データベース構築促進及び技術開発に関する報告書

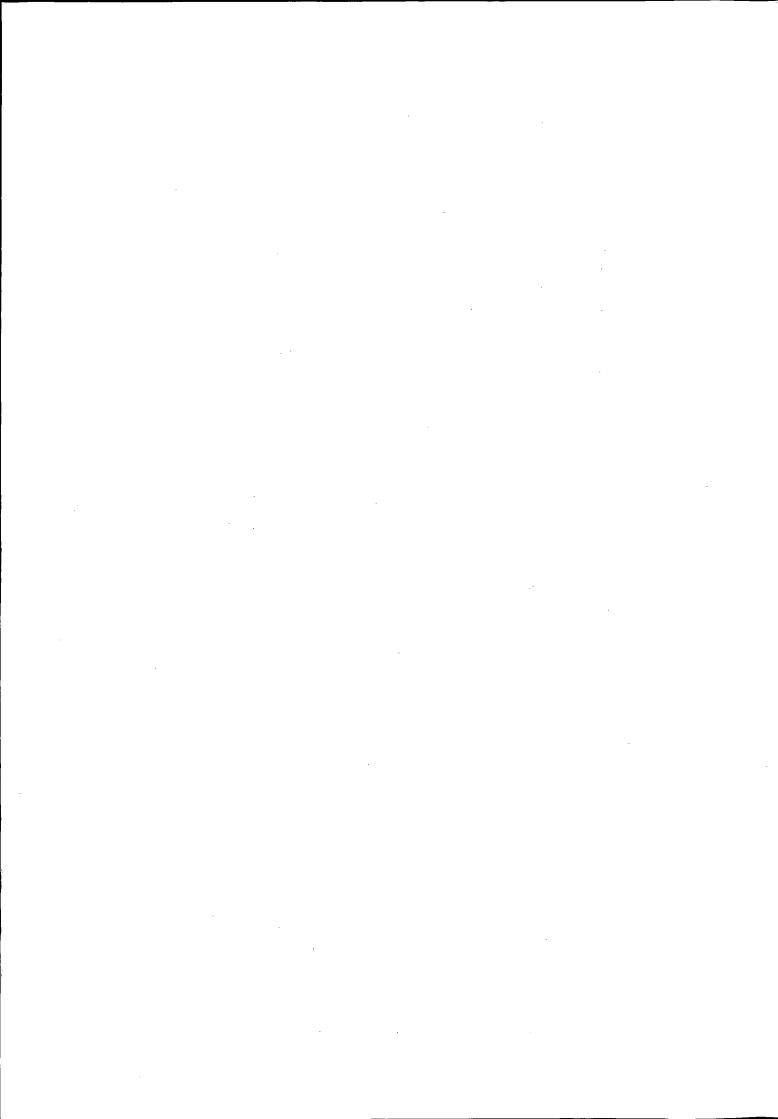
(要 旨)

平成8年3月

財団法人 データベース振興センター

KEIRIN OO

この事業は、競輪の補助金を受けて実施したものです。



			•	
		•		
,				

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及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及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。

このような要請に応えるため、側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及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、また地域及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 前山梨学院大学教授 蓼沼良一氏)を設置している。

この「データベース構築促進及び技術開発に関する報告書(要旨)」は平成7年度のデータベースの構築促進及び技術開発促進事業として、当財団が民間企業、団体等に対して委託実施した課題についてそれぞれの成果の概要をとりまとめたものである。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7年度データベースの構築促進及び技術開発促進事業で実施した課題と委託先は次 表のとおりである。

平成8年3月

平成7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 法的データベースにおける多分野データベースの統合一体的管理とオフ・オンラインの融合化に関する調査、研究	㈱日本法律情報センター
社 会	2 新聞記事分類キーワードの標準モデル構築と自動付与に関する調査研	(㈱エレクトロニック・ラ
	究	イブラリー
	3 パソコンを用いた地図データベースの基礎構築	(前朝日データサービス
	4 景観シミュレーション用樹木のデータベース構築	(㈱ストゥディオサカイ
	5 包装機械データベースの構築	做日本包装機械工業会
中小企業振興	6 新規事業創出支援のためのデータベース構築に向けた基礎調査	(㈱日本インテリジェント
地域活性化		トラスト
	7 高効率化先進材料ファクトデータベースのパッケージ化	(脚次世代金属・複合材料
		研究開発協会
	8 阪神・淡路大震災の情報デジタル・アーカイブ	神戸マルチメディア・イ
		ンターネット協議会
	9 Mosaicの利用によるマルチメディアデータベース検索システムの構築	日本電子開発㈱
	10 大規模データベースにおける構造化情報抽出方式の調査	(株)日本総合研究所
	11 モーバイルデータベースシステムに関する調査研究	(㈱イフ・アドバタイジン
技術		グ
	12 情報収集ロボットによるInternetでのWWW所在検索データベースの 構築	日外アソシエーツ(株)
	13 インハウスデータベース用CGI作成の調査研究	日本電子計算(株)

目 次

1.	法的デー	- タベースにおける多分野データベースの統合一体的管理とオフ・オンラインの融	蛤
	化に関す	する調査,研究	1
	1.1 多分	分野データベースの統合一体的運用	1
	1. 1. 1	目 的	1
	1. 1. 2	法令データベースの構築	. 1
	1. 1. 3	判例, 法令, コメントデータベースの統合一体的運用	3
]	1.2 才	フラインとオンラインの融合化	4
	1. 2. 1	目 的	4
	1, 2, 2	融合化の方法	5
	1. 2. 3	各方式の検討	7
]	1.3 おれ	うりに	8
2.	新聞記事	事分類キーワードの標準モデル構築と自動付与に関する調査研究	9
,	2.1 概	要	9
	2. 1. 1	調査研究の背景と目的	9
	2. 1. 2	調査研究の内容と結果の概略	9
	2.2 実が	歯結果	11
	2. 2. 1	アンケート調査	11
	2. 2. 2	分類キーワードの「標準モデル」作成 ····································	12
	2. 2. 3	分類キーワードの自動付与の実験と結果	15
1	2.3 調金	YMRの評価と今後の展望	17
	2. 3. 1	評 価	17
	2. 3. 2	展 望	17
3.	パソコ:	ンを用いた地図データベースの基礎構築 ····································	21
;	3.1 概	要	21
	3. 1. 1	目 的	21
	3. 1. 2	実施内容	21
;	3.2 実施	拖結果	22
	3. 2. 1	地図データベースの基本仕様	22

	3.2.2 地図データベースシステムプロトタイブの基本機能	25
	3. 2. 3 データ作成 ····································	20
	3.2.4 プロトタイプ作成上の問題点	22
	3.2.5 地図データ作成上の問題点	24
	3.3 今後の課題	25
	3.4 まとめ	2E
4.	1. 景観シミュレーション用樹木のデータベース構築	27
	4.1 目 的	27
	4. 2 実施内容	27
	4.2.1 プロトタイプのコンセプト	28
	4.2.2 データの可視化による視認性の向上	28
	4.2.3 代理機能による操作の簡略化	29
	4. 3 実施結果	29
	4.3.1 基本画面の構成	30
		· · · · · ·
5.	. 包装機械データベースの構築	33
;	5.1 開発の背景と目的	
	5.1.1 開発の背景	33
	5. 1. 2 開発の目的	
į	5.2 包装機械データベースの検討	33
	5.2.1 日本包装機械便覧の構成	33
	5. 2. 2 包装機械のデータ特性	34
	5.2.3 要求される検索の方法	34
	5. 2. 4 データベース作成の作業手順	35
į	5.3 包装機械データベースの基本設計	35
	5.3.1 システムの動作環境	35
	5. 3. 2 データ構造	
	5.3.3 開発したプログラム ····································	
	5.4 包装機械データベースの内容	
	5. 4. 1 メインメニュー	
		2.

	5.	. 5	5	結	論				<i>.</i>	•••••	• • • • • •	• • • • • •	• • • • •			• • • • •	• • • • • •	• • • • •			•••••	••	41
		5	. 5.	1	今後(の課題	題 …	•••••				••••										••	42
		5	. 5.	2	まと	: め		•••••		• • • • • •		···				•••••	• • • • •			•••••	•••••	••	42
6.		彩	f規	事			りため																43
	6.	. 1	-	概	要	• • • • •	•••••	•••••		• • • • • • •		••••						• • • • •		•• •••		••	48
		6	. 1.	1	Ħ	的						••••				• • • • • •				•••••	•••••	••	45
		6	. 1.	2	実施	既要	******	••••••	·• • • • •			••••				• • • • • •			•••••		•••••		43
	6.	2		実施	結果							••••		•••••	• • • • • • • • • • • • • • • • • • • •	• • • • • •	• • • • • •		•••••	• • • • • •	•• • • • •	• •	45
		6	. 2.	1	ニー	ズ調査	查結果	のまと	: W							• • • • • •			•••••		•••••		45
		6	. 2.	2	当該	データ	タベー	スにす	えめ	られる	る姿	•••		• • • • • •			*****				,	•	46
		6.	. 2.	3	事業化	'と検言	寸のま	とめ		· · · · · · · ·		• • • • • • •	• • • • •			• • • • • •	• • • • • •		•••••			•	48
	6.	3		今後	の展	開おる	よび予	想され	いるタ	効果	•••	•••••		• • • • • • • • • • • • • • • • • • • •		• • • • • • •				•••••			49
7.		恒	勃	率化	′先進村	才料;	ファク	トデー	- タ /	ベース	スの	パッ	ケー	- ジ化	í	••••;••			. 				51
	7.	1		概	要						••••	• • • • • •	••••	• • • • • • • • • • • • • • • • • • • •	• • • • • •		• • • • • •	• • • • •		•••••			51
		7.	. 1.	1	目	的		• • • • • • • •									• • • • • • • • • • • • • • • • • • • •						51
		7.	. 1.	2	実施網	圣緯							••••			· · · · · ·	• • ; • • •			•••••			51
	7.	2		実施	結果4	の概要	更				•••••								· · · · · ·				51
		7.	. 2.	1	シスラ	テム権	構成と	ファイ	, , , , , ,	の作原	戏 ·	• • • • • •					• • • • • •			•••••		-	52
		7.	. 2.	2	プロク	グラム	ムの作	成	.				• • • • •			· · · · · · ·	• • • • • •	• • • • •		•••••	,, 		54
		7.	2.	3	CD-	- R 0	D試作						• • • • •										57
		7.	. 2.	4	CD-	- R 7	での D	B 運用	評価) —						 .	• • • • • •						59
	7.	3			の課題																		59
		7.	. 3.				等のH																59
		7.	. 3.	2	CD-																		60
		7.	3.	3	インタ	タラク	フティ	ブWW	√W+	サーィ	ヾのぇ	運用	とR	DB	M S				· • • • • ·	·· · · · · · · · · · · · · · · · · · ·		. ,	60
8.		阪	神	· 淡	路大震	要災の)情報	デジタ	' ル・	・ア-	- カ・	イブ	•••			· • • • • • •	• • • • • • • • • • • • • • • • • • • •			· · · · · · · · · · · · · · · · · · ·		. (63
	8.	1		概	要								• • • • •				• • • • • • •					. (63
		8.	1.	1	Ħ	的	•••			•• •• •		•••••		• • • • • • •			*****					• (63
		8.	1.	2	実施内	内容				•••••			• • • • • •				• • • • • •					- (65

	8.	2		実施	結果																								••••			67
		8.	2.	1	プロ	٢	ター	イブ	'の1	作成	の	指	針			· • • ·		••••	• • • •	•••	••••			••••	••••	• • • •				• • • • •	-	67
		8.	2.	2	プロ	۲	夕~	イブ	'のf	作成		•••	· · · ·	•••	•••	· • • •			•••	•••				· · · ·				• • • • •				68
	8.	3		今後	の課	題	と届	展望	<u>.</u>	•••••		•••					• • • •	• • • • •	•••	•••	••••		, ,				••••	• • • • •			•	69
		8.	3.	1	デー	夕	ベ-	ース	作	成上	<u>の</u>	問	題,	Ħ,			•••	••••	•••	•••	••••	••••	• • • •	••••	••••		••••	 .		••••	•	69
		8.	3.	2	今後	の	デー	- タ	べ・	ース	<u>න</u> ්	利	用	• ‡	是(共	Z-	つしい	て		• • • •	••••	<i>.</i>					 .	· . · · · ·		•	71
9.		M	0	s a	i c	の	利月	用に	よ	るマ	ル	チ.	メラ	ディ	1.	ア:	デー	- 夕	べ	—	ス杉	食索	シ	ステ	F 1	の	構築	Ę.	· • • • • • • • • • • • • • • • • • • •	••••	•	73
	9.	1		概	要		••••				•••	••••							•••	•••	••••		• • • •	····	••••	• • • •	• • • • •		·····			73
		9.	1.	1	目		的	• • •	••••							•••			•••	•••	••••	• • • • •					· · · · ·			••••		73
		9.	1.	2	実施	内	容	•••	••••	•••••	•••	•••	· · · ·			···			•••	· · · ·	••••	• • • •	•••		••••	• • • •	••••	• • • • •		••••		73
	9.	2		実施	結果		••••					• • • •		•••			•••	• • • •	•••	•••	••••	• • • •	•••			· · · · ·		• • • • •		• • • • • •		76
		9.	2.	1	マル	チ	メラ	ディ	アラ	デー	夕·	べ・	— 7	ス札	食?	索:	シフ	ステ	ム	のオ	構翁	A.	•••				••••					76
		9.	2.	2	НТ	M	L <u>z</u>	上成	機能	造の	開:	発	•							•••				••••	••••	· · · ·		· · · · ·				78
		9.	2.	3	SQ	L	生反	戈機	能	の開	発				•••	•••							.:.	••••								82
		9.	2.	4	マル	チ	メラ	ディ	アト	青報	管.	理相	幾旬	色	ひ 月	用多	毪			•••			•••	•••	••••		· • • • •					83
		9.	2.	5	予想	3	れる	5効	果		•••					•••				· · · ·	••••	• • • •	•••	••••						• • • • • •		86
		9.	2.	6	マル・	チ	メラ	ディ	アラ	デー	夕·	ベ-	- >	ス柞	负氢	衣:	ンフ	くテ	ム	のF	問題	点		••••	••••					• • •		86
	9.	3		今後	の課	題		· · · · ·	••••		•••		•••		•••		••••	• • • •					•••				· • • • •		•••••			87
		9.	3.	1	機能の	の;	広引	Ę			•••			· · ·			••••			· • • •	••••		•••	• • • • •		· · · · ·	••••	<i>.</i>	,.		į	87
		9. :	3.	2	ΑP	I	の摂	提供		••••	 .	- • • •					···•			•••		• • • •	•••	••••	• • • •	• • • •	• • • • •	• • • • •		•••	i	87
		9. :	3.	3	マル	チ`	プラ	フツ	トフ	フォ	— .	ム 1	'Ł		•••		.							• • • • •							;	87
		9. :	3.	4	将来位	の	医皇	9	••••				•••						••••	· • • •				• • • • •	• • • •	• • • •	••••		•••••	•••••	;	88
l0	• .	大	規模	摸デ	ータイ	べ・	ース	くに	おり	ナる	構	造化	比情	青幸	長打	由占	出力	式	の	周	查	•••	•••	••••		• • • • •			•••••		;	89
	10.	. 1	١	問題	と目的	钓		• • • •	••••			. 	••••	•••		• • •			••••		••••			••••					•••••		ŧ	89
	10.	. 2	1	青報	検索:	シ)	ステ	<u>-</u> _	のÿ	見状	. •	. 				•••		• • • •		•••		• • • •	• • • •	••••	• • •						į	89
	10.	3			模化の																											90
	10.	4	,	ΚD	D≀= a	よん	る樟	造	化情	事報	発」	見											••••	••••							Ç	90
		10.	4.	1	GL:	Sē	発見	方	法訴	帝 ·	· • • •	· • • •				• • • •							•••			••••					Ç	91
		10.	4.	2	テキン	ス	トラ	<u> </u>	タ^	: —,	ス(の柞	黄追	51	比二	プロ	コセ	:ス				• • • •				••••					ç	92
	10	_		_	29 . 1	.,	L 1. 1	E1645	ഹദ	÷ 000																					,	20

10.6 マルチモダリティデータ処理	94
10.7 マルチモダリティデータの構造化	94
11. モーバイルデータベースシステムに関する調査研究	97
11.1 概 要	97
11.1.1 目 的	97
11.1.2 実施内容	97
11.2 実施成果の要旨	98
11.2.1 システムモデル	98
11.2.2 検索方式	98
11.2.3 トランザクション管理方式	99
11. 2. 4 データの配置方法	99
11.2.5 モーバイルステーションによるデータベースシステム	99
12. 情報収集ロボットによるInternetでのWWW所在検索データベースの構築	101
12.1 概 要	101
12.1.1 目 的	101
12. 1. 2 実施内容	102
12.1.2.1 ロボットのベースシステム	102
12.1.2.2 日本語情報に対応するための拡張	102
12.1.2.3 得られるデータのフォーマット	104
12.1.2.4 データベースの日本語化	106
12. 1. 2. 5 WWW検索用インターフェース	106
12.2 検索結果	107
12.2.1 ロボットによるデータの収集	107
12.3 今後の課題	111
12.3.1 Harvest の問題点	111
12.3.2 漢字コードの文字化けについて	111
12.3.3 動作環境の増強	111
12.3.4 今後の検索データベースに求められるポイント	111

13.1 目 的	
13.2 実験システムの概要	
13.2.1 設計方針	
13. 2, 2 システムの概要	
13.2.2.1 文書情報の形式	
13.2.2.2 ハードウェア・システムの構成	
13.2.2.3 ソフトウェア・システムの構成	
13.3 実験システムの詳細	
13.3.1 データ整備	
13.3.2 検索システム	
13.3.2.1 検索ロジック	
13.3.2.2 画面の構成	
13.4 成果と課題	
13.4.1 成 果	
13 / 9 選	

1. 法的データベースにおける多分野データベースの 統合一体的管理とオフ・オンラインの融合化に 関する調査、研究



1. 法的データベースにおける多分野データベースの統合一体的 管理とオフ・オンラインの融合化に関する調査、研究

1.1 多分野データベースの統合一体的運用

1.1.1 目 的

近時の社会の複雑化にともなう法的紛争の増加により、わが国もいよいよ訴訟社会への移行 の兆しを見せはじめるとともに、法的データベースの需要も飛躍的に増加している。

ところでわが国の場合、法的データベースは、従来、判例関係のデータベースの構築、利用 に限定されていた。しかし、判例だけ単体の利用では法的データベースとしては限界があり、 さまざまな場面で生起する法的紛争を適切に解決するためには、不十分といわざるをえない。

すなわち、真の法的データベースとして、その高度利用を図るためには、判例にとどまらず、 法令データや判例のコメント・解説等(以下「コメント」という。)も法的データベースの内 容とするとともに、データの構造・整理、キーワード等が一つのソフトの下に統合化され、一 体として利用できることが不可欠であると思われる。

そこで、今回、判例データベース、コメントデータベースに加え、未構築の法令データベースを構築することとした。

また、真の法的データベースとしてその高度利用を図るために、これらのデータベースにつき、それぞれの関係を検討するとともに、統合一体的運用について調査、研究を加えることとした。

1.1.2 法令データベースの構築

(1) はじめに

法的紛争の解決手段の中心である裁判は、具体的な事実に法を適用して行われる。

したがって、実務家にとり、適用すべき法を選択し、その内容を解釈したうえで事実に適用して結論を導き出すという作業は、その思考の中心をなすものといえる。

とすれば、法令に関し、その存在ないし内容を正しく理解することは、適切な結論を得る ための必須の前提になる。

しかし、法令を網羅した法令集は、全体 100冊を超える膨大なものであり、簡易・迅速に 法令を検索することができない。これに対し、法令のデータベースも提供されているが、こ れはオンラインのデータベースであり、料金体系が時間従量制のため、必ずしも、十分に活 用できるものとはいえない。 その中で、近時、これまで中央行政官庁で使用されていた、総務庁保有の法令データを収録したMTが、社団法人行政情報システム研究所を通じて民間に開放された。そこで、当社では、法令データベースの重要性を考慮し、法的データベースの一環をなすものとして、CD-ROMによる法令データベースの構築に着手することとした。

(2) 検索プログラムの開発

① 検索機能

【一覧表検索方式】

法令の一覧表から法令名を選択して表示させる方式。法令の一覧表の表示方法は複数用 意し、ユーザがその中から最も適当な方法を選択して一覧を表示させる。一覧表の表示方 法は以下のとおりである。

○法令名50音順一覧

法令名を50音順に並べて表示する。

- ○法令名JISコード順一覧 法令名をJISコード順に並べて表示する。
- ○法令番号順一覧法令が制定された際に付された番号の順に表示する。
- ○法令公布年月日順一覧 成立した法令が公布された日の順に表示する。
- ○法令施行年月日順一覧 法令が効力を生じた日の順に表示する。
- ○任意語検索一覧

