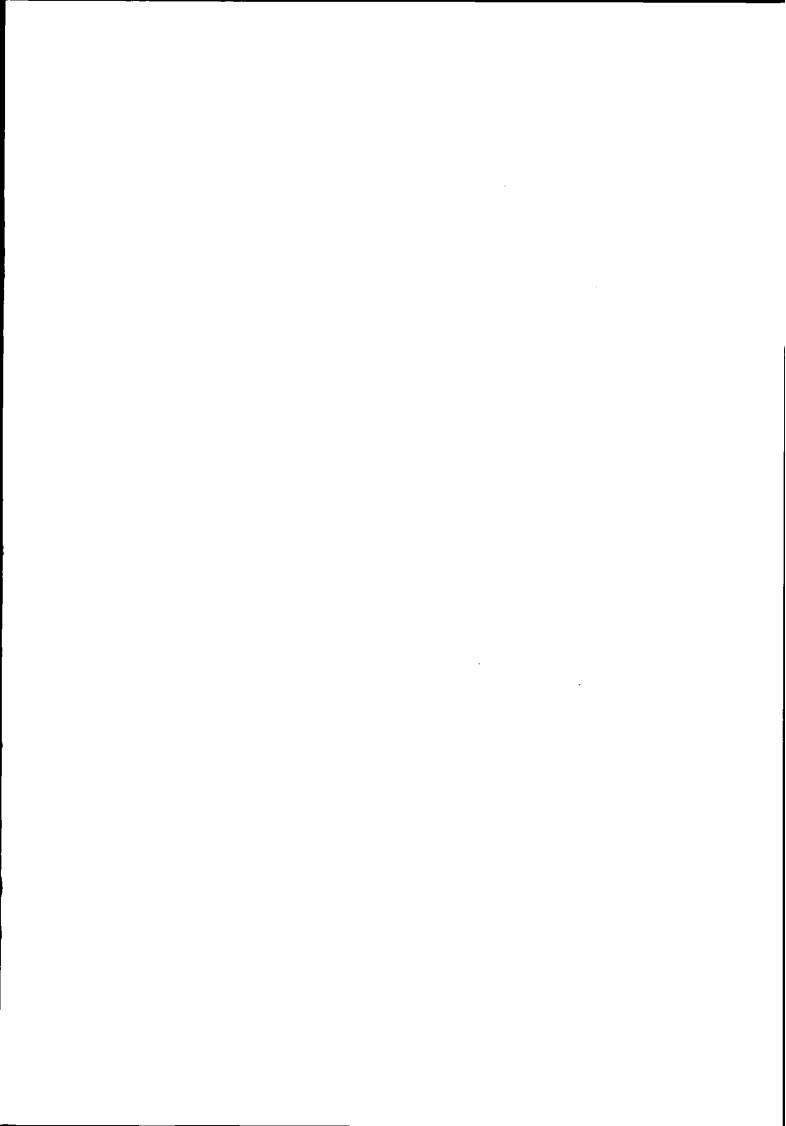
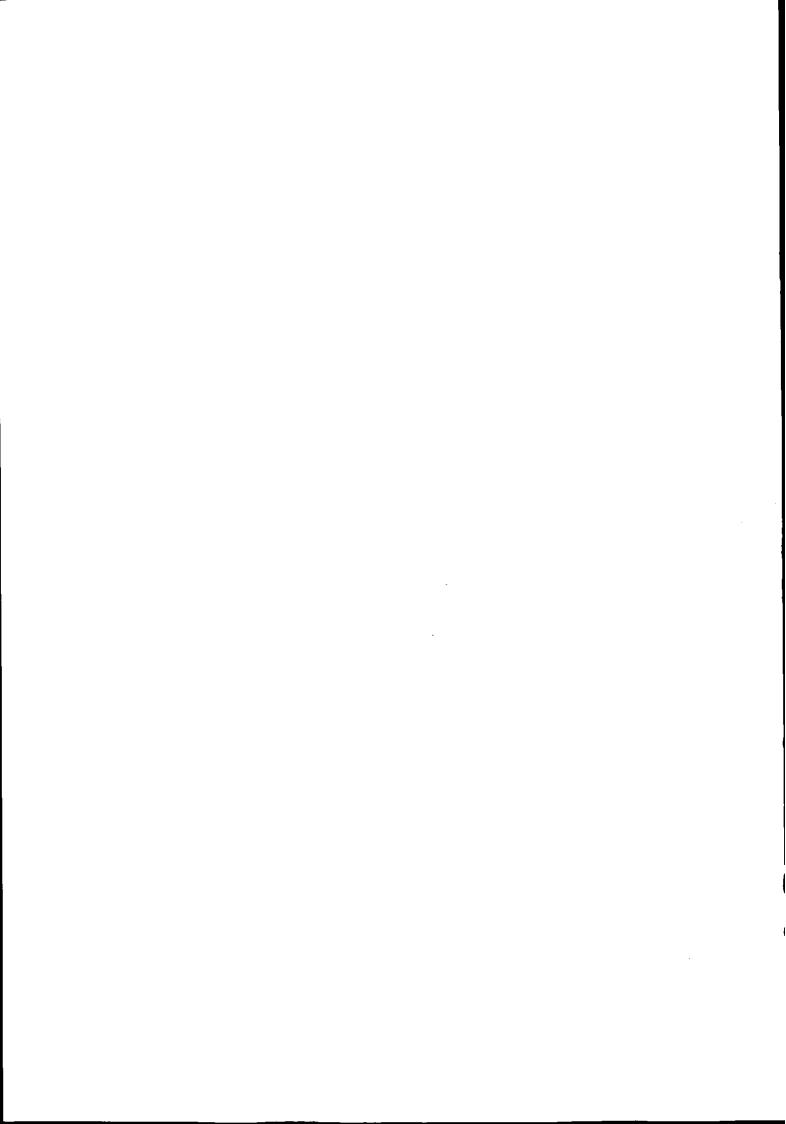
の1-開一19



本報告書は、日本自転車振興会から競輪収益の一部である機械工業振興資金の 補助を受けて作成したものである。





データベースは、わが国の情報化の進展上、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 / 4 にすぎないのが現状であり、わが国データベ ースサービスひいてはバランスある情報産業の健全な発展を図るためには、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強力に促進し、データベースの 拡充を図る必要がある。

とのような要請に応えるため、例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の促進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 東京工科大学教授 西野博二氏)を設置している。

この「異種データから構成されるデータベースの総合的処理技術に関する調査研究」は平成元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、当財団がリソースシェアリング㈱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立てば幸いである。

なお、平成元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表のとおりである。

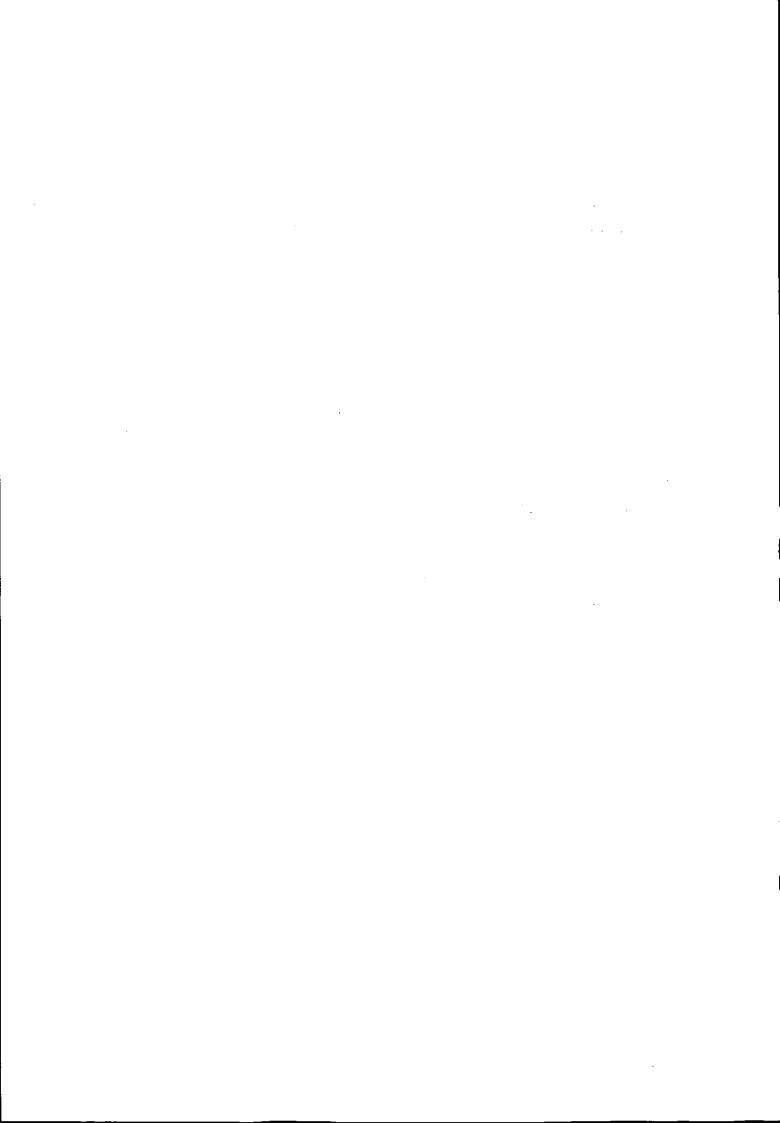
平成2年3月

平成元年度データベース構築・技術開発促進委託課題

分!	野		課	題	名
		1 気候情報	データベース	の構築	
	ļ	2 電磁波環	境障害に関す	·るデータベース	マの構築
		3 災害情報	シソーラスの	構築	
		4 意味情報	を中核とした	医療評価データ	スペースとコミュニケーシ
社	会	ョンシステ	ムの構築		
		5 ハンディ	キャップバー	ソンの情報ニー	- ズに即したライフサポー
		トデータベ	ースの構築		
		6 博物館情	報データベー	スシステムの構	孝 樂
		7 中央省庁	での電子計算	機利用に関する	る報告書のデータベース化
		8 沖縄地域	における文化	:情報データベー	- スの構築
 地 域 活 性	+ 11	9 九州地域	の人材情報デ	ータベース構築	奏に関する調査研究
中小企業物		10 高岡市商	圏データベー	スの構築	
了了。 	*	11 地域の物	産・人材・文	工化情報のデータ	タベース構築と新しい地域
		間交流推進	に関する調査	研究	
地	図	12 マルチメ	ディア型地区]データベース構	ちりゅう ちゅう ちゅう ちゅう ちゅう はいま はいま はいま はいま はい
エネルギー・	資源	13 燃焼技術	と燃焼装置設	計のデータベー	- ス作成
	14	14 技術支援	システムにお	·ける産業機械部	邓品データベースの構築
部品• 杉	才料	15 マイクロ	コンピュータ	のプログラマフ	プル周辺デバイスのデータ
		ベース構築			
┃ ┃ ┃ 標 準	化 化 17	16 イオンク	ロマトグラフ	'ィー・データ^	ベースの構築
12x +4=		17 CD-R	O M マルチメ	ディアデータン	フォーマットの調査
海	外	18 データベ	ース構築のた	こめのターミノロ	コジーの調査研究
		19 異種デー	タから構成さ	[れるデータペー	- スの総合的処理技術に関
技		する調査研	究		
		20 バイナリ	モデルに基つ	がく先端的文書を	検索システムの開発

		·	
	·		

I		査研究の概要	
	1.	調査研究の目的	1
	2.	実施内容および成果	1
ľ	蓚	[合的データベース・システムの全体像	
	1.	マンマシン・インターフェイス	5
		推 論 機 能	
	3.	ハイパーメディア表現	2 2
Ш		D-ROM用検索システム	
	1.	C D - R O M 検索ソフトの機能調査	3 9
	2.	CD-ROM用検索システム機能の高度化	5 9
IV	· .	う後の課題	
	1.	複合的データベースの C D - R O M 上での実現にむけて	7 0
	2	データペース的アプローチによるシステム構築の問題点	7 2



Ⅰ 調査研究の概要

1. 調査研究の目的

データベース処理に係るソフトウェア諸技術は、オンライン環境でのデータ量の増大化とデータ処理の高速化という要求に応えるべく、ハードウエアとしての演算装置や記憶装置の進歩と相まって、大きな発展を遂げている。同時に、官公庁や民間企業におけるデータベースの利用分野も大いに拡大し、データベースの利用方式に係る要求も「メーカーから提供されたデータベース操作手順をそのまま利用する」というシステム指向型から「業務に適した手順でデータベースを操作する」という業務指向型に移行している。

このような状況に対処するためには、データベース処理技術も、単種データに対して定型的な処理を行う従来の処理方式(同様の処理を大量に繰り返すというコンピュータが本来得意とする処理方式)だけではなく、種々の業務で発生する多種多様な表現形式を持つデータを複数の分散したデータベースに格納し、随時発生する業務目的に応じた臨機応変な総合的データ処理に対応していくことが必要になる。このような異種データから構成される分散データベースに対して、データ間の関係をも考慮した総合的かつ複合的なデータ処理を実現するためには、これまでに開発されてきたデータベース処理技術のほかに、

- ① 多種多様なデータをデータベースに蓄積し検索する技術
- ② データベースの中の多様なデータを総合的に利用する技術等の技術開発が重要な課題となっている。

このような状況に鑑み、本調査研究は、データベースの利用形態と処理技術に関する現状と要望を調査し、業務指向型の複合的データベース処理に係る技術開発を行うことを目的とすると同時に、今後ますます普及していくことが予想される上記のようなデータベースの高度利用に対して大局的かつ技術的な指針を示すことを目指している。

2. 実施内容および成果

上記目的に沿った「異種データから構成されるデータベース」を有効かつ高度に利用するためのシステム体系モデル、さらにはそのモデルに基づいたソフトウエア・パッケージとしての複合的データベース・システムは、時間的空間的な処理能力あるいはデータの表現能力というような面から、最終的には、大型汎用コンピュータの上で稼動することを想定しているが、当面の目標としては、パソコン上で、CD-ROMを含めたような多様な

入出力媒体を総合的に制御するようなシステム構成を想定している。そのシステム・イメージとしては、次のような機能と特徴とを持つことを期待している。

- (1) 各種統計データ、技術文献データ、社内通知データ、会議スケジュール等のような多様な業務で発生する異なる種類のデータを、統一的な分散データベースに格納し利用することができる。
- (2) 単種データに対する定型的業務処理だけでなく、随時発生する業務目的に応じた臨機 応変な非定型データ処理に対応することができる。
- (3) 種々のデータを、数値情報、文字情報、文章情報、図表情報、文書情報、画像情報、イメージ情報、音声情報、などのデータの性質に応じた種々の表現形式を用いて記述し、データベースに記録し、利用することができる。
- (4) データベース中に格納されているデータから、個々のデータの内容に基づいて所定の情報を検索(内容検索)したり、アクセス・パスとしての利用者ビューに基づいて所定の情報を検索(構造検索)したりすることができる。
- (5) リレーション型データベース、演繹データベース、オブジェクト指向型データベース、マルチメディア・データベース、など新しいデータベース処理技術とその機能を採用する。
- (6) 異種データを操作するための複合的データベース・システムを、ひとつのデータベース・モデルとして参照することによって、データベース高度利用体系の導入にあたっての指針を得ることができる。

これらの機能を実現するためには、現行のデータベース処理技術を越えたいくつかの技術的問題を解決していくことが必要となろう。このために、本調査研究は、昨年度を初年度として、3段階からなる3年間の期間を想定している。第2年度にあたる今年度は、昨年度の複合的データベース・システムの有すべき機能の検討を踏まえて、本システムの全体像とパソコン(CD-ROM)上での構築に係る問題点とについて調査研究を行った。

具体的には、今年度、「異種データから構成されるデータベースの総合的処理技術」 に関して調査研究を行った項目は次のとおりである。

(1) 複合的データベース・システムの構成

異種データから構成されるデータベースの総合的処理技術体系としての複合的データ

ベース・システムにおいては、「多種多様なデータをデータベースに蓄積する技術」及び「データベース中の多様なデータを総合的に利用する技術」等が、重要な技術要素となることは言うまでもないが、これらの技術を実現していくためには、リレーション型データベース・モデルをはじめとして、意味データベース、オプジェクト指向型データベース、あるいはマルチメディア・データベース等の新たなデータベース処理技術を採用していく必要があろう。そこで、これらの新しいデータベース処理技術の調査を行い、複合的データベース・システムの構成要素としての位置づけについて検討を行い、本システムの全体像の明確化を図った。

との調査検討の成果として、次の事項が明らかになった。

- ① データ入力、データ検索、データ出力というような、本システムにおける各種機能モジュールごとのマンマシン・インターフェイスの有すべき機能
- ② A I 技術とデータベース技術との統合体としての演繹データベースの役割と複合的 データベース・システムの中での位置づけ
- ③ 文書情報のように、種々のデータ表現(データ項目)から構成される事象(オブジェクト)に対しての、ハイパーメディア表現の必要性と、オブジェクト指向型データベースの考え方の意義

との成果は、「Ⅱ、複合的データベース・システムの全体像」として取りまとめた。

(2) 複合的データベース・システムのシステム設計

本調査研究で対象としている複合的データベース・システムは、将来的には、パソコンン上での構築を目指している。そこで、最近、その高密度性、経済性、量産性から、パソコンの記憶媒体として脚光を浴びているCD-ROMに注目し、CD-ROM用検索ソフト機能の現状調査を行い、その調査結果を踏まえて、検索機能の高度化に係る検討を行った。

この調査検討の成果として、次の事項が明らかになった。

- ① 国産CD-ROMに関する検索ソフトの現状と特長
- ② CD-ROMによるデータベース化の特性
- ③ CD-ROM検索機能の、従来型データベース検索機能に比べての優位性 この成果は、「Ⅲ. CD-ROM用検索システム」として取りまとめた。

3. 複合的データベース・システムの開発にあたっての留意点

「複合的データベース・システムの構成」に係る調査検討の成果として得られた本システムの特化した機能を踏まえつつ、パソコン(CD-ROM)上でシステム構築する場合の問題点と技術的課題について検討を行った

この検討において、次の事項が「今後の課題」として明らかになった。

- ① CD-ROMの特性を生かすことを前提とし、複合的データベース・システム構成の中での、その記憶メディアとしての位置づけ
- ② システム設計において、データベースをシステムの中心に位置づける、いわゆるデータペース的アプローチを行う場合の、各種データの表現と、オプジェクト指向型データベース採用の意義

