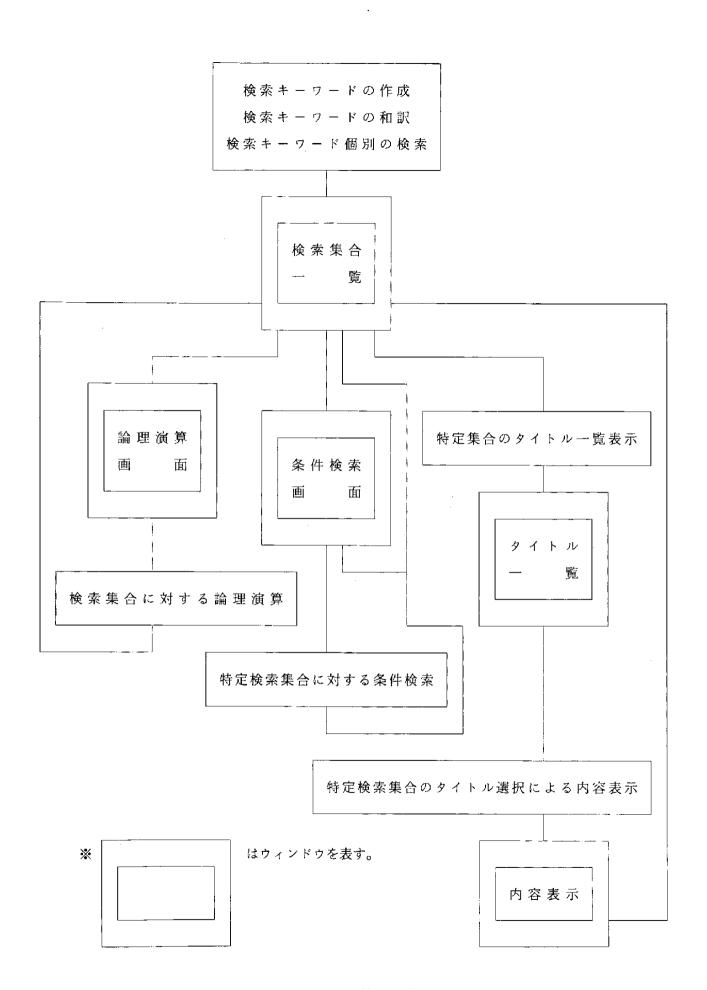


図4-18 検索の流れ

(1) 検索集合一覧

検索集合一覧は、検索キーワードを個別に検索する一次検索や、検索集合同士での論理演算や特定の検索集合に対する絞り込み検索といった二次検索の結果作成された検索集合と検索件数を表示する。 一連の検索手続きは、検索集合一覧を起点にして行う。検索集合一覧を表示するダイアログボックスは次のとおりである。

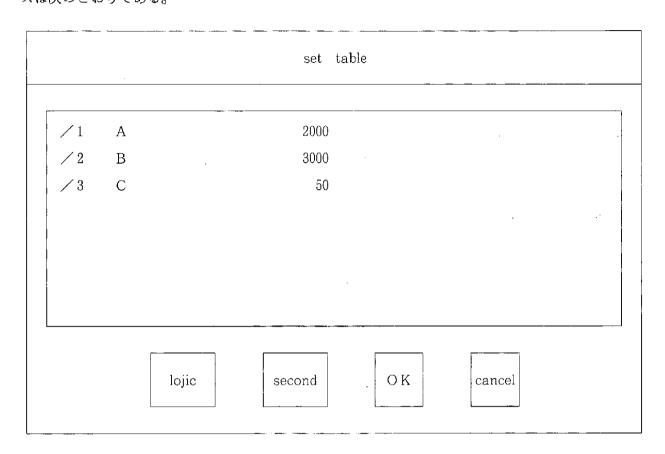


図4-19 検索集合一覧ダイアログボックス

上記リストボックスに表示されているのは、検索集合番号、検索集合名(入力時のキーワードまたは論理式)、検索件数である。「/1 A 2000」の場合は、「/1」が検索集合番号、「A」が検索集合名、「2000」が 2 0 0 0 件を意味している。

初期の表示は、入力したキーワードを個別に検索した結果が表示される。検索自体は、和訳された キーワードで行われるが、検索集合名には、検索キーワード作成機能体験で作成したキーワードをそ のまま表示する。したがって、検索集合名は、通常、英語で表示される。

各ボタンの機能は、次のとおりである。

1 lojic

論理演算ダイアログボックスに処理を移す。

② second

選択した検索集合を絞り込むための絞り込み検索ダイアログボックスに処理を移す。

③ OK

選択した検索集合のタイトルの一覧を表示するためのダイアログボックスに処理を移す。

4 cancel

基本ウィンドウに戻る。

(2) 検索集合に対する論理演算

和訳後の検索キーワードで個別に検索すると、最大件数が、キーワードの件数分の検索集合がセンター側に作成される。もちろん、ヒットしなかった検索キーワードの検索集合は、作成されない。ユーザが最終的にデータを見る段階では、このような、複数の検索集合は、検索手続き中、一つに集約していかなければならない。そのためには、検索集合同士のANDやORの論理演算を行えば、両方検索集合の意をくんだ新しい検索集合が作成される。特に、ANDを使えば件数を減らすことができる。検索集合に対して論理演算を行うためのダイアログボックスは次のとおりである。

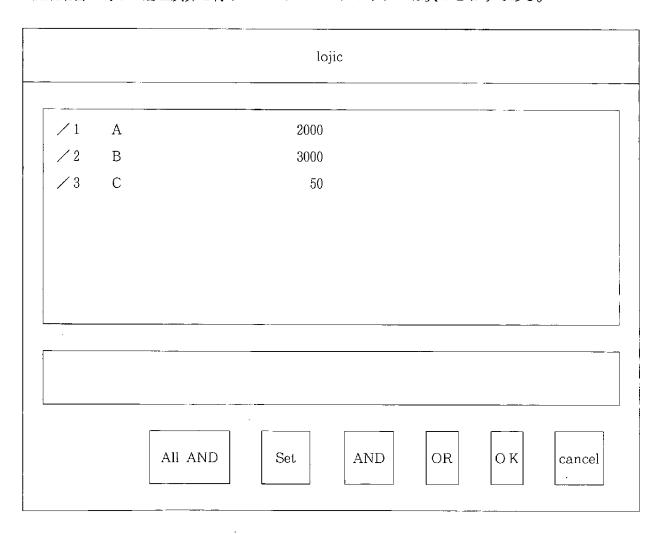


図4-20 検索集合論理演算ダイアログボックス

検索集合の一覧を表示したリストボックスに下にある矩形の部分に論理演算式を入力する。キーワードからでも、下のボタンをクリックしても論理演算式を入力することができる。

各ボタンの機能は、次のとおりである。

① ALL AND

検索集合一覧に表示されているすべての検索集合の AND条件の式を生成する。この場合は、 $\lceil /1 \text{ AND } /2 \text{ AND } /3 \rceil$ となる。

② Set

検索集合一覧の中から一つ選択して、このボタンをクリックすると、この検索集合番号が表示される。既に式が入力されている場合は、その式の後に追加される。式の入力状況が「/1 AND」のときに二行目を選択しこのボタンをクリックすれば、「/1 AND /2 」となる。

(3) AND

式にANDを追加する。式の入力状況が「/1」のときにこのボタンをクリックすれば「/1 AND」となる。

4 OR

式にORを追加する。式の入力状況が「/1」のときにこのボタンをクリックすれば「/1 OR 」となる。

(5) OK

検索集合一覧の中から一つ選択して、このボタンをクリックすると、その検索集合内に記事タイトルを表示するためのダイアログボックスに処理を移す。

6 cancel

検索集合一覧ダイアログボックスに戻る。

以上①から④が論理式を作成するためのもので、⑤、⑥は、処理を制御するためのボタンである。 ここで、すべての検索集合のANDをとる一例を示す。検索キーワードA、B、Cがあり、それが 個別に検索された結果、次のような検索結果が得られたとする。