ユーザが指定した任意語(フリーキーワード)で、法令名、条文全文を検索し、該当する法令の法令名を表示する。

任意語の検索では、全文検索システムを採用した。

また、同義語検索機能により、指定した任意語とともに、同義語でも検索できるようにした。

これらの条件に加え、「法令の種別」及び「法分野」を掛け合わせてさらに一覧表を絞 りこむことができるようにした。

【法令条文検索方式】

法令名または、法令名と条文を直接指定して検索する方式。

条文まで指定した場合、条文の表示は、その指定した部分が表示される。ただし、スクロールすることにより、当該法令の全部分を参照することができる。

② その他の機能

○参照条文表示機能

各条文の参照条文を表示する機能

主要45法令について、各条文に参照条文のデータを付加し、「条文表示画面」中の [参照条文] ボタンをクリックすることによって、参照条文を表示できるようにした。

○付せん機能

ユーザが頻繁に参照する法令や条文にマークを付け、いつでもその法令や条文を呼び 出せるようにする機能

また、マークした法令や条文にメモを書くことも可能にした。

- ○縦書表示機能
- ○アンドゥ(やり直し)機能 操作を間違えた場合, 1回前の操作に戻ることができる機能
- ○ロギング機能 検索経過を表示する機能
- ○印刷機能
- ○ダウンロード機能
- ○ヘルプ機能
- (3) 問題点
 - ① (社)行政情報システム研究所提供のデータは、憲法・法律・政令・勅令の4種類であり、 規則・省令・条例・条約等が含まれていないこと
 - ② 図表や数式などを法令の原本どおりに再現できなかったこと
- (4) 今後の課題
 - ① 法令の改正経過の追従機能の設置
 - ② 判例データベース、コメントデータベースとの相互参照機能の付加
 - ③ 最新データのオンラインによる提供

1.1.3 判例, 法令, コメントデータベースの統合一体的運用

- (1) 判例データベースと法令データベース
 - ① 判例データベースを利用しながら、条文を参照する場合
 - ○法令条文入力画面

判例データベースを「法令・条文」で検索する場合、検索キーとして判例を特定できるのは、法令名より条文である。したがって、「法令・条文」を入力する際に、法令データ

ベースで条文の内容を参照できれば、効率的にキーワードを設定できる。

○検索結果詳細表示画面

検索結果詳細表示画面は、検索した判例の内容そのものを参照する画面であり、条文の 内容を参照する必要性が最も高い。

しかし、参照条文のデータのない判例をどうするのか、判例中の参照条文以外の法令も 表示させるのかという問題がある。

② 法令データベースを利用しながら判例を参照する場合

法令はその性質上規定のしかたが抽象的であり、その内容を理解するためには、判例の 理解が必須になる。したがって、判例の参照条文のデータを逆にたどって、条文内容表示 画面から判例を参照できるようになっていることが必要である。

しかし、判例の参照条文の中には法令の解釈に役立たないものも多く、それらをいかに 選別するかという問題がある。

(2) 判例データベースとコメントデータベース

コメントは特定の判例の解説としてつけられるものである。したがって、判例を検索しながらそのコメントを参照し、またその逆にコメントを検索しながら判例を参照するという使い方ができることが必要である。このように、判例とコメントは対応関係がはっきりしているので、随時相互に往来できるようにしておくことが必要である。

またコメント中に引用されている、過去のコメントおよびそのコメントに対応している判例をたどっていける機能も有効であるが、何階層までたどれるようにするかを慎重に検討する必要がある。

(3) コメントデータベースと法令データベース

コメントデータベースと法令データベースの関係については, コメントが判例を解説した ものであることから, 判例データベースと法令データベースに関する議論が概ねあてはまる。

1.2 オフラインとオンラインの融合化

1.2.1 目的

オフラインデータベースには、提供時期からくるタイムラグの問題が宿命的に発生する。しかし一方でオフラインデータベースは、いつでも、時間・料金を気にすることなく利用できるというメリットがある。そこで、オフラインとオンラインの融合化をはかり、双方の短所を補ったうえで、その長所を活かせないかを検討することとした。

1.2.2 融合化の方法

(1) JLIC提供方式(その1)

当社(以下「JLIC」という。)にサーバを置き、前回のCD-ROM発売以後のデータをオンラインで提供する。ユーザは、以前のデータをCD-ROMで検索し、最新のデータはオンラインにより検索するという方法。

[メリット]

- ① 検索ソフトの方で必ずオンラインを検索にいくので、ユーザは従来と同様の操作で常に最新の判例まで入手できることになり、オフラインとオンラインの融合化という目的に最も近いものになる。
- ② データの形式がいずれもJLIC形式なので、データのマージ等が行える。 [デメリット]
- ① サーバ管理プログラムの開発等、初期投資に要する費用が大きい。
- ② サーバを常に稼動させなければならず、そのための保守・管理が必要になる。
- ③ アクセスポイントがないので、地方のユーザにとり通話料の負担が大きくなる。
- (2) 商用ネットワーク方式(その1)

民間のネットワークを利用して、最新判例のデータを送信する方式。ユーザは、旧判例についてはCD-ROMで参照し、新判例はネットワークを利用して参照する。

[メリット]

- ① 初期投資の額が前記JLIC方式(その1)に比較し少ない
- ② アクセスポイントがあるので、通話料が低額ですむ
- ③ JLIC側でシステムのメインテナンスを考える必要がない

[デメリット]

- ① 検索の方式が、オフラインとオンラインで全く異なるので、1回の操作で両者を検索 することができず、「融合化」のイメージからはずれることになる。
- ② オンラインで検索した結果は、ひとつずつ独立した単なるテキストデータのため、そのままではCD-ROMから検索した結果とマージすることができない。マージしようとすれば検索したデータにあらたに検索キーを付さなければならないが、そのようなシステムを構築するにはかなりの時間と費用が必要になる。
- ③ 検索の際のキーワードの数および文字数が限定されているため、適切に分類すること ができない判例が生じる可能性が高い。
- ④ 初期投資は低いが、ランニングコストがかかるため、必要とされる費用が数年で上記 の独自方式とかわらなくなる可能性がある。

(3) JLIC提供方式(その2)

最新判例のデータをJLICからユーザのハードディスクに転送し、ユーザはハードディスクとCD-ROMで検索をする方式。

検索はユーザ側のパソコンだけで行い、検索の都度サーバにアクセスするものではない。 転送データの保存形式により、次の(A)、(B)の2種に分かれる。

- (A) 転送データだけでなく、既存のCD-ROMのデータもハードディスクに保存する方式 サーバからデータが転送された段階で、CD-ROMのデータと合わせて、ハードディ スク内に最新の検索データを作成し、検索は、全てこのハードディスク内の最新検索デー タで行う。
- (B) 転送データだけをハードディスクに保存する方式

単一の最新検索データの作成は行わない。

検索は、旧データについてはCD-ROM、最新データについてはハードディスクで行い、両者の検索結果をマージして表示する。

[メリット]

- ① 新判例のデータをサーバから一方的に送信するだけのシステムなので、JLIC方式 (その1)より簡単であり、構築費用がそれ程かからない。
- ② データの形式が同じなので、新・旧判例を自由に表示、編集できる。
- ③ 更新データの提供は、無料でおこなうことを予定しているところ、JLICからユーザにアクセスして送信する方式にすれば、通信料を含めユーザに費用の負担をかけることがない。

「デメリット】

① 新判例のデータは、JLICから送信したものにもとづくので、常に最新判例とはい えない。

送信サイクルの間は、前回送信分のデータが最新判例のデータになる。

次に、上記転送データ保存形式の(A)および(B)について、メリット、デメリットを 検討する。

「A方式のメリット]

- ① 新データ、旧データともハードディスクから検索するので、検索速度が速い。
- ② データ転送時に最新検索データを作成し、検索はそのデータを対象にするので表示時のデータのマージが不要である。

「A 方式のデメリット]

① CD-ROMのデータもハードディスクに取り込むので、ハードディスクに最大 640

メガバイト程度の空き容量が必要になる。

[B方式のメリット]

- ① A方式より遅いものの、検索の際にサーバと交信をしないので、検索が速い。
- ② ハードディスクに保存するのは最新判例分のみのため、ハードディスクに最大50メガバイト程度の空き容量があれば足りる。

[B方式のデメリット]

- ① 検索をCD-ROMとハードディスクの双方で行うので、やや遅くなる。
- ② CD-ROMから検索したデータと、ハードディスクから検索したデータをマージする必要があるので、その間の処理時間がかかる。
- (4) 商用ネットワーク方式(その2)

リーガルベースの形式で構築した最新判例のデータを圧縮し、その圧縮したデータを商用 のネットワークを利用してユーザに送信する方式。

ユーザは、そのデータをパソコンに取り込んだうえで解凍し、CD-ROMのデータと合わせて検索する。

[メリット]

- ① 商用ネットワークを利用するので、システム構築のための初期費用が少ない。
- ② 送信されるデータの形式はリーガルベースのものなので、(その1)と異なり検索結果をマージすることができる。

「デメリット〕

① 本方式と対応する「JLIC方式(その2)」の構築費用が、それ程高額にもならないと思われることを考えれば、ランニングコストの方が、すぐに上回ることになりかねない。

1.2.3 各方式の検討

以上検討した各方式は、大きくわけて「最新判例の検索をオンラインで行う方式」と「オンラインは新判例のデータを送信するだけで、検索はオフラインで行う方式」の2つに分けられる。

このうち、前者は、検索時に必ずオンラインにアクセスするため、常に最新判例が入手できる。したがって、タイムラグの解消を目的とするオフラインとオンラインの融合化という課題に最も適合するものといえる。しかし、それぞれ検討したとおり、システムの構築費用や、ランニングコストなどの負担が大きい。

これに対し、後者の方式では、常に最新判例を検索するものではない点で、前者に比較して

「融合化」の目的から一歩後退するものといえる。しかしこの点は、JLICからの送信頻度を週1回ないし2回とすれば、最大でも7日サイクルでデータが更新されることになり、この程度のタイムラグであれば、それ程の障害にはならないと考えられ、タイムラグの解消という目的も十分に達成できると考えられる。しかも後者の方式には、構築費用の負担が前者に比べて少ないという大きなメリットがある。ただし、後者の方式のうち、商用ネットワークを利用するものは、ランニングコストがかかることを考えれば、負担は必ずしも少ないとはいえない。

結局、タイムラグの解消という目的の達成度合いおよびそのための構築費用という「費用対効果」の点から判断して、JLIC方式(その2)を中心にさらに検討を加えることが適当であると判断した。

1.3 おわりに

以上、法的データベースをより機能的、効率的に利用できるようにし、真の法的データベースと呼べるものにするために、「多分野データベースの統合一体的管理」と「オフ・オンラインの融合化」という2点について調査、研究をおこなった。

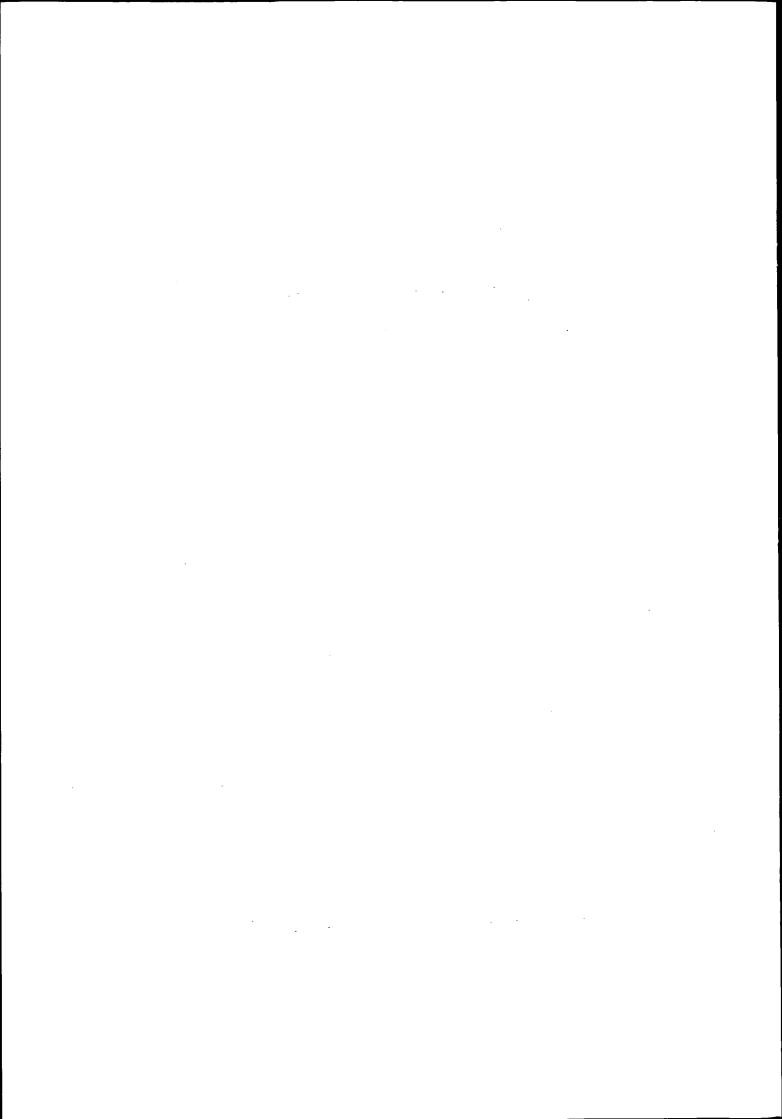
まず「多分野データベースの統合一体的管理」については、未構築の法令データベースの開発および多分野データベースの統合一体的運用について調査、研究を行ったが、前者の法令データベースの開発作業に予想以上の時間がかかり、後者の研究が必ずしも十分であったとはいえない面がある。

・今後も、判例データベースである「リーガルベース」、コメントデータベースに今回完成した法令データベースを加え、実際にリンクづけの実験を行いながら引き続き調査、研究を続ける予定である。

また「オフ・オンラインの融合化」については、オフラインデータベースの宿命であるタイムラグの解消という目的のために研究を行い、各種方式の考案、比較を行って、一応の成果を見たといえる。

融合化を実現し、オフラインとオンラインの長所を活かしたデータベースシステムを構築することは、データベース全体の発展に大きく貢献するものと思われるので、今後は、本調査、研究の結果に、近時爆発的な拡がりを見せているインターネットを利用するシステムの構築を加え、さらに調査、研究を重ねていく予定である。

2. 新聞記事分類キーワードの標準モデル構築と自動付与に関する調査研究



2. 新聞記事分類キーワードの標準モデル構築と 自動付与に関する調査研究

2.1 概要

2.1.1 調査研究の背景と目的

新聞記事データベースに使われている「分類キーワード」(記事の主題を集約的に表わす分類語)は、不人気な検索キーである。データベースを構築する側からいえば、①人手でつけるため、どうしてもばらつきが出る②データベース編集の効率化を阻害する③人員負担が過大になる、などの理由があげられる。ユーザ側からいえば、①記事にどんな分類キーワードがついているか分かりにくい②新聞社ごとに分類キーワード体系がバラバラである③分類表を開いて探すより、通常のキーワード検索で十分だ、などの不満が出ている。

しかし、分類キーワードは通常のキーワードでは検索できない記事を引き出せる強みがあり、 また、キーワード検索と分類キーワードを組み合わせることによって、検索対象の記事を効率 的に絞り込める、などの効用もある。代行検索の「プロ」や検索の「ベテラン」には分類キー ワードを高く評価する人も多い。

われわれの目的は、分類キーワード作成のコストパフォーマンスを上げ、ユーザにもより利用しやすい分類キーワードを作ることにある。このためには、まず分類キーワード体系を、構築側にも検索側にも便利なように簡便なもの(100~200語程度)に作り変え、かつ、ばらつきの少ない分類キーワードの標準化(標準モデルの構築)であり、後者は分類キーワード自動付与の可能性を探る問題である。

分類キーワード自動付与の試みとして、まず12個のサンプル分類キーワードを選んで論理式を作り、実際の新聞記事にあててみて、有効な分類キーワードを自動的に抽出してくるかどうか、実験してみた。株式会社エレクトロニック・ライブラリー(略称 E L = イーエル)は、1994年度(平成6年度)に、(助データベース振興センターの委託事業として「創出キーワードの自動付与に関する調査研究」を行い、テキスト文中に現れない重要キーワード(創出キーワード)を、自動的に付与する方法論を探った。この手法を発展的に援用した。

2.1.2 調査研究の内容と結果の概略

われわれの調査研究はまず、①新聞記事データベースを提供する側と実際に利用しているユーザ側へのアンケート調査およびヒアリング調査②一般紙6社とEL合計7社の分類表を検討・統合化して標準モデルを作成した③標準モデルの中から実用性の高い「分類キーワード」を

選定し、その自動付与の実験を行なうという手順で進められた。

① アンケート調査およびヒアリング調査

新聞記事データベースを提供する側と利用する側で、それぞれが分類に対してどのような問題点を抱え、どのような意見を持っているのか、その実態を把握し標準モデル構築の足がかりとするために以下二つの調査を行った。

<新聞社に対する分類アンケート調査とヒアリング調査>

このアンケート調査では新聞記事データベースを持っている新聞社34社を対象に行われ、 68%の回収率を得た。また一方ではヒアリング調査とし、実際に新聞社を3社訪問し、より 詳しい現場の意見を聴き取り、研究の参考にした。

< ELNETユーザに対するアンケート調査>

EL社の提供するELNETのユーザで、利用頻度が多い 309 I Dを対象にアンケート調査を行った。回答率は62%だった。その半数は分類をまったく使用していないユーザだったが、キーワードだけで十分検索できることを主な理由としている。が、なかには「使用方法がよくわからない」「分類による検索ツールがあることを知らなかった」というユーザもいた。

② 分類キーワードの「標準モデル」作成

分類キーワードの「標準モデル」作成にあたっては、一般紙 6 紙 (朝日新聞、読売新聞、産経新聞、中日新聞、北海道新聞、河北新報)に E L を加え、合計 7 社の分類表を対象とした (中日新聞は「'90 ニュースシソーラス」による。96年現在は分類を使っていない)。

分類キーワード候補の選定には、7社の分類表をマージして大分類レベルから順次検討し、 統合や移動によって整理する、という方法をとった。また、試案の段階で専門家からの評価 をヒアリングという形で受けた。その結果、7社の分類項目をすべて包括できる 129項目の 分類キーワードに絞り込み、簡便で実用的な「標準モデル」とした。

③ 分類キーワードの自動付与実験

分類キーワードの標準モデルの中から主要の12項目を選び、実際の新聞記事を解析し論理 式を作成し、検証してみた。

新聞記事は1項目ごとに 200件, 計 2,400件選んだ。奇数番号の記事は論理式を作るための解析用に、偶数番号の記事は実際に論理式を当てる検証用に使用した。検証用として用意した 1,124記事に対しては、自動付与実験とは別にELデータベースのインデクサーを中心とする実務メンバーら5人がそれぞれ全記事の分類索引を行い、その過半数をもって "正しい分類付与"とした。自動付与による結果の適否は、人手による評価と一致したものを "適合付与"と判定した。ここから「再現率」(ヒットすべき検索対象の総数に対して自動付与

でヒットした割合)ならびに「適合率」(分類キーワードの自動付与記事総数に占める適合 記事の割合)を算出し、再現率80%、適合率70%の同時達成を一応の合格目標として自動付 与実験の精度を検証した。

その結果おおむね以下のような結果が得られた。

- (a) 再現率が80%を超えたものは1項目であった。12項目の平均は63%だった。
- (b) 適合率が70%を超えたものは6項目で、全体の平均では72%だった。
- (c) 再現率と適合率の両方とも目標を超えたものは1項目だった。
- (d) 自動付与での結果が目標を超えなかった11項目を再度検討し、論理式の修正を行ったところ、再現率80%以上が7項目、適合率70%以上も7項目となり、再現率と適合率の両方が目標を超えたものは4項目に増えた。全体の平均では再現率82%、適合率72%と目標をクリアした。
- (e) 第1回検証で自動付与の再現率があまり良くなかった原因としては、データの分析不足や論理式記述のミスや論理式作成者と分類キーワード判定者との意識のずれなどがあげられる。しかし、再度の実験による結果の向上から判断して、比較的簡便な変更によって精度を高めることが可能であることが確認されたといえる。

2.2 実施結果

2.2.1 アンケート調査

新聞社とユーザに対するアンケート調査は以下の要領で行った。まず新聞社の対象は新聞記事データベースを持っている以下の新聞社34社となった。

朝日新聞社 愛媛新聞社 河北新報社 北日本新聞社 岐阜新聞社 京都新聞社 熊本日日新聞社 高知新聞社 神戸新聞社 佐賀新聞社 産業経済新聞社 静岡新聞社 上毛新聞社 中国新聞社 中日新聞社 西日本新聞社 日本経済新聞社 北海道新聞社 北國新聞社 毎日新聞社 南日本新聞社 読売新聞社(一般紙五十音順)

化学工業新聞社 交通新聞社 住宅新報社 全国新聞情報農業協連 鉄鋼新聞社 電気新聞社 日刊建設工業新聞社 日刊工業新聞社 日刊スポーツ新聞社 日刊木材新聞社 日本食糧新聞社 薬業時報社(専門紙五十音順)

(このうち京都新聞社、中日新聞社、北海道新聞社の3社はヒアリング調査も行った。)

質問内容はできるだけ簡便にし、分類についての基本的な点のみにした。また、回答は複数

回答を可とした。

<調査結果の分析>

回答は34社中23社で、全体の68%だった。そのうち、70%が新聞記事データベース検索用の 分類を持っており、30%はなかった。

分類を作った理由として、多かった回答は以下の通り。

- ① 「特定記事をまとめて検索するのに有効なため」 88%
- ② 「検索の絞り込みに有効な手段のため」 50%
- ③ 「キーワードだけでは検索しにくい記事もあるため」 38%
- 一方、分類を持っていない理由として多かったのは、
- ① 「人的手間がかかりすぎる」 57%
- ② 「検索上のメリットがない」 43%