この成果は、「N. 今後の課題」として取りまとめた。

Ⅱ 複合的データベース・システムの全体像

異種データから構成されるデータベースの総合的処理技術体系としての複合的データベース・システムにおいては、「多種多様なデータをデータベースに蓄積し検索する技術」及び「データベース中の多様なデータを総合的に利用する技術」等が、重要な技術要素となることは言うまでもないが、これらの技術を実現していくためには、リレーション型データベース・モデルをはじめとして、意味データベース、オプジェクト指向型データベース、演繹データベースあるいはマルチメディア・データベース等の新たなデータベース処理技術を採用していく必要があろう。また、データベース利用者が広い層にわたるため、専門的なデータベース操作手順とは異なったいわば一般利用者のための業務指向型(または問題解決型)のデータベース操作手順を備えたようなマンマシン・インターフェイスが重要な役割を果すことになる。

ことでは、このような複合的データベースシステムの全体像を把握するという観点から、

- ① マンマシン・インターフェイス
- ② 推論機能
- ③ ハイパーメディア表現

という、従来のデータベース・システムの考え方から特化した要素について検討した。ことに挙げる要素は、複合的データベース・システムが本来的に有すべき機能であり、実現すべきコンピュータの規模や種類とは独立したものである。したがって、後述するように、本システムは、最終的には、バソコン上でのパイロット・システム構築を目指しているが、その際には、この機能を基本にして規模をバソコン向きに設計する必要があろう。

1. マンマシン・インターフェイス

複合的データベース・システムにおいては、利用者層は一般利用者すなわち業務担当者が、それぞれの業務処理の一環としてデータベースから必要な情報を取り出したり、あるいはデータベースに記録されている情報を更新したりすることを想定している。従って、データベース操作手段としてのインターフェイスも、業務指向型の手続きであることが必要となる。言い換えれば、一般利用者が、それぞれの業務で用いている言語や書式、あるいは処理手順にのつとってデータベースが利用できる環境を設定することが、このインターフェイスに託された役割であると言えよう。このためには、

- ① データベース操作の自動化あるいはロボット化によるユーザ・フレンドリーなシステムの設計
- ② 分散しているデータベース群のネットワーク化によるデータベース・アクセスに対する時間感覚あるいは距離感覚の克服
- ③ データ表現やデータベース操作に対する標準化あるいは知識ベース化によるデータベース利用者間の格差(人種、文化、年齢、性別、等)の克服
- ④ オフィス・オートメーション化による(データベース構成をも含む)システム環境の 広域的普及

等の諸点を考慮することが重要な課題となるが、これらを実現するための技術的課題として「データの自動解析技術」、「情報表現の自動変換技術」および「情報内容の自動解析技術」がある。

第1の「データの自動解析技術」は、主に、数値データから構成されるデータベースに 対して、種々の数値解析技法を用いてデータの解析をおこならものであり、官公庁におけ る各種調査票の集計や統計処理、あるいは企業における売上等の集計や時系列解析、多変 量解析などがこれにあたる。このように、データベースを中心とした最近の情報システム においては、従来のファイル処理システムと違って、データとソフトウェアとの総合的な 結合を図る傾向にある。言い換えれば、従来のファイル処理システムでは、個々のデータ はそれぞれ独立した形で(データ間の関係が存在しない形で)ファイル中に格納されてお り、データ検索においては所定のデータをファイルから取り出して利用者に提供するとい **う形態をとっていた。これに対して、情報システム設計へのデータベース的アプローチと** も呼ばれる方式で構築されている最近の情報システムにおいては、ファイル操作とデータ 処理とを分離して把えるのではなく、あくまでも業務処理体系を構成する要素として有機 的に結合した形で構成することになる。このような環境の中では、情報システムの利用者 は、業務処理の種別あるいは手順を表現することによって、関連するデータ群とソフトウ ェア群とが自動的かつ有機的に組み合わされ、所定の業務が遂行されることになる。さら には、最近のオフィス・オートメーション化によって種々のハードウェア、例えば図形表 示装置、画像表示装置、三次元ディスプレイ装置などの入出力装置や、端末装置用固定デ ィスク、光ディスク、フロッピー・ディスクなどの記憶装置が普及しており、先に述べた .データを入力したり表示・出力したりするための各種入出力装置と組み合わされることに なる。言い換えれば、コンピュータ・システムに係る3大資源であるハードウェア、ソフ

トウェアおよびデータベースを、それぞれの業務に適った形で自由にかつ有効に組み合わせることを可能にすることが、今後の情報システム設計の重要な要件となろう。これによって、例えば「ある商品の最近5年間の売上げと来年と再来年の売上予測が知りたい」というような業務要求に対して、データの取り出し、時系列解析、解析結果のグラフ表示という3段階の処理を独立して行うのではなく、先の業務要求「ある商品の最近……売上予測が知りたい」を、そのままの形で情報システムに入力し、所定の編集形式で結果を得るための環境が整うことになろう。しかし、このような環境が完全に実現するには、以下に述べるような知識ベースを中心とした人工知能技術の発展に期するところも大きい。

第2の「情報表現の自動変換技術」は、データの入出力あるいは情報要求に際して、情報の表現形態を自動変換することによって、使い勝手のよい利用者インターフェイスを提供することである。とくにデータベースの利用に係る利用者のインターフェイスのための情報表現自動変換手続には次のようなものがある。

① 業務データのシステム内部表現への変換

業務で発生した情報(原票)を基に、所定のデータをデータベースに蓄積したり、あるいはデータベースの内容を更新したりする場合に、メニュー方式やコード化技法を用いながら出来るだけ原票に近い形でシステムに入力できるようなインターフェイスを作成し、一方システム内部では、データベース・システムの特性やソフトウェアとの結合形態を考慮したシステムの内部表現に入力データを変換する。

② 情報要求言語の変換

データベースを中心とした情報システムに対する情報要求(帳票や資料の作成要求等)は、通常、業務処理の内容に沿った形での日常言語に近い業務処理言語やメニュー方式によって表現される。この情報要求言語は、情報システムの中で、データ検索手続、データ加工手続、ソフトウェア結合手続と言ったような手続言語に変換され、コンピュータ内部での情報要求は処理手順を規定した手続き型言語であることが多く、情報要求言語の変換においては、非手続き型言語から手続き型言語への変換という形態をとることになる。これは現在のコンピュータがノイマン型であることに帰因する宿命的なデータ処理技術とも言えよう。

③ 回答結果表現の変換

データベースに対する情報要求に基づいて検索された情報は、情報の蓄積形式(記憶構造)そのままの形で利用者に提供されることのほかに、情報の表現形式を変換し、利

用者にとって利用し易い形式あるいは業務帳票の様式に適った形式で表示したり印刷し たりすることがある。とくに最近の情報システムに対する要求として、業務指向型ある いは問題指向型といった体系が強調されており、後者のような情報表現変換機能が必要 不可欠となっている。この情報表現自動変換技術には、文字セットの変換、定型的デー タ項目値の変換、非定型的データ項目値の変換というようないくつかのレベルがある低 かに、数値データからグラフや図形のような2次元の表現形態に変換する手法も含まれ よう。文字セットの変換技法は、例えばJISコード体系を他のコード体系に変換する というような変換技法であり、通常、一文字(1バイトあるいは2バイト)対応での変 換を行うものである。また、定型的データ項目値の変換技法には、ENCODE -DECODE 処理のように各種コード(産業分類コード、都道府県コード、企業コード、商品コード 等) をそれぞれの名称に変換するための技法、数値データに対して桁記号や正負符号を 付すような編集技法、YYMMDDというような6桁の数値で表現されている年月日項目 に対して「元号」、「年」、「月」、「日」を付すような編集技法などの変換技法があ る。このレベルでの変換技法は、データベース・システムの記憶構造(システム内での 効率性に基づく表現形態をとっている)と、利用者側あるいは業務に即した情報の表現 形態とのギャップを埋めるための、言わば形式的データ変換と言えよう。このように、 コードと言うような一つの情報単位の形式や処理単位があらかじめ定まっている定型的 データに対して、化学構造式や文章のように、要素の組み合せ方(文法)が意味をもつ ような情報が非定型的データである。従って、このようなタイプのデータを処理したり、 他の表現形式に変換するような場合には、情報の処理単位は独立した要素(化学構造式 の場合には原子や基が、また文章の場合には単語がこれに対応する。)だけではなく、 それらの要素間の構造をも加味した情報処理が必要となる。例えば、日本語と英語と言 うような、自然言語間での自動(機械)翻訳機能などがこの典型的な例であり、次に述 べる「情報内容の自動解析技術」と密接な関連をもつととになる。また、最近では図形 やグラフを表示するための出力装置(あるいは図形処理用端末ソフトウェア)が普及し ており、データベースから取り出した検索結果を、単に数字で表示するだけではなく、 棒グラフや円グラフというような各種のビジネス・グラフを用いた視覚的な表現にする ことが可能となっている。

第3の「情報内容の自動解析技術」は、先にも述べたように第2の技術が言わばデータ値の 形式的処理であったのに対して、データ値の意味内容をも考慮した情報処理技術と言える ものである。言い換えれば、非定型的データとして表現されているデータ値の意味あいを解析したり、あるいは、データベースに格納されている複数の情報を総合的に判断したりすることがこの技術の特徴と言えよう。このような情報処理技術は、文章の意味、複数情報からの総合的判断、あるいは情報間の連想、あるいは情報に基づく推論、と言ったような人間のもつ知的活動に類似した情報処理であり、従来、AI(人工知能)の分野で研究が進められてきた技術であるが、最近では、多種多様な情報を格納し、管理するデータベース処理技術とリンクすることにより、データベースの利用体系を高度なものにしている。このようなタイプのデータ処理においては、記憶媒体上に格納されている情報(レコード)を独立したものとして扱うだけでなく、業務処理に対比してレコード内に有機的な関係を付与し、その関係(利用者ビュー)に基づいて総合的な判断を含むデータ処理を行うこととなる。

とのような複合的データベースシステムのマンマシン・インターフェイスをシステム構成モジュール別にみるとつぎのようになる。

(1) データ入力インターフェイス

データベース構築にあたっての第一歩はデータ入力ということになるが、先にも述べたとおり、このデータ入力機能も多様化している。ところで、データベースに蓄積するためのデータ入力機能を実現するツールとしてのデータ入力インターフェイスの発展には、大別して2つの流れがある。一つはOSIを背景とした業務指向型のデータベース利用システムの一環としてのデータ入力インターフェイスである。これは業務処理手続きの中で文書を作成したり、他の部所へ転送したり、あるいは文書を保存したりする事務手続きを、このコンピュータを利用したデータ入力かインターフェイスで実現するものである。従って、ここでのデータ入力手続きは、事務手続きの内容に即していること、コンピュータに関して特別の知識を有しない業務担当者が操作可能なこと、などが基本的な要件となる。このためには、

- ① わかり易い入力画面の設定
- ② 仮名(ローマ字)漢字変換機能
- ③ データ(全部または一部)のコピー機能
- ④ 入力エラーのチェック・再入力機能

などの機能を用意する必要がある。ととで重要なことは、①、②、③のように、データ

入力の簡便性を実現するための、言わばワープロ型の機能のほかに、④のように、データベースの整合性を維持すること、すなわち、データベース中に不正なデータが入り込んだり、データベース中のデータに対して不正な修正や更新が行われたりすることも防ぐための機能を用意することである。

データ入力インターフェイスの流れの中でのもう一つは、高度なデータ入力技術を背景としたものである。これは、多種多様な表現形態をもつ一次情報を、正確かつ簡便な方法でデータベースに入力するための技術であり、イメージ・スキャナーやOCR装置のようなハードウェアに依存する方法と、ダウン・ローディング(アップ・ローディング)や自動索引のようなソフトウェアに依存する方法とがある。とくに後者の自動索引技術は、情報の自動分類技術などと共に、情報の意味内容に係る処理であり、人工知能の分野とも密接な関係をもつことになろう。

そもそもデータベース処理技術は、後でも述べるように、人工知能技術と大きな関連をもっており、データベース中に蓄積されたデータの意味解析(内容分析)、データ表現の自動翻訳、データベース中に蓄積されたデータ群(種々の情報)に基づく情報の推論処理などが脚光を浴びている。このような技術が実現すれば、日常言語によるデータの入力や問合せ、データベース(システム)と会話をしながら種々の問題解決にあたるOAシステムなどが現実のものとなろう。

」(2) データ検索インターフェイス

データベースに蓄積されている大量かつ多様なデータから、情報要求に合致した情報 (データ)を取り出す技術としてのデータ検索技術は、データ入力技術やデータ編集技術と共に、通常データベース処理技術の歴史点発展の中で中心点な役割を担ってきたと言っても過言ではない。通常、データベース処理技術の発展段階は、次に挙げるような 4つの世代に分けて把えられている。これはコンピュータの発展過程を第一世代、第二世代、第三世代、第四世代、そして現在研究中の第五世代コンピュータと、その中心的 な技術(真空管、トランジスタ、IC、LSI)で特徴づけて分類しているのと同じような考え方に基づいている。

① 第一世代(IOCS(入出力制御システム)の時代

利用者がデータベースから、必要なデータ(レコード)を取り出すにあたって、そのレコードの記憶媒体上での物理的位置(絶対アドレス)を指定する方法であり、そ

の絶対アドレスに基づいているレコードを取り出すツールがIOCSと呼ばれる基本 的なデータ操作ルーチンとなっている。従って、利用者はデータ検索にあたって、必 要とする個々のレコードの絶対アドレスを知っている必要があるということになる。 これは言わば、図書館(データベース)で読みたい本(情報)を探すにあたって、そ の本の保存されている棚位置を知っていることが前提となるようなデータ検索方法で あり、棚の構造や本の並べ方を知らない一般利用者が利用するには、多少難かしい方 法と言えよう。

② 第二世代(アクセス手法の時代)

SAM(遂次アクセス方法)、DAM(直接アクセス手法)、ISAM(索引付遂次アクセス手法)、VSAM(仮想アクセス手法)などのアクセス手法を用いて、データペースから所定のレコードを取り出すのが、この第二世代のデータ検索技術である。この手法を利用することによって、利用者は相対的なレコード位置、あるいは、データ値に基づいてデータを取り出すことが可能となっている。前述した第一世代の側で言えば、図書館で読みたい本を探す場合には、その本のタイトルや発行所を指定することによって、あるいは、著者名を「アイウエオ順」に取り出すことによって、所定の本を見つけることになる。このアクセス手法は現在でもよく使われているデータ検索手法であり、利用者が、データペース中でのレコードの記憶位置(データペースの物理構造)を知っている必要はなく、タイトルとか発行所とか著者名とか言うようなデータの属性(値)に基づいてデータ検索が行える点で、ある程度利用者向の手法と言える。ところで、これまでに述べたデータペース処理技術の第一世代と第二世代においては、データペースに蓄積されているレコードはそれぞれ独立した(相互の関係をもたない)ものと看做している。

③ 第三世代(アクセス・バスの時代)

通常、情報源としてエンティティ(例えば「本」)はいくつかの属性(例えば、「本 に対する属性としての「タイトル」、「発行所」、「著名者」がある)によって表現 されそれぞれの属性に対する属性値の列(上の例で言えば「データベース入門:日本 出版社:日本太郎」となる)がレコードと呼ばれ、データベースに蓄積される。とと で、データベース構築の対象となる情報源(エンティティの集合)に対して定義され る「属性の組」は業務によって異なるレコードと関係づけて(一つの組レコードとし て)処理されることがある。例えば、「本の発注」と言う業務では、「本レコード」 と「発行所レコード(発行所に係る属性から構成される)」とを組として処理することになろう。このように、データベースを利用する業務を主体として考えた場合の、業務処理で必要とするレコードに対する属性の組、およびレコード間の関係をデータベースの論理構造と呼んでいる。もちろん、このデータベースの論理構造は、データベースを利用する業務の種類の数だけ用意されることになる。言い換えれば、データベース利用者は、それぞれの業務処理に対応する論理構造を選択し、その論理構造に沿ってデータ検索を行うことになる。このようなデータベースの論理構造に沿ったデータのアクセス手段がアクセス・バスと呼ばれるものである。現在、このアクセス・バスに基づいた第三世代のデータベース処理技術が最も普及しており、多くのデータベース・システムはこの技術(設計思想)に基づいて構築されている。この考え方によれば、データベース利用者は、それぞれのデータベース利用にあたって、論理構造のみ注目することになり、そのような観点からすると、データベースの物理構造と論理構造とが独立したものとなることから、この世代は「データ独立の時代」とも呼ばれる。

④ 第四世代(リレーションの時代)

第三世代のデータベース・システム環境においては、先にも述べたとおり、データベースの論理構造という概念を導入することによって、「データの独立」の考え方を実現しているが、この論理構造は、データベースを利用する種々の業務目的あるいは業務処理手続きに沿った形で、実際のデータベースを利用する業務処理の実行に先だって、あらかじめ定義しておく方法がとられる。多くの場合は、データベースの生成時にこの論理構造が定義される。ところが、このような定義方法では、データベースの設計時、あるいはデータベースの生成時に、そのデータベースを利用する業務、あるいはデータベースの利用形態を想定し設定しなければならない。もちろん、データベースを利用する業務が定常的であり処理内容も定型的なものであれば、この方法で充分であり、かつ便利な方法であろう。しかし、あらかじめ作成されているデータベースを多目的かつ汎用的に利用していくような場合には、このように、データベースの論理構造をあらかじめ固定した形に設定する方法は、必ずしも適していないことは明らかである。そこで、データベースを種々の目的で、言わば臨機応変に利用できるようにすることを目的として提案されたデータベースシステム(モデル)が、この第四世代のリレーショナル・データベース・システムである。このモデルの考え方は、

リレーションと呼ぶフラットなデータ構造(レコードの集合)に対して、リレーショ ン代数言語とリレーション論理言語という2種類のデータ検索(データ操作)言語を 用意している。前者の代数言語は、レコードの集合としてのリレーションに対して、 集合稿、集合差などの集合演算のほかに、射影処理(プロジェクション)、結合処理 (ジョイン)、制限処理(レストリクション)、および商(デビジョン)という4つ の代数演算を用意している。この代数演算では、個々のレコードに対するデータ検索 処理のほかに、複数のレコードを結合したりするデータ操作処理が含まれており、こ のデータ操作によって、データベーズ中のレコード群が臨機応変に結合されることに なる。ここでのデータ検索は、これまでの世代のデータ検索が単にデータベース中の レコードを取り出すと言う操作であったのに対して、複数のレコード(情報)から新 たたレコード(情報)を(推論や加工をして)作り出すという操作になっている。ま た後者の論理言語は、一階述語論理系を構成する言語であり、人工知能の分野におけ る言語処理、推論処理、QAシステム等で利用されているものである。ところで、こ こでとのモデルにとって大変重要な点は、この2種類の言語系が「意味的に同値」と なっていることである。言い換えれば、リレーション代数言語をもちいて行うデータ 検索処理によって得られる情報は、リレーション論理言語をもちいても得られ、また 逆も成り立つということである。このことによって、データベースを日常言語(例え は日本語)で検索したり、データベース中に蓄積されている情報群を総合的に分析・ 推論して、所定の情報を得るというようなデータベース利用環境実現への道が大きく 開けたと言えよう。