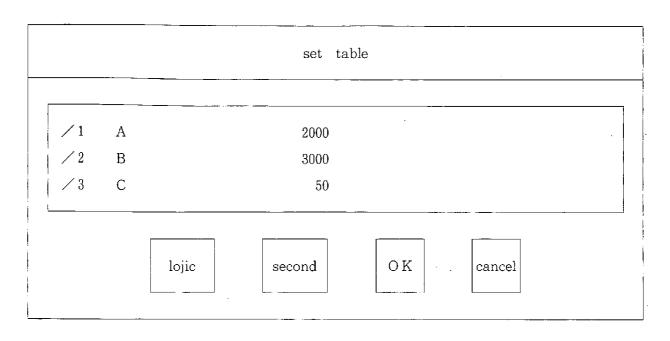


図4-21 検索集合一覧ダイアログボックス

ここで、ユーザが/1、/2、/3のAND条件の検索結果を得たいとき、 $\lceil lojic \rfloor$ ボタンをクリックする。このとき、次のダイアログボックスが表示される。

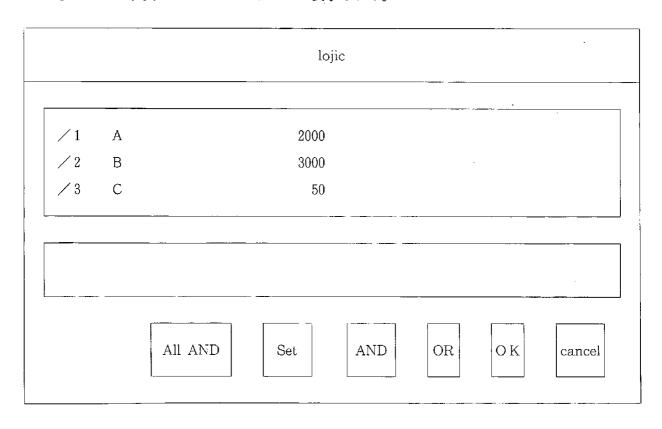


図4-22 検索集合論理演算ダイアログボックス

ここで、「ALL AND」ボタンをクリックすると、次のように自動的に論理式が生成される。

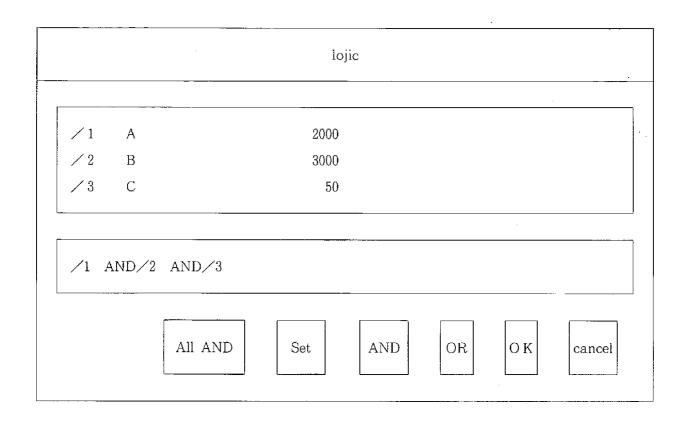


図4-23 検索集合論理演算ダイアログボックス

式を確認したら、「OK」ボタンをクリックするとAND条件の検索が開始され、結果として、検索集合一覧のダイアログボックスに新しい検索集合が表示される。

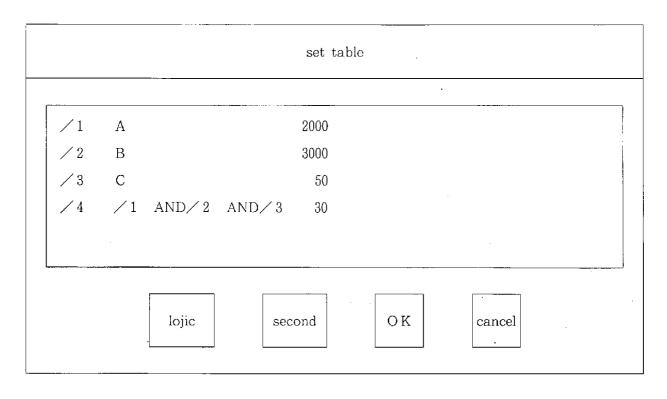


図4-24 検索集合一覧ダイアログボックス

(3) 特定検索集合に対する条件検索

論理演算でかなり絞り込まれている検索集合に対し、検索条件を与えて、件数をさらに絞り込む必要がある。本プロトタイプ・システムでの対象データベースは、新聞記事情報であるので、検索条件は、収録年月日と収録媒体にする。

条件を入力するダイアログボックスは、次のとおりである。

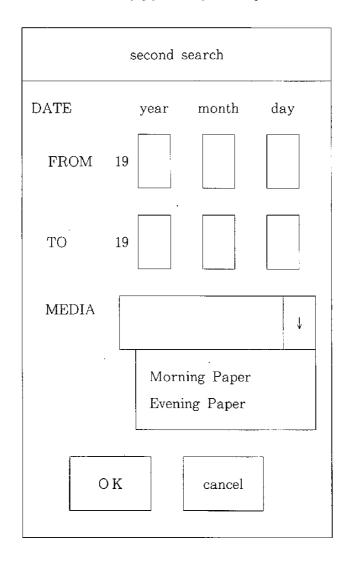


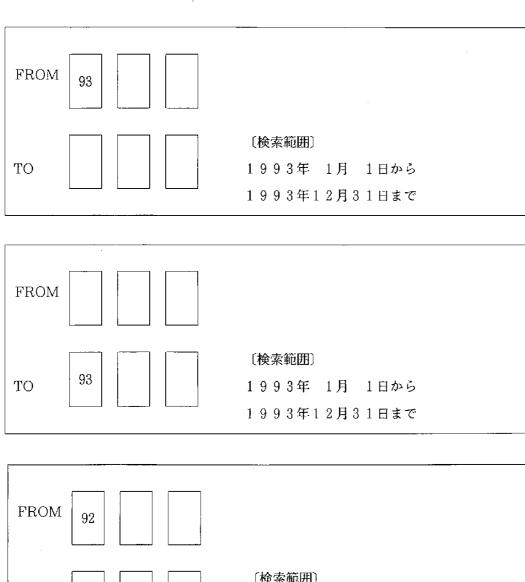
図4-25 絞り込み検索ダイアログボックス

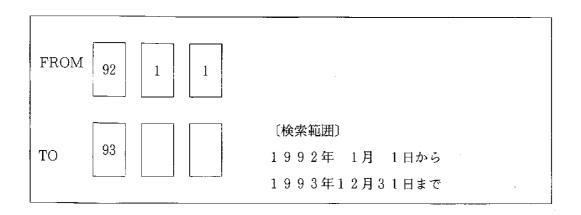
このダイアログボックスでDATEは、データの収録範囲でFROMからTOまでを西暦で指定する。MEDIAは、データの収録媒体でコンボボックスという一種のリストボックスから朝刊とか夕刊が選択できる。データの収録範囲と収録媒体の両者を同時に入力した場合は、AND扱いになる。それぞれの入力規則は、次のとおりである。

① データの収録範囲

- 年は西暦下2桁を指定する。
- FROMとTOの両方に入力があったときの検索範囲は、次のような関係式になる。 FROM ≤ 検索範囲 ≤ TO
- 年を指定して月日は指定しなくてもよい。この場合、その年がFROMに入力されていて、TO が入力されているときは、その年の1月1日とし、TOが入力されていないときは、その年全体 を範囲とする。その年がTOに入力されていて、FROMが入力されているときは、その年の 12月31日とし、FROMが入力されていないときは、その年全体を範囲とする。