だった。

また、ELNETユーザに対して行ったアンケート調査は、次のような結果が出た。対象ユーザは、日頃分類を使用しているユーザと、逆に使用していないユーザの意見を比較するために、以下3種のユーザに限定し、合計 309 I Dに絞り込んだ。

- ① MOR(早朝クリッピングサービス)の登録検索式に分類を使用しているユーザ
- ② オンラインDB検索で分類を利用したユーザ
- ③ 検索の利用回数が多いユーザ (最低月10回以上)

回答率は61.5%だった。

記事分類を使用したことがあるかという質問に対しては

① 「全くない」 51%

② 「過去に 1, 2回使用したことがある」 22%

③ 「ときどき使用している」 21%

④ よく使用している6%

という結果になった。

分類を全く使ったことがないユーザにその理由を聞いてみたところ,

① 「キーワードでの検索だけで十分」 51%・

② 「分類による検索機能があることを知らなかった」 31% などが目立った。

2.2.2 分類キーワードの「標準モデル」作成

分類キーワードの「標準モデル」作成にあたっては、データベースで分類項目を持っている

新聞各社の分類表を素材として利用した。その際、専門紙のデータベース分類はどちらかといえば狭い分野に限られているために除き、一般紙6紙(朝日新聞、読売新聞、産経新聞、中日新聞、北海道新聞、河北新報)にELを加え、合計7社の分類表を対象とした(中日新聞は「'90ニュースシソーラス」による)。

分類キーワード候補の選定には、7社の分類表をマージして大分類レベルから順次検討し、 統合や移動によって整理する、という方法をとった。第1次案の段階で専門家の評価を受け、 いくつかのアドバイスを取り入れながら、さらに修正した。

最終的には、7社の分類項目すべてを包括する 129項目の分類キーワードに絞り込んで標準 モデルとした。(表 2-1)

(1) 大分類「総類」の排除

新聞記事データベースの多くは、日本新聞協会の『新聞切抜・写真分類表』を参考にしている。この分類表は十進分類法を採用しているため、データベースの分類にも「総類」という大分類項目が残っている。総類はいわば「一般の一般」として、他のどの分類にも入らないものをまとめてある。他の具体的な内容の大分類に帰属しない項目を、この「総類」に入れるため、「何でも屋」の様相を帯びてしまう。表としての整合性は取れるが、実用性に欠ける。従来「総類」に入っている項目は、できるだけ他の具体的な大分類に移動させ、「総類」自体をなくした。

(2) 「国際」の取り扱い

「国際」の分類はELを除く6社が使っていた。各社の「国際」の使い方はまちまちで、 主題であったり、国名・地域名であったりしている。国内と国際での分類の使い分けが難し い、との議論もあったが、ここでは紙面によく登場する項目に絞り込んだ。

(3) 「産業」の細分化

データベースのユーザは、現状はビジネス面の検索が多いことを配慮して、「経済」から 「産業」を分離させ、実用に耐えるようできるだけ細かく分割した。

(4) 記事種別分類の一部使用

記事主題分類としての分類キーワードというのが原則だが、実用性、簡便性を考えて、例外的に記事種別(形式)としての「訃報」(死亡記事)と「人事」の項目を作り、あえて表の整合性を損ねるのを覚悟しつつ、「社会」の中に入れた。

表 2 - 1 分類キーワードの標準モデル

1			
【政治】		【社 会】	
政治一般	司法・裁判	社会一般	市民運動
国会	財政	生活	社会保障•福祉
選挙	外交	家族	余暇
政党	安全保障	世代	行事・イベント
行政	軍事	性差	計報
地方自治	皇室	世相・風俗	人事
		環境保護	
【経済】			
経済一般	知的所有権	【事件・事故】	
経済・金融政策	産業・経済団体	事件・事故一般	交通事故
経済協力	企業経営	凶悪犯	航空機事故
景気・物価	貿易	経済犯	火災
公正取引	H201	公安事件	海•水難事故
五儿秋汀		→ 人質犯	山岳事故
「产 业 】		コンピューター犯罪	産業災害
【産業】	自動車工業	外国人犯罪	自然災害
金融業	日 男 早 工 来 電 子 ・ 電 機	青少年犯罪	コンピューター事故
農林漁業		最少年紀年 暴力団犯罪	原子力事故
土木建設	諸工業	麻薬・覚せい剤	凉 1 刀 尹 以
食品工業	情報・通信	麻祭・見せい削	
繊維工業	新聞・通信社	TIN MAT	
紙パルプ	放送 tulk rask	【科学】	地球
資源・エネルギー	出版・印刷	科学一般	
鉄鋼	広告	医学・医療	気象・
非鉄金属	不動産	生物	原子力
鉱業	運輸交通	バイオテクノロジー	理工学
化学工業	卸・小売業	宇宙	
医薬品製造業	飲食業		
機械工業	レジャー産業	【文 化】	
精密機械	各種サービス業	文化一般	文学
輸送用機器		美術	文化財
		→ 音楽	文化施設
【労 働】		映画	宗教
労働一般	労働組合	演劇	人文・社会科学
雇用	職業	舞踊	学校教育
労働条件		演芸	社会教育
		- 伝統芸能	言論報道
【国際】			
国際一般	戦争・紛争	【スポーツ】	
国際政治機構	核兵器	スポーツ一般	水上競技
国際経済機構	軍事同盟	総合競技	冬季競技
南北問題	軍縮	球技	ギャンブル競技
国連		陸上競技	武道•格闘技

2.2.3 分類キーワード自動付与の実験と結果

分析に使用した新聞記事データは、(㈱ジー・サーチ社を通じ磁気テープで入手した。対象の新聞は朝日、読売、日刊工業新聞の3紙とした。サンプル分類キーワード12項目の1項目に200記事ずつ抽出した。

2,400件の新聞記事データは、NTTデータ通信(株)に依頼して、日本語解析ソフト「IND EXER」にかけて、単位単語に切り出したうえ、キーワード候補ごとに出現位置、段落、名詞分類の情報を付加した。このあと、さらにEL社内でELキーワード辞書と照合して合致したキーワードを抽出したり、単位単語を結合して複合語を作成、ELキーワード種別も追加した。

記事の収集は各分類キーワード別に行ったので、 2,400記事全体の中には同一記事がダブっていたり、「INDEXER」の制約で1記事の文字数が長大なものは処理できないこともあって一部を処理対象から除いた。このため対象記事は最終的に 2,264記事に落ち着いた。

2,264記事は、収集時の通番により奇数記事を解析用ファイル、偶数記事を検証用ファイルと二つに分けた。解析用は論理式をたてるために記事を解析し、式を実行して自己検証を繰り返し、本番に備えるための練習用ファイルである。検証は本番テスト用のものである。解析用ファイルは 1,124記事となった。

論理式の具体例は次のようなものである。

CD-ROM; 段落0-3

?ソフト/CTGK+ソフトウエア+OS:段落0-3

出版+販売+発売+開発+価格+画像:段落0-3

(S1+S2)*S3

ADD 情報通信

第1行:第3段落までに「CD-ROM」

第2行:第3段落までに「ソフトウエア」「OS」または語尾に「ソフト」をもつ統制語の一般キーワード

第3行:第3段落までに「出版」「販売」などの指定キーワード

第4行:第1行あるいは第2行のいずれかと第3行の論理積

第5行:第4行の条件を満たす記事に分類キーワード「情報・通信」を付与

自動付与の論理式は,解析用ファイルで何度もプリテストを繰り返し,精度を高めた。こう

して作られた論理式を, 1,124記事の検証用ファイルにあて,分類キーワードを自動付与した。 自動付与された分類キーワードが,ヒットした記事にとって本当に正しい分類付与であるかは 人による評価と比較する方法で判定した。

ELデータベース作業で実際に分類を付与しているインデクサーら 5 人それぞれが全記事に分類索引を行った。人による分類づけは多少のばらつきは生じるので、3 人以上が同じ分類キーワードを付与した場合、その記事は該当分類にあたる"正しい付与"と判定した。3 人以上の評価が一致した分類キーワードは713件あったので、自動付与の621件と図のように上下に重ね合わせた結果、"正しい自動付与"は448件となった。また機械が取り損ねた「漏れ」は265件、余分に取った「ノイズ」は173件となった。(図2-1)

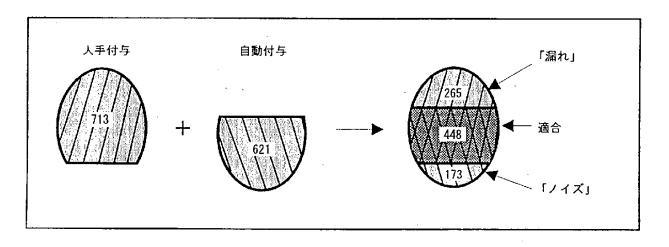


図2-1 第1回検証実験の結果

これにより、448件が全正解件数 713件中に占める割合を示す「再現率」は63%、また全自動付与件数 621件中に占める割合を示す「適合率」は72%となる。われわれが目標にしていた再現率80%以上、適合率70%以上に対して、適合率はクリアしたものの再現率は手が届かなかった。

分類キーワード別に実験結果をみると、再現率、適合率ともに目標を突破したのは「司法・裁判」の1項目で、適合率のみ合格点に達したのは「情報・通信」「バイオテクノロジー」「気象」「公正取引」「社会保障・福祉」だった。

再現率が低かった原因究明のため、目標に達しなかった11項目について修正した論理式を使って、狙い通りの結果がえられるかどうかを実験した。その結果、「表 2 - 2」で示したようにかなりの改善がみられ、再現率は82%、適合率72%で目標をクリアした。

再現率80%以上が7項目,適合率70%以上が7項目,再現率と適合率の両方が目標を超えた ものは4項目となった。再現率は「経済協力」,「バイオテクノロジー」は変わらなかったが, そのほかの 9 項目で向上した。適合率は「安全保障」,「情報・通信」,「気象」で下がったが、そのほかの 8 項目は変わらないか、向上した。

第2回目の検証では、再現率を高めることを主眼として論理式を修正したもので、「漏れ」は第1回目の 265件から 115件に大幅に減った。(図2-2)

2.3 調査研究の評価と今後の展望

2.3.1 評価

分類キーワードの「標準モデル」については、(㈱朝日新聞ニューメディア・サービスに依頼、1996年2月15日付朝刊を対象に、「標準モデル」の使い勝手を評価していただいた。実際に対象としたのは、1面、2面、第1外電面、第3経済面、第1家庭面、第1社会面の計54記事。記事データベースの編集や代行検索に従事している社員7人が、標準モデルの分類キーワード付与にあたった。産業分類が細分化され便利になった半面、慣れないと迷う面もあるなどの指摘も受けたが、全般にはぶれが少なく付与できる簡便なものであるという評価を受けた。

分類キーワード自動付与の実験については、第1回の検証ではやや再現率が悪かったが、論理式の簡単な修正で精度を高められることが実証できた。主題に絡む重要キーワードの選定など辞書面でのメンテナンスと論理式の精度を高める作業を続ければ、実用化への展望は開ける評価した。

2.3.2 展望

分類キーワードの「標準モデル」を作成するというのは、ある意味では大変無謀な試みである。なにをもって「標準」とするのか。新聞記事データベースを作成する新聞社にはそれぞれ事情があり、独自の分類表を持っている。極端な言い方をすれば、分類は作成者の目的・対象次第ではどのようにも作り得る。なにが「正しく」「標準である」とは、一概には言い切れないのである。あえて「標準化」を提起したのは、従来の記事分類法が急速な時代の変化についていけず、古くなりつつあるのではないかという危惧と、これからますます盛んになるであろうオンライン検索に向けて、分類キーワードでも横断検索できる道を探っておきたいという願いからである。

分類キーワード自動付与の実験は、データベースを製作する新聞社側に対する要員負担を軽減し、付与される分類キーワードの均質化を狙ったものである。

特に今後の課題として、検索の全文検索方式(フルテキスト・サーチ)の問題がある。われ われのこれまでの議論は、キーワード検索を前提としている。全文検索方式は、テキストから 切り出されたキーワードを必要としない。テキスト中の文字列との完全一致で検索する新しい 方式である。キーワードの読みのチェックや保守は必要としない。データベースを製作する側 には大変経済的メリットがある。

しかし、問題がないわけではない。文字列の完全一致だけが検索の決め手になっているため、「東大」は「東大阪」、「東大寺」の中の「東大」と一致し、「東大阪」や「東大寺」まで検索してしまう。この雑音封じにいまのところ有効な手段が見つかっていない。検索技術のレベルでいうと、キーワードのAND検索で絞り込むことや期間を限定して雑音が出るのを少なくする、などが考えられる。分類も対象の絞り込みには大変有効な手段だ。全文検索との分類の組み合わせは、魅力ある検討課題といえよう。

表 2 - 2 論理式修正前・修正後比較結果

(修正前)

項番	分類KW	M (人手のみ)	B(阿 方)	A(自動のみ)	人手計	自動計	再現率	適合率
1	安全保障	14	14	9	28	23	50.0	60. 9
2	司法・裁判	2'	71	13	73	84	97.3	84. 5
3	経 済 協 力	22	59	27	81	86	72. 8	68. 6
4	公正取引	19	42	13	61	55	68. 9	76.4
5	情報·通信	39	63	10	102	73	61.8	86. 3
6	レジャー産業	17	11	13	28	24	39. 3	45, 8
7	気 象	16	7	2	23	9	30. 4	77.8
8	バイオテクノロジー	13	46	9	59	55	78. 0	83, 6
9	環境保護	36	37	27	73	64	50.7	57.8
10	社会保障・福祉	49	27	11	76	38	35. 5	71.1
11	学 校 教 育	30	40	19	70	59	57. 1	67.8
12	文 化 財	8	31	20	39	51	79.5	60.8
	合 計	265	448	173	713	621		
	平 均	22, 1	37. 3	14. 4	59. 4	51, 8	62. 8	72. 1

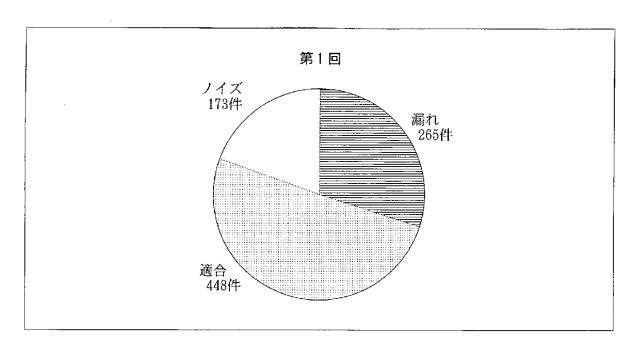
〔修正後〕

項番	分 類	K W	M(人手のみ)	B(両 方)	A(自動のみ)	人手計	自動計	再現率	適合率
1	安全	保『	章 5	23	17	28	40	82. 1	57. 5
2	経済	協	22	59	8	81	67	72. 8	88, 1
3	公 正	取	13	48	12	61	60	78. 7	80. 0
4	情報	• 通 (15	87	40	102	127	85. 3	68. 5
5	レジャ	一産	€ 5	23	6	28	29	82, 1	79, 3
6	気	4	₹ 13	10	6	23	16	43. 5	62. 5
7	バイオテ	7/09	- 13	46	9	59	55	78. 0	83. 6
8	環境	保言	隻 13	60	38	73	98	82. 2	61, 2
9	社会保障	章・福祉	L 2	74	28	76	102	97.4	72.5
10	学 校	教育	<u>†</u> 8	62	22	70	84	88. 6	73.8
11	文(Ł ,	4 6	33	14	39	47	84. 6	70, 2
	合	計	115	525	200	640	725		
	平	均	10. 5	47. 7	18. 2	58. 2	65. 9	82.0	72. 4

[修正後再現率順]

							-		
項番	分類	K W	M (ኢ ៛ወង)	B(両 方)	A(自動のみ)	人手計	自動計	再現率	適合率
1	社会保障	賃・福祉	2	74	28	76	102	97.4	72, 5
2	学校	教育	8	62	22	70	84	88. 6	73.8
3	情報•	通信	15	87	40	102	127	85, 3	68. 5
4	文 化	2. 財	6:	33	14	3 9	47	84.6	70.5
5	環境	保 護	13	60	38	73	98	82. 2	61. 2
. 6	レジャ	一産業	5	23	- 6	28	29.	82. 1	79, 3
7	安 全	保障	5	23	17	28	40	82. 1	57. 5
8	公 正	取引	13	48	12	61	60	78. 7	80, 0
9	バイオテク	ノロジ ー	13	46	9	59	55	78. 0	83. 6
10	経 済	協力	22	59	8	81	67	72, 8	88, 1
11	気	象	13	10	6	23	16	43, 5	62, 5
	合	計	115	525	200	640	725		
	平	均	10. 5	47.7	18. 2	58. 2	65, 9	82, 0	72. 4

	漏れ	適合	ノイズ	再現率	適合率
第1回	265	448	173	63%	72%
第2回	115	525	200	82%	72%







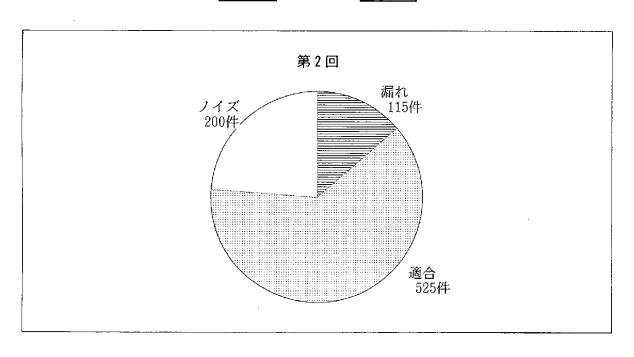


図2-2 自動付与実験結果の比較

3. パソコンを用いた地図データベースの基礎構築

(有)朝日データサービス

	·			
·				

3. パソコンを用いた地図データベースの基礎構築

3.1 概要

3.1.1 目 的

地図を用いて分析すればより有効に利用することが可能と考えられる災害時の危険区域, 住宅分布, 道路網, 上下水道網, 高齢者分布, 土地利用状況などのデータベースは大都市圏については整備されつつある。その他の地域に関していえば普及が遅れている。

この理由としては、地図を使うシステムがワークステーションと呼ばれるコンピュータ、またはそれ以上の高機能コンピュータ上で構築されていて高価なシステムであったことと、地図データベースシステムに必要な電子地図データが大都市圏は整備されてきているものの、それ以外の地方の地図データは主要都市以外はほとんど整備されていないことが原因と考えられる。

そこで今委託課題では、地図データベースが比較的安価に構築できるパーソナルコンピュータを用いて、地図データベースの基礎となる地図データを整備し、地域における地図データベース普及をはかり、地域におけるデータベース拡充の一助となることを狙いとした。

3.1.2 実施内容

本課題では、以下の内容を実施した。

(1) 地図データベースの技術研究

まず、地図データベースのシステム、及び地図データの基本を専門家にヒアリングをうけたり、資料を集め研究した。

- (2) 地図データベースに関する調査 地図データベースのシステム及び地図データの現状がどうなっているのか、専門家にヒア リングを受け、把握した。
- (3) 基本設計 本課題で作成する地図データベースの基本的仕様の設計を行った。
- (4) 詳細設計 (3)の仕様に基づきさらに詳細なデータの項目、プロトタイプの機能、画面の設計を行った。
- (5) 地図データベースの基礎構築 (3)(4)の仕様に基づき実際に地図データベースの構築, プロトタイプの作成を行った。

3.2 実施結果

3.2.1 地図データベースの基本仕様

作成する地図データベースの基本的仕様を以下のように決定した。

- (1) 使用ソフト MapInfo for Windows Ver. 3.0 (三井造船システム技研株式会社)
- (2) 適用範囲 佐賀県
- (3) 精 度 元となる地図の縮尺は2万5千分の1
- (4) 形 態 地図データベースソフトウェアMapInfoで扱える形態
- (5) 地図データ項目 表 3 1 に示す
- (6) 文字データ項目 表 3 2 に示す

表3-1 地図データ項目

区	分	項	B	内容
線デー		道路	路	高速道路,一般国道,一般県道,地方道,農道
	h	甩	坦 哈	(幅員3m以上の道路)
	-9	鉄	道	JR, 私鉄
		河	Ш	河川
点デ-	-タ	県庁	, Γ	5町村役場,病院,学校,公共施設,老人用施設
面デ-	ータ	行吗	女界,	県界, 湖, 沼

表3-2 文字データ項目

区分	内容
県	1. 行政コード, 2. 名称, 3. 面積, 4. 世帯数, 5. 人口, 6. 男子人口
乐	7. 女子人口,8. 6 5 歳以上人口,9. 人口密度
市町村	1. 行政コード 2. 名称, 3. 面積, 4. 世帯数, 5. 人口, 6. 男子人口
	7. 女子人口,8. 6. 5 歳以上人口,9. 人口密度
道 路	1. 道路区分, 2. 道路区分名称, 3. 道路名
役 所	1. 行政コード, 2. 市町村役場名称, 3. TEL, 4. FAX, 5. 所在地
鉄道・駅	1.区分コード, 2.区分名称, 3.鉄道区分, 4.鉄道区分名称, 5.名称
数理 数	6. 所在地, 7. TEL, 8. FAX
河 川	1.区分コード, 2.区分名称, 3.名称
公共施設	1.区分コード, 2.名称, 3.所在地, 4.TEL, 5.FAX
医療機関	1.区分コード, 2.名称, 3.所在地, 4.TEL, 5.FAX, 6.備考
老人施設	1.区分コード、2.名称、3.所在地、4.TEL、5.FAX、6.備考
学校	1.区分コード、2.区分名称、3.学校名称、4.所在地、5.TEL
- f* - f 	6. FAX, 7. 備考
農業統計	各市町村別の農業統計を入力