このように、リレーション型データベースは、データベース・システムのインプリメンテーションとは独立した、いわばひとつの論理的なデータモデルとして体系をはなしている点で、従来型の、いわば経験的データベース・システムに比べて優れていると言えよう。このような観点からすると、リレーション型データモデルはデータベースの体系(データ構造とデータ操作)に対する理論的基礎を提供しており、ある意味で一時代をなしたと言えよう。ところが、これに対して、現実世界を自然な形で表現しようという試みがあり、それには現実世界の事象をデータ構造とデータ操作の組(オブジェクト)として記述し、それに対するデータ処理を可能とするような手続きが提案されており、オブジェクト指向型データベースと呼ばれている。このほかに、データベースキーに記述し、その記述に従って一種の推論をおこなうような意味データベース、あるいは、推

論規則と命題群とから、主に、一階の述語論理子に沿ってデータの演繹をおこなり演繹データベースなどが提案されている。これらのデータベース・モデルは、現実世界を自然な形で記述したり、推論機能を導入したりしている点で、それぞれの長所があるが、それぞれの基本になるモデル(体系)が異なっていることから総合的にモデル化することは、必ずしも容易ではないが、オブジェクト指向型マルチメディア知識ベースというような試みもある。複合的データベース・システムにおいても、順次、上記のようなそれぞれのモデルの長所を取り入れていくことに留意していくべきであろう。

このようなデータベース処理技術の発展段階を見ると明らかなように、その大きな方向は、利用者中心型データベース利用環境への指向、言い換えれば業務処理の一環としてのデータベース処理であり、日常の活動の中での(自然言語を用いた)データベース処理であり、人間と同じような知能をもったデータベース処理であると言えよう。この第四世代のリレーショナル・データベース・システムは、最近いくつかが実用化されてきているという段階であるが、今後、増々普及することが考えられる。

(3) データ蓄積インターフェイス

コンピュータ処理の適用分野は、官公庁、産業界、研究機関、学校等ばかりではなく、家庭や個人にまで広がっている。この理由としては、ハードウェアの多様化や低廉化、ソフトウェアの多様化、コンピュータ・コミュニケーション技術の発展などが挙げられるが、その中で、情報の表現形態としてのデータ・タイプの多様化も見逃すことのできない要素である。これによって、コンピュータ処理の対象となる種々の業務分野で扱う情報(エンティティ)の表現形態(データ属性)に対応するデータ・タイプが豊富になり、次のようなものが処理可能となっている。

- ① 数值(固定小数点、浮動小数点、等)
- ② コード(各種コード、商品コード、住所コード、等)
- ③ 文字(住所、氏名、企業名、文献名、商品名、会議名、キーワード、等)
- ④ 文章(資料内容、法令条文、文献抄録、会議録、等)
- ⑤ 表(各種統計表、日程表、財務諸表、各種対応表、税率表、等)
- ⑥ 図(地図、設計図、図柄、等)
- ⑦ 画像(航空写真、顕微鏡写真、等)

このほかに、音声表現、色彩表現などもコンピュータ処理の対象として研究され、実

用化されつつある。このような豊富なデータ・タイプを利用することによって、統計情報、文献情報、企業情報、人事情報、不動産情報、商品情報、等々の情報が適格な形で表現されデータベースに蓄積され、利用されるわけである。もちろん、一つの情報(エンティティ)を表現するためのデータ属性群についてみると、ある属性は数値として、またある属性は文章として、そしてまたある属性は画像として表現されると言うように、複数種類のデータ・タイプの組合せによって表現される。例えば、不動産情報としてマンション(エンティティ)を例にとると、住所は文字タイプ、価格は数値タイプ、居住地は文章タイプ、外観は画像タイプ、付近地図は図タイプと言うように表現されよう。複合的データベース・システムでは、このようなデータ属性に対応するデータ・タイプの特性に応じて、記憶媒体と操作手続きを用意することになる。このような要求を満たしたデータベース体系としてマルチメディア・データベースやハイバーメディア・データベースがある。

(4) データ出力インターフェイス

複合的データベース・システムの基本的考え方として、業務処理に密着した利用者指向型ということが挙げられ、データの入力機能、蓄積機能、検索機能においても、この考え方に基づいたツールが用意されることは先に述べたとおりであるが、データの表示機能においても例外ではない。このデータの表示機能を実現するためのデータ処理は編集処理と出力処理とから構成される。複合的データベース・システムにおいては、レコード中の項目値(たとえば、人事レコードの「社員名」や「所属」など)が、質問条件と合致しているものを取り出し表示するだけでなく、特定の関係にあるレコード群を取り出したり、場合によってはデータベース中のレコード群から特定の命題(質問)を推論したりしながら、質問に対する回答(情報)を得ると同時に、その回答を編集し、適当なメディアに数値、文字、文章、表、図、画像、音声などの表現で出力するほかに、中間ファイルをインターフェイスとして各種のソフトウェアとの結合を可能とする。

① データの編集

複合的データベースから検索したデータ群を、メディア表現あるいはファイル表現 に適った形で編集する。前者のメディア表現のための編集は、検索結果として得られ た汎用形式のレコードをディスプレイや印刷装置などの表示装置や、データ通信やFAX 転送などの通信装置のインターフェイス形式に編集する。 この 場合 この 編集 方式は これらの装置のインテリジェント性能に負うところも大きい。たとえば表示装置では、ホストコンピュータで最終的な図表表示形式に編集をおこなう場合と、ホストコンピュータでは表示装置に組み込まれた図表表示ソフトウェアに対する入力データ形式に編集するにとどめ、最終的な図表形式には、表示装置側で変換する場合とがある。またデータ通信では、とくにOSI通信インターフェイスを意識して、ホスト側でOSI形式変換をおこなう場合と、FEP(フロント・エンド・プロセッサ)などの装置で形式変換する場合とがある。また、後者のファイル表現のための編集は、検索結果データをファイル経由で各種ソフトウェアに結合させることを目的としている。この中間ファイルの形式としては、大別して2つのものがある。ひとつは、利用するアブリケーション・プログラムの入力形式に合わせた形式に変換する方法であり、もうひとつはつぎのような抽象データ形式

- (i) スカラー・データ
- (ji) ベクトル・データ
- (前) 3次元データ
- (V) 4次元以上の高次元データ

に変換し、アプリケーション・プログラム側で個有のデータ形式に変換する方法である。

② データの出力

データの出力機能は、言うまでもなく、複合データベースからの検索結果を、何らかの出力媒体(メディア)上に記録することである。したがって、この機能は、前述した「データの編集」機能とも密接に関連している。通常、この「データの出力」機能においては、出力対象となるデータの内容に係る編集は行わずに、出力媒体に依存する制御データを付与して、所定のメディアにデータを伝送することになる。言い換えれば、複合データベース・システムの検索結果に対する「出力制御」を行うことである。ここで重要なことは、同一のデータ内容(「データの出力」機能では関知しない)であっても、この出力制御処理によって、ある場合は、数値が印刷されたり、表形式になったり、さらにはグラフ表示で出力されたりすることである。

③ CTS編集

最近は、各種情報の流通機能としての(オンライン)データベースと印刷物とが密接に関連している。たとえば、新聞記事、技術文献、特許情報、統計情報など各種の

情報がデータベース化されてオンライン提供サービスの対象となっていると同時に、印刷物でも提供されている。このことの背景には、コンピュータ印刷(CTS)の実現がある。すなわち印刷物を作る際の基本となる版下作成をコンピュータでおこなうことから、この過程でマシン・リーダブルな形式で表現される情報をブログラムをもちいて自動的にオンライン・データベース化することから、言わば印刷過程の副産物としてデータベースが作成されたことによる。ところが最近では、この流れは逆転し、新しい情報を随時データベースに入力することによって、データベースを更新していき最新情報をオンラインで提供することが中心となり、そこで作成されたデータベースから必要な情報を取り出して版下を作成するという流れになっている。この過程では、レイアウト構造と呼ばれる印刷ベージ上での文字の形式、位置あるいは改行などに係る情報を付与することになる。すなわち、ワープロ作成文書のようなレイアウト構造をもつ情報と、「文書名」や「日付」などのようなデータ項目から構成されるデータの論理構造と変換手続きを用意することが、複合的データベースとCTSとのインターフェイスになる。

2. 推 論 機 能

データベースの高度利用を目指す複合的データベース・システムにおいては、データベースをひとつの世界(知識の集積)として扱い、そこから利用者の要求に適った情報(知識)を取り出すことが期待される。このような期待に答えるためには、前述したマンマシン・インターフェイスのほかに、データベース操作の一環としての推論機能の実現が望まれる。複合的データベース・システムにおけるデータ操作(処理)の一環として、この推論機能を用意することによって、データベース利用の高度化が飛躍的に発展することになる。すなわち、従来型のデータ検索機能のように、データベース中の情報(レコード)を独立した(相互に関連をもたない)事象として把え、個々に質問条件に適うレコードを独立に取り出すというデータ操作に対して、データベース中の情報(知識)全体から、質問に対する回答を連想、あるいは推論して、適確な表現で利用者に提供することになる。

コンピュータによる推論機能の実現は、当初、人工知能の分野で研究され、「定理証明」、「ゲーム」、「事実検索」、「Q-A システム」等として実験的に構築されてきた。この推論機能に対する理論的な基礎としては、「解決原理 (Resolution Principle)」や「意味ネットワーク (Semantic Network)」によるものが中心であった。ところで、

このような人工知能分野における推論体系は、対象となるデータの量すなわちデータベースに難点があり、大容量データベースとの結合を模索していたのに対して、データベースの分野では、データ操作の一環としての推論機能に難点があった。そこで、これらの人工知能分野の技術と、データベース分野の技術とを取り込んだ形でのデータ操作モデルが提案されてきた。このような人工知能分野(推論)とデータベース分野(データ検索)の統合体の例として事実検索システムがある。

事実検索システムはデータベース中に蓄えられた種々の情報から必要な情報を取り出すという意味において情報検索の一分野と考えることができる。事実検索システムが情報検索システムの中で特徴的なことはその推論機能にある。すなわちデータベース中に記憶されている情報からある事実を推論して質問に答えるわけである。この推論機能に注目すると定理証明(Theorem Proving)や問題解決(Problem Solving)などの分野と深い関連をもつことになる。実際、人工知能の分野ではQAシステム(Question and Answering System)という形で推論機能をもった情報検索システムの研究がなされてきた。これから述べる事実検索システムもQAシステムと同義の概念である。

事実検索システムは情報検索システムの中で推論機能をもつことにその特徴があるが、 これまでに提案された事実検索システム、あるいはQAシステムはこのほかに自然言語に よる会話型の入出力を指向している。

ここでは事実検索システムにおけるデータ構造とデータベースについて述べるにとどめ、 自然言語の扱いについては立入らない。 もちろんこのデータ構造あるいはデータベース中 での情報の表現形態が推論機能と深い関連をもつことは言うまでもない。 事実検索システムの概念図は図1のようになろう。

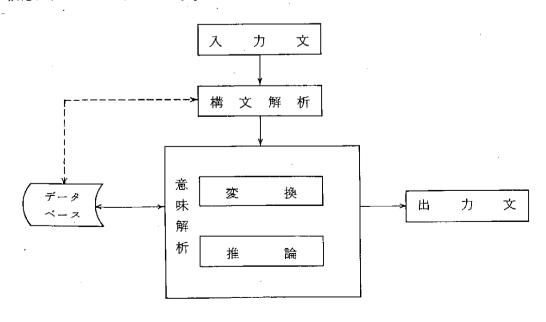


図1 事実検索システム

事実検索システムは広義の情報検索システムにおけるようにデータペース中に情報を記憶しておき、質問に対してはそのデータペースを操作することによって回答を得る。しかし事実検索システムではデータペースの操作が推論機能をもっており、操作によって得られた回答がデータペース中の情報に推論を施して得られた結果と同じ効果をもっている。この節ではデータペース中にある表現形態で記憶されている情報とそれに対する操作によってどのような推論がおこなわれるかについて述べる。

データペース中の情報は質問に対しての回答を得るための推論がおこない易いような形で記憶されている。事実検索システムでは一般的に情報はいくつかの事象の間の関係としてとらえられ、推論はこの関係にもとずいておこなわれる。したがって関係を情報の表現あるいは操作でどうあつかうかによっているいろな事実検索システムが特徴づけられるとも言えよう。

ひとつにはデータベース中に構造を陽に導入し、その構造を事実情報の中にあらわれる 関係と対応づけるようなシステムがある。この言わば構造型データベースを採用している システムでは構造をたどるという操作によって推論が実現される。たとえば家族構成に関 する情報をあつかり場合、家族構成に特有な情報構造を用いて情報を表現している。すな わち図2のような情報構造がとられている。

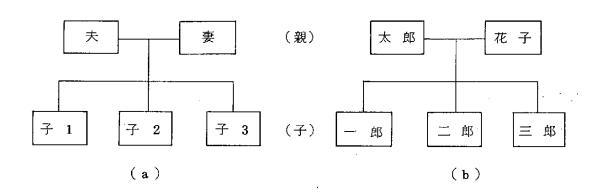


図2 家族構成の情報構造(a)とその具体値(b)

この家族構成をあらわす情報構造の一般的規則として、

- (1) 子供同士は兄弟である。
- (2) 夫の子供は妻の子供である。
- (3) 夫と妻は子供の親である。

を与えておけば

- (1) 二郎は花子の子供である。
- (2) 一郎と三郎は兄弟である。

などの事柄が構造をたどることによって導ける。しかし、このように情報構造を固定的なものにすると家族構成などでは有効であっても、一般的な事実情報をあつかうためには必ずしも有効ではない。そこである事実情報をそこに含まれる関係にもとずいてリストを作る方法もある。すなわち関係にある性質(推移性、対称性など)をもたせ、その性質を利用しながら構造をたどることによって質問に対する回答を得る。いいかえればリストをたどる過程がそのまま推論の過程になっている。たとえば「(aはbの)東にある」という位置についての関係を考えてみよう。

- (1) 東京は静岡の東にある。
- (2) 静岡は名古屋の東にある。
- (3) 名古屋は大阪の東にある。
- (4) 大阪は岡山の東にある。

という事実情報から図るのようなリストが構成できる。



図3 関係「東にある」のリスト

ここで「東にある」という関係は推移性がある。すなわち「xがyの東にある」をExyであらわせば(Ez)(Exy ∧ Ezy)→Exyという推論規則が成り立っている。この場合推移性に関する推論はリストを順にたどっていくことによって実現される。たとえば「静岡は大阪の東にあるか?」という質問に対しては、この「東にある」という関係リストをたどり「静岡」と「大阪」がこの順に出現すれば「真」でありそうでなければ「偽」(場合によっては「回答不能」のこともある)という解答が得られる。また「名古屋の東にある都市は?」という質問に対しては、やはり「東にある」という関係のリストをたどり「名古屋」の前に出現するノード(都市)の集合{東京、静岡}を得ればその集合が質問の回答となっている。このように構造型情報表現を採用しているシステムではデータベース中の情報構造をたどることによって推論がおこなわれる。

これまでに述べてきた構造型の情報表現に対して集合型による情報表現を採用しているシステムがある。Relational Data File やRelational Data Baseなど

がこれにあたる。この集合型情報表現ではひとつの事実情報をn-tupleで表現し、この tuple の集合としてデータベースを構成する。したがってこのデータベース中には陽に は構造が導入されていない。このようなデータベースに対する操作は集合演算によってな され、この集合演算によって推論が実現される。しかし事実情報や質問文は集合操作とい う形で記述されるわけではない。Relational Data FileでもRelational Data Baseでもデータベースの要素となっているひとつの事象はn-tupleであるが、質問文 は通常の一階述語論理系の式に近い形で与えられる。この質問文は意味的に同値な集合演 算に変換されるわけであるが、この変換が集合型システムの大きな問題点になっている。 たとえば「xはyを持っている」をHxyであらわし、「xはyを欲している」をWxyで表 すと図4のような集合ができる。

Н	a	別荘
Н	a	車
Н	а	テレビ
Н	b	車
Н	ь	テレビ
Н	С	車
Н	С	テレビ
Н	d	テレビ

W	b	別荘
W	d	別在
W	d	車

図4 事実情報の集合(事実ファイル)

このデータベースに対する「テレビを持っていて別荘を欲している者は?」という質問はHxテレビへWx別荘のような形で記述される。この式の中でxは自由変数となっており、このxを満たすような個体定数の集合が質問の回答となる。この式と意味的に同値な集合操作はテレビを持っている者の集合{a,b,c,d}と別荘を欲している者の集合{b,d}とのintersectionをとることであり、回答としては集合{b,d}を得る。また「車を持っている者はテレビを持っているか?」という質問は(∀x)(Hx車→ Hx テレビ)というような形で記述される。この式には自由変数が含まれていないので「真」あるいは「偽」が回答となる。回答を得るには「車を持っている者」の集合と「テレビをもってい

る者」の集合の包含関係を調べることになる。このような論理式を意味的に同値な集合演算に変換するにあたっては、通常用いられるような集合演算子だけでは必ずしも充分ではない。集合型情報表現を用いて推論をおこなうシステムについては後述する。

事実検索システムは構造型データベースを持つにしろ集合型データベースを持つにしろデータベース中の事実ファイルだけを用いて推論をおこなうこと、言い換えれば必要な事実情報をすべて事実ファイルに記憶しておくことは時として冗長である。そこでいくつかの関係から新たな関係を定義することができるようにしておけば、質問文を作る際に有効であるばかりでなく事実ファイルの容量の節約にもなる。たとえば「xとyは兄弟である(Bxy)」という関係と「xはyの親である(Pxy)」という関係を用いて「xはyのおじである(Uxy)」は(Uxy)~((⇒Z)(Bxz∧Pzy))のように定義できる。Relational Data Fileでは事実ファイル中の関係を用いて新たに定義された関係は内包ファイルとでもよぶべきファイルに登録しておき推論の際に利用している。上に挙げた例ではUxyに対応するような事実情報は事実ファイルには入れず、内包ファイル中に定義式を登録しておく。