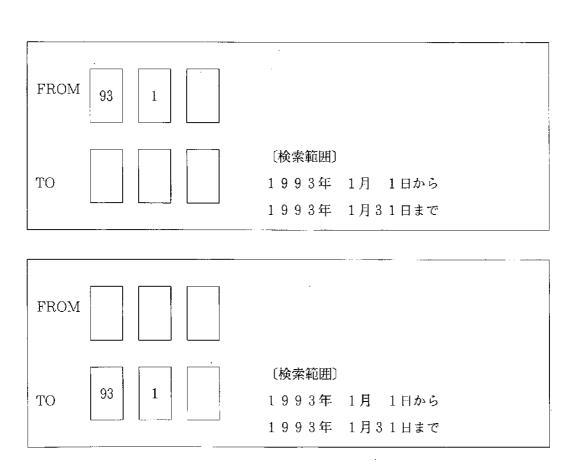
例)

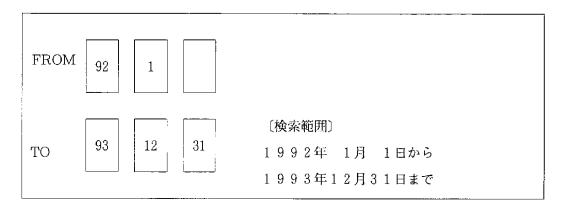


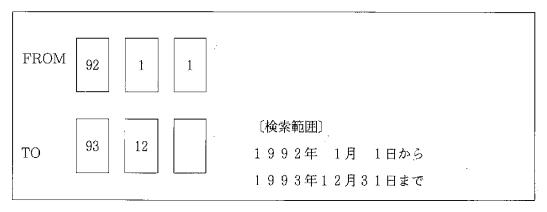


年月を指定して日を指定しなくてもよい。この場合、その年月がFROMに入力されていて、TOが入力されているときは、その月の1日とし、TOが入力されていないときは、その月全体を範囲とする。その年月がTOに入力されていて、FROMが入力されているときは、その月の末日とし、FROMが入力されていないときは、その月全体を範囲とする。

例)







- ・データの収録媒体の選択があったときは、入力しないでもよい。
- ② データの収録媒体
- ・データの収録範囲が入力されているときは入力しなくてもよい。
- ここでは、絞り込み検索の一例を示す。

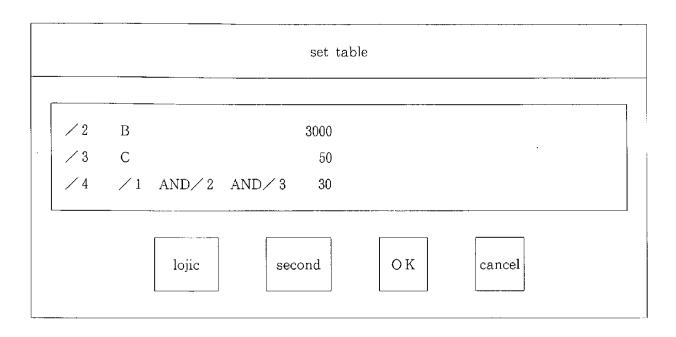


図4-26 検索集合一覧ダイアログボックス

ここで「/4」の検索集合を絞り込みたいとき、リストボックスの4行目をクリックして「second」ボタンをクリックする。

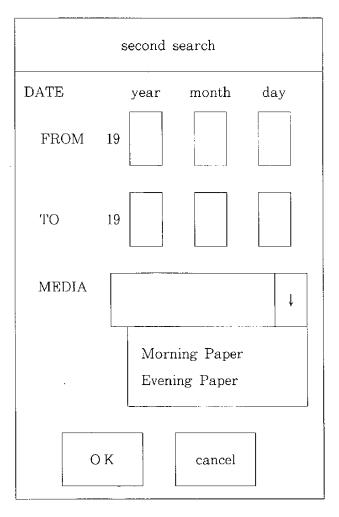


図4-27 絞り込み検索ダイアログボックス

ここで、ユーザが、「1993年10月の夕刊」という条件検索を行いたいときは次のように入力する。

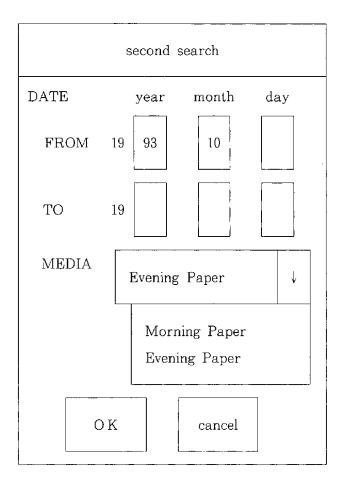
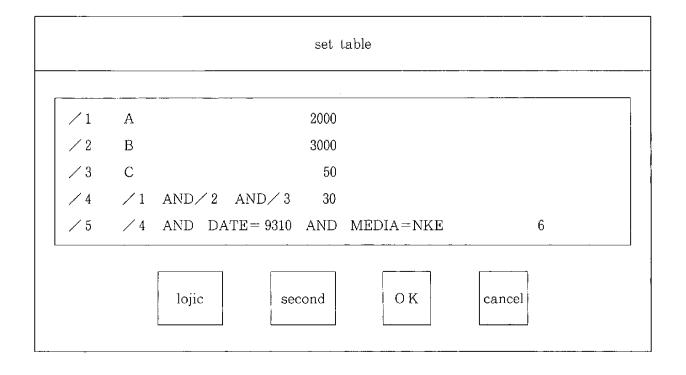


図4-28 絞り込み検索ダイアログボックス

ここで、「OK」ボタンをクリックすれば、次のように検索結果が出る。



(4) 特定検索集合のタイトル一覧表示

ユーザが、これ以上検索集合の絞り込む必要がないと判断したら、絞り込んだ検索集合に含まれている記事のタイトルの一覧を表示する。タイトル一覧は、内容を表示したい記事を選択するときに使用する。タイトル一覧ダイアログボックスを次に示す。