3.2.2 地図データベースシステムプロトタイプの基本機能

作成する地図データベースシステムプロトタイプの機能仕様を以下のように決定した。

- (1) 佐賀県の基本全体図を表示する。その上に道路のみ、河川のみ、市町村界のみの地図データを重ねて表示し、画面に表示されているイメージを印刷もできる。
- (2) 佐賀県の市町村の行政区界地図と基本的データを表示し、印刷もできる。
- (3) 佐賀県の市町村の行政区界地図と農業統計のデータを表示し、印刷もできる。
- (4) 検索, 画面の拡大縮小などは使用した地図データベースソフトMapInfo for Windows Ver.3.0 の基本機能を使用する。

例として図2-1に画面表示の一例を示す。

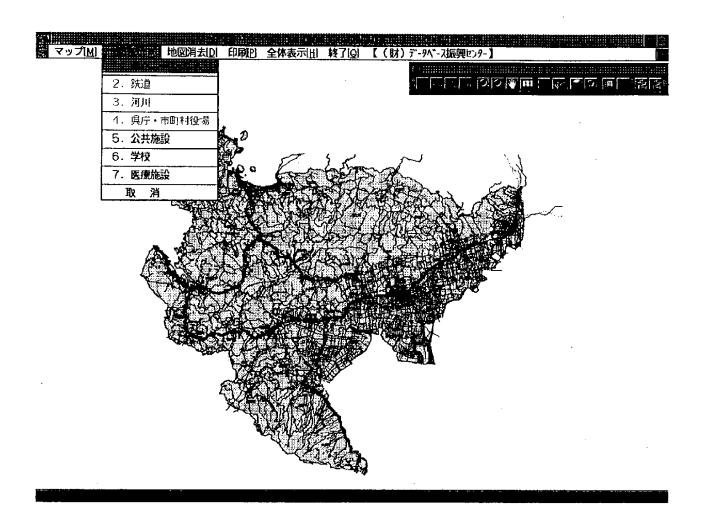


図3-1 プロトタイプ表示一例

3.2.3 データ作成

基本となる地図データは建設省国土地理院(国土地理院)発行の数値地図を基盤とし、不足データは同じく国土地理院発行の2万5千分の1地形図を MapInfoのディジタイザ入力を用いて行った。

3.2.4 プロトタイプ作成上の問題点

プロトタイプを作成する上で次のような問題点が発生した。

(1) 地図データベースシステムの機能への不満

今回は市販の地図データベースソフト MapInfoを使用したが、市場に現れ始めて間がないためか、ユーザインターフェースの面、処理スピードなどやや不満な点があった。

3.2.5 地図データ作成上の問題点

地図データを作成する上で次のような問題点が発生した。

(1) 地図データ入力の方法

紙の地図からディジタイザよるデータ入力は非常に手間がかかり、今までの文字データのようなスピードはとても望めない。これは入力の対象が図形であり、文字データのようにキーボードを使うわけではなく、カーソルを用いて図形をなぞっていくように入力することが原因であり、地図データを作成する場合において人件費によるコスト高を招くことになる。

3.3 今後の課題

地方における地図データベースの地図基礎データの構築を行った。今後さらにデータベース の充実を計るためには次のような様々な課題が残る。地図データベース作成上の課題, 地図デ ータベースシステムの課題, 地図データの提供の課題などである。

(1) 地図データ入力方法の再検討

今回は紙の地図からディジタイザによる入力を行ったがその場合に時間がかかりすぎ、コスト高を招きかねない。紙の上の地図データをスキャナーで取り込みデータの変換を行う等、他の方法の研究が必要であろう。

(2) 地図データベースシステムの機能の強化

今回、市販の地図データベースソフトウェアを使用したが、ユーザインターフェース、スピード等の点でまだまだ不満が残る点も少なくない。地図のデータに関してはそれぞれのソフトウェア用に変換が可能なので使用する地図データベースソフトウェアに対しての調査、研究が必要となろう。

(3) 地図データの変換の充実

データベースは汎用性がなければ普及しない。そのためにはデータを汎用的な形でつくることが望まれる。また、汎用的でなくても相互変換プログラムが存在すればデータの有効利用が可能である。そのための地図データベースソフトウェアのデータ形式の研究が必要であるう。

3.4 まとめ

地図を使ったデータベースシステムは今までの文字だけのデータベースシステムとは違い、 データの地図上の位置関係や分布の状態などを視覚的にとらえることができ、今までの紙の地 図の上で手書きでやっていたことをコンピュータ上で行うことにより繰り返しシミュレートし たり、検索したりした結果を地図の上に即座に表示したりといったことが簡単にでき、都市計 画や土地管理といったシステムへ利用すれば非常に魅力あるシステムであると我々は確信している。

しかし、地図データベースシステムには、データの整備状況の遅れなどまだまだ課題も多い。 今後も我々はその普及のためにも地図データベースの研究開発とその拡充に努めることが重要 であろう。 4. 景観シミュレーション用樹木の データベース構築



4. 景観シミュレーション用樹木のデータベース構築

4.1 目 的

現在,生活環境の充実,なかでも防災計画,暮らしやすい都市,美しい都市景観に対する期待が高まっている。

経済成長を推進するなかで、生活者個人としては物質的欲求のみならず、質的な向上や潤いある生活文化を求める志向が強くなる傾向にある。このような背景のもと、これから行われる環境計画の策定作業に対して、最も効率的な方法を論ずるとともに、システムとして運用できるファクトデータベースの可能性を探った。

今回の委託課題では、景観設計を行う際に多く用いられる3次元CADを操作する者を対象としている。3次元を前提とする理由は、景観設計において完成イメージを立体で捉える必要からである。また、従来の図面などでは表現できない奥行き、立体感を確認できることが可能となり、CGにおけるコミュニケーションが、単なる表現にとどまらず合意形成と問題解決を支援する手段、すなわちデザインツールとして幅広く活用される可能性を示唆しているからである。

これらのシステムに効率的なファクトデータベースを連動させることにより、今までにない設計手法を探り、実用に向けての詳細な利用形態を検証する。

4.2 実施内容

以下の方法によりプロトタイプ作成を通じて本研究を実施した。

(1) 景観シミュレーションの現状把握

景観設計に対する全般的な取り組み、3次元CADの活用はどのようになっているのか、また現状でどのような問題点があるのかを関係業界の専門家よりヒアリングおよび説明を受け把握した。

(2) 樹木データベースのあり方を討議

ファクトデータベースとしてどのようなソフトが供給できるのか、また、その構成の望ましい姿はどのようなものか討議した。

(3) 樹木データベース供給の完成イメージ作成

実用段階に表現される樹木の形態について、実際に生育している樹木との対比について検討 を踏まえ、樹木データプロトタイプの完成イメージを作成した。

(4) システム設計

(3)の完成イメージを個々から抜粋し、システムとして使用できる内容に構築する。実際に操作しても違和感のない構成とする検討を行った。

(5) 実地検証

完成したプロトタイプシステムを用いて景観シミュレーションを行う。ここでは、既に計画 されている景観に対して樹木データベースの効果を検証する。

(6) 今後の計画

プロトタイプを通じて得た課題とシステム評価を参考にこの研究を今後どのように発展させるか検討した。

4.2.1 プロトタイプのコンセプト

データベース検索には、専門の知識と機器を必要とする。本委託課題では、専門知識を有しなくとも、簡単にわかりやすい操作ができる環境設計を念頭に設計を進めた。また、本プロトタイプは景観シミュレーション用ファクトデータベースの幅広いジャンルの中で、今後都市部などで必要とされる景観要素である「樹木」を題材に進めた。

基本コンセプトとしては、「誰もが簡単に操作できる」を実現するためにはどのような方法をとればいいのか、また、どのような考え方で進めればいいのか検討を重ねた。ここでは、このような問題点を以下の3つのテーマに絞り研究を試みる。

- (1) 操作性の向上(3次元CADデータを可視化)
- (2) 検索されたデータを代理機能(エージェント)により CADアプリケーションまで移行させる。
- (3) パソコン操作を前提とする。

4.2.2 データの可視化による視認性の向上

本プロトタイプでは、内容物(樹木の形態)の表示方法を最優先に検討した。従来ファクト データベースは、構成する要素の「量」を重視している傾向にある。そこでは検索名称と内容 イメージとの相違は問題にあがらない。

一方、景観シミュレーション用となると完成イメージを検索段階から要求するためよりリアルな表示が望まれる。そこで本プロトタイプでは、検索画面を表示する段階で、樹木を構成するデータの全形、ディテール、樹木全形回転体などを同時に表示することができないかを検討した。

その結果、視認性の向上が認められた。

4.2.3 代理機能による操作の簡略化

この度のプロトタイプにおいて、データベース検索ソフトに代理機能(エージェント)を付加することで、今まで以上の操作性の向上が計れるか研究を試みた。

最近のパソコンの多くは機種に依存しないオペレーティングシステム(OS)を使用している。今回はその代表である Windowsを使用している状態でエージェント機能の可能性を検証してみることとする。

ここで示す「エージェント機能」とは、Windows上で起動している本プロトタイプを使用して樹木を検索している際に、目的とするファイルを直接CADアプリケーションに受け渡す機能のことを指している。このエージェント機能を使用することで景観シミュレーション作業時に、その都度ごとにCADアプリケーションを起動する手間を省き本来の業務に専念できるようにした。

今回実施する機能を受け取るCADアプリケーションは、建築設計で定評のあるインターグラフ社のマイクロステーションを使用し実地検証した。実際に操作した結果、所期の目的は達成することができた。特に、登録されているデータを視覚的に確認した後CADに移す操作は今までにない効果がみられた。また、同時にいくつかの問題点も発生し今後の研究を推進するうえで貴重なデータを収集することができた。

4.3 実施結果

(1) 樹種, 用途, 樹齢による検索機能

本データベースを使用したデータ検索をする際、検索条件を設定する手段として、簡単な操作で効率よく設定できる方法で検索できることを考慮、樹木の分類及び選択の判断基準の観点から樹種、用途、樹齢の3項目に分類、これらの項目を設定することで該当樹木を検索可能とした。

(2) 樹木名入力による検索機能

既に目的とする樹木がわかっている場合,この機能を利用することで素早く検索できるようにした。

(3) 樹形形状ファイルの保存形式選択機能

3次元コンピュータグラフィックスの世界において標準的なDXF形式による保存機能だけでなくCADと合わせて本データベースを利用することも考慮、アプリケーションファイル形式も含める。

(4) シーケンシャル,ランダム検索の切り替え機能

樹種,用途,樹齢のみの設定で検索を行うと該当樹木本数が多いため,通常はシーケンシャルに閲覧する設定になっているが、検索後の検索条件該当樹木名の一覧を表示させることにより、ランダムに閲覧することも可能とする切り替え機能を加える。

4.3.1 基本画面構成

検索画像を表示するベース画像を元に、ボタン、検索結果の画像、動画を配置。

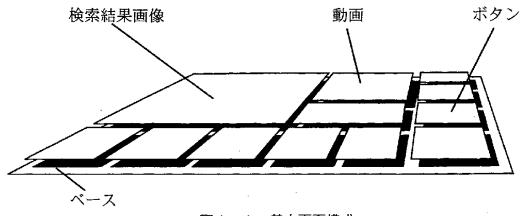


図4-1 基本画面構成

(1) ボタン

視認性と操作性を考慮,80×80のピクセルのボタンに,25ポイントゴシック太文字で表記。 また、ボタンをアイコンにすると、初心者では操作時の混乱を招くと判断、全てのボタンに 文字を表記。

(2) 動 画

本データベース内で使用されている動画は Cinepak圧縮でCD-ROM上での再生のため 毎秒 266KBの転送レートとしてある。再生時には全てメモリー上に読み込み、滑らかに再 生するよう考慮した。

(3) 検索結果画像

検索後の樹木画像は当初静止画像を使用していたが、検索後の表示処理高速化のため 1 fps (フレーム/秒) の動画像を一時停止の状態で表示することにより、 1 枚のグラフィックとして表示していたときに比べ、トータルで 3 割程度の速度向上がみられた。

(4) 表示文字

画面に現れる文字サイズは、ゴシック体14ポイントで表現する。画面に触れた感覚を出す ために、触れたボタンが瞬間に奥へスライドする表示がされている。

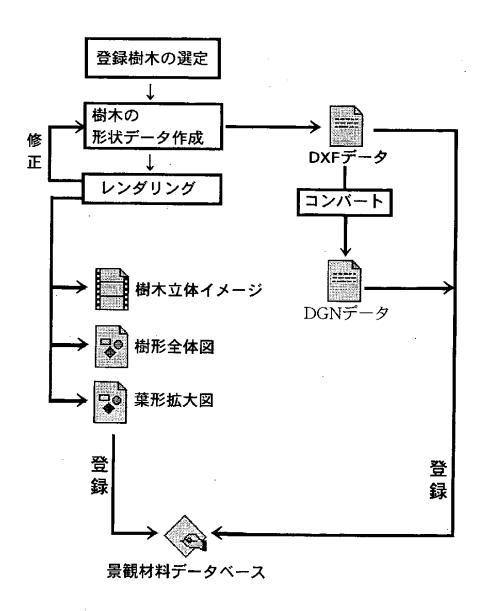
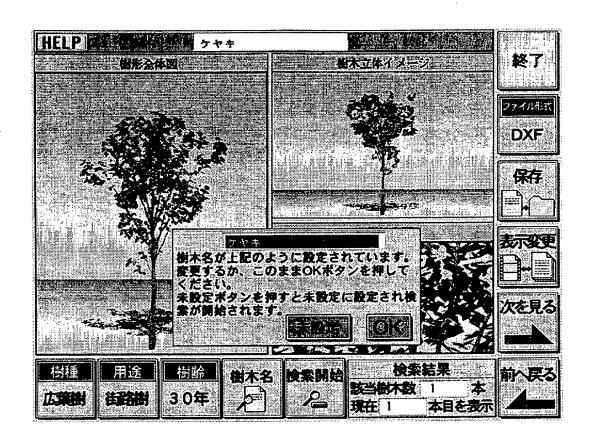


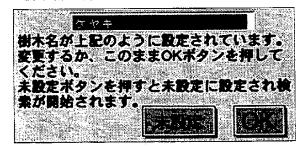
図4-2 登録データ制作行程の概要と流れ

画面名称	樹木名継続入力確認画面
画面No	0 0 1



機能説明

樹木名を設定後、同じ樹木名で検索を行う際の確認及び入力の変更。





検索条件の樹木名を「未設定」にし、検索を開始する。



検索条件の樹木名を継続して適用させ、検索を開始する。 リターンキー入力はこちらのボタンに割り振られている。 5. 包装機械データベースの構築

紐日本包装機械工業会

·			

5. 包装機械データベースの構築

5.1 開発の背景と目的

5.1.1 開発の背景

包装機械工業は、食品、薬品、各種工業製品などを消費単位の大きさに包装する機械の製造・販売を行っており、生産される包装機械の種類はきわめて豊富である。包装機械のユーザは、食品加工業から医薬品工業など多岐にわたっている。これらのユーザが必要な包装機械の導入を行う時に、各種の包装機械に関する情報を効率よく引き出すデータベースが必要になっている。

5.1.2 開発の目的

包装機械データベースの開発の目的は、包装機械のユーザがコンピュータを利用して、包装機械に関する各種の情報を自由に検索して、その希望する情報を簡単に効率よく引き出せるようにすることである。

これまで、 1,000機種以上の包装機械の情報を掲載したカタログ集「日本包装機械便覧」が機械選定の参考資料として利用されてきた。しかしこれは、紙製であり、A 4 判 1,200ページ、厚さ 5~cm、重量 2.8~kgと大部のものであり、誰もが簡単に、的確に目的の機械を捜し出すにはかなり面倒な作業を必要とした。

コンピュータを利用して各種の検索方法が可能になれば、目的の機械を簡単に高速で検索して、探し出せるようになる。包装工程の計画を行うときに、このような包装機械データベースを利用すれば、各種の包装機械に関する情報を効率よく引き出すのに役立つと期待される。

5.2 包装機械データベースの検討

ここでは、包装機械データベースを作成するにあたって、重要な事項を検討する。まず、既に ある紙製の「日本包装機械便覧」の構成を検討し、包装機械のデータ特性の把握、検索方法の検 討、そしてデータベースの作成作業の検討を行った。

5.2.1 日本包装機械便覧の構成

日本包装機械便覧は以下のように包装機械および関連する機器・材料を分類している。

(1) 計量機/計数機

- (2) 充塡機/びん詰機械/かん詰機械
- (3) 製袋充塡機
- (4) 容器成形充塡機
- (5) ラベル貼機
- (6) 小箱詰機/製箱機
- (7) 上包機
- (8) シール機
- (9) 収縮包装機
- (10) 真空包装機/ガス置換包装機
- (11) バンド掛け機/ひも掛け機
- (12) ケース詰機/封函機/製函機/その他の外装荷造機械
- (13) 包装プラント・システム
- (14) 包装材料加工機械(製袋機/スリッター等)
- (15) 包装関連機器
- (16) コンポーネント
- (17) 食品機械
- (18) 包装資材

以上の他に、日本包装機械工業会の会員(正会員、賛助会員)のリストがあり、これも様々 に利用されている。

5.2.2 包装機械のデータ特性

包装機械のデータとしては、ひとつの機械を記述するための最低限の情報であって、以下のようなものが上記の便覧には、機械の名称、機械の型番、機械の分類名称、特徴、仕様などが記述されている。これらは文字情報であるが、この他に機械の写真、必要に応じて包装された品物の写真がつき、具体的にその機械の特徴を伝えるものになっている。

5.2.3 要求される検索の方法

この包装機械データベースを検索するユーザは、一般に次のような検索をしたいと考えている。以下は、包装機械の専門家から提出された検索の要望である。

- (1) 包装対象からの検索
- (2) 包装形態からの検索
- (3) 包装機能からの検索

- (4) 包装対象品の性状からの検索
- (5) 情報の文字列などからの検索

5.2.4 データベース作成の作業手順

包装機械データベースを作成するにあたっては、包装機械のデータのキーボードからの入力、 写真のスキャナーによる入力、検索ソフトの設計と開発、出来上がった包装機械データベース のCD-ROM化などの作業が必要になる。これらについて検討を行った。

5.3 包装機械データベースの基本設計

ここでは、包装機械データベースの基本設計を行っている。

まず、包装機械データベースの動作するシステムの動作環境、データ構造、プログラム構成などを検討した。

5.3.1 システムの動作環境

本システムが稼動するための環境条件は、最近普及しはじめている WINDOWSパソコンを対象 にするものとした。

5.3.2 データ構造

包装機械のデータの形式としては、既存の便覧の記述をベースにした。しかし、それぞれの 機械がどのような検索に対応するかの情報は、便覧には記載されていない。そこで、ひとつひ とつの機械について、検索に対応した情報のデータ構造を設計する必要がある。

ここでは、つぎのようなデータの形式を考えて、包装機械に詳しい専門家により、データシートに記入してもい、これをコンピュータに入力した。

- (1) 整理番号
- (2)包装機械の機種・型式名
- (3) 会社名
- (4) 包装対象品の分野
- (5)包装対象品の性状
- (6) 包装資材(該当するもののみ)
- (7)包装形態別-袋詰め形態
- (8) 包装形態別-各種の容器詰め形態

- (9)包装形態別-カートン詰め形態
- (10) 包装形態別-紙, 贈答用紙などによる上包み形態
- (11) 包装形態別-紙以外の上包み形態
- (12) 包装形態別-真空包装形態
- (13) 包装形態別-収縮(シュリンク) 包装形態
- (14) 包装形態別-結束, 結さつ形態
- (15) 包装形態別-ラベル貼り
- (16) 包装形態別-外装·輸送包装形態
- (17) 包装形態別-上記以外の包装形態
- (18) 機械の機能別-計量・計数機能
- (19) 機械の機能別-充塡・瓶詰・缶詰機能
- (20) 機械の機能別-製袋充塡・ピロータイプ・袋詰め機能
- (21) 機械の機能別-容器成形充塡・ブリスター・PTP・ストリップ包装機能
- (22) 機械の機能別-ラベル貼り機能
- (23) 機械の機能別-小箱詰め・製箱機能
- (24) 機械の機能別-上包み機能
- (25) 機械の機能別ー収縮(シュリンク) 包装機能
- (26) 機械の機能別ーシール(封かん)機能
- (27) 機械の機能別-真空包装・ガス置換・ガスフラッシュ機能
- (28) 機械の機能別-バンド掛け・ひも掛け
- (29) 機械の機能別-製函・封函・ケース詰めその他の外装荷造機能
- (30) 機械の機能別ー包装材料加工機能
- (31) 機械の機能別-コンポーネント(機器)
- (32) 機械の機能別-包装関連機器
- (33) 機械の機能別-食品加工・製菓など
- (34) 機械の機能別-上記以外の機能
- (35) 卓上型・簡易型の機械
- (36) 使用可能な包装材料別-紙・板紙
- (37) 使用可能な包装材料別-木・プラスチック・セロファン
- (38) 使用可能な包装材料別-ラベル
- (39) 使用可能な包装材料別-金属製
- (40) 使用可能な包装材料別ーガラス製