これまでには構造型情報表現によるシステムと集合型情報表現によるシステムについてだけ述べてきたが、このほかにも定理証明でよく用いられる解決原理(Resolution Principle)を利用して推論を行うシステムもある。このシステムでは事実情報は公理、質問文を定理とみなして推論をおこなう。

3. ハイパーメディア表現

これまでに、データベースの高度利用を目指した複合的データベース・システムの、従来型データベース・システムと比べての特徴として、マンマシン・インターフェイスと推論機能について述べてきたが、もうひとつの重要な要素としてハイバーメディア表現のための機能がある。このハイバーメディア表現はマルチメディア表現ともいわれ、2つの意味合いをもっている。

ひとつは、種々の形式、たとえば数値データ、文字データ、文章データ、図表データ、 画像データ、音声データというような論理的なデータ形式をもつデータを、記憶メディア (磁気ディスク、光ディスク、など)上の形式に変換して表現(データベース化)したり、 あるいは、データの入出力形式として、種々の入出力メディア上の形式に変換して表現し たりすることである。もうひとつは、オプジェクト指向的な把え方で、ひとつの事象(オ ブジェクト)を表現するための属性(データ項目値や処理ロジック)をひとまとめにして取り扱うこと、いい換えれば、種々の表現形式で記述されているデータ(情報)の集合(オブジェクト)として事象を把えると同時に、それをクラスとしてオブジェクト指向論理系の性質にもとづいてデータ操作をおこなりものである。

入力データの表現 出力データの表現 はカデータの表現 は 値 表 現 数 値 表 現 文 字 表 現 文 字 文 萱 表 親 文 賁 図 表 現 表 図 表 表 現 イメージ表現 ジ表現 像 画 表 現 画 像 表 声 音 現 表 声 表 現 オブジェクト オブジェクト 1 2 オブジェクト

ハイパーメディア・データベース . (データの内部表現)

図 5 ハイパーメディア表現

ことである事象に係る「入力データの表現」は、内部表現に変換されてデータベースに格納され、一方、データベース中の「内部表現」は「出力データの表現」に変換されて、所定の出力メディア(ディスプレイ、印刷装着、グラフィック・ディスプレイ等)上に出力される。通常、ハイバーメディア表現と言った場合、データの論理表現(論理構造:データ項目に関しては入出力データの表現に対応していると見ることができる。)と、その結合(オブジェクト化)として把えているが、複合的データベース・システムでは、ハイバーメディア・データベース中でのデータの内部表現を中心にデータ構造を把える方が便利である。ことで、「データの内部表現」といのうは、いわゆるデータの記憶構造(記憶メディア上での表現)そのものである必要はなく、より論理的な構造であってもよい。また、この「内部表現」と「入出力データの表現」との関係は、データ変換処理として位置づける。ことで注意すべきことは、「入力データの表現」と「出力データの表現」とは必ずしも一致しないことである。たとえば、これらの表現の間には、つぎのような関係が考えられる。

- ① 数值表現 → (内部表現) → 文字表現
- ② 音声表現 → (内部表現) → 文字表現
- ③ 文章表現 → (内部表現) → 図表現
- ④ 数值表現 → (内部表現) → 画像表現
- ⑤ 文章表現 → (内部表現) → 音声表現
- ⑥ 画像表現 → (内部表現) → 数値表現

したがって、ハイパーメディア・データベースにおいては、基本的なデータ操作として 内部表現に対するデータ操作を用意することになる。このデータ操作は、データの内部表 現の形式に依存することになるが、複合的データベース・システムで用意するデータ形式 としてはつぎのものが考えられる。

(1) 数値データ型

数値の入出力データとしての表現は、桁記号、小数点、単位などを付して、いわゆる 編集形式として記述されるが、内部表現としては、固定小数点表示または浮動小数点表 示のいずれかであり、編集方法は当該データ項目(人口、売上、雨量、年令、など)の 定義情報として設定されることになる。たとえば、あるものの金額をあらわすようなデ ータ項目は、つぎのように定義され、表現される。

① 入力データの表現

(定 義): xx, xxx(円)

(表現値):12,345円

② 内部表現

(定義):5桁の固定小数点表記

(表現値):12345(実際はビット列)

③ 出力データの表現

(定 義):××(ドル)

(表現値):87ドル

この例の数値データはスカラー値であるが、このほかに時系列データを表現するためのベクトル表現や3次元(高次元)マトリックス表現などがある。たとえばベクトル表現となっている数値データは、棒グラフ表現して記述されることもあろう。ここで重要なことは、他のデータ型についても言えることであるが、内部表現の定義は、事象ごとに独特の意味あいをもつデータ項目に依存しない、抽象的なデータ型であるということである。

とのような数値データ型に対しては、四則演算、比較、グラフ化、クロス集計、時系列 解析、多変量解析などのデータ操作が対応する。

(2) 年月日データ型

文献情報における発行年月日、統計情報における調査時点、企業情報における決算期、人事情報における生年月日、等々、各種情報を構成する属性として「年月日項目」は重要な役割を果している。この年月日項目に対する入出力データの表現には、西歴年、和歴年(慶応、明治、大正、昭和、平成)、アラビア数字による月表現(1月、2月、3月、4月、5月、6月、7月、8月、9月、10月、11月、12月)、和名による月表現(睦月、如月、弥生、卯月、皐月、水無月、文月、葉月、長月、神無月、霜月、師走)、さらには十二支による表現、など多様なものがあり、データの利用環境(業務)に合わせた形での表現をとっている。一方、年月日データ型の内部表現は、これら種々の表現形式をもつ年月日データを抽象的に表現し、かつ年月日処理に必要なデータ操作を可能にしておくことが要請される。

① 入力データの表現

(定義):和歴年および和名月による漢数字表記

(表現值):明治二年如月十日

② 内部表現

(定 義): 「日」単位の絶対表記(紀元元年1月1日を基準とする)

(表現値):682310(実際はビット列)

③ 出力データの表現

(定義):西歴年アラビア数字による年月日表記

(表現値):1869年2月10日

この例での内部表現は、西歴1年1月1日を「整数の *1"」とした、1869年2月10日の日数であるが、「時間」単位の絶対表記においては、「西歴1年1月1日午前0時」を基準とした経過時間によって表記することになる。このような内部表現をもつ年月日データに対しては、検索条件として、絶対表記に対して年月日時の相対表記も利用される。たとえば、昭和50年10月28日の前後100日間というような範囲の指定である。この絶対表記と相対表記に対してのデータ操作は、四則演算、範囲の比較、絶対表記間の比較などが用意される。

(3) 文字データ型

文字データ型の情報は、人事情報における「氏名」、「住所」、「所属」、「趣味」、企画情報における「企業名」、「住所」、「電話番号」、「業種」、「取扱商品」、等 4 のように、語句やコードで表現されるデータ項目であり、それぞれのデータ値はデータ操作にあたっての最小単位となっている。この文字データ型の情報に対する「内部表現」と「入出力データ表現」との関係は、主々、コード変換である。たとえば、国(または地域)を例にとると、JISの国コード(X-0304)では、「英国」はつぎのように表記されている。

(日本語による国名・地域名):グレートブリテン及び北部アイルランド連合王国 (英語による簡略名・正式名):UNITED KINGDOM

United Kingdom of Great Britain and
Northern Ireland

(2文字コード):GB

(3文字コード):GBR

(3数字コード):826

複合的データベース・システムにおける国コードの内部表現としては、2文字コード または3文字コードまたは3数字コードが適当であるが、ここでは、3数字コードを採 用すると入出力表現の例はつぎのようになろう。

① 入力データの表現

(定義):国名のローカルの略称

(表現値):英国

② 内部表現

(定 義):JIS X-0304にもとづく国の3数字コード

(表現値):826

③ 出力データの表現

(定義): JIS X-0304にもとづく英語略称

(表現値):UNITED KINGDOM

最近ではデータ処理の国際化が進んでおり、米国に進出している日本企業、あるいは 日本に進出している米国企業も多く、それぞれの業務処理において複数言語、たとえば 日本語と英語の文書を必要とする場合がある。とのような環境での文書管理手法のひと つとしてこのコード化がある。すなわち、データベース中では、データ値をコード化し てもち、入出力データの表現としては、日本語や英語で表記する方法である。このよう な特徴をもつ文字データに対するデータ操作は、それぞれのデータ値を最小単位とする 比較操作が中心となる。

ところが、商品コード(JISX-0405)のように、コード体系が構造化されているものがある。すなわち、ひとつのコードの中に、大分類、中分類、小分類、細分類、細々分類、の情報が包含されているような場合であり、たとえば、つぎのような例がある。

大分類コード : 3 加工基礎材

中分類コード : 23 化学薬品

小分類コード : 231 基礎無機薬品

細分類コード : 2311基礎無機酸

細々分類コード: 23111塩 酸

このような体系をもつコードの操作においては、ひとつのコードを最小単位とするのではなく、その内部構造、たとえば先頭2桁が23(化学薬品)のものを考慮した比較操作が必要となろう。これらのデータ操作は、文献情報などに付与されるキーワードの

比較についてもよくおこなわれる操作で、「前方一致」比較操作、「中間一致」比較操作、「後方一致」比較操作とよばれており、ひとつのデータ値を最小単位とするような操作は、とくに、「完全一致」比較操作とよんでいる。

(4) 文章データ型

各種情報がデータベース化される中で、情報を文章で表現する場合、すなわち文章情報が多く扱れてきている。たとえば、文献情報、特許情報、法令情報、判例情報、などをはじめとして、企業情報における企業概要、統計情報における注釈文、等々多岐にわたっている。との文章データによって表現された情報は、文字列としての文章そのものを入力しデータベース化しておき、また文字列として出力するという基本的なデータ操作に対して、文章のもつ意味内容を解析(内容解析処理)して、その内容にもとづいてコミュニケーションをおこなうという重要なデータ処理が期待されている。これは、とくに自然言語による文章のもつ本来的な性質、すなわち、世界の種々の事象を表現するには便利(慣れている)であり、ある特定の事象(意味)に対して、多様な表現が可能であるということである。したがって、文章データに対するデータ操作には、つぎの2つの側面が必要とされることになる。

- ① 文字列としての文章そのものをデータベース化しておき、その文章をひとつの単位 として処理するという側面
- ② 文章のもつ意味内容を解析し、同じ意味をもつ文章同士(たとえばデータベース中の文章データと質問文データの意味が同じであると識別する)を認識するという側面 ①のタイプのデータ操作は、言わば、ワープロによるデータ管理に似たものであり、ワープロで作成した文書に文書名(番号)を付して登録し、一方で、その文書名(番号)を指定することによって既存文書を参照するというものである。一方、②のタイプのデータ操作は、コンピュータ技術として必ずしも確立されていないが、データベース中には入力データとしての文章そのものを記憶しておき、データを処理する時点で内容解析をおこなう方式と、データ入力時にあらかじめ文章の意味内容を解析して、意味内容をしめす情報を付与、あるいは、表現形式を変換して、内部表現として記憶する方式とがある。

前者の「意味内容をしめす情報を付与」する方法では、文ごと、あるいは文の集合と しての文章ごとに、意味コード(分類コード)やキーワードを付与して、データベース に登録する。一方、データ検索の場面では、質問文中から意味コードあるいはキーワドを抽出し、データベース中のコード(キーワード)と照合することによって、所定の文章情報を表示するという方法がとられる。この方法では、文または文章そのものを加工して内部表現データを作成することはしない。ここで扱う文章の形態としては、つぎのようなものがあろう。

the state of the s

- ① 通常の文
- ② 文献や通知のタイトル
- ③ 法令の条文
- ④ 議事録の質問文または答弁文
- ⑤ 文献の抄録
- ⑥ 新聞や雑誌の記事
- ⑦ 文献中の編・章・節
- ⑧ 文献の全文(フル・テキスト)

この方式での、入出力データの表現、およびデータベース中の内部表現はつぎのよう になろう。この例では、文章として「ひとつの文」を扱い、キーワードを付与している。

① 入力データの表現

(定義):日本語の文

(表現値):(i) W社はA商品の開発に成功し、来年から、販売をおこなう。(1985年10月21日、出所α)

(ii) X社は、昨年、B商品のシェアを20%から35%に拡大した。(1986年2月1日、出所β)

. .

- (ii) Y社は、来年から、B商品の生産中止を決定した。(1986年11月3日、出所α)
- (V) W社は、B商品の販売分野で、Z社と提携する予定である。 (1987年1月10日、出所β)
- ② 内部表現

(定義):キーワード付与

(表現値):(I)②W社はA商品の開発に成功し、来年から、生産販売をおこなう。

(1985年10月21日、出所α)

⑩1985年10月21日:出所α W社、A商品、開発、販売、生産

- (ii) ② X 社は、昨年、B 商品のシェアを20%から35%に拡大した。(1986年2月1日、出所β)
 - ®1986年2月1日:出所β X社、B商品、シェア
- (iii) ② Y 社は、来年から、B 商品の生産中止を決定した。

(1986年11月3日、出所α)

⑥1986年11月3日:出所α Y社、B商品、生産中止 ·
(Ⅵ) ② Z社は、B商品の販売分野で、W社と提携する予定である。

(1987年1月10日、出所β)

●1987年1月10日:出所β Z社、B商品、販売、W社、提携上記のような内部表現において、図は入力データとしての文そのものであり、⑩は、それぞれの情報(文)の発生年月日と出所、および、それぞれの意味をあらわすキーワードである。通常、発生年月日は年月日データとして、出所は文字データとして、また、キーワードは繰返しのある文字データとして定義されることになるが、このようにひとつの入力データ(この例では「文」および「年月日+出所+キーワード」で構成されている。)に対して、複数のデータ項目が対応することがある。このような例の典型的な情報として、後述する文書情報がある。

ところで、上記のような内部表現をもつ4つの文章データから構成されるデータベースに対する検索結果(出力データの表現)は、それぞれの質問文の内容に応じて、つぎのようになろう。

③・出力データの表現

- (i) 「1985年以降のW社の動向は?」という質問に対して、「年月日≥1985年、キーワード=W社」という条件式が作成され、つぎの回答が得られる。形式そのものは入力データの表現と同じである。
 - W社はA商品の開発に成功し、来年から、販売をおとなう。(1985年10月 21日、出所α)
 - 2社は、B商品の販売分野で、W社と提携する予定である。(1987年1月10日、出所β)
- (ii) 「B商品の動向は?」という質問に対して、「キーワード=B商品」という条件 式が作成され、つぎの回答が得られる。
- ・X社は、昨年、B商品のシェアを20%から35%に拡大した。(1986年2

月1日、出所β)

- ・ Y 社は、来年から、 B 商品の生産中止を決定した。(1986年11月3日、出所α)
- ・ Z 社は、 B 商品の販売分野で、W社と提携する予定である。(1987年1月10日、出所β)

また、この「意味内容をしめす情報を付与」する方法に対して、後者の「内部表現に変換」する方法では、「文」を「一階述語論理系における式(命題)」による中間的な表現に変換して、データベースに記憶する。この中間的な表現としては、たとえば述語として、つぎのような式

Sxyvw: x 社は y 商品を v 時点でw 量販売(S)している。

Cxyzv: x 社と y 社は z 商品 に関して v 時点で販売提携 (C) している。

Pxvvw: x 社は y 商品を v 時点でw 量生産 (P) している。

を定義することによって、先に挙げた文章はそれぞれつぎのような内部表現でデータ ベース化されることになる。

(i) W社はA商品の開発に成功し、来年から、販売をおこなう。

(内部表現)S(W)(A)(1986年)(一)

(ii) X社は、昨年、B商品のシェアを20%から3:5%に拡大した。

(内部表現)S(X)(B)(1984年)(シェア20%)

S(X)(B)(1985年)(シェア35%)

- (iii) Y社は、来年から、B商品の生産中止を決定した。(一は否定をしめす。)(内部表現) S(Y)(B)(1987年)(一)
- (V) Z社は、B商品の販売分野で、W社と提携する予定である。

(内部表現) C(Z)(W)(B)(将来)

このほかに、述語式中に変数を含むことも可能であり、たとえば全称記号(\forall x :「すべての変数 x について」と読む。)を導入することによって、「M 商品は生産者のみが販売を許可される。(M商品を販売しているすべての会社(x)は、M商品を生産している。)」という文は、つぎのような中間形式で表現される。(「 $A \rightarrow B$ 」は「A ならばB」と読む。)

(内部表現) ∀x(SxM→PxM)

このような表現形式で文を記述しデータベース化するメリットとしてはつぎのよう な点が挙げられる。

- ① 自然言語表現による文の意味を、比較的保存した形で、表現することが可能である。
- ② 解決原理(Resolution Principle)など、コンピュータによる推論(演繹)処理の応用に適している。
- ③ リレーション型データベース・モデルの論理型データ操作言語(Relational Calculus)とも類似しており、集合型データ操作言語(Relational Algebra)を介して、通常のデータベース操作手順とリンクすることができる。
 - ④ 機械翻訳のテクニックとして、この一階述語論理系の内部表現を介して、多くの言語(日本語、英語、露語、仏語、中国語、等々)に変換可能となることが期待できる。 このように、一階述語論理系の式として文章情報を表現することのメリットは多いが、 その全体像を図示すると図6のようになる。

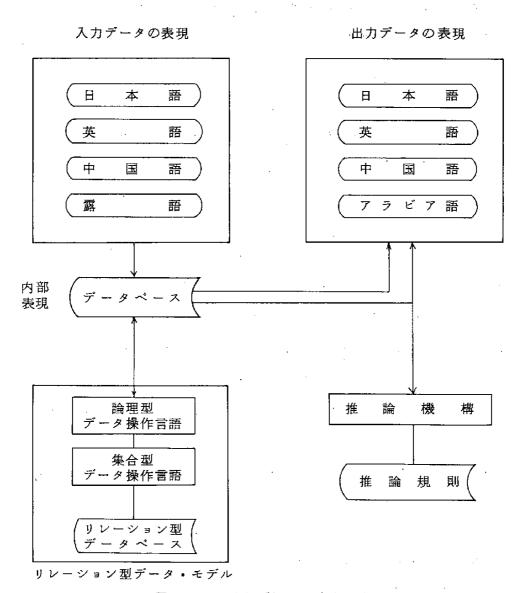


図6 述語論理系による文章の表現

(5) 画像データ型

資源衛星や気象衛星によるディシタル画像、医療分野での各種撮影写真、物体認識ロボットによる物体画像、等々の分野における画像処理技術は、コンピュータによるデータ処理技術の一環として、最近大きく発展している分野である。したがって 処理技術も日進月歩でありデータ構造も模索中(研究中)の部分も多い。このような分野でのデータベースの構造については、データベースに係る一般的な考え方、あるいはシステム開発場面でのデータベース的アプローチとでも言うべき考え方を適用するのも一方法である。