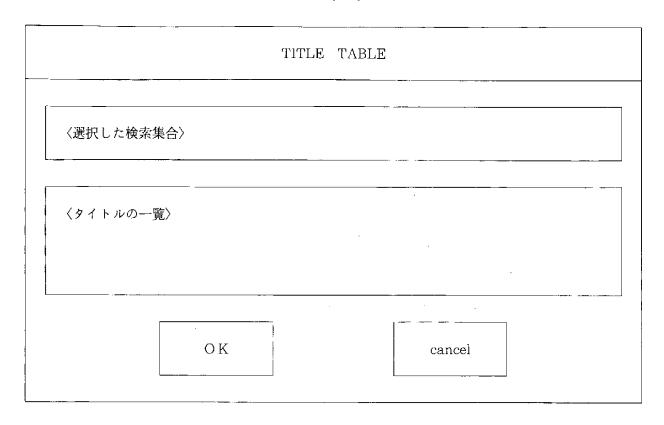


図4-29 タイトル一覧ダイアログボックス

検索集合一覧で内容を見たい検索集合を検索し、「OK」ボタンをクリックすると上記ダイアログボックスが表示される。

ここで、タイトルを選択して、「OK」ボタンをクリックすると内容が表示される。

(5) 特定検索集合のタイトル選択による内容表示 タイトル一覧で選択した記事の内容が、次のようにダイアログボックスに表示される。

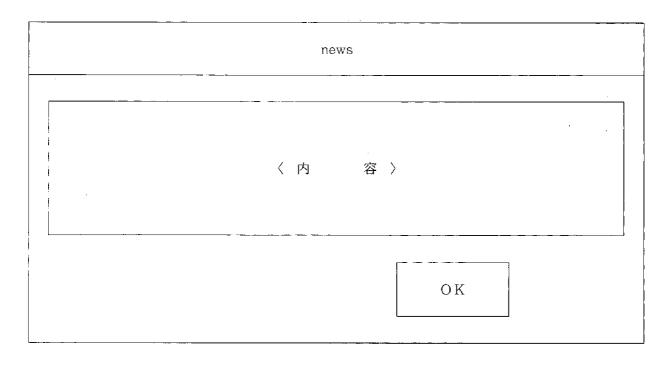


図4-30 内容表示ダイアログボックス

- 一連の検索手続きの操作例は次のようになる。
- 検索キーワード個別の検索の結果に対して論理演算を行う。

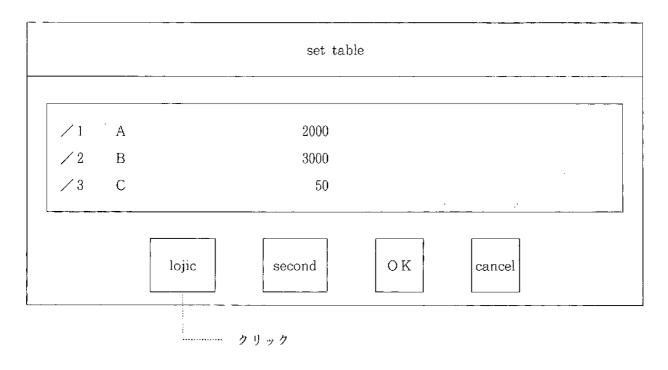


図4-31 検索集合一覧ダイアログボックス

・ 全てのANDをとる。

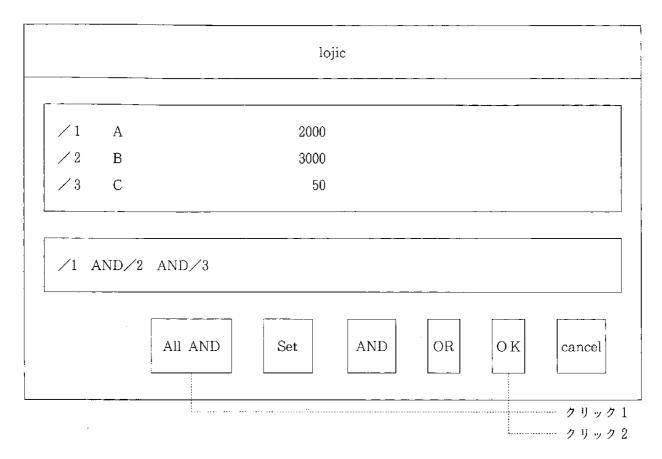


図4-32 検索集合論理演算ダイアログボックス

絞り込み検索を行う。

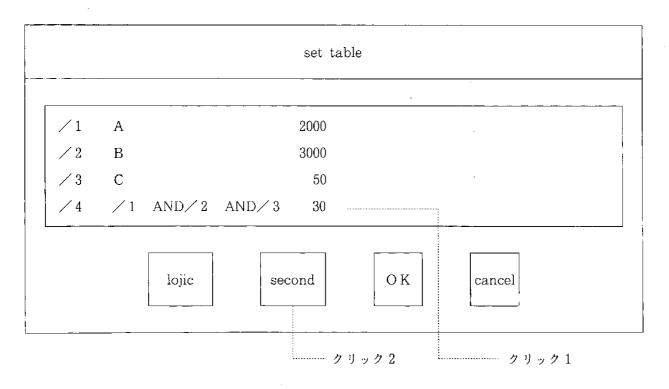


図4-33 検索集合一覧ダイアログボックス

• 絞り込み条件を加える。

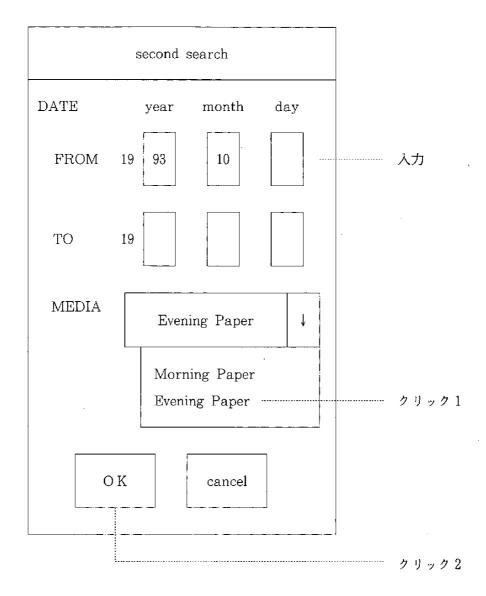


図4-34 絞り込み検索ダイアログボックス

タイトル表示を指示する。

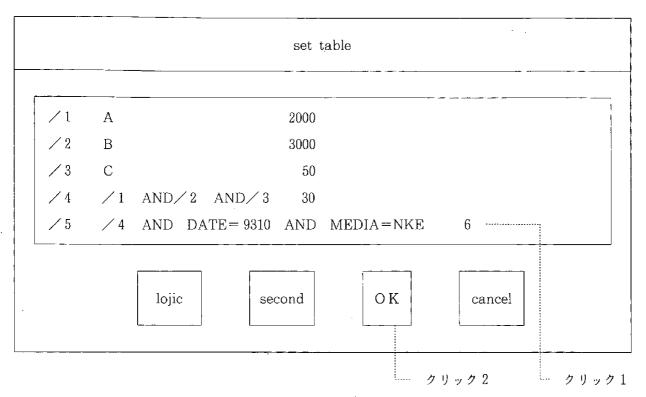


図4-35 検索集合一覧ダイアログボックス

記事1の内容表示を指示する。

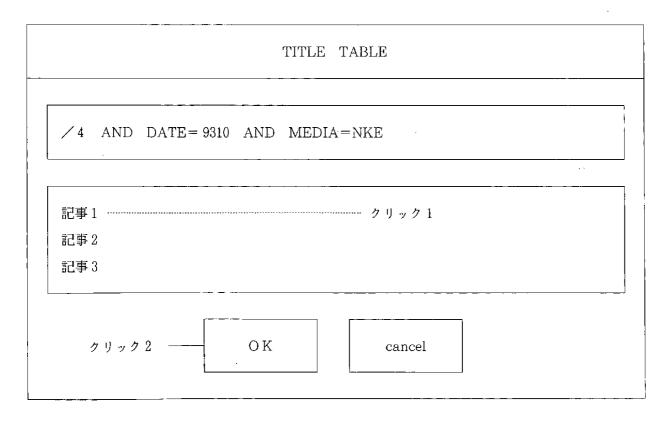


図4-36 タイトル一覧ダイアログボックス

・ 内容を表示する

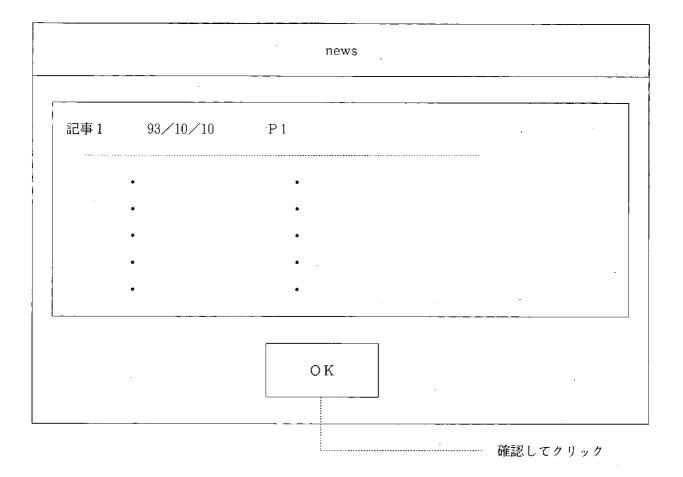


図4-37 内容表示ダイアログボックス

4.5 検索コマンド生成機能

従来の、データ検索は、ユーザの判断でコマンド入力し、検索手続きを進めてきたが、本プロトタイプ・システムは、MS-Windows のユーザ・インターフェースを利用して、コマンド入力をしなくてもよい形式にしている。