- (41) 使用可能な包装材料別 結束・封緘材
- (42) 使用可能な包装材料別-ゴム・コルク類
- (43) 使用可能な包装材料別ー包装副資材
- (44) 使用可能な包装材料別-上記以外の包装材料

またこのデータ入力と並行して 1,607枚の写真をスキャナーによって取り込む作業を行った。 この写真はそれぞれの機械が呼び出されてきたところで、文字と同時に画面に表示するように 計画した。写真は全ての機械について機械そのものの写真が1枚あり、ものによっては包装さ れた品物の写真が1枚追加されるようになっている。

5.3.3 開発したプログラム

開発するプログラムとしては、データ入力用のプログラムと検索用のプログラムがある。データ入力プログラムは準備段階で使われるだけであるが、将来はバージョン・アップなどにも使われると考えられる。検索用のプログラムは、実際にユーザによって動作し、CD-ROMに格納して配布されるものである。

5.4 包装機械データベースの内容

プログラムがスタートするとまずメインメニューが表示され、そこで目的の作業を選択する。 メインメニューから、それぞれのルーチンに分岐してゆくが、目的の作業が終了すると必ず最初 のメインメニューに戻るように構成されている。

5.4.1 メインメニュー

システムを立ちあげると、次のようなメインメニューが表示される。

196日本包裝機械便監

'96日本包装機械便覧

社団法人 日本包装機械工業会

検 東 包装対象分野、包装形態、包装機能、対象 品の性状などから、該当する会社と製品を 探して紹介します

(社)日本包袋機械工業会 正会員一覧

(社)日本包装機械工業会 贊助会員一覧

包装資材コーナー・機器リースコーナー

#37

ここで以下のようなコマンドが利用できる。

1)検索

このボタンをクリックすると、各種の検索が行えるようになっている。検索の方法は、包装対象分野、包装形態、機械の機能、卓上型・簡易型の機械、使用する包装材料、対象品の性状などがある。検索結果は該当する機械とそれを製造・販売している会社が表示され印刷出力ができる。

- 2) 正会員一覧
- 3) 賛助会員一覧
- 4) 包装資材コーナー、機器リースコーナー

5.4.2 検索

メインメニューで検索のコマンドボタンをクリックすると、検索画面に移行する。検索画面 は次のようになっている。

一 96日本包装機械便覧
包装・加工対象品の分野
包轴形態
機械の機能
卓上型・簡易型の機械
使用可能な包装材料名
包装・加工対象品の性状
会社名
機械型式(記号)
特定の用途・目的・対象品
検索件数 = 0000
·
検索開始 ・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・

検索の方式は以下のような特徴を持っている。

- 1) 各種の条件に同時に該当するもの(AND検索)を探し出す。
- 2) 一度検索したら、該当件数が表示されるので、さらにそこから対象を絞りこむために条件を追加して、再度検索を行うことが出来る。

ここでの検索の方法は以下のような各種が用意されている。

1) 包装・加工対象品の分野

ここでは、食品、医薬品、化粧品、化成品、一般化学製品、工業製品、その他の内のいず れかから選択して、検索できる。

2) 包装形態

ここでは、袋詰め、容器詰め、カートン詰め、上包み、真空包装、シュリンク包装、結束、 ラベル貼り、外装、その他の内のいずれかを指定して検索できる。

3)機械の機能

ここでは、計量・計数機能、充塡・瓶詰め・缶詰め機能、製袋充塡機能、容器成形充塡機能、ラベル貼り機能、小箱詰め・製箱機能、上包み機能、シール機能、シュリンク機能、真空包装機能、バンド掛け・ひも掛け機能、製函機能、包装材料加工機能、コンポーネント、

食品加工・製菓、その他のうちからひとつを指定できるようになっている。

4) 卓上型・簡易型の機械

ここでは、卓上型・簡易型の機械でその機能としては3)で指定するのとほぼ同様の分類 にもとづいて、機械の検索を指定することができる。

5) 使用可能な包装材料名

ここでは、包装材料を指定する。紙・板紙、木・プラスチック・セロファン、ラベル、金 属、ガラス、結束材・封緘材、ゴム・コルク類、包装副資材、その他の内のいずれかを指定 することができる。

6) 包装・加工対象品の性状

ここでは、包装・加工する対象物品の性状を、液体、粘体、粉体、顆粒、固体、その他の うちの何れかから選択する。

7) 会社名

ここでは会社名の一部を入力して、該当する会社を探し出すことができる。

8)機械型式(記号)

ここでは機械の型式番号を入力して、その機械の名称と製造している会社を探し出すこと ができる。

9)特定の用途・目的・対象品

ここでは、包装対象品の文字を入力して、その文字が機械の包装対象品欄にあるものを探 し出すようになっている。

10) 検索開始

このコマンドボタンをクリックすると検索が開始される。検索結果は検索件数として画面 の左下に表示される。これを見ながら検索条件を絞りこんでゆくことができる。

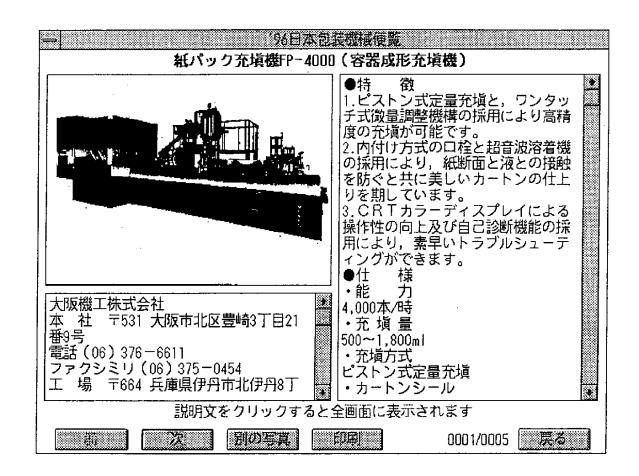
11) 一覧表示

このコマンドボタンをクリックすると検索結果の該当するものがリスト表示される。

12) 条件初期化

このコマンドボタンをクリックすると検索条件を白紙にもどして最初から入力することができる。

次には検索結果として呼び出された画面の例を示す。



これは、該当する機械の画面である。 1,042の項目がこのような画面に納められている。最上部にタイトルとして、機械の名称、型番、機械の分類名が表示される。左側上部には機械の写真、その下には製造会社名、住所、電話番号などが表示される。右側には、機械の特徴、仕様などが表示される。

5.5 結論

「日本包装機械便覧」をベースにして、包装機械のユーザが効率的に利用できる包装機械データベースの開発を行った。

各種の検索に適したデータ構造を検討し、データ入力の方法を設定して、1,042種の包装機械に関するデータの入力を行った。同時に1,607枚の写真をスキャナーにより入力して、検索結果と同時に画面に呼び出せるようにした。 WINDOWS上で動作する検索プログラムを開発して、検索機能と速度を確認した。

開発した包装機械データベースを関係者に試用をお願いして、表示画面、検索性能、使い易さなどを評価していただいた。いまのところ、大きな問題点はなく、包装機械関係者からは非常に使い易いという声を聞いている。

5.5.1 今後の課題

今後の課題としては、以下の各項があげられる。

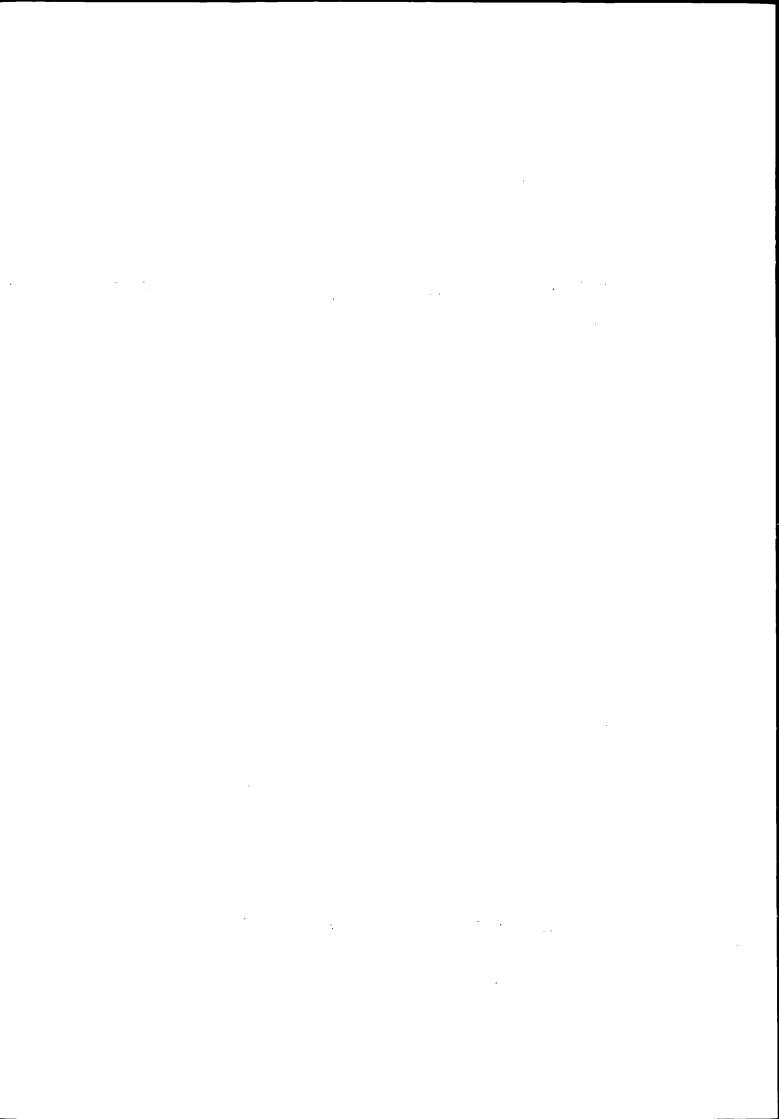
- (1) 検索項目の強化
- (2) 写真および動画像の扱い
- (3) 英語版の開発

5.5.2 まとめ

包装機械の使用者が効率的に目的とする包装機械を見つけ出せるようにするため、包装機械データベースの開発を行った。現状の「日本機械便覧」をベースにして、各種の検索に適したデータ構造を検討した。データ入力の方法を設定して、1,042種の包装機械に関するデータの入力を行った。同時に1,607枚の写真をスキャナーにより入力して、検索結果と同時に画面に呼び出せるようにした。検索方法としては、広範に普及しつつあるパソコンで動作し、WINDOWS上で使える検索プログラムを開発して、検索機能と速度を確認した。

6. 新規事業創出支援のためのデータベース構築に向けた基礎調査

(株)日本インテリジェントトラスト



6. 新規事業創出支援のためのデータベース 構築に向けた基礎調査

6.1 概要

6.1.1 目 的

我が国経済を取り巻く環境は、昨今の急激な円高とアジア各国の急成長、米国製造業の「復活」などにより、ダイナミックな国際的競争の時代に突入している。このような状況の下、我が国の産業は既存産業の成熟化と新規産業展開の遅れ、製造業の海外展開の加速と産業の空洞化の懸念、本格的なリストラクチャリングの進展に伴う雇用不安など、大きな曲がり角にある。このような現在の閉塞状況を克服し、我が国の産業の更なる発展を実現していくためには、今後高い成長が期待される先端技術分野や地場資源を活用した産業分野などにおいて新規事業の創出を促し、海外競争力の高い高付加価値製品を核に新たな需要を創造する独創的製品の開発および新たな需要を喚起するニュービジネスの創出を図る必要がある。政策的にも通産省が「総合的新規事業支援施策」を平成7年度の重点施策として掲げていることに代表されるように、その重要性が強く認識されつつある。

当調査研究テーマは、国および地方で展開されつつある新規事業支援施策の一環として、また起業家および新規事業展開を図る中小起業の経営資源の高度化、技術力の向上等の一助となる、有益にして有効なデータベースの構築を目指すものである。

6.1.2 実施概要

当調査研究では、以下の内容を実施した。

- (1) 提案課題の必要性の検討 ~新規事業創出の必要性~
 - (内容)・新規事業創出が求められる背景と、わが国における新規事業支援施策の動向 ・情報面での支援施策の考え方
 - (概要)企業に密着した支援を行っていくには、各地域における支援機関の充実が不可欠である。そこで全国各地でベンチャー財団が設立されるなど、資金支援面については施策が充実しつつある。ただし新規事業の創出を図るためには、資金面の支援措置だけでは不十分であり、様々な障害を総合的に支援すること、また更には成長段階ごとに長期的に一貫した施策により支援していくことが重要である。情報発信、情報提供、情報交流の促進により、より多くの今後の課題に対処していくことが可能である。
- (2) 関連データベースのニーズ検討

- (内容)・アンケート調査、ヒアリング調査の実施
- (概要) 当該データベースのニーズ把握のため、アンケート調査を実施した。対象は起業家 層ならびに起業家予備群である。また、アンケートでは拾えきれない項目については、 別途地方圏のベンチャー企業を対象に、個別ヒアリングによる意見収集を行った。

アンケートの結果からは、商用データベースのニーズ層としてベンチャーに的を絞ることは、情報提供者にとって市場性の高いテーマであると認識した。また情報提供手段については、オンラインネットワークによる全国規模での事業展開について十分な可能性が見い出せた。さらに情報内容については、ベンチャー企業に望まれている最大の関心事が「交流」であることが明らかとなった。ヒアリングにおいても、ウェットな情報を求め情報交換が行える交流環境の整備が強く期待されていることが分かった。

- (3) 地域におけるデータベース化への取り組み
 - (内容)・地方公共団体における地域情報化の動向および事例調査
 - ・関連機関のネットワーク化の検討
 - (概要)事例調査からは、全国各地において地域産業の活性化もしくは中小企業の高度化を主目的とした、地域独自の様々な情報提供サービスが実施されていることが明らかとなった。しかしこの様な貴重な情報収集、情報提供サービスも、独自のシステム、独自のノウハウ、独自の活動において実施されているため、いずれの認知度もあまり高くなく、利用状況も良好とはいい難い。また今後の課題も明らかとなった。

そこでこのような時期, 関連情報提供者がネットワークを組み, 一つの窓口を設けて情報発信を統合・強化することによって, 非常に効率的かつ利用者側にとってもメリットが多い事業展開が可能になると考えた。ただしその構築には幾つかのステップが必要となる。

- (4) 新たな情報インフラの利用と動向把握
 - (内容)・インターネットの拡大とインターネットを利用した先行事例の調査を実施
 - (概要) インターネットの急激な拡大は、情報提供者側にも今後の新たな事業展開の可能性 を提供するものである。また米国の事例では、バーチャルインキュベーターと呼ばれ る起業家支援事業がみられ、非常に参考となった。
- (5) 当該データベースに求められる姿
 - (内容) ・ 当該データベース (DB) に望まれる情報内容および今後の整備方針の検討
 - (概要) 当該 D B の情報内容には、単に制度を知るためのものだけではなく、 "人"と "実績" と "評価"のデータを盛り込みたい。また画像情報を取り入れることも是非必要

である。そしてネットワーク上に交流の場を設け、エンジェルやメンターと呼ばれる 起業化促進のためにキーパーソンとなるような人を紹介できるシステムづくりも期待 される。

(6) 事業化の検討

(内容)・事業形態およびシステムについての検討

・事業採算性の検討

(概要) データベース運用のためのインターネットシステムについて、4つの事業形態を想定した。また採用するデータベースシステムを検討し、概算費用などを含むビジネスプランを策定した。結果は、今後の検討課題は残るものの、全国関連機関とのネットワーク構築によるデータベース構築費用の軽減や安定的な会員収入を前提に、当該データベース構築・運営について、成立可能性が認められた。

6.2 実施結果

6.2.1 ニーズ調査結果のまとめ

アンケート調査結果ならびにヒアリング結果から得られた視点について、以下に概括する。

① 対象者についてみた場合

ベンチャー起業家に対するアンケート結果によると、起業家層の多くは企業に勤めるサラリーマン層である。但し過去の絶対層は技術職が多かったが、昨今では事務職がニッチ分野や新業態で起業家を目指す形態が増加しており、その数は逆転していると思われる。後者の起業家層は情報収集力に長けており、情報存在場所にも敏感である。一方の技術職であるが、起業時の相談件数は少なく、技術力に集中するあまりマーケティング等周辺情報の重要性があまり認識されていない。今後わが国のリーディング産業としての本当の意味でのベンチャー企業を育てていくためには、この層の発掘が重要であると思われるが、データベース提供者からみた場合にはより公的支援が必要となろう。ただし情報機器の使用度は高く、また創業後のインターネットを用いた事業展開などの割合も高く、新たなネットワークを利用しての情報活用が期待される。

一方、既存企業の新規事業展開に当たっての情報提供では、企業の情報の収集先が多様化してきており、今後商用データベースや外部調査機関・コンサルタントに対する需要の伸びが期待できる。既存企業の新規事業展開の際に求められる情報は、既に広範な企業取引が存在するため企業情報の需要は少なく、業界動向、マーケティング情報、調査研究情報などであり、当社の有するノウハウの活用が十分発揮される可能性が高い。

結論として、商用データベースのニーズ層としてベンチャーに的を絞ることは、情報提供者にとって市場性の高いテーマと思われる。

② 情報提供手段についてみた場合

ネットワークビジネスが注目される今日、インターネットは第二の産業革命とまでいわれ だした。企業の情報に対する関心は日に日に高まっており、特に若い企業の多いベンチャー 企業のニーズは強い。また情報提供のメディアとして、ベンチャー企業ではパソコン通信よ り機関誌との声が若干多かったが、起業家予備群には圧倒的にパソコン通信が支持された。 オンライン通信の利用は地方のベンチャー企業にとって特に有益であり、利用度も先行して いる。したがって全国規模での事業展開も十分な可能性が見い出せた。

③ 情報内容についてみた場合

ベンチャー企業に望まれている最大の関心事は「交流」である。ヒアリングにおいても、ウェットな情報を求め情報交換が行える交流環境の整備が強く期待されていることが分かった。また、アンケートにおいて望まれる情報内容を尋ねたところ、「支援機関情報」、「支援施策情報」が上位を占めた。しかし一方で、資金調達の手段を幾つか知ったところで資金を調達できるわけでもなく、実際にはそれが活用可能かどうか、担当者と会って会話を交わして初めてそこから始まる。つまりドライな情報は目的に達するまでに辿る"道しるべ"でしかない。アンケートで、この二つの次に「専門家との相談室」が支持されたのはこのためであると思われる。

ドライな情報は多ければ多いほど不確実性に対処可能となるが、ウェットな情報は少なく てもそれ自体に価値が生まれる。したがって当該テーマに望まれる本当の意味での情報内容 は、「交流」を最終目的とする手段を見つけるための情報であり、人と人とのネットワーク を形成することが第一であろうと考えられる。

6.2.2 当該データベースに求められる姿

ニーズ調査や事例調査を踏まえ、また当該調査の実施に当たって組織した検討委員会の委員 の皆様からいただいた様々な意見をまとめ、当該データベースに望まれる情報内容についてま とめた。

① 情報価値を高めること

新規事業創出のために利用価値のある既存のデータベースもいくつか把握できた。しかしその検索結果だけでは、判断基準が明確にされない点が指摘される。当該データベースには、その制度を提供している機関のデータベースとの連携を図り、さらには"人"と"実績"と"評価"のデータを盛り込みたいと考える。

人とは、その制度が実際に利用可能かどうか、またより詳しく知りたいと思ったときに窓口となる人、どの制度を利用すべきかの選択をアドバイスしてくれる人などである。行政機関の縦割り組織の弊害は指摘されているとおりであり、目的の窓口に達するまでに時間と費用がかかってしまう問題点が指摘される。特に地方の企業が中央機関の制度を利用しようとした場合には、問題が生じる可能性が高い。また実績については、客観的なデータとして盛り込むことが可能であろう。評価については、利用実績のある企業のコメントを収集し、総合的に評価しなくてはならない難しさはあるが、情報価値としては重要である。これらの情報を盛り込むことで、利用者が必要な機関にたどり着くまでの機会費用を大幅に削減することが可能である。

② 画像情報の導入

事例調査からは、全国各地において地域産業の活性化もしくは中小企業の高度化を主目的とした、地域独自の様々な情報提供サービスが実施されていることが明らかとなった。しかしこの様な貴重な情報収集、情報提供サービスも、テキストベースが主体であること、独自のシステムであるため操作性が共通でないなど、利用者側の使い勝手の問題点が指摘される。例えば技術交流や受発注目的の利用のケースを想定した場合、絞り込み作業には相当の慣れが要求されること、また判断材料とするためには膨大な情報量が必要となり、上記のデータベースの情報内容だけでは不十分な点である。ある企業の人から、発注が可能かどうかはその工場をみれば大体分かるとの意見が聞かれた。画像情報が持つ情報量はテキストベースとは比較にならないほど膨大であり、当該データベースには是非画像情報を取り込みたいと考える。

③ オンライン交流の場の提供と人の情報の重視

アンケートの結果から明らかになったことは、ネットワーク上に交流の場を設けることが強く求められている点である。電子メール、電子掲示板など顔の見えない相手同士のコミュニケーションもなかなか有効なときがある。オンライン会議室で得られる情報は、相談事の回答より同じ問題を抱える人同士が知り合えるだけという指摘もあるが、これもまた、ネットワークの構築面から考えると、将来的には異業種交流や人材のマッチング機能を果たし、新たな新規事業の萌芽を促進するものと評価される。交流ネットを構築することで、リアルタイムにウェットな情報を盛り込むことが可能となる。具体的な交流テーマとしては、「人」「技術」「資金」などが利用面で期待されよう。