古典的なファイル設計技術は、データ処理手法(業務処理プログラム)に沿ってファ イル構造も設定していく方法がとられていたが、最近のデータベースは、設計時におい て必ずしもその利用形態が確定しているとは限らない。したがって、データベース構造 は、できるだけ利用場面(プログラム)を拘束しないような形にしておくのが得策であ る。リレーション型データ・モデルもとのような動機によるところが大きい。このよう に、利用手法が必ずしも確定していないようなデータベース分野では、データベース・ システムの中にそれぞれの(画像)処理ルーチンを組み込むことも困難であるため、言 わゆるOPEN型(利用者出口ルーチン型)のインターフェイスを用意し、適当な利用 者プログラムと随時結合可能な状態にしておくという方法がとられる。そのためには、 画像データの内部表現は、原情報(一次情報)に忠実な形式にしておくことが、へたな 加工をするよりも、適当であると言える。したがって、この画像データ型には、これま でのデータ型のように比較、照合というような対応するデータ操作を用意することはせ ず、いわば、無定義(不定型)データとして、直接的にデータの内容に関与することは しないものとして定義しておくことになる。このようなタイプのデータは、将来の技術 発展を考慮して、言い換えればシステムの拡張性を重視して、とにかくデータベース化 しておくということに意味があると考えられ、そのような立場からOPEN型データ構 造が得策と思われる。

(6) 図表データ型

図表データは、主に、数値データをテーブル状に表現したり、グラフ化(棒グラフや 円グラフ)したりして表現される。したがって図表データは、基本的には、「図表名」、 「表頭」、「表側」、「定数」というようなデータ項目の集まりとして表現することが できる。ことで「図表名」、「表頭」、「表側」というようなデータ項目は、文章データ型として定義され、「定数」は数値データ型として定義される。図表データをこのような形態で内部表現しておくことによって、表作成ルーチン、あるいはグラフ作成ルーチンを介して、出力データとして表やグラフで表現することが可能になる。一方、原情報(一次情報)が図表の形式で表現されているような場合は、そのまま直接的に、図表データ型に変換してデータベース化することは難かしい。そこで、通常は、データ入力の時点で、上記のようなデータ項目に分解して、位置情報(ページ上でのレイアウトを決めるための情報)と共に、入力する方式をとっている。

この図表データは、後述するイメージ・データ(FAX伝送用データと同様の表現をもつ。)として記憶することも可能であるが、つぎのような点において、図表データ表現の方が優れていると言える。

- ① 「図表名」、「表頭」、「表側」を文字データとしてデータ操作の対象とすることができ、そのデータ操作によって、所定の図表データを検索することができる。
- ② 2次元(あるいは高次元データ、たとえば企業情報における「地域別・規模別」や「業種別・規模別・商品別」)のデータをクロス集計して、この図表データ型表現に変換することができる。(通常、このタイプのデータ処理は、調査票データの集計業務とその結果の表出力とかグラフ出力というように、非常によく利用されている。)
- ③ ひとつの図表データ型をもつデータから、所定の表形式データを作成したり、各種グラフを作成したりすることができる。

この図表データのように、複合的データベースの内部表現として、いくつかのデータ型をもつデータ項目に分解されるという考え方は、オプジェクト指向型データベースにおけるオプジェクト(クラス)の概念を構成している。このクラス概念は、オブジェクトを階層的に構成することを可能にする手法であり、複合的データベース・システムにおいても重要な考え方になっている。

(7) イメージ・データ型

イメージ・データ型のデータは、文章、図、表、グラフ、等々のデータをイメージとして把えた表記法である。このイメージ・データはいわばFAX伝送用のデータ・イメージであり、紙面上の記録(文章であったり、図であったりする。)を、紙面を細分化(G3規格でミリ当り3.85本または7.7本)し、白点または黒点の列として表現する。し

たがって、A4サイズの紙面をイメージ表現するのに、多少の圧縮技法を利用しても、数十キロ・バイトが必要となることがある。さらに、そのデータは、ビット列として扱われ、内容に立ち入ることはしない。すなわち、紙面(イメージ・データ)は、変換機によってビット列に変換され、データ名(光ディスクアドレスなど)と共に内部表現として記憶される。一方、出力データの表現としては、記憶されているビット列を、紙面上の白点と黒点として表記することになる。通常、このようなデータは、CDーROM(読み込みのみ可能)や光ディスク(一回だけ書き込み可能)というような高密度の媒体に記憶され、検索用の書誌データ(磁気ディスクのような読み書き可能な媒体に記録される。)とリンクされる。たとえば判例データを例にとると、磁気ディスクと光ディスクとに格納されるデータ項目はつぎのようになろう。

- ① 磁気ディスクに格納されるデータ項目
 - 事件名と事件番号
 - 原告と被告
 - 裁 判 所
 - 判決年月日
 - · 主 文
 - 判例要旨
 - 光ディスク・アドレス(本文)
- ② 光ディスクに格納されるデータ項目
 - 本 文

従来、このようなイメージ・データは、入力データ表現と出力データ表現とが一致しているものであったが、最近はOCR機器やCOM(computer output microfilming:コンピュータで処理した大量の文字図形を直接マイクロフィルムに高密度に出力する装置)機器の発達によって、種々のデータ表現の変換が可能になっている。

- ① 文書をOCRで読み込むことによって、文字データとして認識(内部表現)する。 文書中に文章と図が混在するような場合には、文章の部分(データ変換対象範囲)を ディスプレイ上で囲むことによって、部分的に文字データとして記録されたり、イメ ージ・データとして記録されたりすることが可能である。
- ② データベースからの検索結果を加工・編集し(文章データや表データとして表現されている。)、それをイメージ・データに変換して、FAX伝送する。

(8) 文書データ型

これまでは、データベースに対するデータ操作にあたって、(図表データは多少性質を異にしていたが)基本的な単位となるようなデータ型を扱ってきたが、ここで扱う文書データ型は、種々のデータ型をもつデータ項目の集りとして事象(文書)を表現するためのものである。文書データ型のオプジェクトは、通常、つぎのようなデータ構造(データ項目の集まり)として定義される。この文書データ型のオプジェクトの例として会議情報が挙げられよう。ここでは、会議情報は、「開催通知オプジェクト」と「議事録オプジェクト」とから構成されているものとすると、「開催通知オプジェクト」はつぎのようなデータ構造として定義することができよう。

- ① 会議名(文字データ型)
- ② 主催者(文字データ型)
- ③ 開催日時(年月日データ型)
- ④ 参加人員(数値データ型)
- ⑤ 会場(文字データ型)
- ⑥ プログラム(文章データ型)
- ⑦ 参加要領(文章データ型)
- ⑧ 会場地図(イメージ・データ型)

この例からも明らかなように、当概事象は複数のデータ項目から構成されており、このデータ構造が、言わば、開催通知オプジェクトの内部表現になっている。従来は、このようなデータ構造に則して、データを入力する方法がとられてきた。ところが、最近は、ワープロの普及がいちじるしく、このような文書はワープロで作成されるのが常識になっており、データベースへのデータ入力の原資料は、このワープロ文書であることが多い。これは、言わば、文書作成とデータベース化という2つの場合で、同一の内容をもつデータを2度入力していることになる。

このようなワープロ入力文書は、開催通知オブジェクトに限らず、論文、規格、特徴、各種報告、等々多岐にわたっている。また、コンピュータ印刷技術(CTS)においても、版下をコンピュータをもちいて作成しており、その際のデータ(文章)入力も、コンピュータの端末を介しておこなわれている。そこで、(オンライン)データ提供サービスのためのデータベース化、および、文書化(印刷)のためのデータ作成、という2つの場面で利用するデータを一括して管理していくという考えが出現し、最近は多くの

場面で、この体系が利用されるようになっている。この体系の概略は図7のようになる。

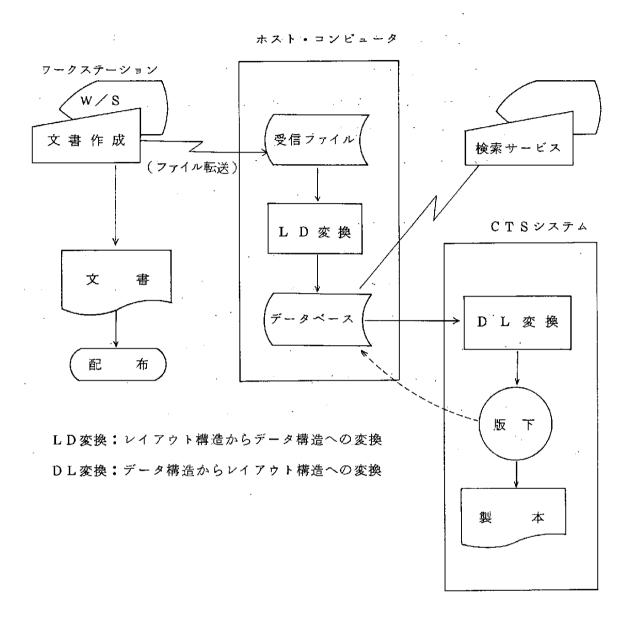


図7 文書データ型の処理体系

このシステムにおいては、業務処理の一環として作成したワープロ文書を配布するほかに、データ提供サービス用のデータベースに蓄積し、サービスに供する。一方、データベースに蓄積されているデータについて、定期的あるいは随時、データ抽出処理をおこないCTSシステムを利用して版下作成をおこなうという処理の流れになっている。従来は、CTSはデータ提供用データベース(システム)とは独立して開発されたという経緯もあって、CTSシステムの中で独自にデータ入力をおこない、この入力データ(CTSデータ)ファイルを変換して、データ提供用データベースにデータを投入して

いた(図 7の点線)が、最近では、データ提供用データベースをマスタ・データとして 位置づけるようになっている。

ここで問題となるのは、文書(版下)の構造(レイアウト構造)とデータベースの構造(データ構造)が異なるという点である。すなわち、前者のレイアウト構造は紙面(ページ)上での文字の配置に係る情報が中心になるのに対して、後者のデータ構造は構成要素としてのデータ項目との対応情報が中心になっているからである。また、データ構造が、データベース・システムのデータ定義言語をもちいて定義されるのに対して、レイアウト構造は、SGML(ISO規格)というような組版指示言語によって規定されるが、通常、このレイアウト構造の文書データから、先に挙げたようなデータ項目を抽出して、データ構造に変換することは困難である。そこで、ワープロ文書は、ホスト側で(人手を介して)データ構造に係る情報(文書データ中でデータ項目を識別するための記号)を付すことによって、当概事象のデータ構造と対応づけ、データベースに格納することになる。一方、CTSシステム側では、データ構造に沿って表現されている文書データに対してSGMLのような組版用記号を付すことによって、組版上での配置(レイアウト)を指示しながら版下を作成することになる。最近は、この2つの構造のデータ変換をコンピュータによる自動処理で実現しようという試みがある。

Ⅲ CD-ROM 用検索システム

本調査研究は、将来的には、複合的データベース・システムのパソコン上での構築を目指しているが、前章でも述べたように、種々のデータ表現形式をもつ情報を統一的なデータベースとして管理することが必要となる。とくに、最近では、パソコン用の大容量記憶媒体としてCD-ROM装置が販売されており、比較的手軽に導入できるような環境になっている。そこで、本システムのパソコン上でのパイロット・システム構築にあたっては、このCD-ROMを大巾に採用していくことが得策と思われる。ここでは、CD-ROM採用にあたっての基本調査として、

- ① CD-ROM検索ソフトの機能調査
- ② CD-ROM用検索システム機能の高度化 について、調査検討をおこなった。

CD-ROM検索ソフトの機能調査

(1) NECの電子広辞苑

(岩波書店、日本電気ホームエレクトロニクス:定価60,00円(税別))

総項目数20万、挿絵2,000に付録を加え、信頼できる国語+百科の辞典として親しまれているのが「広辞苑」である。『電子広辞苑』は、「広辞苑」のデータをすべて収め、CD-ROMの特性を活かした様々な検索を可能にしている。探したい語句をそのまま入力する入力語検索、画面上の語句をカーソルで指定する画面語検索、また、語句の前方または後方部分からの検索のほか、複数のキーワードから項目を探したい時など、電子辞書特有の検索ができる。検索結果を、図版を含めてプリントアウトでき、フロッピーディスクへの保存も可能である。さらに、234種類の色見本表示、60種の鳥の鳴き声を表現するなどの機能も充実している。

	· ·		
総容量	3 5 2 Mバイト/	図版データ	1 9 3 M バイト
	+19分33秒	外字フォント	1 0.5 M バイト
本文データ	3 0 M パイト	色 デ ー タ	0. 5 M バイト
索引情報	B 118Mバイト	音 色	音19分33秒

① 特 色

ア.図 版 表 示

「広辞苑」に含まれている約2,000枚の図版を説明文とともに表示する。

イ. 色見本表示

234種類の色を説明文と一緒に表示する。「今様(いまよう)色」など文字を 見ただけではわからない色をディスプレイ上で見ることができる。

ウ. 音 表 現

「アホウドリ」「ホトトギス」など 6 0 種の鳥の鳴き声を説明文を読みながら聞くことができる。

エ. マルチウィンドウ

画面上に最大5つまでウィンドウを開くことができるので、画面を切り替えることなく複数の検索結果を並べることができる。ウィンドウの拡大、縮少、移動も自由に切り替えられる。

オ、ヘルプ機能

操作途中に困ったときは、ヘルプ画面に切り替えてガイドメッセージを見ることができる。

カ. 簡 単 な 操 作

キー入力以外はすべてマウスを使って簡単に操作できる。

キ. 複写・図形複写機能

検索結果をプリントアウト、またフロッピーディスクに保存できるので、ユーザ ー独自の情報ファイルが作成できる。

②検索機能

ア. 入力語検索

• 見出し語(かな)検索

普通に辞書を引くように「ひらがな」「カタカナ」で見出し語を入力して検索する。

・表記形(漢字・アルファベット)検索

見出し語の「漢字」「アルファベット」を入力して検索する。

≪例≫「適宜」「identity」

前方、後方一致検索

語句の語頭あるいは語尾の部分だけを入力することで、それを構成要素とする語を すべて検索できる。検索結果(項目一覧)表示後は、キー入力の必要がないので連 続的に検索することができる。

せいしん‐ぶんか

…クワ 【 精神文化 】

せん-じょうもん-ぶんか …クワ 【先縄文文化】

せんどきーしんせっきーぶんか …セキ…クワ【先土器新石器

せんどきーぶんか

・・・・クワー【 先土器文化 】

イ. 画 面 語 検 索

画面上にさらに検索したい語句があるときは、その語句をカーソルで指定するだけ で検索できる。キー入力の必要がないので連続的に検索することができる。

しゅうーぶん

シゥ… 【 秋分 】

二十四節気の一。太陽が秋分点に達した時刻。秋分を含まな い。「太陽暦」では九月二三日頃に当る。秋の彼岸の中日で

たいようーれき

...ャゥ... 【 太陽暦 】

地球が太陽の周囲を一公転する時間を一年とする暦すなわち の単位とする暦。代表的なものはグレゴリウス暦で、三六五日

ウ.条件検索

いくつかのキーワードを入力すると、その語句を解説文中に含む見出し語を検索す ることができる。キーワードは、AND条件で最大5個、さらにAND条件1個につ き各3個までのOR条件を入力することができる。

≪例≫ 検索語

①ドイツ、イタリア ②作曲家 ③歌劇 ➡ プッチーニ、ベートヴェン、モーツ アルト、ロッシーニ、ワグナーなど 入力

検索語① : ドイツ イタリア

検索語② : 作曲者

検索語③ : 歌劇

エ. 履歴情報(ヒストリー)検索

過去に検索した項目(最新の項目から32個前まで)の一覧表示からカーソルで指 定するだけで再度検索することができる。

ルネサンス

レオナルドダヴンチ

*文化

フィレンツェ

メディチ

オ.付録検索

付録にある「分野別図版一覧」、「色の見本一覧」や「鳥の鳴き声一覧」などの項目をカーソルで指定すれば、図版、色見本、鳥の鳴き声だけを連続的に見る、あるいは聞くことができる。

(2) CD-ROM on CO-ROM

(日本電気ホームエレクトロニクス:定価48,000円(税別))

美しい映像や音楽に、文字・画像データを組み合わせて、1枚のCDに「CD-ROMの基礎から製造まで」のすべてを収録。

(3) CD-WORD

(三修社:定価100,000円(税別))

① 概 要

世界7か国の出版社の協力のもとに、各社から提供された対象言語8か国語(日本語

/フランス語/ドイツ語/スペイン語/イタリア語/オランダ語/中国語)、収録辞書数13種類17冊、収録語数500万語ものデータを1枚のCD-ROMに収録し、専用の検索プログラムで辞書のように引けるようにしたものである。

検索のための入力言語・出力言語は、8か国語すべての任意の組み合わせが可能で、 その対応関係は56とおりにも及び、さらに類義語検索、主・副見出し語検索、文例検 索、逆引検索、複合語検索などCD-ROM特有のデータ検索が可能である。

また非常に簡単機能として、スクリーンエディタが組み込まれており、文章を構成しながらの検索、検索結果の整理などが行える。作成したテキストはMS-DOSのテキスト形式でセーブできるので、高機能エディタやワープロに吸い上げて活用することができる。

② 操 作

検索ソフトが入っているフロッピィと辞書CD-ROMをFM TOWNS(富士通)に挿入して、リセットスタートすると、簡単な起動画面が表示される。ことでファイル名の入力を求めてきたときに、キーインすると簡易エディタが起動される。この簡易エディタが基本画面になっていて、テキストの入力などを行いながら、必要に応じてPFキーを押して検索モードに移って辞書を引き、検索結果を整理しては次の言葉を検索する、ということを繰り返すようにして利用する。なお、これらの操作は FM TOWNSOSをどのようにマウスで行うものではなく、すべてキーボードで行うようになっている。

検索は PF1 キーで検索システムのメニューを呼び出して行う。検索モードは、 3種の参照形式、4種の範囲指定、8か国語指定、13種の辞書指定の組み合わせでか なり多様なものになっている。