コマンド入力は、メニューやボタン・コントロールをマウスでクリックする操作が、コマンドのパラメータの入力は、リストボックスに表示されている項目を選択する操作が、それぞれ、ほぼ対応している。主に使用するコマンドは、日経テレコン・コマンド体系の「FIND」コマンドと「SHOW」コマンドである。

(1) コマンドの構文

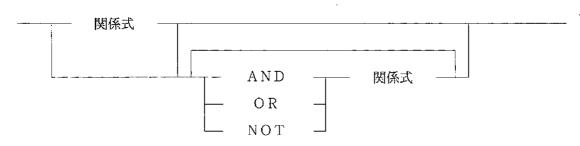
「FIND」コマンドは、検索を行う機能をもち、「SHOW」コマンドは内容を表示する機能をもつ。 使用しないパラメータを排除した構文は次のとおりである。

① 「FIND」コマンド

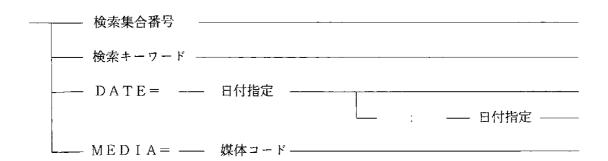
〔構 文〕

FIND —— 検索式 ———

〔検索式の構文〕



〔関係式の構文〕



〔日付指定のパターン〕

年指定 YY(西曆下2桁)

•年月指定 YYMM

•年月日指定 YYMMDD

〔媒体コード〕

日経4紙

NKS

(日本経済新聞、日経産業新聞、日経流通新聞、日経金融新聞)

• 日本経済新聞朝刊 NKM

• 日本経済新聞夕刊 NKE

• 日経産業新聞 NSS

• 日経流通新聞

NRS

日経金融新聞 NKS

② 「SHOW」 コマンド

〔構 文〕

一 出力形式 —— 検索集合番号 SHOW -

番号指定

〔出力形式〕

- 見出し一覧 P1
- 記事本文出力 P 3

(2) 検索キーワードの個別検索・

ユーザが最初に行う検索は、検索キーワードによる検索である。検索キーワード作成支援体系で作 成した検索キーワード(英語)を和訳したものでの検索である。本プロトタイプ・システムでは、後 で行う複数検索集合の論理演算を自由に行うために、複数の和訳したキーワードに対して、1回ずつ 検索を行う。例えば、和訳後のキーワードが J1、J2、J3、J4、J5 の 5 件あるときは、

FIND J1

FIND J2

FIND J3

FIND J4

FIND J5

を連続して発行する。この場合、1回、1回、検索がヒットしたキーワードの検索集合が作成される。

すべて、検索がヒットしたときは、/1、/2、/3、/4、/5、という 5 件の検索集合が作成され、あとの論理演算のときにそのままの記号で使用される。

(3) 検索集合同士の論理演算

和訳後の検索キーワード J1、J2、J3、J4、J5を個別に検索したときに作成される検索集合/1、/2、/3、/4、/5 は、検索手続きを進めていく上で、1つの検索集合に集約していくことになる。通常は、すべての検索集合に対して AND をとる。この場合、

FIND /1 AND /2 AND /3 AND /4 AND /5 という構文になる。言葉にすると、 「キーワード J1、J2、J3、J4、J5全てをもつ記事を検索する」または「検索集合/1、/2、/3、/4、/5全てに存在する記事を検索する」といった表現になるだろう。

しかし、もし/2 と/3 が/1 の関連語である場合のように、単純に AND をとることが適当でないとき(検索集合/1、/2、/3、だけは、OR をとりたい)は、次のような構文も考えられる。

FIND /1 OR /2 OR /3 AND /4 AND /5

という構文になる。言葉にすると、「検索集合/1、/2、/3 を合わせた集合と/4、/5 に共通に存在する記事を検索する」といった表現になる。

(4) 特定検索集合に対する条件検索

検索集合同士の論理演算の結果、/6という新しい検索集合が作成されたとする。このとき、検索集合/6に含まれる記事の件数がまだ多いときは、再度絞り込まなければならない。記事を絞り込むために与える条件は、収録日付の範囲と収録媒体の種類(新聞の種類)である。絞り込み検索の構文は、次のようになる。

① 日付で絞り込む例

FIND / 6 AND DATE = 9310 (1993 年 10 月に掲載された記事だけを検索する)

② 媒体で絞り込む例

FIND / 6 AND MEDIA = NKE (日経の夕刊に掲載された記事だけを検索する)

このような、条件検索を繰り返すことによって、ヒット件数を絞り込む。

(5) タイトルの表示と内容の表示

データベースのユーザは、欲しい情報を速く手に入れることが目的である。すなわち、自分が見た

い新聞記事を見ることが検索の最終的な目的ということになる。特定検索集合に対する条件検索をユーザが納得がいくまで繰り返すことによって得られた検索集合の内容を見たいとき、集合に含まれている記事のタイトルをすべて表示し、その中から見たい記事のタイトルを選択するという手続きが行われる。

ユーザは、最終の検索集合から記事の内容を見るまでに、タイトルの表示、内容の表示という 2 段階のコマンドの発行を行うことになる。タイトルの表示と内容の表示は、意味は違うが、同じ「SHOW」コマンドを使用する。

例えば、最終的に/7という検索集合が作成されたとき、タイトル表示の構文は、次のようになる。

SHOW P1 /7

P1は、タイトル出力を指示するパラメータである。

その結果、

記事タイトル1

記事タイトル2

記事タイトル3

の3件のタイトルが得られたとする。ユーザは、この中から内容を見たい記事を選択することになる。 ここで、「記事タイトル2」の内容を見たいときの構文は、次のようになる。

SHOW P3 /7 2

P3は、内容出力を指示するパラメータで、「2」という番号指定は、記事タイトルの2番目という意味である。

4.6 通信制御機能

オンライン・データベース・サービスを利用するうえで通信は必要不可欠である。通常は、パソコンで通信ソフトを稼働させて検索を行う。本プロトタイプ・システムは、ユーザ・インターフェースに重点を置いているので、MS-Windows を採用している。MS-Windows アプリケーションにも通信ソフトは、いくつかあり、そのままでも利用できる。しかし、本プロトタイプ・システムは、キーワード作成支援機能も含んでいるので通信機能を内蔵すべきであり、しかも、通信ソフト並みの機能が要求される。

MS-Windows で通信をサポートすることは、簡単ではない。通信は、パソコンのシリアルポートにモデムを接続して、行われる。特に、受信データは、転送速度がクロックスピードに比べると遅く、いつデータが着信するか分からない。通信は、割り込み制御なので、MS-Windows アプリケーションより処理が優先される。 しかし、MS-Windows は、複数アプリケーションが同時に実行できるという構造原則がある。受信データのチェックを単なるループ構造にすると他の実行中のアプリケーションの処理が停止してしまう。こうした問題を解決するために、MS-Windows のタイマー制御を

使用するのが一つの方法である。タイマー制御とは、指定した間隔(通信の場合は 0.1 秒程度)に1度アプリケーションにメッセージを返す機能である。すなわち、タイマーのメッセージが帰ってくるタイミングで受信データのチェックを行い、他のアプリケーションに迷惑をかけないという思想である。こうした思想で、データベース検索での受信を行う。また、通信ソフトにもある機能で、オートパイロット機能、通信条件設定機能を用意している。

(1) オートパイロット機能

オートパイロット機能は、回線の接続からデータベースの検索を開始するまでの間を自動運転する 機能である。ユーザが入力する部分は簡易ファイルに納められている。

処理の流れは、次のとおりである。

- ① DCB(データ・コントロール・ブロック)作成通信条件設定機能で入力されている通信条件をDCB(データ・コントロール・ブロック)にセットする。
- ② モデムの初期化ATコマンドでモデムを初期化する。
- ③ 回線の接続 ATコマンドに電話番号を付加して送信する。
- ④ ログインログイン名、パスワード、チャージ・コードを簡易ファイルから取り出し送信する。
- ⑤ データベース・メニューの選択番号入力 データベース独自でオンライン表示するメニューの選択番号を簡易ファイルから取り出し送信する。