特に人の面では、日本にもベンチャーを個人的に支援するエンジェルと呼ばれる人達が増えてきており、起業家との交流の場を組織するなど、様々な支援活動を始めている。またエンジェルの他にも、起業化促進のためにキーパーソンとなるような人、例えばメンターと呼

ばれる相談者(多くは大学の先生)を紹介できるシステムも期待される。

6.2.3 事業化検討のまとめ

「新規事業創出支援のためのデータベース」運用のためのインターネットシステムについて、 以下のような事業形態を想定した。

- (1) 情報提供型の事業
- (2) プロバイダ型の事業・
- (3) システム提供型の事業
- (4) 情報提供型とシステム提供型を組み合わせた事業

また、システム構築に当たっては、以下に示すシステムを前提とした。

(1) データベースシステム (約 1,300万円)

(2) インターネットシステム (約 3,200万円)

(3) その他付随するソフトウェア開発(約 700万円)

上記の初期投資額 5,200万円の他,運営費,および事業の内容について以下に示す。なお, 当事業は当社の付帯事業とし、人件費に係わる会社負担その他諸経費、事務経費等については ここでは見込まない。

表 6-1 年間運営費

	·				
1	人件費			-	1,920万円
	内訳	①システム管理者	1名	50万×12	
		②データベース管理者	1名	50万×12	
		③営 業	1名	30万×12	
		④サポート要員	1名	30万×12	
2	回線使用	月料			804万円
	内訳				
	·	②INS1500		13万×12	
3	スペース	720万円			
4	その他費	費用			2, 160万円
	内訳	①広告宣伝費		100万×12	
		②データベース更新費	(外注)	60万×12	
		③雑 費		20万×12	
		· 合 : 言	t		5, 604万円

表 6-2 事業内容および料金体系

1	データベース会員サービス	2 万円/月	
	会員にのみ詳細な情報をサービスする	4 /3 1/ /3	
2	ダイヤルアップ接続サービス		
	インターネットに参加していない個人向けの接続サービス	6 万円/年	
	(公衆回線とISDN回線)		
3	専用線接続サービス	20万円/月	
	専用線(64Kbps)による接続の企業向けサービス		
4	レンタルサーバサービス		
	当Webサーバ上に広告を中心としたホームページ掲載サービ	5 万円/月	
	X		
5	ホームページ制作サービス	50万円/件~	
	ユーザ(企業)ホームページの提案,設計,制作をサービス		
6	サーバ構築サービス	800万円	
	専用線接続する場合に必要となるサーバの構築サービス	/YZテム~	

初期投資については5年間で償却することとし、運営費については年度毎に処理することを 前提にビジネスプランを立案した。

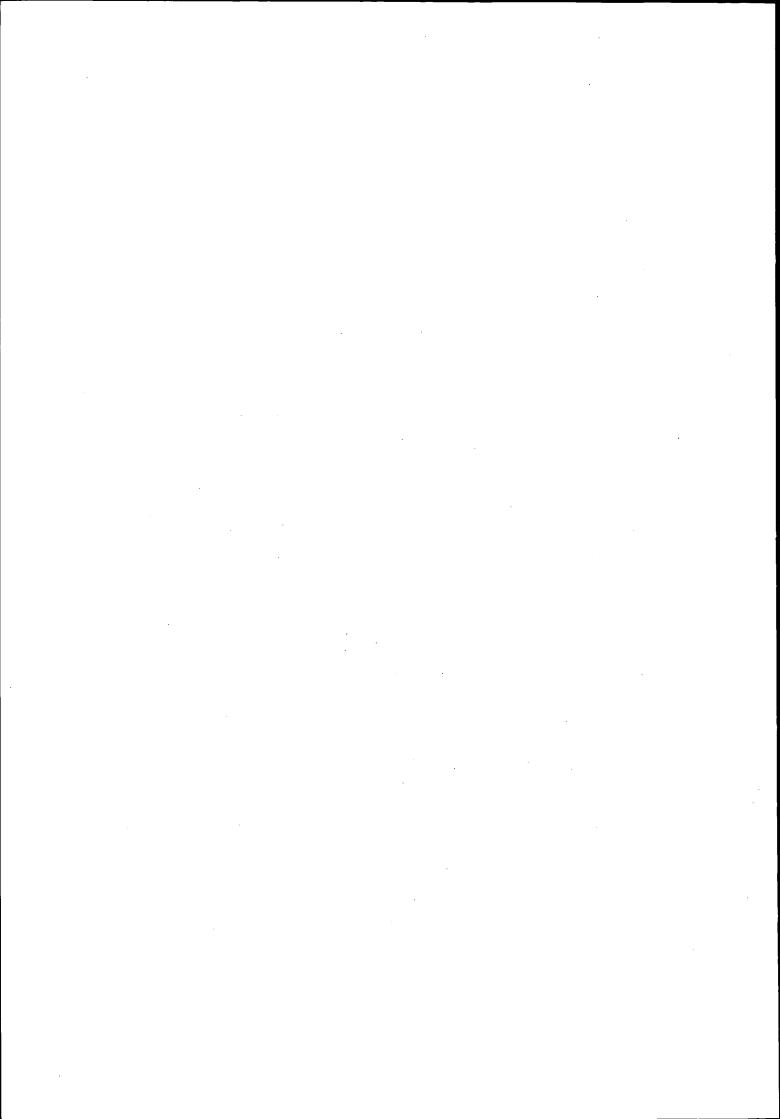
結果は、今後の検討課題は残るものの、全国関連機関とのネットワーク構築によるデータベース構築費用の軽減や安定的な会員収入を前提に、当該データベース構築・運営について、成立可能性が認められた。

6.3 今後の展開および予想される効果

本年度の基礎調査と事業化の実証を踏まえ、次にはプロトタイプの構築、事業化へと展開していきたいと考えている。

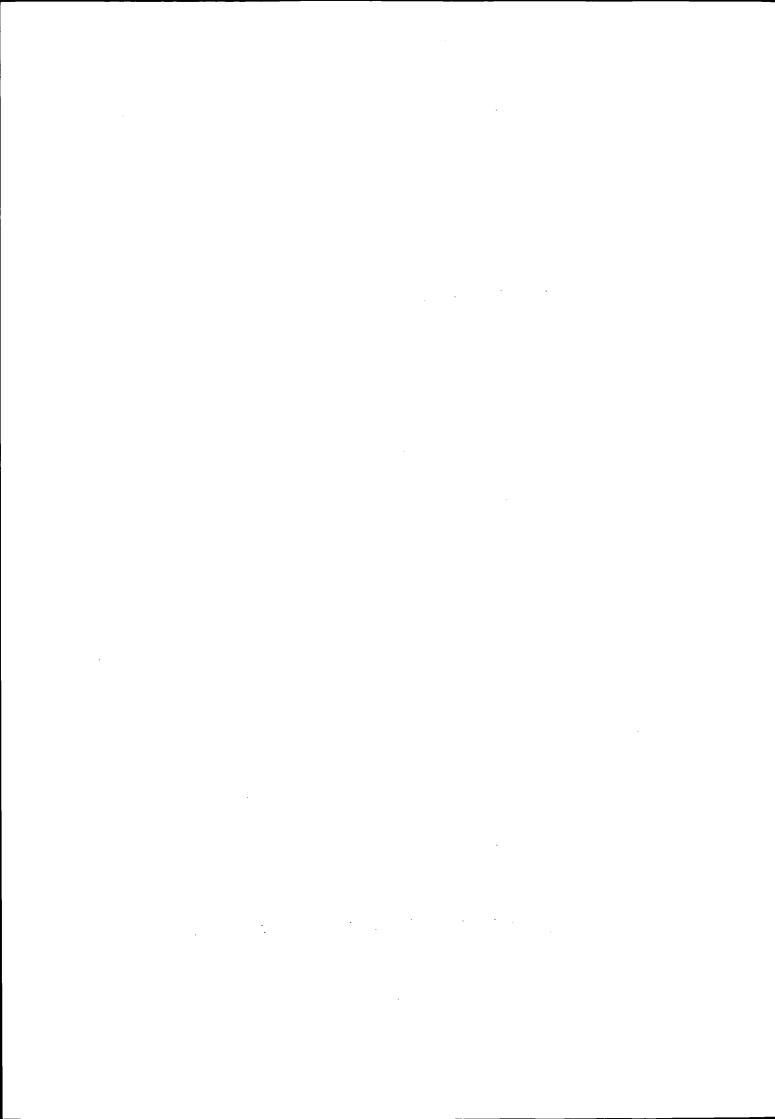
今後予想される効果については、当データベースの構築を通じ、起業家層の底辺を拡げ、また 併せて創業間もないベンチャー企業や中小企業に新規事業展開を図るに必要な"人の情報"や資 金面及びその他の支援情報を提供することで、わが国における新規事業創出の促進を図る。

支援機関側にはより有効な支援措置の照会や、被支援者の発掘に積極的に利用してもらいたい。 また、地方自治体をはじめ全国各地の関連支援機関に対して積極的な提案活動を行うことで、支 援施策の一環としてデータベースを利用した支援事業への取り組みが促進されると期待される。



7. 高効率化先進材料ファクトデータベースのパッケージ化

脚次世代金属·複合材料研究開発協会



7. 高効率化先進材料ファクトデータベースのパッケージ化

7.1 概要

最近、コンピュータの使い易さに対する要求が非常に高くなってきており、コンピュータ同士の相互依存による迅速で効率的な対応処理が求められる。このためのデータベース(DB)を利用する補完技術として、大容量電子媒体(CD-R等)に大量データの保管整理し利用することが模索されている。ここでは先端産業(航空、宇宙、エネルギー等)に必要な有機、金属及びセラミックス等の先進複合材料に関する論文集やファクトデータを、その検索ソフトやアプリケーションソフトおよび利用マニュアルと共にマルチメディア化し、CD-Rにワンパッケージ化して利用する技術の開発を行った。

7.1.1 目 的

- 1) 当協会が管理する先進複合材料関連研究情報をより使い易いデータベース(DB)へと進展させ、当該業界におけるその利用普及を図る。
- 2)特にファクトDBを利用する場合、構造が輻輳しているので、GUIおよびヘルプシステムを充実し、当該DBを扱い易くする。
- 3) これらのコンテンツをHTMLにより記述し、CD-Rにパッケージ化し、利用環境の拡大を図ると共にインターネットWWWサーバでの情報発信が出来るようにする。

7.1.2 実施経緯

当該事業は、当協会内の主任研究員の下に専任研究員1名を置き、委員会形式の審議機関 (平成7年度ACMDB委員会)を設けた体制で実施した。委員には当該テーマが関与する材料DB関連の大学および国立研究機関研究者、ソフトシステム大学研究者等4人にお願いした。システム仕様設定のための事前検証作業は専門技術者に依頼し、仕様設定後、プログラムの作成、データ入力と変換作業等の一部は外注作業とした。

ROM化パッケージ実施前後に、当該ACMDB委員会委員によるDB動作の評価を願った。

7.2 実施結果の概要

当該DBデータとして、当協会が所有する先進複合材料既存データ 4,171件を市販RDBソフトにデータ変換移植入力した他、先進材料開発に係わる最新論文50件の情報データを新規入力した。

また、上記データを利用するアプリケーションとして、索引メニュー、DBアクセスプログラム、SQL検索、統計解析/単位変換、ヘルプシステム等のプログラムを開発して付加した。

更に、これらのデータとアプリケーションをCD-Rへパッケージ化する前にWWWブラウザを利用して運用状況を評価・確認した。

7.2.1 システム構成とファイルの作成

図1に今回制作したRIMCOF DBの全体構成を示す。また、図2にこのDBのホームページ、図3にDB選択メニューを示す。

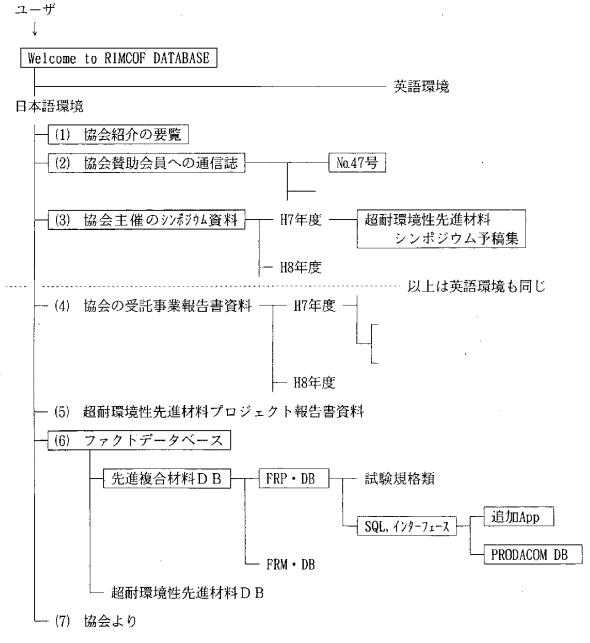


図1 RIMCOFデータベース全体構成(注:□部は今回試行部)

WELCOME TO RIMCOF DATABASE



- English
- Japanese

R&D Institute of Metals and Composites for Future Industries Bridgestone Toranomon Blg., 25-2 Toranomon 3-Chome Minato-ku, Tokyo 105 JAPAN E-Mail: rimcof@bnn-net.or.jp

TEL: (+81) 3 3459 6900 FAX: (+81) 3 3459 6911

図2 DBホームページ

(財) 次世代金属·複合材料研究開発協会

データベース内容 メニュー

- (1)協会紹介の要覧
- (2)協会賛助会員への通信誌
- (3) 協会主催のシンポジュウム資料

平成七年度 第六回超耐環境性先進材料シンポジュム予稿集

- (4)協会の受託事業報告書資料
- (5) 超耐環境性先進材料プロジェクト報告書資料
- (6) ファクトデータベース先進複合材料DBプラスチック基FRP・DB金属基FRM・DB

超耐環境性先進材料DB

(7)協会より

図3 DBメニュー

7.2.2 プログラムの作成

(1) httpを用いたデータの送受信とデータベースの駆動アクセスプログラム

クライアント側の webブラウザからhttpに基づいてドキュメントの要求が webサーバに送られると、サーバ側からデータベースの検索のための初期設定に必要なHTMLドキュメントが送信される。データベースへのアクセスは、はじめにこの設定用のページにより表示や単位などの設定を行い、データベースアクセスのページより httpdを介して VisualBasicで作成された実行ファイルを起動して、データベースへのアクセスが行われる。検索や設定の結果も種々の実行ファイルの出力が httpdを介して、クライアントに送信される。

WWWブラウザよりhttpdへ接続すると、まず一連の作業手順などを含んだホームページが表示される。このページより、概要説明、表示項目の設定、表示単位の設定、データベーステーブルのリレーションの表示、一般検索、SQL検索のページに移動する。このとき、サーバ側には検索条件のファイルが作成される。

SQL検索のページでは、SQLにより記述された検索条件をそのまま実行ファイルの入力として、単位設定ファイルを読み込み検索結果を表示する。また、絞り込み検索のため一つ前の入力条件はそのままSQLの入力フィールドに表示されるため、これを編集することにより柔軟に絞り込み検索を行うことができる。

データベースアクセスのアプリケーションは、VisualBasic により作成されている。一般 検索もSQL検索も同様に、検索条件ファイルを基にACCESSを用いて検索が行われ、検索結 果は VisualBasicで作成された実行プログラムによりHTML形式でクライアントに送信される。

(2) 統計解析/単位変換プログラム

構造材料を用いて機械構造物を設計する場合、最も重要な点は設計における材料強度の許容値の設定である。これが低いと構造は少しの過荷重で破損し、高すぎると破損しない代わりに過剰設計となり、航空機や宇宙構造あるいは自動車などでは重量増加が極めて大きな影響を及ぼす。このため、材料強度の許容値の決定は通常、材料試験で得られた多くのデータを基に統計解析を行って決定される。

許容応力の基準としてA値、B値、およびS値が以下のように決められている。

A値:母集団の99%の値が95%の信頼水準で入る値

B値:母集団の90%の値が95%の信頼水準で入る値

S値:規格や仕様書で定められた最低試験値

ここではこうした材料許容値を求めるために、次の統計解析を行う。

- 1) 基礎統計量(平均値,標準偏差,変動係数)の算出
- 2)2母数ワイブル分布への当てはめを行い、形状係数αおよび尺度母数βを求める。これ

らの母数の推定は線形法による。

- 3) 正規分布N(μ , σ 2) への当てはめを行う。
- 4) 各分布の適合度検定を行う。
- 5) A, B基準値の算出
- 6) ワイブル確率スケールを用いてデータのプロットを行う。
- 7) 正規確率スケールを用いてデータのプロットを行う。

ここではこのためのプログラムを VisualBasicを用いて開発する。このとき,入力データとして強度特性に関するサブファイルが必要であり,web のブラウザから統計解析に必要な強度特性値を指定できるようにする。これにより検索されたデータにおける指定強度値のサブファイルが生成され,統計解析の起動を指定すると,そのファイルを基に上で述べた統計解析を実施し,結果をブラウザに戻す。結果は数値のみでなく,2種類の統計解析にかかわるグラフも gifファイルとしてクライアント側に送信され,ユーザにとって利用価値は極めて高い。

(3) データ変換プログラム

すでに完成している複合材料特性データベース(PRODACOM: PROperty DAtabse of COmposite Materials)はdBASE III Plusを用いて開発されたPC用のデータベースである。データと検索および出力のプログラムはすでにフロッピーディスクにて日本規格協会から市販されている。しかしながら、このデータベースを用いるには別途 dBASEコンパチブルのデータベースソフトを購入する必要があり、これが欠点となって、これまで広く利用されてこなかった。今回開発する web上のデータベースはクライアント側にはブラウザ以外何も必要なく、こうした欠点が取り除かれた。

今回の開発ではデータベースにACCESSを用いている。その理由は、dBASE が標準的なデータベース問合せ言語である SQLをサポートしていないからである。SQL が利用できない場合にはユーザの期待する検索をすべて当該データベースのマクロ言語でプログラム化しなければ webのクライアント側から利用することはできない。そこで、データベースファイルは、これまでに dBASEで作成されたデータベースファイルをACCESSにアタッチした。

SQL を用いない GUIho方式の一般検索では全てのテーブルのリレーションを持ったテーブルを検索するが、全てのテーブル (15個のテーブル) のリレーションをSQLにて記述するとSQL文が複雑になり、VisualBasic でもエラーが生じる。そこで、15個のテーブルを2つの大きなテーブルにわけ、その2つのテーブルにリレーションを持たせたものに対して検索を行う。この2つのテーブルは dBASEの15のテーブルをACCESSでアタッチし、ACCESS上で新たな2つのテーブルを作成し、それを dBASE形式でエクスポートし、それをアタッチして

用いる。なお、SQL を用いた検索では直接 dBASEの15のテーブルをアタッチして用いる。このため、17のテーブルを用いている。

これら一連の作業を行うためのデータ変換プログラムを作成した。

(4) GUIを基本とするクライアントインターフェースプログラム

クライアント側ではできる限りマニュアルフリーですべての作業が行えるように、GUIを用いてユーザの負担を軽減した。GUI部分は、表示項目の設定、検索条件の設定、および単位の変換である。

表示項目の設定ページでは、データベースファイルの全てのテーブルの全てのフィールドが表示される。各テーブル、各フィールドには、チェックボックスがつけられ、そのチェックボックスをクライアントからマウスでチェックし、ページの最後の表示設定ボタンを押すことにより、表示項目の設定が行われる。

検索条件の設定は6条件のANDを3組設定できるようにした。これでも多すぎると思われるが、最大限考慮して決定した。3組にした理由は、本データベースが基本的に素材、試験、および特性値の3組から構成されているためである。

データベースファイルの検索結果の表示を行うとき単位変換を行うものについては単位の 設定が行えるようにした。単位ごとに変換対象となるフィールドが表示され、一つの変換対 象に対して表示される一つの単位を選ぶことができる。

なお、こうしたGUI方式の検索とは別に SQL式をそのまま記述する検索も可能なようにした。これは、ユーザの多様な選好に応えることを目的としている。

(5) ヘルプシステム

PCの Windows環境ではヘルプシステムが構築可能である。ヘルプシステムはヘルプコンパイラ,ホットスポットエディタ,およびマルチプルリゾリューションビットマップコンパイラなどのツールを用いて構築する。しかしながら、このヘルプシステムは同一マシンで動作するアプリケーションに有効であり、web システムでは有効ではない。このため、ここでのデータベースではCD-Rによるスタンドアロンシステムインターネットを用いたwebシステムでは別のヘルプシステムを考えなければならない。

スタンドアロンシステムでは検索用のアプリケーションに Windowsのヘルプシステムを直接組み込む。この場合にはヘルプ情報を構造化し、インデックスをつけ、リンク情報を持たせてハイパーテキストとし、その文書をヘルプコンパイラでコンパイルすることによってヘルプシステムが完成する。一方、web システムでは Windowsのヘルプシステムは利用できないため、HTMLドキュメントにヘルプ情報へのリンクを組み込むことによってヘルプシステムが実現する。

しかしながら、こうした2種類のヘルプシステムを実現することは労力の点からも、ヘルプ情報の一貫性の維持という点からも問題が多い。幸い、webのブラウザは本来的にハイパーテキストとなっているため、スタンドアロンシステムでもHTMLドキュメント上でヘルプシステムを考え、同一の方法とした。