ア. 辞書検索モード

参照形式の1つ、辞書検索モード (Look Dictionary)では、13種の辞書から1つ 辞書を選び、その辞書に対する検索を行う。

イ. 対訳語検索モード

対訳語検索モード (Display Translation)は、8 か国語の対訳を行うモードで、このモードでは辞書の指定を省略することができる。たとえば、入力語を英語に指定し、対訳対象を日本語に指定すると、1 3種のうち「英→日」で検索可能な辞書をすべて検索し表示するので、その中から適当なものを選べばよいことになる。

またこのモードでは、対象となる言語間に直接相当する辞書がない場合でも対訳が可能になる。たとえばスペイン語から中国語への翻訳などは、英語を経由して翻訳をする。さらに対訳語の言語種指定は省略できる。こうすると残りの7か国語すべてに対する対訳が可能になる。しかし入力言語種の指定までは省略できない。単語のスペルはわかっているが、どこの言葉かわからない、などというときには、手当りしだいに検索するしかない。

ウ. 同義語検索モード

同義語検索モード(Display Synonym)は入力語と同じような意味をもつ単語を表示するモードである。出力言語種、対象辞書、範囲指定はすべて意味をもたなくなるので、指定できない。英文などではできるだけ同義語をちりばめたほうがよい文章とみなされるので、英語で文章を書くときには同議語検索モードは重宝する。

(4) CD-Ward 8

(三修社:定価168,000円(税別))

『CD-Word 8欧文ワードプロセッサ』は、世界8ヶ国・500万語の単語を1枚のCD-ROMに収録した電子辞書『CD-WORD』を利用した、マルチ言語対応欧文ワードプロセッサである。

6ヶ国語による入力・表示・印刷・単語の翻訳機能を持ち、文書編集作成中にとの『CD - Word 』の辞書を随時に検索、また検索参照したものを取り込むことや、日本語のローマ字入力により7ヶ国間の対話検索を可能にした、まったく新しいタイプの欧文ワードプロセッサである。

① 特 長

ア. 多国語フォントをサポート

「英」、「独」、「仏」、「スペイン」といった各国語フォントを、ソフト的にサポート。通常のタイプライタの場合のように、タイプボール(あるいはデイジーホイール)等を取り換えることなく、マルチ・リンガルなタイプライタとしても使用できる。

イ. 各国語対応のキーポードをサポート

IBM/PCが各国向けに対応モデルを揃えている各種キー配列キーボードを、 PC-9801(日本電気)のキーボードでソフト的に対応。慣れている好みのキー 配列キーボードを指定すれば、プラインド入力もできる。

ウ. CD-ROM収録の辞書検索機能

通常の欧文ワープロの使用以外に「CD-ROM」に収録された12種類・16冊の辞書から、瞬時に、目的用語の検索をするといったことも可能。対訳、同義語、辞書内容参照といったことを自由に行え、編集中の文書への取り込みもできる。

エ、マルチウィンドウによる操作

マルチウィンドウ、プルタウンメニュー、ダイアログウィンドウといった最先端の ユーザーインタフェースを実現。

オ. 豊富なワープロ機能を装備

自動インデント、タブ機能、文字かざり、ワードラップ、ソフト・ハイフン、検索・置換、スペルチェック機能、罫線といった豊富な機能を用意し、表現力に優れた文書を作成することができる。

カ. 1カ国語に対して7カ国語の検索

収録されている 7 カ国語のどの言語からでも、他の言語の対訳語を検索することができる。例えばスペイン語からイタリア語の検索も、ドイツ語から英語の検索も可能である。また検索は双方向から可能なので合計 4 2 通りの組み合わせで行える。

② 機 能

- ア. 多機能検索
 - 辞書内容表示機能【 Look Up 】
 - ・同義語検索機能【Synonym】
 - 対訳語検索機能【Translation】
 - 主見出検索機能【 Head word 】
 - 副見出検索機能【 Sub-Head word 】
 - ・逆引き検索機能【Reverse 】
 - 複合語下成分検索機能【 Last-Compound 】
- イ. 多国語検索
 - 対象言語

日本語、スペイン語、英語、イタリア語、フランス語、オランダ語、ドイツ語

- ム. 収録辞書
 - Harrap's Shorter (英↔仏)

- Harrap's Science (英↔仏)
- · Harrap's Date Processing (英↔仏)
- · Harrap's Business (英↔独)
- Brandstetter's Science & Technology (英↔独)
- · Groupe Anaya's (英↔西)
- Nicola Zanichelli's (英↔伊)
- Nicola Zanichelli's (英・仏・伊・西)
- · Wolter's Noordhoff's (英↔蘭)
- Sansyusya's Science & Technology (英・独・和)
- · Sansyusya's (英↔和)

エ. 入 力 方 式

• 6 種類 5 カ国対象キーボード

UNIVERSAL FRENCH

ENGLISH SPANISH

GERMAN ITALIAN

オ. 多国語フォント(ソフトウェア対応)

日本語(検索表示のみ)

フランス語(aec)

ドイツ語 (a o u)

スペイン語(n)

- カ、字 体
 - ・サブスクリプト【Subscript】
 - ・スーパースクリプト【 Superscript 】
- キ. 文 字 修 飾
 - アンダーライン【 Under on/off 】
 - •強調【Bold on/off】
 - ・右寄せ【Right】
 - 左寄せ【 Left 】
 - 中央揃え【 Centered 】

ク. コマンド選択

- ブルダウンメニュー方式
- ・ショート・カット・コマンド

ケ.編集機能

- 新規文書作成 【 New 】
- ・既存文書の読込み【 Load 】
- ・ファイル一覧[日付/名前/サイズ/昇順/降順]
- ディレクトリ指定
- ・自動インデント … 自動機能
- ・強制インデント 【 Indent 】
- ・他ファイルへの書出し 【 Write 】
- •他ファイルからの読込み 【 Insert 】
- ・書式の設定 【 Page Setup 】
- カット 【 Cut 】
- ・ペースト 【 Paste 】
- 複写 【 Copy 】
- ・単語の削除 【 Delete Word 】
- •行の削除 【 Delete Line 】
- ・罫線/枠の描画、削除 【 Draw Box, Erase Box 】
- 改ページ 【 Pagc Break 】
- ・ソフト・ハイフン 【 Hyphen on/off 】
- ワードラップ … 自動機能
- ・タブのセッティング 【 Tab Setting 】
- コ.編集付加機能
 - •検索/置換機能 【 Seach 】/【 Change To 】
 - カーソル移動機能
 - サ. ウィンドウ
 - プルダウン・メニュー・ウインドウ
 - ■辞書検索ウィンドウ〔検索語入力 / 辞書選択 / 多機能設定 / 検索語内容表示〕

シ. 印 刷

- ページ番号設定
- ・行端揃え印刷 【 Justify 】
- ・印刷部数の設定
- プリンタの設定 [機種設定/文字間ピッチ/用紙サイズ/カットシートフィーダの 設定]
- ス. システムセットアップ機能
 - 文書ドライブ
 - CD-DRIVE指定
 - 辞書設定(12種類)
 - ・入力言語設定(7ヵ国語)
 - 多機能検索設定(8種類)

(5) 模 範 六 法

(三省堂:定価120,000円(税別))

- ① 多採な検索機能
 - フリーキーワード検索
 - 事項索引からの検索
 - 判例・先例の検索
 - 法令の改正年・法律番号による条文からの検索
 - 被参照条文の検索
- ② 簡単な操作法

キー操作は↑↓←→などのキーと、1,2,などのテンキー(数字キー)の操作のみでできるように、特段の工夫を凝らしてある。文字キーの操作は、キーワードの入力等だけに限定。「かな」「ローマ字かな」「ローマ字カナ」「英数字」の4通りの方式が可能で、どんな言葉でもスピーディに入力できる。また、漢字に変換する辞書も充実。これまでコンピュータを利用したことがない人でも短期間に利用法をマスターできる。

③ 見やすい縦書き表示

パーソナルコンピュータの画面表示はほとんど横書き表示であるが、この「模範六法 CD-ROM版では、文字表示技術を駆使し、条文・判例等はすべて見やすい縦書き表 示とし、画面をスクロール(連続移動)する場合も、縦書きのまま左右にスクロールする。さらに、法令等が表示されている画面は、本のページを繰るように画面 1 ページ単位でも移動できるよう工夫されている。

④ 3 通り選べる画面表示

条文だけに目を通したいときは条文のみを、それに参照条文の表示を加えたいときは 条文+参照条文の表示を、判例までを見たいときはさらにその表示を加えることができ、 その時の使用目的に応じてコンピュータライクに画面表示を選択することができる。

⑤ 画面分割表示

ある条文に参照すべき条文が示されている場合、画面を2分割し、もとの条文を右に、 参照条文を左に示し、同一画面上で両者を読むことができる。また、この2つの画面を それぞれ左右にスクロール(連続移動)することもできる。

⑥ 充実したシソーラス(類義語)検索

フリーキーワード検索をする場合、「損害賠償」と入力しても「損害の賠償」と入力しても、双方を類義語と判断するシソーラス機能をもっているので、この二者のどちらか一つを含む条文や判例をすべてリストアップする。また、「預け入れ」「預入れ」をど違った送り仮名で入力される用語についても、前記と同様にどちらかを含む条文・判例をすべて出力できる。このほか、言葉としてはまったく違う2つ以上の用語で、法律上類似性の強いもの(建物 - 家屋 - 建造物 - 建築物など)についても同様である。また、この機能は、同義・近類似した法律用語を学習するのにも役立つ。

⑦ 収録内容(法令・判例・参照条文を全て収録)

- ア. 憲法・民法・商法・民事訴訟法・形法・刑事訴訟法・労働基準法等379件の法令
- イ, 関連する条文を示した参照条文
- ウ、民法・形法・商法等主要法例31件の条文ごとの判例要旨。

⑧ プリント機能

パソコンにプリンタを接続することによって、検索した結果や、条文と参照条文、被 参照条文ともとの条文など、画面に表示された全部、または範囲を限定した一部を COPYボタンを押すだけで印刷できる。

⑨ 多彩な検索項目

- ア. 法令名・条数からの検索[分野別・五十音別]
- 「分野別法令名一覧」に↑↓←→でカーソルを合わせて呼び出す方法。

- 「五十音順一覧」選択画面を呼び出し、↑ ↓ ← → キーで法令名の最初の一音を選択して呼び出す方法。
- ・ 法令名を直接入力して呼び出す方法。
- イ. キーワードによる検索

法律用語、法律関連用語だけでなく、自由なキーワードを入力し、条文や判例要旨 を検索できる。

ウ. 事項索引からの検索

「模範六法」の巻末にある「総合事項索引」を収録してあり、スピーディに利用で きる。

エ. 判例・先例の検索

裁判所・年月日・出典からの検索と事件名からの検索ができる。

オ. 法令の改正年・法律番号による条文からの検索

「法令名一覧表」で法令を特定し、改正年月日を入力し、その法令のどの条文が指定した期間または年月日に改正されたのかの一覧を素早くリストアップし、順次目を通せる。

カ.被 参照条文の検索

ある条文が、他のどの条文の参照条文になっているかを検索できる。これも人間に はほとんど不可能な、コンピュータならではの機能である。参照するよう示している 条文に遡って、それらを通覧することにより、条文同士の関係を効率よく理解できる。

(6) オックスホードの英語大辞典

(紀伊國屋書店、日本電気ホームエレクトロニクス;定価250,000円(税別))

OED初版の本巻全12巻を1枚のCD-ROM盤に収めている。冊子体では不可能だった幾多の検索方法が、CD-ROMを利用することにより可能となり、OEDの利用範囲の拡大が期待できる。

- ① 用 途
 - ア. 古語・外来語の逆引き検索
 - イ. 同義語検索
 - ウ. 用例文の典拠資料の出版年次、作者、作品名からの検索
 - エ. 引用例文中の単語の網羅的な検索

オ. 分野・地理区分の一覧と、それによ検索・限定

(7) バイブルズ CDーHIASK

(紀伊國屋書店、日外アソシェーツ:定価85年版100,000円/86年版100,000円/87年版120,000円/88年版120,000円(税別))

朝日新聞の全文記事データベースである。毎年1年分(1月1日~12月31日)がディスク1枚に収録されている。検索も一つの記事を、主題別、記事別、国別、年月日など、いろいろな角度から捜すことができる。

(8) バイブルズCD-BOOK

(紀伊國屋書店、日外アソシェーツ:定価500,000円(税別))

本の内容がひと目でわかるCD-ROMである。昭和61年1月より刊行の一般流通本のほか、頒布可能な非流通本も収録している。本の目次、要旨も収録し、新刊図書情報としては国内唯一のものである。

(9) バイブルズ人物情報

(紀伊國屋書店、日外アソシェーツ:定価80,000円(税別))

科学技術、医学、技術評論家から政治、財界、芸術、マスコミ、作家のほか、各分野の専門家、実務家など約26,000人を収録し、略歴、昭和50年~61年までの主要著作、研究論文、本人に関連する紹介・人物研究記事約35万件をディスク1枚に収めている。出身地、生年月日、出身校、職業、受賞名、さらには専用分野、研究主題等からの多採な検索が行える。

(10) バイブルズ世界各国情報

(紀伊國屋書店、日外アソシェーツ:定価70,000円(税別))

世界100ヶ国の統計を中心として、1ヶ国につき70~130項目(経済、貿易、投資情報等)のマクロ指標、政治、文化などのプロフィールを収録する日本貿易振興会の"JETRO ACE"と、機械振興協会経済研究所が昭和34年から作成している海外の産業・企業情報の"KSKSCANNER"の2データベースをCD-ROMディスク1枚に収録したものである。

(11) バイブルズ新聞・雑誌記事情報

(紀伊國屋書店、日外アソシェール:定価60,000円(税別))

JOINT(経済文献研究会)、KSKSCANNER (機械振興協会経済研究所)、Computer World Japan (コンピュータ・ワールド・ジャパン)、 Techgram (日商岩井)、マスコミ(日外アソシェーツ)の5データベースをCD-ROMディスク1 枚に収録したものである。特定企業や産業動向、新製品、新技術情報などが5種のソースから横断的に得られるほか、逆に各ソースごとの検索が可能なため、コンピュータ製品だけに絞り込みたい場合は、Computer Worldだけを検索して、必要のない情報・データを最初から除く — などの利用も可能である。

(12) CD-ROM 版有価証券報告書く1部・2部・地方上場会社>財務データ

(大蔵省印刷局:定価515,000(本体500,000円・税15,000)

① 約2,000社に及ぶ報告書の財務データを1枚のディスクにコード入力

証券取引法第24条に基づいて大蔵省に提出された、第1部・第2部及び地方単独の有価証券報告書の内、1988年12月期~1989年11月期に決算の行われた約2,000社の財務データをCD-ROM化している。

収録内容 — 会社の紹介、会社の沿革、貸借対象表、損益計算書、各種明細書、利益 金処分計算書、重要な会計方針、表示方法の変更、注記事項、1株当たり情報、重要な 後発事項。

② 対話形式による簡単検索

会社及び報告書の内容選択は、対話形式でメニュー画面により簡単に検索できる。会社名の選択は、証券コード、大蔵コードによる検索およびメニュー画面の会社名一覧(五十音順)、業種別、条件等による検索により行なわれる。

- ③ ハード環境は、ハイシェラ(High Sierra)フォーマット対応機種。 対応パソコンは、日電PC-9801VM以降の各機種及びIBMPS / 55シリーズ。 ハイシェラ・フォーマットの採用により、MS-DOS CD-ROM Extensions を使 える全てのドライブが使用できる。
- ④ LOTUS 1-2-3やMS-MULTIPLANでユーザーが独自に財務データを分析・グラフ化できる機能。

全てのデータをSYLK形式で入力しているので、表計算ソフト等によりユーザーが

自在にデータ分析やグラフ化ができる。

(13) 平成元年版<1989>全官公庁の機構と人事・職員録

(大蔵省印刷局:定価/CD-ROM ディスク盤 8 0,000円/CD-ROM ディスク盤 + 検索ソフト 2 0 0,000円/CD-ROM ディスク盤+ 宛名印刷機能付検索ソフト 5 0 0,000円)

- ① 50万人に及ぶ全官公庁の職員名鑑
- ア、「職員録(上・下巻)」に掲載されている約50万人の職員データを収録。
 - イ. 国及び地方公共団体等の係長(同等職を含む)以上の職位と氏名を、また課長以上 の職員は原則として住所と電話番号も収録している。
 - ウ、収録されている内容は、昭和63年7月1日現在のものである。
 - エ. 編 集

中央官庁・特殊法人	立法・行政、司法の国の機関、特殊法人の職員。官署の所在地、電話番号等。
都 道 府 県 ・	都道府県、市町村の職員。
市 町 村	官署の所在地、電話番号等。

- オ. 書籍版職員録には、上記の他に、国の各機関の掌握業務を簡記した機構図と所掌事 務要覧、定員表等を掲載している。
- ② ハイシェラ (High Sierra)フォーマットの採用によりハードの選択を不用化。 ファイル形式には統一規格であるハイシェラ・フォーマットを採用している。日電、 東芝などハイシェラ・フォーマット対応のCD-ROMドライブが使用できる。
- ③ 必要な機器
 - ア、パソコンは、日電PC-9801、VM、VX、RAの各機種
 - イ. CD-ROMドライブ
 - ウ. CD-ROM盤
 - エ、CD-ROM検索ソフト
 - オ. その他 MS-DOS(Ver 3.1以上PRODUCT 0 0 1以降)のシステム・ディスク、MS-DOS Extensions および CD-ROMドライバールーチンの各ソ

フトウェア

④ 検 索 機 能

ア. 完全一致・前方一致による氏名のキーワード検索

原則として漢字により入力するが、データに読み方が付されている場合はカナ入力も 可能である。

イ. コード入力による検索。

所属コードを指示することにより、該当する職員の一覧を表示する。また、組織・ 個人別の住所コードを指示することにより必要な職員のデータが検索できる。

ウ. メニューによる検索

コードがわからなくてもメニュー方式で検索ができる。中央官庁・地方公共団体別の分離方式で、所属先から必要な職員のデータを検索できる。更に、組織・個人別の住所は一覧表で都道府県を選択することにより検索できる。

⑤ 利 用 方 法

CD-ROM盤、検索用ソフト(FD)をそれぞれセットし、パソコンのディスプレイ(画面)に示されるガイドに従って操作すると、自動的に検索結果が表示される。

(14) 判例MASTER

(新日本法規出版株式会社:定価1,170,000円(税別))