簡易ファイルには、本来、通信ソフトウェアを使用したデータベース検索処理では、手で入力する項目をテキスト形式で格納している。

(2) 通信条件設定機能

データベースによって、1秒間に転送できる情報量の差や、コード体系の違い等、通信に関する情報をファイルに格納する機能である。ここで扱う項目は、次のとおりである。

- センター名
- ② 電話番号
- ③ コメント
- ④ 簡易ファイル
- ⑤ リダイアル回数

- ⑥ リダイアル間隔
- ⑦ 転送速度
- ⑧ キャラクタ長
- ⑨ ストップビット
- ⑩ パリティ
- ① Xon/Xoff制御

以上のような、情報を、ダイアログボックスから設定し、ファイルに格納する。これらの情報は、 オートパイロット機能の実行しているときに DCB (データ・コントロール・ブロック)にセットされる。

5. 実用システムに向けての課題

本課題で、作成したプロトタイプ・システムは、「海外運用システム」と名乗っていて、海外からわが国のデータベースを検索することを目的にしているが、今後のシステムの拡張によっては、わが国から海外のデータベースを検索することも可能である。また、日本語、英語以外も利用できる可能性がある。これは、本プロトタイプ・システムがウィンドウズ・アプリケーションである点が大きく、関連アプリケーションを追加することもできるし、画面やメッセージなど視覚に訴える部分の修正が、アプリケーションのプロセス部分から独立しているので比較的容易である。本プロトタイプ・システムを便利な検索サポート・ツールにするために、次の切り口から実用システムに向けての課題を考察した。それは、ハードウェア、ソフトウェア、データベースの現在状況及び将来について検討した「外的環境に関する課題」、システム自身の機能について検討した「システム的な課題」、英日内部シソーラスの内容、拡張性、用途について検討した「内部シソーラスについての課題」である。

5.1 外的環境に関する課題

現在、パーソナル・コンピュータは、大きく分けて二つの潮流がある。一つは、IBM-PCから始まった MS-DOS を基本 OS とした DOS 系マシン勢力、もう一つは、アップル社のマッキントッシュの勢力である。マッキントッシュ系のマシンは、初めから、 GUI (グラフィカル・ユーザ・インターフェース) を基本とした OS で画面上のアイコンをマウスでクリックすることで動作し、ユーザ・インターフェースに非常に優れている。しかし、アップル社のマシンでしかプログラムは動かない。オブジェクト互換性、データ互換性がともにない。 DOS 系のマシンは、別メーカーの互換機が台頭してきている。

本プロトタイプ・システムで使用するハードウェアは、DOS/V が標準である海外製のパーソナル・コンピュータであり、低価格化がもっとも進行している。オープン化が既に UNIX の代名詞という時代ではなくなり、価格の低いパーソナル・コンピュータ上で WINDOWS を稼働しても充分実現できるようになった。ハードウェア、ソフトウェアの高性能化により、メモリ容量の問題、ディスク容量の問題、実行速度の問題などの心配がなくなってきている。反対にユーザの立場からは、選択する幅が増えたと言うことで、ハードウェアとソフトウェアのコーディネイト能力が重要な位置を占めてきたと言えるだろう。

次に、わが国のオンライン・データ通信サービス産業の問題にも触れなければならない。民間がデータベースを構築する際には、経済性が最優先される。そのために、コストに見合わないからといって、必要と思われても構築されないデータベースも数多いのではないだろうか。公共的にも必要あると思われるデータベースは、政府が政策的な助成をする方針をとっているが、これが商業ベースにのる保証はない。また、ユーザ側も、データベースから得られる情報の価値基準がはっきりしていない。このような背景から、現在、わが国のデータベース・サービス産業は、普及の状況、情報の質と量、検

素の仕組み、価格など、どれをとってもアメリカに遅れをとっている。特に価格については、非常に深刻で、いくらハードウェアやソフトウェアの低価格化が進んでも、肝心のデータベース使用料が高ければ、意味がなくなる。今後のディストリビュータのコスト削減の努力が待たれるところである。

5. 2 システム的な課題

現在のデータベース・サービスのほとんどが、通信ソフトウェアによるオンラインのメニュー選択とコマンド入力の混在型である。MS-Windowsのアプリケーションの役割は、メニューを選択する行為とコマンド入力する行為をダイアログボックスで行う。すなわち、ユーザの操作が介入する部分を可能な限りマウスのクリックで行い、キーボード入力は、単語入力程度の最小限にとどめるという考え方である。これらを、踏まえて、MS-Windows上で稼働するために作成したアプリケーションの開発の問題点とシステムの今後の方向性について考察する。

(1) 通信の制御

MS-Windows 用の通信ソフトウェアは、市販されているが、データベース・サービスを使用する場合、検索の手続きに従って、操作を選んでいくので、すべてが手入力になり、MS-DOS 用の通信ソフトウェアの操作と変わらない。しかし、コマンドの入力は、ある程度データベースのホスト側の機能を限定(使用するコマンドを最小限にする)すれば、ダイアログボックスのボタンをマウスで選択する方法がコマンド入力の代わりになり、ユーザにとっても使いやすい。したがって、MS-Windows操作性を充分利用するためには通信制御機能を内蔵した形態のシステムを構築するべきであろう。

(2) コマンドの自動生成

本課題で開発したシステムは、プロトタイプ・システムの立場から対象とするデータベースを一つ にしているが、ユーザは、複数のデータベースを利用することが想定される。

どの、データベースも検索の考え方は、似ているが、検索コマンドと検索手続が微妙に異なっている。通信ソフトウェアは、簡易ファイルをカスタマイズすることでログインまでのコマンド入力の自動化が可能だが、データベースを検索する段階になると、ユーザ独自の判断が必要な場面が多くなってくるため、データベース毎に異なった作り方をしなければならない。外に見える入出力のインターフェイスをとる部分は、それぞれのデータベースの特長であるデータソースの種類によって変化するし、内部的なコマンド生成は、コマンド体系によって変化してくる。また、データベースによっては、ホスト側の設定をコマンド入力によって行えるものがある。この場合、ユーザがどのような設定を行っているかを判定することは、非常に難しい。

このように、複数のデータベースの相手をするときの拡張性がこれからの課題である。

(3) 今後の方向性

本プロトタイプシステムは、DOS/V、MS-Windows 環境上で開発しており、AT 互換機や PC98 系のハードウェアを対象としている。これらは、世界的なシェアをもち、今後、ますます高性能化、低価格化が進むことが期待される。実際、ここ1、2年の間に発売された AT 互換機は、処理速度の高速化(CPU のバージョンアップ)、メモリ容量、ディスク容量の大容量化には、目を見張るものがある。また、従来の、デスクトップ型パーソナル・コンピュータの他にノートブック・パーソナル・コンピュータも普及してきてユーザの好みにより色々な組み合わせが可能になった。

ソフトウェア面から見ると、MS-Windows を使用しているので他の Windows アプリケーションと併用して使用することができる。このことから、検索結果のデータをワープロ・ソフトや表計算ソフトさらには機械翻訳ソフトで扱うことができるようになる。

現段階では、扱う検索結果データは、テキスト・データだが、将来的には、データがマルチメディア化(画像、音声、動画等)に向かうことが予想され、MS-Windows の特長である GUI (グラフィカル・ユーザ・インターフェイス)を高度利用した形態が実現することになるだろう。