ユーザは操作上必要な情報をマウスのクリックによって容易に取り出すことができるよう に設計した。

7.2.3 CD-Rの試作

CD-Rへの書き込みは、ISO9660 フォーマットで行うので、書き込み対象データを一度ハードディスクに書き込み可能データ(トラックイメージ)として展開している。普通この段階で高速化シミュレーションを行い、最適なファイル配置をするためのトラックイメージ修正を行うが、ここではHDD記録ソート順にDisc at once方式で直ちに書き込んだ。この方式はディスクの最内周から外周に向かってデータを一気に書き込む方式であり、大量生産方式と似ている。

また書き込みエラー等を避けるため、頻繁に割り込みがかかるソフトを起動しないようにした。生ディスク(書き込み用のブランクディスク)には故意に低価格品を採用し、当該書き込み技術の難易度把握を試みた。

上記要領での試作テスト結果ではファイル呼び出し速度等の実用上問題点は見あたらなかった。

次頁図4に今回テストに用いた装置全体図と仕様概略を示す。

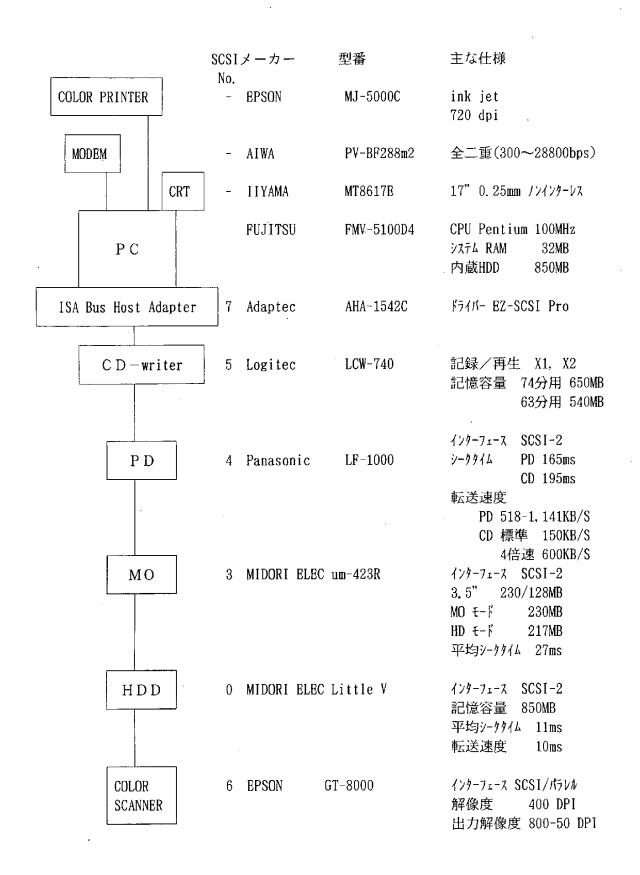


図4 テスト装置全体図と仕様

7.2.4 CD-RでのDB運用評価

- 1)特定の指定材料データに対してはその詳細資料を抽出できた。また、データ値範囲指定の データ群も抽出した。いずれもインターネットのサーバでのマークアップ記述言語によるイ ンターフェースを通じてRDBを制御し実行した。
- 2) 統計処理や分布の検定等の統計処理判定の付加ソフトにより、異常値の判定ができる。
- 3) 実験ファクトデータに必要な図面・写真及び主要特性値結果グラフ等の検索や表示が可能 となった。
- 4) 紙媒体主体マニュアルでなくオンライン方式とし、これにより DBの利便性と効果対コストが向上した。
- 5) データ様式は国際機関等の規格・基準類の提案に沿い、単位系の変更によるデータ換算表示ができた。
- 6) この ROMでインターネットの研究情報発信ニーズへの対応が容易となった。
- 7) このパッケージ化したオールインワン材料 DBの成果は、今後の全ての複合材料分野での DB構築手法に適用できるので、DBの活用・普及に大きく貢献する。

7.3 今後の課題

7.3.1 論文資料等のHTMLの記述過程

HTMLの記述は下記の手順で行った。

- 1) 使い慣れたワープロ(一太郎)で記述・保管した。
- 2) また、テキスト形式 (XXX, TXT) でも保管する。
- 3) ワープロまたはエディターの検索置換機能で参考文献番号を検索し、HTML言語に置換記入する。また、同様にしてFig., Table、Photo、等のリンク関連HTMLを検索置換記入、FONTのSIZEの検索置換記入をする。
- 4) netscape 1.1版で表示テストし、不都合箇所は txtファイルで訂正し再テストする。
- 5) 画像入力をスキャナーで取込 gif保存する。
- 6) netscapeで組み合わせテストした。
- 7) HTMLエディターで txt文を呼び込みHTML文に置換(xxx.htm) 保存した。

上記過程での課題・問題点およびノウハウを3種類に分離し把握した。

7.3.2 CD-R化について

計画当初一番懸念されたCD-R化について慎重に作業した結果であろうか、特に大きな問題となることはなかった。

今後、今回テストしたDisc at once書き込み方式だけでなく Track at once (マルチセッション)でフォーマットの組み合わせで書き込みを行うことを考慮することが必要である。ここでセッションとは一回の書き込み操作で書かれる部分をいうが、マルチセッションとは複数回の追加書き込みを行う技術をいう。

7.3.3 インタラクティブWWWサーバの運用とRDBMS

PCの Windows環境を利用して WWWによるデータベースのサーバに関しては基本的に必要な性能を発揮させるプログラムを開発することができた。

(1) 本システムの利点

- 1) クライアント側に webブラウザ以外, 何のソフトウエアも必要ない。
- 2) データベースの更新や改良などのメンテナンスはサーバ側で随時可能である。これはデータベースそのものをCD-Rなどでリリースする場合に比べてユーザへのバージョンアップなどのメンテナンスをする必要がなく、また、ユーザにとっても損失が少なく、極めて利点が多い。
- 3) GUIを多く用い、しかも統計解析結果の図まで提供できることにより、極めてユーザフレンドリである。

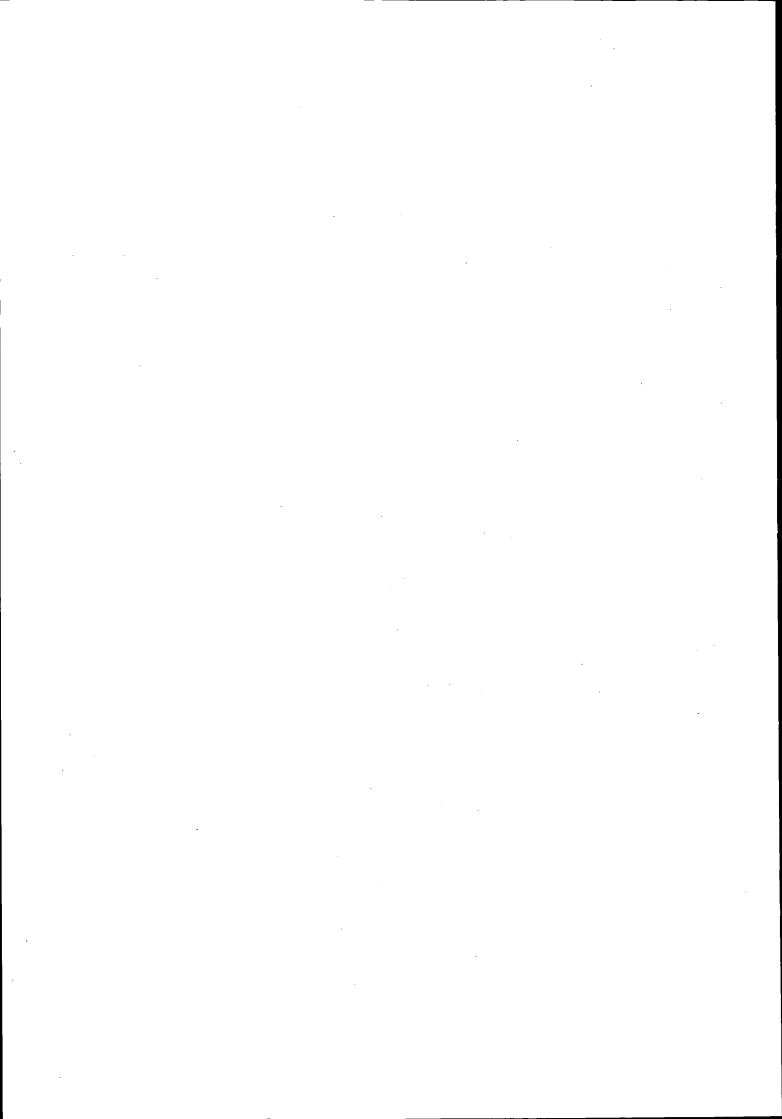
(2) 今後の検討事項

まず第一に、本システムではユーザ管理を行っていない。このため、データベースサーバ側で生じる状態、すなわちサーバ側に生成される検索条件ファイルと統計解析用の強度ファイルにはユーザのIDがない。したがって、あるユーザが本システムを利用している間に別のユーザが本システムをアクセスした場合に問題が生じる。これを解析する必要がある。

次に、ACCBSSの動作速度は若干遅い。これはACCBSSがかなり重いソフトウエアであるためであるが、これによってユーザ側の待ち時間が長く、特にリモートで用いている場合にはサーバの状態がわからず、不安要因となる。このため、高速のデータベースを用いるか、あるいは動作中に何らかの状態表示を行う必要がある。しかしながら、ACCBSS動作中の情報提供は容易でない。

最後に、本システムで主要な動作を制御しているプログラムはVisualBasic Ver3.0 (これは日本語版はない)で開発された。これはデータベースとのリンクができる始めてのヴァージョンであり、これを用いることによって webデータベースが可能になった。しかしながら、

これはACCESS Ver. 1.1にしか対応しておらず、今回の開発においてもACCESS Ver. 2.0との関係に問題が生じた。最近Ver. 4.0 が日本語版で発売され、これを用いることによりこうした問題は回避されると考えられる。



8. 阪神・淡路大震災の情報デジタル・アーカイブ

神戸マルチメディア・インターネット協議会

8. 阪神・淡路大震災の情報デジタル・アーカイブ

8.1 概要

8.1.1 目的

平成7年1月17日午前5時46分に阪神・淡路地域を襲った阪神・淡路大震災は最大震度7の 大規模な地震で、神戸を中心とした地域に潰滅的な打撃を与え、関東大震災以来の未曾有の被 害を市民生活及び文化に与えるとともに都市の繁栄を奪っていった。

その姿は50年前の終戦後の神戸を彷彿させる状況であり、戦後、市民・企業・行政が一体となって築いてきたまちが一瞬にして焼失していくさまをわれわれは映画を見るように目の当たりに目撃し、また永年住み慣れた住居および培ってきた文化を失った。

さらに被災者・行政ともに正確な情報が把握できずに不足したことおよび情報が伝達されないこと等が事態の収拾を遅れさせていたことは明らかである。

千人近い死者を出すなど大きな被害を受けた神戸市長田区および須磨区と山を越えた地区にある神戸市外国語大学との間では地下鉄,電話等の交通・情報手段が不通になり一切の情報の入手・伝達が不可能であったが、神戸市外国語大学に設置されたインターネットを通じて被災状況の発信,世界からの問い合わせ、援助・ボランティアの申し出等の情報の交流は行われいた。

そのような背景からライフラインとともに今後の都市の防災面の情報インフラ網の整備の見直しが認識され、新たな情報伝達手段としてのインターネットを中心としたコンピュータ・ネットワークの活用が今後の市民生活、都市づくりの基盤として欠かせないものであることが一般にも認識されるに至ったと思われる。

インターネットの普及、啓蒙および神戸の防災・復興の側面から情報インフラ網の構築を促進する目的で設立された神戸マルチメディア・インターネット協議会(委員長 北村新三 神戸大学工学部教授 設立:平成7年3月)では、このようなコンピュータの処理能力を活用して震災の記録のデータベース化を図るため、側データベース振興センターの委託を受けて阪神・淡路大震災の情報デジタル・アーカイブの構築を行った。

震災の発生当時から被災地の被災状況等について神戸市の広報課及び他の部局等で収集した膨大かつ貴重な写真・映像の記録は散逸するおそれが強いため、早急にその防止を図るとともに、20世紀最後の悲劇である阪神・淡路大震災のビジュアルな記録を、デジタル化することによりコンピュータにデータベースとして保存し、災害発生時の対応等の基礎的データベースとして活用できるよう構築を図ったものである。

デジタル化して保存するメリットは、容易に、かつ恒久的に保存が可能であり、またコンピュータ上でそれらの資料の加工が容易にかつ多様にできること、さらにインデックスの付与により目的別の検索が可能になる点などがあることは周知のことであろう。

このような大地震は事前に予想されたものでないため、また映像および記録写真等について も当然データベースの作成を目的として収集されているわけではない。

そのため、データベース作成本来の目的にそぐわない資料も含まれているが、震災の被災状況・復旧等にあわせて記録されていった事実の記録であり、今後とも引き続き整備を続けていくことが可能となるように作成に努めた。

このような経験を経た結果,災害に対応していくために,震災発生等から蓄積された映像・写真記録を散逸することなく,すべての映像・写真記録を保存していくことに務め、さらに今後の研究・調査等に生かせるよう分類を行い,まちづくり,救援活動の構築の一助になることを目的に作成したが,データベースとしてはすべての面で充実しているとはいえない点もあり、今後も引き続き一層の拡充につとめていくことが必要と思われる。

《兵庫県南部地震による被災状況》 (神戸市災害対策本部調査より)

(1) 概 要

① 発生日時 平成7年1月17日(火) 午前5時46分

② 震 源 地 兵庫県淡路島 北緯34.6° 東経135.0°

③ 震源の深さ 14km

④ 規 模 マグニチュード 7.2

⑤ 津 波 なし

⑥ 各地の震度 震度 6 神戸、洲本

震度 5 彦根,京都,豊岡

震度 4 福井, 敦賀, 岐阜, 四日市, 津, 上野, 奈良, 舞鶴, 大阪, 高野山, 和歌山, 姫路, 徳島, 鳥取, 境, 津山, 岡山, 高松, 多度津, 高知, 福山, 呉

(注 震度4以上のみ記載)

(2) 神戸市の被災概要

① 避難者数 236,899人

(586避難所, 1月24日最多)

② 火災発生件数 175件(1月17日~27日)

③ 焼損述べ面積 約82ヘクタール

8.1.2 実施内容

データベースの構築にあたっては、神戸市市長室広報相談部広報課(以下神戸市広報課という)、神戸市長田区役所まちづくり推進課(以下神戸市長田区という)および京都大学防災研究所、神戸大学工学部、神戸市外国語大学の協力を得て、神戸マルチメディア・インターネット協議会内に阪神・淡路大震災情報デジタル・アーカイブ研究会を設置し、情報デジタル・アーカイブにつき討議を重ねたが、全てのデータの把握が困難であり、基本的なの点も含めて試行錯誤の繰り返しであった。

(1) 入力データの資料

当初は、入力データとして神戸市広報課が大地震発生時から収集していた映像記録および記録写真を対象としていた。

入力対象の資料の量は、当初の予定では神戸市広報課の写真資料が約1万枚、映像資料が家庭用ビデオで約2~3時間余のテープであったが、その後神戸市長田区の写真資料が約7,000枚、最終的には実施結果の項で触れるがさらに増加した。

神戸市広報課の写真資料は、震災発生から被災地の現況のありのままの姿をとらえていったものであり記録写真として神戸市が市全体の記録として蓄積していくものであった。

また神戸市長田区の写真資料は震災当初から復興ボランティアとして協力していた京都大学防災研究所の指導・助言のもと、地図に対応した倒壊家屋および家屋の状況を撮影した資料であり、火災等による家屋の倒壊という被災規模の大きかった長田区の特性をあらわしている。

(2) 入力方法

① 写真資料

神戸市広報課および神戸市長田区で収録された写真資料はその保存形態が異なるためそれぞれにあわせて処理を行い、その処理状況を見ながら最終段階でアーカイブとしての整合性の確保および互換性についての検討を行うことにした。

神戸市広報課の資料はネガフィルムで保存されており、神戸市長田区の資料は紙焼きで保存されているこめ、スキャナ入力の方法が異なり、機種についてもそれぞれ使用機種が異なる。また記憶媒体についても前者はPD650MBを使用したが、後者はMO230MBを記憶媒体として利用した。

なお、プロトタイプとしてこれらのPD、MOに格納されたデータベースのうちから、 実施結果の項でも触れるがCD-ROMに収めていきプロトタイプを作成した。

② 映像資料

当初の入力の予定では大地震発生直後から2日間(1月17日~18日)の映像記録のデータ入力(約2~3時間程度)を予定していたが、事実経過として震災後の8月末までの映像記録を入力した。

その結果,すべての映像の収録時間は60時間超にわたり,スキャナ入力にあたり動画を 採用して入力しているため膨大な入力時間および記憶容量が必要となっている。

映像資料については、撮影段階から整理が充分にされており、写真資料のように整理の 側面での煩雑さは解消されていたが、技術的にはその膨大な量の多さとともにかなりの処 理時間を要した。

No.	映像記録入力データ(当初30分)	月日	地区	開始	終了
11	須磨高倉台より東側を望む	0117	須磨	0:00:00	0:01:03
12	堀池町,戸政町,妙法寺川公園	0117	須磨	0:01:03	0:05:10
13	山陽板宿駅,戎町の火災状況	0117	須磨	0:05:10	0:08:55
14	長田御屋敷通,西代通	0117	長田	0:08:55	0:10:50
15	JR新長田駅前,若松町	0117	長田	0:10:50	0:13:50
16	大正筋商店街,久保,駒栄,大橋町の火災	0117	長田	0:13:50	0:17:30
17	若松町,JR線高架上	0117	長田	0:17:30	0:21:10
18	神楽町,細田町,水笠通,川西通	0117	長田	0:21:10	0:23:10
19	山陽西代駅付近,大道通,長田区役所	0117	長田	0:23:10	0:27:10
110	一番町付近の火災	0117	長田	0:27:10	0:29:50

③ インデックスの付与について

神戸市広報課において収録した写真等につき、部外者を含む複数の撮影者に依頼していることおよび定点観測等の手法を利用していないため、大部分の写真資料が大地震の影響を受けて未整理な状況であり、その後も日々資料数が増加し整理に手間取るといった状況が連日続いた。

インデックスの付与については日本図書目録の手法によるが、日本においてはいわゆる マルチメディア関連の資料に対する標準的な目録規則が未だ十分に整理されておらず、さ らに今回のインデックスの付与は資料の外形式にとどまらず被写体および将来の二次使用 への対応を考慮して写真資料の撮影者のデータをも付加した。

④ インデックスとしてのキーワード

キーワードとして今回の大地震のアーカイブでは、通常考えられるキーワードとしての

時間軸を優先するとともに、GISに基づく地域軸および被写体としての対象(人、物)、写真・映像の撮影者等をキーワードとして検討し、可能な限りのキーワードによる分類に努めたが、現実に使用に耐えうるものかどうかについては今後の課題となり、またそれらの意見に対応して改善に努めていくことがアーカイブとしての価値を高めるとともに本来の防災の見地等からも要請されるものと考えている。

キーワードの選定については、いろいろ討議を行ったが、その点については、今後の課題の項で触れていく。

8.2 実施結果

8.2.1 プロトタイプの作成の指針

(1) 入力データの見直し及びデータの重要性について

研究会の意見として時間的な制約があるという背景のもとで入力資料については量的には満たされているが、資料の質的側面については現状で良いのかとの意見が提示された。

すなわち、神戸市広報課の写真資料はどちらかといえば報道写真のイメージであり、一方、神戸市長田区の写真資料は倒壊家屋に偏りすぎているが、阪神・淡路大震災はさまざまな方面に被害や影響をもたらしており、アーカイブとしては多くの側面からの写真資料の整備が必要であるとの見解が研究会の席上提起された。

この指摘は、アーカイブ構築の根幹にかかわる問題であるが、すでに事業として入力作業を 進めていたため、入力データの資料としてとりあえず神戸市広報課および神戸市長田区の写真 ・映像資料についてはすべての入力作業を継続して行い、上記の提案については徐々に付加し ていくことで対応することとされた。

本事業の目指すアーカイブは、データ入力について作成側として(研究会)の意図を反映せずに、入力されたデータをそれぞれの利用者の観点から必要である写真・映像資料が多量の資料の中から選択可能なように構築していくことが肝要であり、そこには写真の質として常に高水準なものは必要ではなく、逆に一般に広報されていない資料、あるいは数量的な点が価値を有するのではないかと検討を重ねて、データの重要性をそこに位置づけて構築した。

(2) プロトタイプ作成とデータ

写真資料および映像資料の膨大な量から、入力にきわめて時間がかかり、最終的にプロトタイプとして選択する対象および構成の決定にも時間が十分さけなかった。

また本事業が阪神・淡路大震災を対象にしたものであるため、写真資料についても研究会が プロトタイプ作成に必要とする資料が常にあるわけではなかった。 被災者サイドに立脚すれば、被災を受けた自宅、あるいは被災を受けた本人・家族の写真を 撮影されるのはいかに報道のため、記録等のためといわれても心情的には納得できない面があ るのは仕方なく、研究会としても今回の写真資料およびデータベース等の取り扱いについては 細心の心配りが必要と思われる。

被災者の救済等の参考になると考えられるのは、本来ならば被災者の避難所における生活の 日々の暮らしをカメラで追跡した資料等であろうが、震災の発生当時、避難所でそのような趣 旨でカメラ撮影等を行うことは被災者の心情を慮ると好ましいものとはいえないことは明白で あり、アーカイブとして不満な点も残るがこのような背景を考慮すると研究会としては許され ると考えている。