① 41,000判例をCD-ROMに収録

判例MASTERは、国内の主要法律雑誌を網羅。民事判例約41,000件の要約と、詳細な書誌情報を1枚のコンパクトディスクに収録。行政事件・労働事件・無体財産権関係も充実している。

② 検索が簡単、22,000をキーワード設定 必要な情報をスピーディーに的確に引出すために、検索キーワードを22,000 語用意 している。

③ シソーラス機能で正確に自動検索

入力キーワードの同意語・類似語・関連語や上位概念・下位概念を自動的に検索対象とする、シソーラス機能が付加されている。

④ 高速処理機能

世界最速のCDドライブ(東芝XMシリーズ)を活用。判例MASTERは、指定の条件に対し瞬時に反応する。

⑤ 検索結果をダウンロード可能

標準の印刷機能の他、検索結果をフロッピーディスクなどに書き込むことができる。 ワープロなどを使えば、希望の書式に変えることも、文書を加工することもできる。

⑥ 年2回情報更新

最新の判例を加え、年2回更新版を発行。

⑦ 検 索 方 法

ア、AND条件入力

キーワードをタテに並べると、AND条件入力となり、例えば「交通事故」「交差点」「自転車」とすれば、数多い交通事故関係の判例の中から、交差点で自転車が関係するものを自動的に絞り込むことができる。

イ. OR条件入力

キーワードをヨコに並べると、OR条件入力となり、「原付自転車」「自動二輪車」とすれば、原付自転車又は自動二輪車に関するすべての判例が検索できる。

ウ. NOT条件入力

建物の無断増改築に関する判例の中で借地上の自己所有建物については不要の場合、NOT条件として「借地」を指定すれば、除外して検索ができる。

(15) J-BISC

(村日本図書館協会:定価142,200円(税込み))

J-BISC(Japan Biblio Disk)は図書館業務の効率化をめざして開発された。 簡単な操作で効率よく、しかもマイベースで図書の検索ができる。

① 概 要

国立国会図書館作成の国内刊行図書の書誌情報をCD-ROM化したものである。1枚のCD-ROMには、約10年分、50万件の書誌データを記憶させることができ、高速度のアクセスも可能である。

② 検 索

書名、著者、出版者、分類、件名、刊行年、JPナンバー、ISBN、官公庁コード

などからも検索できる。また、検索項目を組み合せて、きめ細かな検索もできる。

③ データの追加

TAG表示画面のユーザエリアを利用し、それぞれの図書館の請求番号、登録番号の入力ができる。フロッピーディスクにダウンロードして、データを必要に応じて追加・加工することもできる。

(16) 現代用語の基礎知識

(自由国民社:定価30,000円(税別))

① 特 色

ア. マルチウィンドウ機能

画面上に最大5つまでウィンドウを開くことができ、画面を切り替えることなく複数の検索結果を並べることができる。

ウィンドウの拡大、縮小、移動も自由にできる。

イ、ヘルプ機能

操作途中に因ったときは、ヘルプ画面に切替えてガイドメッセージを見ることができる。

ウ. 簡 単 な 操 作

キー入力以外はすべてマウスを使って簡単に操作できる。

エ. 複 写

検索結果をプリントアウト、またフロッピーディスクに保存でき、ユーザ独自の情報ファイルが作成できる。

オ、キーワードセレクト機能

キーワード検索のレディーメイド版であり、新しい角度からながめた現代社会の縮 尺図として活用できる。

カ. 複合語検索

新語には多くの複合語が含まれている。とのCD-ROMでは複合語に含まれる単語単位をキーワードとして抽出、条件検索を利用して中間一致検索ができる。

② 検 索 機 能

ア. 入力 語検索

(1) 見出し語検索

普通に辞書を引くように「ひらがな」「カタカナ」で見出し語を入力して検索する。

(2) 表記形(漢字・アルファベット)検索

見出し語の「漢字」「アルファベット」を入力して検索。

≪例≫ 「電子出版」「AI」「レトロ」「AIDS」 ……

(3) 前方·後方一致検索

語句の語頭あるいは語尾の部分だけを入力することで、それを構成要素とする語をすべて検索できる。検索結果(項目一覧)表示後は、キー入力の必要がないので連続的に検索することができる。

イ. 画面語検索

画面上にさらに検索したい語句があるときは、その語句をカーソルで指定するだけ で検索できる。キー入力の必要がないので連続的に検索することができる。

ウ. 条 件 検 素

いくつかのキーワードを入力すると、その語句を解説文中に含む見出し語を検索することができる。キーワードは、AND条件で最大5個、さらにAND条件1個につき各3個までのOR条件を入力することができる。

エ. 履歴情報(ヒストリー)検索

過去に検索した項目(最新の項目から32個前まで)の一覧表示からカーソルで指 定するだけで再度検索することができる。

ォ. 分野別検索

書籍版と同様に、各分野から用語を検索することができる。ある分野の全貌を眺めたい時に便利である。

(17) 最新医学大辞典

(医歯薬出版株式会社、定価:60,000円(税別))

臨床・基礎・社会医学系をはじめ歯学・薬学等に関する「最新医学大辞典」の全項目をCD-ROMに収録。医療スタッフの業務・研究用、また医学生の学習用として利用できる。

≪収録内容≫

医歯薬出版「最新医学大辞典」の全項目 本文 約40,000語 凡例・出筆者一覧・医療関係組織住所録・挿絵も含む

(18) 角川類語新辞典

(角川書店、定価:30,000円(税別))

「角川類語新辞典」は、言葉のニュアンスの確認や適切な単語選択ができるよう、日本語を意味別に分類・体系化した辞典でありその全項目がCD-ROMに収録されている。

≪収録内容≫

㈱角川書店「角川類語新辞典」の全項目

本文 約50,000 語 ; 凡例・略語・分類メニューも含む

(19) ニューセンチュリー英和辞典・新クラウン和英辞典

(三省堂、定価24,00円(税別))

「ニューセンチュリー英和辞典」と、「新クラウン和英辞典」の2冊の内容を、一枚の CD-ROMに収録している。

≪収録内容≫

㈱三省堂「ニューセンチュリー英和辞典」の約30,000語

英文法のまとめも含む

㈱三省堂「新クラウン和英辞典」の約40,000 語慣用語も含む

2. CD-ROM用検索システム機能の高度化

--- 検索機能を中心に ---

CD-ROMは、現時点に於いて、最も高密度な、しかも軽量・小型な情報蓄積媒体である。そのために、前章の調査結果に於いて示されているように、大量流通・個人使用の利便性から、フル・データの提供媒体として、多くのデータベース(以下、この章ではCD-ROM-DBと記す)の提供に利用され出している。

CD-ROM-DBの提供形態は、現時点では、データベースそのものをCD-ROMに、その為の検索ソフトウエア(検索システム)をフロッピー・ディスクに収録し、この2つをセットにして提供している。このことは、メイン・フレームとしてのパーソナル・コンピュータとCD-ROMドライバーを用意すれば、その時点で、必要な情報検索ができ、しかもフル・データを得られる便利さを提供する。

さらに、最近では、メイン・フレームとしてのパーソナル・コンピュータとそれにCD-ROMドライバーを一体化したコンピュータ(AVマシン)の出現により、CD-ROM-DBにデータベースと共に検索ソフトウェアを組み込み一体化した形式のCD-I型СD-ROM-DBが出現し出してきており、より多様な利用が可能になってきている。

また、最近のコンピューは、例えば、音声、色彩等による多様なデータの処理が可能になってきている。同様に、CD-ROMにも、例えば、音響、写真等のデータの蓄積が可能になってきており、また多様な形態のデータ形式による情報の提供が可能になってきていることから、より高度な検索システム機能の検討が求められてきている。

そこで、この章では、CD-ROM-DBの特性について検討し、従来のデータベースに比べCD-ROM-DBを対象とする検索システム機能の高度化機能について検討する。

2.1 CD-ROM-DBの特性

2.1.1 検索結果として得られる情報形態

人間の五感は、情報受容器官である。故に、人間は、5種類の情報表現(視覚に対し文字・図形・色彩による情報表現、聴覚に対し音声による情報表現、触角に対し熱による情報表現、味覚に対し味による情報表現、臭覚に対し匂いによる情報表現)を持つ。マルチ・データベースとして、現時点のCD-ROM-DBは、その内の2種類の情報の蓄積を可能にしている。故に、CD-ROM-DBからは、文字表現、図形表現、色彩表現および音声表現による情報を検索結果として得ることが可能である。このことは、

文字表現だけのデータベースに比べ、情報入手の目的から、以下の様なメリットが認められる。すなわち、文字情報(文章情報)では表現しきれない、もしくは文字情報では理解しきれない情報について、図形表現、色彩表現および音声表現により、いわゆる"説明をしなくても分かる"情報の提供ができるというメリットがある。例えば、

- ① 『リンゴの形は?』という検索要求に対して、『丸い形』と出力してもイメージは様々に別れる。このことに対して、リンゴの"絵"もしくは"写真"を出力できれば一目瞭然である ……… 図形情報による出力。
- ② 『紫はどんな色?』という検索要求に対して、これは説明しにくい。『輝度何度』と出力しても何にもならない。このことに対して、"紫色"を直接出力できれば一目瞭然である ……… 色彩情報による出力。
- ③ 『ウグイスの鳴き声は?』という検索要求に対して、これも説明しにくい。『ホーホケキョ』と出力しても何にもならない。このことに対して、"鳴き声"を直接出力できれば一目瞭然である ……… 音声情報による出力。
 - ④ 『コマの回る様子は?』という検索要求に対して、これも説明しにくい。『回転する』と出力しても何にもならない。このことに対して、"回る様子"を動画像として直接出力できれば一目瞭然である ……… 動画像による出力。この様に、CD-ROM-DBは、情報提供の幅を広げたことになる。

しかし、前述のように、人間には残る触覚、味覚、臭覚による情報認識感覚があるが、現在のCD-ROM-DBでは、というよりも現行のコンピュータでは、触覚、味覚、臭覚に訴える情報、すなわち、"熱"、"味"、"匂い"を直接出力はできない。

例えば、『快適な温度は?』、『リンゴの味は?』、『サンマの焼く匂いは?』という検索要求に対して、"熱"、"味"、"匂い"を現行のCD-ROM-DBから直接出力はできない。この3つの情報提供が可能になれば、データベースは、より多様に利用されるようになる。

一方、原情報は、文字、図形、音声、色彩により表現されるマルチ、データといえる。そこで、原情報のフル・データを蓄積できるCD-ROM-DBの検索結果出力に於いて、文章による説明的なデータを出力するよりも図形、音声、色彩により出力するほうが直観的に理解しやすい情報については、図形、音声、色彩により出力するほうが効果的であることから、蓄積する原情報の特性を良くとらえ、類別し蓄積することが重要と

なる。また、これらの図形、音声、色彩表現のデータ出力には、最近のコンピュータでは、さらに、AV型コンピュータの出現により処理が可能となってきている。そこで、マルチ・データベースとしてのCD-ROM-DBの特性をより効果的に生かし、多様な情報検索結果が得られるよう原情報のデータベース化と、その特性を生かす検索システムの機能化を計る必要がある。特に、文字表現、図形表現、色彩表現および音声表現による情報を扱えることが可能となった以上、CD-ROM-DBを対象とする検索機能は、これら特性を充分生かし、高度化を計る必要があるといえる。

2.1.2 高度データの検索の可能性

高密度情報蓄積媒体であるCD-ROMは、フル・データを対象とするデータベースの提供に利用されだしている。すなわち、CD-ROM-DBに、原情報(一次情報)そのものの内容全でを収録することにより、多様な情報提供を可能にしている。そこで、フル・データそのままの提供に限らず、フル・データの内容解析処理を可能にすれば、原情報からファクト・データ、さらには知識データを抽出し、提供することも可能になる。

例えば、文例として、下記のブル・データ

「日本語の速記タイプライタの発明者である川上は、かな鍵盤を用いて日本文の各文字を2ストロークでコード化するラインプット方式を提案した。」に対して、下記のレベルのデータ(情報)の抽出を可能とすれば、フル・データの蓄積を可能とするCD-ROM-DBの高度利用が図れる。

① 索引データ(内容索引)

シソーラス用語 …… 『仮名漢字変換システム、(ラインプット)方式 』 自然語キー ………… 『日本語、速記タイプライタ、発明者、川上、かな鍵盤、 日本文、文字、2ストローク、ラインプット方式 』

② ファクト・データ

『仮名漢字変換方式の中の一方式であるラインブット方式は、川上が提案した。』 ファクト・データの形式例 …… { (提案した) { (川上)、(仮名漢字変換方式(ラインブット方式)) } }

- ③ 知識データ
 - ②『川上は、ラインブット方式を提案した。』 知識データの形式例 …… {(提案した){(川上)、(ラインプット方式)}}
 - ⑤『川上は、2ストロークで行った。』
 知識データの形式例 … {(行った){(川上)、(2ストロークで)}}
 - ②『川上は、日本語速記タイプライタの発明者である。』知識データの形式例 … {(である){(川上、発明者)、 日本語速記タイプライタ)}}
 - ①『川上は、かな鍵盤を用いた。』知識データの形式例: … {(用いた){(川上)、(かな鍵盤)}}
 - ②『2ストロークは、各文字をコード変換するものである。』知識データの形式例 … {(ものである){(2ストローク)、(各文字、コード変換)}}

2.2 データベース化および検索システム機能の現行システムの問題点

データベースによる情報提供の一連のシステムは、検索機能と関連し、原情報に対していかなるデータを生成するか、またそのデータ生成に必要な機能をどうするかという 3点の観点についての検討を必要とする。

理想的な一連のシステムは、原情報のみをフル・データとして自動的にデータベース 化でき、検索要求内容を文章形式で受け付け、データベース内のフル・データ内容を自 動的に解析・検索し、検索結果を検索者が最も理解しやすい形式で出力する機能である。 それには、文脈解析、文意抽出等の情報内容の自動解析を行う必要がある。しかし、現 時点では、技術的に実現不可能である。

そこで、現時点では、検索要求内容を代表する検索キーを利用者に指示させ、それにより検索を行う方式が採られている。そのために、データベース化する際に、個々の情報に対して、事前に情報内容を分析し、内容を代表する索引キーを付与(インデクシング)している。すなわち、情報ファイルと個々の情報に対し与えた索引キーのファイル(インデックス・ファイル)を作成し、検索ユーザによって指示された検索キーを基にインデックス・ファイルに対して照合処理を行ない、対応する情報を情報ファイルから出力する機能になっている。

その為に、現行の一連のシステムには、以下の様な問題点がある。

2.2.1 データベース構築上の問題点

1) 索引データ生成作業効率上の問題点

索引データの生成は、人手で行なわざるを得ず、データベース化の対象情報量に対し、インデクシング作業の遅れが発生し、データベースによる情報提供の Time lag (遅れ) およびデータベース化経費アップが生じている。 Time lag の問題は、速やかなる情報流通の妨げになり、経費アップは、情報入手の経費高に直結している。

2) 索引データの最適性に対する問題点

索引データの付与には、情報内容の解析を必要とする。その結果として情報内容を表示する索引データを付与しなければならない。このことに対して、人手によることから、例えば、対象情報内容の専門的な理解、記述言語読解、索引キー付与の深度の一定化等に対し難易性があり、最適な索引データの付与、平準化といったことに対して、"揺れ"が生じ、その結果として"検索洩れ"、"ごみ出し"を生じさせ、検索結果精度に影響を及ぼしている。

また、検索キーと索引キーとの照合処理において、両者の形態上の一致性を補償するために、検索キーと索引データの決定にシソーラス内の用語を用いるが、対象となる情報の逐次性に対しシソーラスのメンテナンスの逐次性(更新)に難易性があり、遅れが生じ、検索要求内容および原情報の内容に対して最適な検索キーと索引キーの設定ができないという問題が生じ、最適な検索キーと索引キーの判明ができないという現象がしばしば生じる。

2.2.2 検索キーと索引キーの照合(検索処理)機能の問題点

1) 検索要求内容の表現上の問題点

検索要求内容の表現に、現行のシステムでは、まず、検索要求内容を代表するキーワードを選択し、そのキーワードを基に、検索要求内容の文脈に照らし、配列することにより表現する方式が採られている。配列の方式は、論理演算式表記を用いる。

例えば、

『CD-ROM-DBの情報検索方式について、日本語または英語で記述されている技術論文について、最新のものを知りたい』という検索要求内容に対して、

キーワードとして『CD-ROM-DB、情報検索方式、日本語、英語、最新の技術語文』が選択され、

論理演算式として『CD-ROM-DB×情報検索×方式×(日本語+英語)× 論文』が設定される。

ただし、選択されたキーワードは、論理演算式に設定するときに、検索キーとの 形態的一致性を計るためにシソーラス内の用語を用い、意味的類似性から選択し、 用いる。

上記プロセスに於いて、下記の問題点が指摘できる。

- ① 常に上記プロセスを経なければならず、検索行為に繁雜さが認められる。
- ② 最適なキーワードの設定に困難さが認められる。

例えば、上記例に於いて、『最新』という要求内容について、具体的にどう表現するのが最適であるのか、要求内容の具体性と合わせ問題となる。この場合、『最新』の内容を、例えば『発表年』で指定するのが一般的である。

③ 選択したキーワードに対して、シソーラス用語を確認する際に、意味的類似性から検索キーを選択することになるが、シソーラス内の用語との間に意味理解に不一致性が生じ、検索精度に影響を与える。

例えば、上記例に於いて、『技術論文』は『論文』の下位範疇になり、広く検索 されることになり、要するに"ごみ出し(利用者要求に対するノイズ)"を生じさせる。

2) 形態不一致上の問題点

検索キーと索引キーの照合処理は、現行のシステムでは、キーの表記上の形態ー 致により行っている。この方式には、下記に示す幾つかの問題点を生じさせる。

① 同義異表記キーは照合されず、"検索洩れ"を起こす。 例えば、

短縮系表記キーの場合 … 『電子計算機』に対して『電算機』、『計算機』は 照合されない。

カナ表記キーの場合 … 『コンピュータ』に対して『コンピューター』等、 多表記キーは照合されない。

② EBCDICコード表記キーの使用に於いて、同音異義語キーは検索されてしまい、"ごみだし"を起こす。

例えば、

『記者』に対し『汽車』は照合されてしまい、"ごみ出し"を起こす。

③ 複合キー内の部分組み合わせ検索(トランケーション検索)の場合、"検索洩れ""ごみ出し"を起こす。

例えば、

『図書館業務処理システムについて知りたい』という場合、『図書館業務システム』という複合語索引キーがシソーラス用語に無い場合には、現行のシステムでは、例えば、『図書館×システム(×は and 条件を表す。)』という指示により検索を行うことになり、この場合、『図書館』および『システム』のキーが付く全ての検索キーに対して、前方一致照合、後方一致照合、中間一致照合処理を行い、検索結果を得る。この結果、例えば、『図書館の組織』に関する検索結果が得られ、"ごみ出し"が生じる。