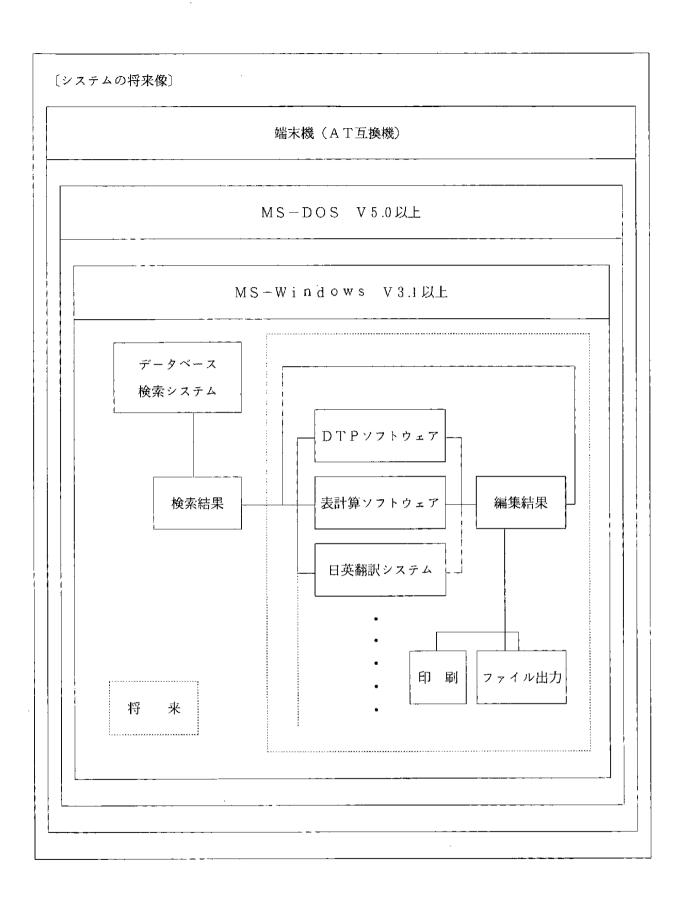


図5-1 システムの将来像

5.3 内部シソーラスについての課題

本プロトタイプシステムで採用した英日シソーラス・パイロット・モデルは、次の条件で作成されている。

- ・ 英語と日本語は1対1である。
- アルファベット順検索できる。
- 階層順に検索できる。
- 用語に含まれている単語で検索できる。

通常、ユーザは階層順に統制語を整理して掲載しているデーベース別の冊子体シソーラスを見ながら、キーワードを入力している。したがって、シソーラスの用語は、上位概念語から下位概念語を階層順に検索する方法が一般的であろう。こうした、階層構造は、カテゴリー別に分類しているので上位概念語を外してしまったら、目的の用語を探し出すことができない。このようなときの補完機能として、用語のアルファベット順検索、用語に含まれている単語のアルファベット順検索を用意している。しかし、データ検索の専門家は、検索コストを安くするとか、必要な情報を正確に取得するといった観点から、色々なアプローチを用意していると思われる。多角的な検索を実現するには、英日シソーラスは、まだ改善の余地がある。

将来的な改善点として、次のような方向が考えられる。

- 英語と日本語の関係は、1対1だけでは無く、1対多、多対1もある。また、日本語に有り英語に無い概念、英語に有り日本語に無い概念がある。この問題をどこまで解決できるか?
- 英語以外の外国語対応。
- 現在、日本語シソーラスを英訳しているが、海外ユーザから見れば、自国のデータベースのシ ソーラスを基軸とした内部シソーラスが欲しいのではないか?
- 用語の利用頻度の統計をとり、ほとんど使われない用語についての存続を検討する機能を用意する必要があるのではないか?
- わが国のデータベースは、半角カタカナのキーワードで検索するものが、まだ多いのでカタカナ変換機能を追加する必要が出る。

5. 4 プロトタイプ・システムの評価

本課題で作成したプロトタイプ・システムの特長は、英語圏のユーザが日本語データベースを利用するための便宜を図ったところにある。通常のデータベースの利用法はメニューの選択またはコマンドを入力することにより行われる。以下に、「海外運用システム調査研究委員会」等で議論されたことについて記す。

(1) キーワードの作成及び変換について

キーワードの新規入力は、コマンド入力と同じなので、当初は、利用頻度が低いことを想定していたが、実際の海外ユーザは、タイプライタを使い慣れている人々が多いこともありコマンド入力に対しての違和感はない。このことから検索キーワードの新規入力はもっと重要視しなければならず、現段階より自由な入力が必要となってくることが分かった。たとえば、エディット・コントロール上でホワイトボードに字を書くような入力ができ、その上で複写、切りとり、貼り付け等の編集を行うといった環境が望ましい。但し、操作の流れとして、このことが適正かどうかは、今後の研究課題である。新規入力の他の検索キーワード作成は、アルファベット順の検索、上位概念語からの検索、単語による検索がある。これらは、冊子体の英日シソーラスの編集方法に合わせたものだが、ダイアログボックスからデータを検索する仕組みは、あまり存在しないもので、一つの提案としては価値のあるものだと思われる。英語キーワードの日本語変換は、現在、見たいときに見れる状態にあるが、常時表示されていた方がいいと言う意見が多かった。

(2) 検索について

第一段階では検索キーワード毎の検索を行い、第二段階では、検索集合の論理演算を行い、最後に 日付やメディア等の条件検索を行うという、一連の検索プロセスは、コマンド入力で行う検索をほぼ カバーしている点については、それほどの意見や要望はでなかったが、将来的に、質問の意味を解析 して質問式を自動生成するといった、一括検索モードの考え方についても研究すべきだろう。

(3) 通信について

本プロトタイプ・システムでは、通信状況を表示しないで通信を行っていた。外的要因による障害 の確率が高い通信は、状況をモニタリングする機能が必要ではないかとの意見が多かった。

なお、本格的な評価については、近い将来予定している、在日外国人による試験使用の結果及び海外の展示会でのデモンストレーションの結果を待つことになる。

6. おわりに

わが国のオンライン・データ通信サービス産業は、不況の現在も拡大傾向にある。しかし、わが国程の経済規模があればもっと普及していても不思議ではない。その一例として「データベース白書 1993」に次のような表が掲載されている。

国・地域	データベース売上高(億円) ※1ドル125円換算	1 人あたりの利用額(円) ※ 1 ドル125円換算		
アメリカ(含むカナダ)	12,761	5, 050		
日 本	2, 160	1, 743		
イギリス	2, 157	3, 760		
フランス	6 1 3	1, 074		
ドイッ	4 8 2	6 0 7		
西ヨーロッパ合計	4, 451	1, 292		

表 6-1 [データベース・サービス市場規模の国際比較 (1991年)]

わが国のデータベース産業の売上高は、上記の表のとおりアメリカ(含むカナダ)についで2位だが、およそ6分の1にすぎない。また、売上高でわが国より下回るイギリスも1人あたりの利用額は大きく上回っている。フランスにしても売上高の割には、1人あたりの利用額が多い。このことからも、データベースが国内での普及状況が欧米と比較してよくないと思われる。

海外からわが国のデータベースを利用することになると、国内の普及不足に加えて、データベース の内容が日本語ということが余計に普及の阻害要因になっている。

現状では、日本の情報を求めている海外ユーザで日本語を理解できる人は少数派であるが、欧米での日本語学習熱や機会増加がみられることから、今後日本語を理解する海外ユーザが増加するものと思われる。この認識をもとに本プロトタイプ・システムでは「日本語を理解できる海外ユーザを対象としている。しかし、これらのユーザは、検索結果は理解できても、その検索結果を得るまでの過程で入力する検索条件の日本語入力(ことによったら半角カタカナ入力)ということになると、利用には、かなり消極的になるに違いない。

欧米アルファベット文化圏で育ってきたコンピュータ文化に日本語という異質なコンピュータ文化を取り入れるために本課題のプロトタイプ・システムは、「キーワードの英語入力」という観点に絞った開発をした。

端末機をDOS/Vパーソナル・ゴンピュータにしたのは、

- 海外製のものが使える
- ソフトウェア依存の日本語入力ができる
- ・低価格であること

などの理由である。MS-Windows を採用したのは、ユーザ・インターフェースをよくするためであるとともに将来のアプリケーション(翻訳ソフト、DTP ソフト等)追加を想定したものである。