3) 意味的不一致上の問題点

索引キーと検索キーの表記上の形態一致処理によることから、下記に示すキーの 意味的な照合処理が出来ず、幾つかの問題点が生じる。

① 類義語キー検索ができない。この結果、"曖昧な検索要求"(関連する情報の検索)に対処できず、"検索洩れ"を起こす。

例えば、

『電子計算機』に対して『コンピュータ』は照合されない。

これら、現行の検索システム機能の利用に於いて、検索システム利用者は、この現象が発生することを知っていて検索する必要があり、例えば、短縮系表記キー、カナ表記キーの全てを、すなわち、可能性のある検索キーの全てを、論理和の形式で指示するとか、検索結果の"ごみ"は、検索結果を見て取り除くことをしなければならない。

2.2.3 検索結果の出力上の問題点

データベース検索システムへの利用者の期待は、情報の網羅性と検索結果(情報) の直接的利用にある。網羅性は、そのデータベースの利用によって、情報入手の充足 性を補償するものでなければならず、また、情報の即時性を補償するものでなければ ならない。検索結果の直接利用を可能にするためには、フル・データを提供できるデ ータベースでなければならない。

このことに関して、現行の多くのデータベースに以下のような問題点がある。

1) 情報の網羅性と即時性についての問題点

情報の網羅性については、世の中に存在する大容量の情報に対して、必要となる 情報をいかに選択し、収録するのが最適であるのかといった難ずかしい問題がある が、即時性については、現行のデータベース構築のプロセスの多くが人手で行われ ていることから、多くのデータベースに於いて問題となっている。最大のネックは、 索引キーの人手による付与プロセスにある。

2) フル・データの提供の問題点

フル・データの提供については、CD-ROMの出現により、解消されつつあるといえる。しかし、フル・データの収集およびCD-ROM化に大変な経費がかかる問題はある。

CD-ROM-DBは、大量のフル・データをコンパクトに提供することを可能にし、同時に2.2で述べたように、文字表現、図形表現、色彩表現、音声表現による情報の同時提供を可能にしているので、最適な形態での検索結果出力表示が可能である。

このことに対して、現行のデータベースは、原情報の紹介を目的に二次データの 提供を目的とするものが多く、これら多様な情報提供に対応できない現状にある。

2.3 高度化機能

2.3.1 データベース構築機能の高度化

個々の情報の認識には、情報そのもの(情報内容)とその情報の属性データ、すなわち索引データ(例えば、表現言語種別等のデータ)が必要である。その証拠に、情報検索に於いて、検索結果の情報利用を意図し、情報そのものと、情報の属性を制限する指示が成される。

例えば、"日本語で記述されている情報 **という要求は、情報の属性を制限している ことになる。そこで、データベースの構築には、この2点の、すなわち、情報内容そ のものと情報の属性データをセットし効率良く蓄積する機能化が必要になる。

データベースの構築に於いて、対象となる情報の全て(いわゆる、フル・データ)を入力することが理想である。この場合、大容量の情報入力の効率化および正確な属性データの生成の効率化が問題になる。特に、CD-ROM-DBの構築に於いては、その蓄積容量から、大容量のフル・データのデータベース化が可能であることから、

原情報の蓄積が可能であり、効率性という側面から、原情報の自動的ダイレクト入力機能および属性データの自動生成機能が必要になる。原情報の自動的ダイレクト入力方式としては、例えば、ドキュメントを対象とする場合には、イメージ・センサーにより入力することを、属性データの自動生成には、イメージ・センサーで取り込んだフル・データの内容を自動的に解析し、各データ項目を自動的に識別し、その内容を自動的に判別処理するシステムの機能が必要になる。

このことに対して、最近では、原情報のダイレクト入力としてOCRが利用できる様になってきている。残る問題として、情報内容の自動解析機能がある。しかし、情報内容の自動解析機能は、未だ研究開発の途上にあり、実現には尚時間がかかる。そとで、このことに対して、最近、コンピュータの性能の向上および自然語処理技術の成果を生かし、索引キー生成および検索機能の高度化システムとして、下記2つの機能のシステムが実現し、実用化されてきている。

2.3.2 索引キー生成機能および検索機能の高度化

1) 人手による索引キーとシソーラスを用い自然語キーを自動的に抽出・設定し、両者を併用する方式

索引キーの欠落回避および最新の語に対応する索引キーの付与を保証するために、 人手による索引キー(この場合シソーラス用語により設定される)と共に、シソー ラス内の用語を基にデータベース化された情報のタイトルもしくは抄録を対象に形 態素解析(単語切り)を行い、人手による索引キー以外の語を索引キーとして自動 抽出し、インデックス・ファイルを生成し、自然語キーによる検索をも可能にした システム。

システム事例としては、例えば、JICSTのデータベース構築に於いて運用されており、また㈱平和情報センターから「HAPYNESS」が、㈱エム・シー・ワードセンターから「Free Base」といったシステムが市販されている。

2) シソーラスを用い自動抽出・設定し、人手による索引キー付与をおこなわない方式

システム内にシソーラスを維持し、このシソーラス内の用語を基にデータベース 化された情報を対象に形態素解析(単語切り)を行い、シソーラス内の用語と形態 的に一致した語を索引キーとして自動抽出し、一致したシソーラス内の用語の下に 配列し、対象となったカテゴリ内の用語のセット全てを索引キーとして、対象の情 報に対し位置付け、検索キーをシソーラス用語を用いず指定させ、検索するシステム。当システムは、最近では、フル・テキスト・データを対象とするCD-ROM-DBの多くの検索システムに用いられてきている。

システム事例としては、例えば、Thunderstone /EPI Inc.から「METAM-ORPH」が、Adovanced Projects Internatinal Inc.から「DARS」といったシステムが市販されだしている。

特に、2)のシステム機能を、最近では、データベース管理システム(DBMS) に組み込むことの開発もなされている。

しかし、

- 1)のシステムは、シソーラス用語のメンテナンスの遅れに対処、および索引深度の保証を図り、検索精度の向上を狙ったシステムといえる。しかし、検索漏れを防ぐために基本的な索引キーは人手により付与しておくことになり、課題を残している。
- 2)のシステムは、人手による検索キーの付与、および類語キー検索機能の保証を 図り、検索漏れを防ぐことを狙ったシステムといえる。しかし、抽出されるキーが、 形態的に抽出され設定されることから、あくまでも表層的な処理結果のキーであり、 情報の意味内容を捕らえ設定されるかという保証はなく、課題は残している。

また、両者とも、検索要求内容に対しての適用は、現時点ではさせていない。 そこで、特に、フル・データの提供を可能とするCD-ROM-DBへの適用か ら、2)のシステムについて、以下の機能が必要になると考えられる。

(1) 検索要求内要に対する適用

いずれにしても、検索時に検索式(論理式)を作成する必要があり、依然煩わしさが存在する。そこで、検索要求内容を文章で表現させ(理想的な方式である)、検索内容について、カテゴリ構造を生成し、一度、検索者がその内容に対して検索キーの更新を行い、その後に、システムに送り、検索を実行させるよう機能化することが考えられる。この方式によれば、検索キーの洩れもカバーでき、検索洩れの改善に結び付く。

例えば、下記のようなカテゴリが生成されたとする。

『KW: 電子計算機

RT : コンピュータ

電算機、計算機

NT : パーソナル・コンピュータ

大形電子計算機 』

(ただし、 KW … 主キー、

RT … 主キーと同義語キー、

NT … 主キーに対する下位概念語キーを表わす。)

このようなカテゴリーに対して、検索実行前に、

この場合の検索キーは、つぎのようなものになる。

『KW: 電子計算機

RT : コンピュータ、コンピューター、電算機

NT: パーソナル・コンピューター、PC』

(2) シソーラスの増殖機能化

常に、システムが維持するシソーラスを基本的に使用するのではなく、上記 1) によって更新されたカテゴリーをシソーラス内に取り込み利用できるようにする。 ただし、フル・データから更新されたものは取り込まない。何故ならば、いわゆる "ごみキー"の発生が起るからで、あくまでも検索利用者による更新カテゴリを利 用する。このことにより、より最新の、かつ充実したシソーラスを維持、利用でき、 検索効果に寄与することになる。

フル・データを蓄積可能とするCD-ROM-DBの検索システムとして、2)のシステムに上記機能を付加し利用することが、現時点において最も最高なレベルのシステム機能となろう。

Ⅳ 今後の課題

本調査研究は、昨年度からの3年計画で実施しており、最終的には複合的データベース・システムのパイロット・システムをパソコン上に構築することを目指している。そこでは、最近脚光をあびているCD-ROMの採用を考慮する必要があろうと思われる。すなわち、本パイロット・システムの構築にあたっては、「複合的データベース・システムの全体像」を基本におきながら、「CD-ROM用検索システム」の構成を踏まえて、その設計をおこなうことが必要となろう。

そこで、今年度の調査研究における「今後の課題」として、

- ① 複合的データベースのCD-ROM上での実現にむけて
- ② データベース的アプローチによるシステム構築の問題点
- の2点について問題を提起することによって、今後の調査研究の指針としたい。
- 1. 複合的データベースのCD--ROMトでの実現にかけて

現在、パソコン上で稼動する外部記憶媒体で、比較的安易に利用できるものとして、

- ① ハード・ディスク(10~100Mバイト)
- ② フロッピー・ディスク(1Mバイト程度)
- ③ CD-ROM(500~600Mバイト)

があるが、この中で、CD-ROMはその高密度性、量産性、経済性という観点から、大容量データの記録に適していると言えよう。従来、CD-ROMは、あらかじめ記録されているデータの読み込みのみが可能(書き込みは不可)であったが、最近では、その発展形として、追記形光ディスクや書換形光ディスクも登場している。(CD-ROMも広い意味では光ディスクであるが、通常、書換え(追記)可能なものを光ディスクとよび、書換え(追記)不可のものをCD-ROMとよんでいる。)

パソコン上に、複合的データベース・システムを構築するにあたっては、このCD-ROMのもつ特長、とくに大容量性を利用することが得策であろう。すなわち、複合的データベース・システムにおいては、数値データや文字データのように比較的少ない容量で記録できるデータ型ばかりでなく、フル・テキストを前提とした文書データ、さらにはイメージ・データや音声データのように、その記録と多くの容量を必要とするデータ型にいたるまで、多種類のデータを総合的に管理することが必須の条件となっていることからも、

当面の大容量記憶装置としてCD-ROMの採用が考慮されるべきであろう。ところで、 このCD-ROMとCD-ROMドライブは、国内でも多くの企業で販売しているが、従 来はハイシェラ形式という多少共通的な形式はあるもの、CD-ROMの標準形式が存在 しなかったために、CD-ROMごとにソフトウェアが用意されており、現在もこのこと は尾を引いている。しかし、1988年にCD-ROMのファイル形式としてISO96 60が制定され、この標準形式によるCD-ROMやCD-ROMドライブが普及しつつ あり、今後は、ハイシェラ形式とISO9660形式とが主流になるものと思われる。こ のような標準化の動きと相俟って、国内においても、このISO9660形式に対応して、 日本語対応CD-ROM標準データベースの構築がおこなわれているが、CD-ROMの 標準化という観点からすると、単にファイル構造の標準化だけではなく、文書データ型、 図表データ型、画像データ型、音声データ型など各種のデータ型に対応した表現形式の標 準化は、現時点で定まったものはなく、実験の段階にあると言えより。いずれにしても、 CD-ROM上での各種データ型の表現に依る論理構造(アプリケーション・レベルでの ファイル形式)は、標準化されていくことになると思われるが、CD-ROMの普 及とCD-ROM処理技術の発展を考えた場合、その標準化(規格化)を待つので はなく、積極的に独自の手法でシステムを構築していくのが得策と考えられる。こ こで、注意が必要なことは、CD-ROMのデータ操作にあたってのデータ表現を、で きるだけ論理的な(アプリケーション・サイドの)形式として定義しておくことである。 これによって、言わば、CD-ROM(装置)の種類とは独立した形でのアプリケーショ ン・システムが構築でき、実機への移植も比較的容易におこなえることが期待できるから である。

ところで、わが国におけるCD-ROMおよびCD-ROMドライブの普及は、当初の期待に反して伸び悩んでいる状況にある。その理由は、海外製CD-ROMの普及を困難にしていると考えられるつぎの問題によっているという指摘がある。(平成元年3月「CD-ROM調査研究報告書」刷データベース振興センター)

- ① CD-ROMソフトを積極的に使用しようという市場環境が整っていない。
- ② ①の大きな要因となるが、エンドユーザが強く要望する魅力あるソフトが、わが国には乏しい。
- ③ 海外製CD-ROMは、ほとんどハード的には、IBM-PC(またはコンパチ機) 対応で作られているが、日本ではIBM-PCパソコンは $2\sim3$ 万台と普及数が少ない。

④ IBM-PC(APPLE製でも同様)用のCD-ROM検索プログラムの日本製P Cへのプログラム変換は、ソフト所有権がソース・コードを開示しない、仮に開示して も開発費が高額である、などの理由で製品化が難しい。

しかしながら、これらの理由はCD-ROMに係る、言わば、外的環境の不備に帰因し ているものであり、CD-ROMの複合的データペースへの適用に関しては、何ら障害に なるものではない。しいて、複合的データベース・システムへのCDーROMの採用にあ たって、多少、障害になるとすれば、それは②の魅力あるソフトの不足を挙げることがで きようが、このことは、利用者のデータ利用に対する目的意識にもよるのではないかと思 われる。よく、データペースの高度利用は、データ処理やデータペース操作に係る高度な 技術的背景の必要性もさることながら、利用者の目的(問題)意識、言いかえれば、利用 者のデータペースの把え方によって、利用者の問題解決に最適な情報が獲得(データベー スを高度に利用)できたり、うまくいかなかったりすると言われている。本調査研究にお ける複合的データベース・システムも、データベースの高度利用をはかることをその主目 的としていることからすれば、個々のCD-ROMソフトに拘わることたく、ハードディ スクやフロッピー・ディスク上の情報等とも広く関連をもたせ得るような体系を用意する ことが重要な点となるように思われる。このような体系(複合的データペース・システム) に沿って、利用者が問題の解決にあたると言うことは、「〇〇辞書CD-ROMから×× ×単語を引く^注」という意識ではなく、たとえば、先に述べた文章データ検索の例のよう に、「 B 商品の最近の市場動向は? 」というような問題意識をもって、各種データベース 操作の一環としてひとつのCD-ROMを参照すること(データベース・ナビゲーション の一環としてCD-ROMデータを参照する)、またはそれを可能にしておくことが重要 であり、このような体系の中で各種CD-ROMソフトの需要も喚起されるのではないか と思われる。

- (担) データベース利用者の中には、まさに、「○○辞書から単語を引く」ことを目的としている業務も考えられるが、そのような利用者にとっては、○○辞書CD-ROM利用に対する需要は大きいであろう。
- 2. データベース的アプローチによるシステム構築の問題点

システム設計におけるデータベース的アプローチとは、システム構成の中心にデータベースを配置し、それを管理するデータベース・システム(具体的には、そのデータベース

に対して当該データベースシステムのデータ定義言語をもちいて設定したデータベース構造)をインターフェイスとして、各種のデータ処理プログラムを結合していく手法である。とくに、最近のように、データベースの利用形態も定型的な業務処理だけでなく、随時発生する非定型的な処理に臨機応変に対応していくことが必要となっている。このためには、データベース構築の時点では、必ずしもその処理形態が定まらないような業務に対しても、データ操作がおこなえるような状況にしておく必要がある。本調査研究での対象となっている複合的データベースシステムも、このことは当初からのシステム目的となっている。

ここで一番問題になる点は、データベースと、業務処理プログラムとのインターフェイスとなるデータベース構造(基本的なデータ操作も含む)である。リレーション型データベース・モデルは、このようなデータベースに対する(あらかじめデータのアクセス・パスを設定できないような)多目的利用に対する有力なモデルを提供していることはよく知られている。言い換えれば、リレーション型データベース・モデルは、システム設計におけるデータベース的アプローチの基本的な考え方になっているとも言える。ところが、このモデルは、事象をデータ項目の列で表現(リレーション)する点では明解であり、それにもとづくデータ操作(集合型データ操作と論理型データ操作)も論理的かつ意味的に明解であるが、個々のデータ項目の表現に関しては無関心である。すなわち、複合的データベース・システムでは、データ型として、

- ① 数値データ型
- ② 文字データ型
- ③ 年月日データ型
- ④ 文章データ型
- ⑤ 図表データ型
- ⑥ イメージ・データ型
- ⑦ 画像データ型
- ⑧ 音声データ型
- ⑨ 文書データ型

など、豊富なデータ型と、それぞれに特有のデータ操作を用意しているため、リレーショ 型データベース・モデルの体系を前提としながら、上記のようなデータ属性を追加する必 要がある。この2つの要素を統合的にあつかうための手法として、オブジェクト指向デー タベースの考え方を導入していくことが有力である。すなわち上位のオブジェクト(クラ ス)としてリレーション型のデータ構造を設定し、つぎに複合データ型を配置し、最下位のオブジェクトとして基本的なデータ型から構成されるデータ項目を位置づけるという方法である。このような体系を設定することによって、リレーション型データベース・モデルの特長、オブジェクト指向データベースの特長、さらには、上に挙げたような各種データ型(データ操作)の特長をそれぞれ利用しながら、複合的データベースに対するデータ操作を実現することが可能になることが期待される。しかしながら、現実の複合的データベース・システムとして実現させるためには、それぞれのモデルに沿った論理構造の設定や、多種類の記憶媒体(ハード・ディスク、フロッピー・ディスク、CDーROMなど)間の結合を意識した記憶構造の設定など、残された問題も多い。

---- 禁 無 断 転 載 ---

平成2年3月発行

発 行 財団法人 データベース振興センター 東京都港区浜松町2丁目4番1号 世界貿易センタービル7階 阻 (03)459-8581

委託先 リソースシェアリング 株式会社 東京都江東区潮見2丁目10番24号 カテナビル5階 IL (03)615-3271代

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 阻 (03)5685-3701代