このような、安くて使いやすいパーソナル・コンピュータ端末上で本課題で作成したような、バイリンガルな検索サポート・ツールを利用することにより、わが国のデータベース・サービス産業のますますの普及、振興を期待したい。

6. 1 データベース普及に関する提言

本課題で作成したような検索サポート・ツールが、二つの点で弱点を抱えていることが判明している。一つは、データベースサービス産業の現在の普及状況であり、もう一つは、各データベース・ディストリビュータのコマンド体系の変更への対応である。本課題の目的は国外のデータベース・ユーザにわが国のデータベースを利用してもらうところにあり、さらに突き詰めれば、わが国のデータベースサービス産業の普及振興の一助となることを念願している。このような観点から、わが国のデータベース産業および関連する産業に対して以下のように提言する。

提言1. 検索コマンド体系の標準化

現在、データベースは、独自の検索コマンド体系で運用されている。このことは、扱っているデータの種類も異なることもあり、ある程度やむをえない部分もあるが、このことが実現することにより、

- ① 色々なデータベースを利用したいという要望も多くなる。
- ② ゲートウェイが利用し易くなる。
- ③ 操作性のよい検索サポート・ツールの存在価値が上がる。
- ④ 初心者が利用し易い土壌ができる。

等の効果が見込まれる。

実現する基準の一例として次の項目を挙げる。

- ① 検索操作の共通点を可能な限り洗い出す。
- ② ほとんどのデータベースの検索を可能にすることを目指す。
- ③ 従来の検索コマンド体系も残す。
- ④ 特殊なデータを扱うなどデータベースの性質上、検索手順が特殊なものに関しては、最低限の標準化にとどめる。

- ⑤ データベースの個性を損なわないことを考慮に入れた標準化を行う。
- ⑥ 国際標準に則った標準化を進める。

提言 2. シソーラスの標準化

検索コマンド体系と同様にシソーラスもデータベースによって異なる。このこともユーザを混乱させる原因の一つになっている。検索キーワードにする用語は、同じ意味でも分野によっては違う表現をしているものもある一方、同じ表現でも分野によって意味の異なる用語もあるため、何らかの指標が必要である。現状での改善策としては、シソーラスの標準フォーマットを作成して、個々のデータベースにおけるシソーラスの構成要素を標準化するとともに、各シソーラス間で類義語を相互参照できるようにする等が考えられる。勿論、従来のデータベース独自の統制語キーワードは、残しておいても構わない。

また、カタカナのキーワードしか受け付けないデータベースがあるので、これも早急な改善が待たれる。

提言3. 通信環境の整備

わが国のデータベースの通信速度は、回線の事情から 2400bps でも速い方である。欧米では、9600 bps が普通である。将来データベースは、なお一層のマルチメディア化が進むことが予想され、そのためには、短い時間に大量の情報を流さなければならない。現在の回線網から、21 世紀を見据えた、全国光ファイバー網を構築することが重要である。このような、インフォメーション・ハイウェイ構想は、データベース・サービス産業のみならず他の産業の活性化および新たな産業の発生にも貢献するものであると考えられる。ぜひ、政府主体でこのような事業が行われることを期待したい。

提言4. 流 通 面

個々のデータベース・ディストリビュータは、固有の検索コマンド体系や情報電送手順を公開して、 自社を含め複数の業者に検索サポートツールを作成させ、ユーザに選択させることが望ましい。これ によって、操作性のよいものが出てくれば、コマンド入力を嫌っていた潜在ユーザ層が利用を開始す ることが期待される。

データベース・ディストリビュータ各社は、データの質、量、価格、スピード、サービスや利用法の提案といったソフト面で競争し、ユーザ数の拡大を図ることが望ましい。

6. 2 海外運用システム調査研究委員会

本課題を実施するにあたり、有識者による「海外運用システム調査研究委員会」を設置し、海外のデータベースの事情、開発プロトタイプ・システムへの要望、わが国データベースに対する注文など、3回に渡って審議した。プロトタイプ・システムは、当委員会での意見を重視して開発したものである。

第1回 平成5年9月14日 火曜日

第2回 平成5年10月29日 金曜日

第3回 平成6年1月20日 木曜日

「海外運用システム調査研究委員会」名簿

委員長 安達 淳 学術情報センター助教授

委員 窪田春男 リソースシェアリング株式会社取締役

データベース技術部長

古 関 正 裕 日本経済新聞社データバンク局研究委員

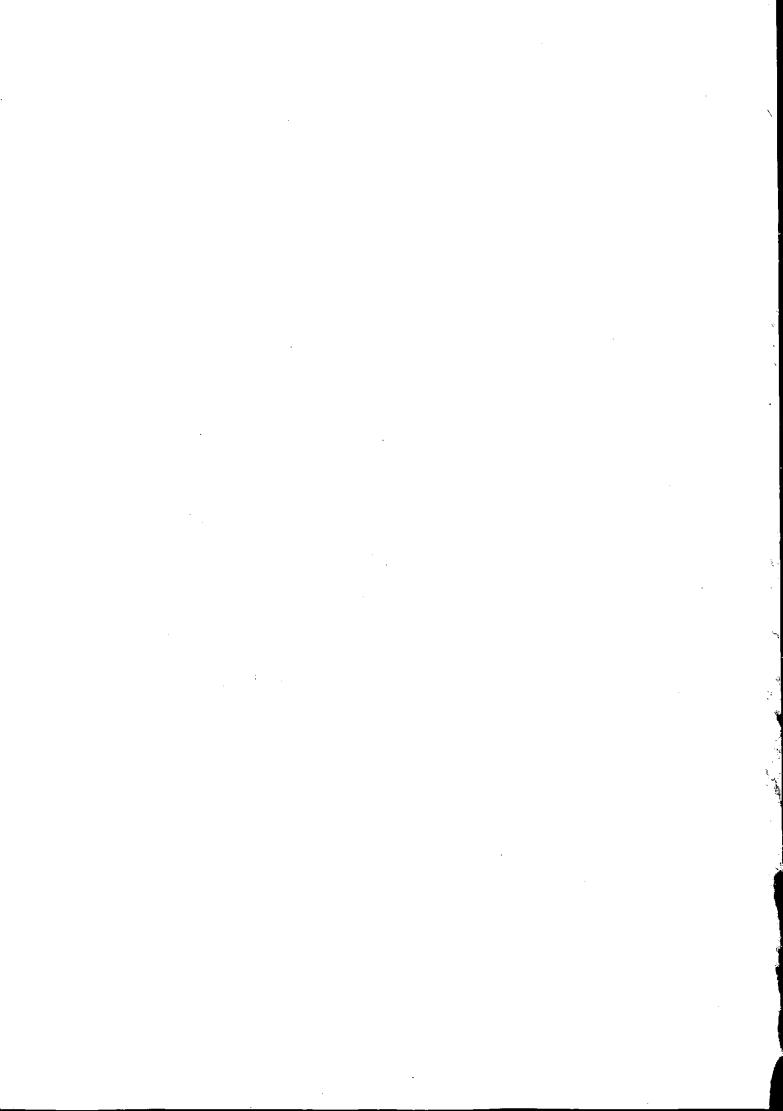
白岩一哉 情報図書館RUKIT副館長

鈴 木 実 財団法人データベース振興センター振興課長

高 浜 忠 彦 カテナ株式会社常務取締役SI事業本部長

三 輪 眞木子 株式会社エポックリサーチ代表取締役社長

事務局 カテナ株式会社



- 禁無断転載 ─

平成6年3月発行

発 行 財団法人 データベース振興センター 東京都浜松町2丁目4番1号 世界貿易センタービル7階 TM (03) 3459 - 8581

委託先 カテナ株式会社 東京都江東区潮見2丁目10番24号 ILL (03) 3615 - 3211

印刷所 株式会社 進 栄 社 東京都台東区谷中3丁目18番4号

	•		