また神戸市の各部局では、大地震の発生後は、当然復旧業務が最優先の職務であり、本事業の対象である写真等の記録の収集は二次的な業務であり、さらにすべての被災現場等における 撮影等は時間的には、人員的には余裕がなかったと思われる。

8.2.2 プロトタイプの作成

それぞれの写真資料は、記憶媒体としてPD、MOの形式で格納され、それぞれからプロトタイプとして必要な資料の引出しを行った。

① 写真資料について

プロトタイプとしてCD-ROMに格納する際の考え方は、被災地神戸のなかで神戸を象徴する地域を中心に地図情報にプロットし、震災前、震災直後、その後の経過の記録を先に述べたインデックスで検索可能となるように作成した。

なお、対象地点は神戸三宮を中心とするオフィス街(5ポイント)を中心におき、ポートアイランド、六甲アイランド、東灘区の酒造業、灘区のJR六甲道駅周辺、長田区のJR鷹取駅周辺、市営地下鉄の駅舎、長田区の民家等を取り上げた。

今後,利用に際して評価を受けながら改善に努めていくが、平成8年1月17日から一部の写真資料について、それぞれ単体の資料として神戸市のポームページの素材としてインターネットで発信を行った。

平成8年度からは、神戸市のポームベージを利用して、より多くの関係者のアクセスにこた えることが可能なように編集を行っていくことによりアーカイブ構築の目的の一部が達成され るものと考えている。

②映像資料について

映像資料については、先に記載した映像記録のとおり当初からかなり詳しく整理されており、資料の整理については、比較的容易であったが、入力の取扱いについては技術的に労力を

費やした。

また量的にきわめて多いためCD-ROMへ入力する対象の選別に苦労を要し、またCD-ROMの容量の制約から動画としてどの程度の分量を入力するかのバランスの点の検討もあった。

プロトタイプとしては撮影地点から動画で数十秒ごとに入力を行い, その撮影地点のサーチ は神戸の俯瞰図から選択可能な形式になっている。

写真記録のプロトタイプと異なり、インデックスはGISに基づく検索手法となっているが、 先に述べたようにCD-ROMには容量の関係から限られた情報のみの入力を行っている。

8.3 今後の課題と展望

8.3.1 データベース作成上の問題点

(1) データ収集および使用について

今回の事業の構築にあたっては当初から神戸市の所有する資料を収集・使用したのであるが、いまだそれぞれの部局には貴重な資料が存在しているのは明らかであり、また本来の目的である散逸の防止が十分に達成されてるとはいいがたい。

さらに事業の進捗にあわせて、民間を含むその他の団体等からデータ提供の申し出が数多く寄せられたが、申し出には応えられずに貴重な写真・映像記録が漏れている点は研究会としては心残りなところを感じている。

これらの点については、当該データの使用につき①著作権、②肖像権、③所有権等の法的な問題が未だ明確にされていない点を研究会としては危惧した面もあり、これは今後、データの入力を拡大する際に検討することになるであろう。

このような人類にとって未曾有の阪神・淡路大震災のデータベース化の趣旨から,法的問題もクリアされるとの思いもあり各権者に対して交渉の余地はあると考えられたが,現実には交渉するにつき時間的ゆとりがなかったことから取り組むことができなかった。

また、神戸市(広報課、長田区、交通局)の記録写真等といえども被災者の心情を配慮してデータとして採択するかを検討することも考えられたが、研究会としてはアーカイブの趣旨から撮影された記録についてはすべて入力処理を行った。

今後の技術革新に伴いこのようなビジュアルな資料を直ちにコンピュータに入力され、かつ 容易にインターネット等で情報が流通していく時代になり、情報を提供する側の倫理観に委 ねられる場合が多くなるが、そのような良識に依存するのではなく、客観的かつ合理的な規範の制定が望まれる。

(2) データ数量について

本研究会では、入力データについての考え方としては、①すべてのデータの入力を研究会として判断を加えることなく入力を行う、あるいは、②データの選別を行ったのち入力を行うとの考え方があったが、後者については選別の判断基準を研究会でどう定めるのが妥当かとの問題が解決されず研究会での判断をさけて前者①の手法により入力を行った。

しかし、現実的に処理を行っていくと資料としては膨大な量となり、業務的にはかなりの 労力を費やし、また個別の資料を確認すると首をひねるような資料も存在した。

しかし、研究会では利用者にその価値を判断してもらうことがアーカイブの姿勢であると のスタンスに立脚し、データ量の確保をコンセプトとしていたため割愛することなく入力作 業に重点をおき処理を行ったが、それ故より広範なデータの入力に取り組みたかったと考え ている。

最終的に、入力データのインデックス付けを処理した後、ソート・分類を行う予定であったが、各PDごとに入力されていたため、この度の事業では行えなかった。

なお、今後、追加データの取扱いと併せて、格納方法につき検討をしていくことが必要で あると研究会では考えている。

(3) 検索におけるインデックスの付与作業について

データの入力にほとんどの時間と労力を費やした結果,インデックスの付与が,現実には パラレルに進むことなく,大幅に作業工程が遅滞していった。

インデックスのうち、時系列のキーワードはデータ入力に併せて入力が可能であったが、 地区(行政区)、被写体の名称(建物、人等)および事項別インデックス等については、それぞれの資料に対する調査が並行処理で行えなかったため、同時入力が不可能であった。

また、インデックスの具体的な項目も、大地震という特殊な事例から、専門用語等が使用され、専門家向きの用語がよいのか、一般向けの用語がよいのかの決定にかなりの時を費やしながら検討を行ったが容易に結論には至らなかった。

例えば、ポートアイランド等の人工島で見受けられた「液状化現象」は、テレビ等の画像で見ると理解できるが用語として一般的に認知されているのか、また支援イベント液状化現象という用語を使用することは適格かなどがあげられよう。

さらに、被災者の立場を考慮して制約を受ける用語もある、被災者、避難者、避難所等は 報道等で一般的に用語として使用されているが、事実上そうであっても使いにくいのではな いかなどと研究会では取り上げられていた。

時系列および地域的な用語としては、日付、行政区を使用する等、比較的制約を受けることなく定まっていったのであるが、とくに現象面、事項面の用語でいろいろ研究会が過度に

神経質になったのか決定までに時を要した。

このような討議の結果、インデックスとして時系列の用語として日付、地区を特定する用語として行政区の名称およびポートアイランド等の一般的に認知されている名称を使用するとともに、建物の名称についても一般的に認知されていると推測される名称を基準にインデックスの付与を行った。

現象面、事項別の用語としては、同じように一般的な用語を使用してできるだけ利用者の 便を考慮したが、さらにフリータームについてもあわせて付与して検索項目を拡大する形式 をとりついれた。

8.3.2 今後のデータベースの利用・提供について

最終的に、阪神・淡路大震災の情報デジタル・アーカイブとして、神戸市広報課及び神戸市 長田区、神戸市交通局の写真記録として入力データは数量的には、統計で約2万5千を超え、 また映像ビデオとしても総量で約60時間をこえた入力データとなっている。

ただし、これが全ての資料でないこと、またインデックス等の付与について十分に機能する 段階まで至っていない点等につき今後引き続き検討を加えていくことが必要である。

(1) 入力データの拡充および改善について

今後は、このアーカイブの入力データとして新たなデータの入力を引き続き行い、充実を 図っていることが必要であり、また検索について技術的により改善を行わなければならない し、また地理情報とのマッチングを一層すすめていかなければならない。

今回の、委託事業ではデータのデジタル入力に重点を置いたが、アーカイブの本来の目的である、資料散逸の防止、今後の防災およびまちづくり、安全な市民生活を考慮するとデータ数および使用方法等につき、より改善が図られなければ目的の達成には機能しない部分もあると思われる。

今後は一層のデータ資料の入力を図り、検索機能を充実するとともに、まちづくり、住宅 建設、震災等の発生時の対応等につき今回のアーカイブの各分野の専門家の方々に活用して 頂くことが本事業の将来的な価値であろう。

(2) データベースの利用と提供について

神戸マルチメディア・インターネット協議会としては施設等を所有しているわけではなく、 あくまでもソフト作成の組織であるため、構築されたデータベース等は基礎的なものとして しかるべき公共的性格を有する機関へ付与していくことを検討している。

なお、その際の条件としては引き続きデータベースの拡充・改善を行い、また多くの人に 資料の閲覧および参照ができるように扱うことを依頼したい。 また、(1)でのべたようにその他の資料の提供・入力について各機関に働きかけ、長期的な 視野での構築を考えいくことが、被災地神戸としての貴務であると研究会では考えている。 9. Mosaicの利用によるマルチメディアデータベース 検索システムの構築

•		

9 Mosaicの利用による

マルチメディアデータベース検索システムの構築

9.1 概要

9.1.1 目的

インターネットの急速な発展に伴い,既存のデータベースシステムや各種のマルチメディア 情報を付加価値として,インターネット上に情報発信することを望む企業や団体等が増加して きている。

インターネットは、高度な情報提供サービスの創造という点で新ビジネスの機会を増大し、 マルチメディア市場を拡大していく方向にあり、社会をより豊かなものにするために、利用者 ニーズを満たす新ビジネスを具体化できる環境整備を行うことによって、ニーズと共に成長し ている可能性を秘めている。

今後も多くの企業や団体、教育機関等がインターネットへの情報発信、インターネットを有効活用したニュービジネスへの参画が加速化してくると予想される。しかし、既存のデータベースシステムやこれからデータベースシステムを構築し、インターネット環境でこれらのデータベースを有効活用するための手段が未整備なのが現状である。

今回の委託課題である「Mosaicの利用によるマルチメディアデータベース検索システムの構築」は、データベースシステムとマルチメディア情報とをリンクし、短期間に低コストでインターネット上に情報発信するためのシステム構築手段を確立することに主眼をおき、インターネットとデータベースとの融合を押し進める技術をより発展させることを目的としている。

9.1.2 実施内容

マルチメディアデータベース構築システムは、WWW(World Wide Web)サーバとデータベースエンジンとをリンクし、異なるデータベース(普及度の高いORACLE、informixを対象とした)に格納されている情報を扱うことが可能なミドルウェアを開発することにより、このミドルウェアを核として、音声、画像等のマルチメディア情報をデータベース化し、利用者にデータベースの違いを意識させることなくインターネットからの容易な検索が可能なシステムである。

今回は、このようなトータル的システムの実現に向けて開発を実施した。図9-1にシステムイメージを示し、以下に本システムの特徴を示す。

(1) WWW技術を使用した情報発信システムを活用することで「利用者を限定しない開放的な

検索環境」を実現できる。

- (2) WWWブラウザをGUI (Graphical User Interface) に使用することにより「マルチメディアによるユーザインターフェース」を実現できると共に、「使用マシンを限定しないオープンシステム環境」を実現することができる。
- (3) 集合光磁気ディスク装置制御機能等「マルチメディア情報の統合管理」が実現できる。

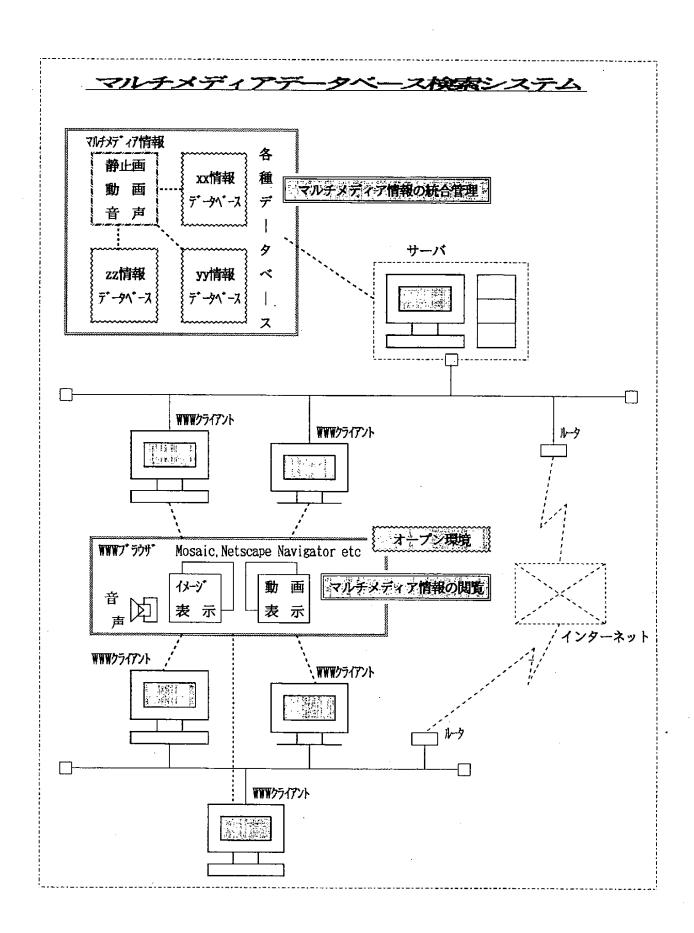


図 9 - 1 システムイメージ図

9.2 実施結果

9.2.1 マルチメディアデータベース検索システムの構築

インターネットに接続されたプラットフォーム上に、音声、静止画、動画、ハイパーテキスト等のマルチメディア情報のデータベースを構築しインターネットの情報検索システムであるWWWとリンクするミドルウェアの開発を行った。開発したミドルウェアは、図9-2の機能構成図に示すように「HTML生成機能」、「SQL生成機能」、「マルチメディア情報管理機能」から構成される。なお、マルチメディア情報の参照には、インターネットで急速に普及しているMosaicやNetscape Navigator等のWWWブラウザを、WWWサーバはCERN httpd (Hyper Text Transfer Protocol Daemon) 3.0を使用した。また、データベースエンジンは、ORACLEおよびinformixを対象とし、ミドルウェアとWWWサーバとのインターフェースは、CGI (Common Gateway Interface)を使用してシステムの構築を行った。このミドルウェアは、同一ユーザで、複数のデータベースエンジンを使用してシステム構築を行うことは少ないと考えられたため、ORACLE用、informix用の二つのミドルウェアに分けて開発を行った。

今回開発したミドルウェアは、WWWクライアントからWWWサーバにアクセスがあると検索条件を設定するための画面情報(HTML)を自動生成し(「HTML生成機能」)クライアントに送信する。利用者は、検索条件を設定しサーバに対し検索のリクエストを送信すると、サーバはこのリクエストをミドルウェアに通知する。ミドルウェアは、このリクエストを解析してSQL文を生成し(「SQL生成機能」)データベースの検索を行い検索結果画面情報(HTML)を編集し(「HTML生成機能」)、サーバ経由で検索結果をクライアントに返却する。また、検索結果とリンクするマルチメディア情報があれば、該当する情報をクライアントに返却する。(「マルチメディア情報管理機能」)なお、「マルチメディア情報管理機能」は、ミドルウェアとしての機能以外にデータベース登録機能(各種マルチメディア情報をデータベースとリンク)を有する。

以上のように、マルチメディア情報のデータベース化とミドルウェアの開発により、トータル的なマルチメディアデータベース検索システムを構築することができた。

<注>CERN……欧州素粒子物理学研究所(European Laboratory for Particle Physics)

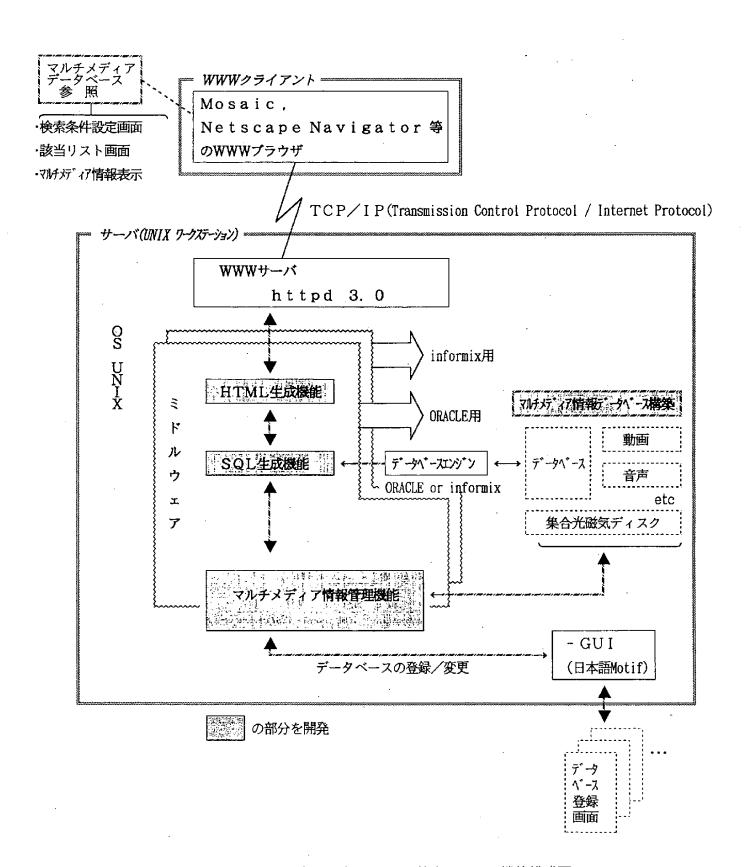


図9-2 マルチメディアデータベース検索システム機能構成図

9.2.2 HTML生成機能の開発

(1) 機能概要

HTMLは、ファイルにWWW用のリンク(関連付け)を指定するための、ハイパーテキスト形式を記述する言語であり、タグ付けまたはマークアップ方式で記述される。ハイパーテキストとは、情報アクセスの単位となるオブジェクト間の参照関係をアンカーで結んで作られる情報表現形態である。

なお、WWWブラウザは、タグ以外の部分を画面表示する。タグの部分は、「表示の方法」を指示するということで、例えば画面上にタイトルを表示する場合は、<TITLE>と
</TITLE>の間に記述された文字が表示される。

今回開発したマルチメディアデータベース検索システムにおいて「HTML生成機能」の対象としたページ(画面)は、「検索条件設定画面」と「該当リスト画面」の二画面である。この二画面を表示するために必要なHTMLを、できるだけ汎用的に自動生成するメカニズムを検討し当機能の開発を行った。なお、HTMLは表操作等の拡張機能が容易に扱える、HTMLレベル3.0を使用した。

「HTML生成機能」の実現にあたり、各種管理ファイルを使用して、画面上で流動的に変更される可能性がある情報(ボタン表示の際のイメージや表示項目当)を管理し、この情報をHTMLで埋め込む方式でHTMLの自動生成を実現する。

なお、各種管理ファイルは、以下に示す3ファイルでありエディタにて容易に変更することが可能である。

① 動作環境情報ファイル

画面に表示する背景等のイメージが格納されているファイルのファイル名、ディレクトリ名や本システムが動作するために必要な情報を管理するファイルである。

③ テーブル情報ファイル

データベースの項目と日本語名に関する情報を管理するファイルであり、データベース項目が変更された場合でも柔軟に対応できるように考慮している。

③ 初期値情報ファイル

各プルダウンメニューに表示する項目の初期値を管理するファイルであり、表示項目 の初期値が変更されても柔軟に対応できるように工夫している。

図9-3にテーブ情報ファイルのファイルフォーマットと、どのようにHTMLを生成するかを例で示す。

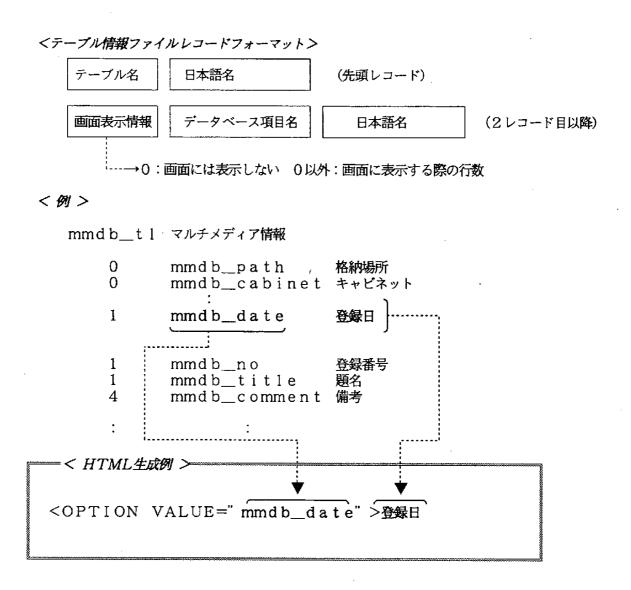


図9-3 テーブル情報ファイル使用例

<OPTION>タグは、画面に表示されるメニュー項目を<SELECT>と
</SELECT>タグの間に定義するタグで、レコードフォーマットに示す画面表示情報
"0"(画面に表示しない項目)以外の全項目をメニュー対象項目等としてHTMLの生成に使用する。

HTML生成例で示す…

<OPTION VALUE="mmdb_date>登録日

…の"登録日"(レコードフォーマットに示す日本語名)が画面上に表示され、VALUE=" $mmdb_date$ "がメニューより当該項目を選択した場合に、後述する「SQL生成機能」に渡され、SQL文自動生成の際の情報として使用される。なお $mmdb_da$ te 等の情報はデータベース項目名と一致している。

(2) 「検索条件設定画面」用HTML自動生成処理の流れ

図 9 -4 に「検索条件設定画面」用HTML生成処理の流れを示し、図 9 -5 にHTML 生成例を示す。

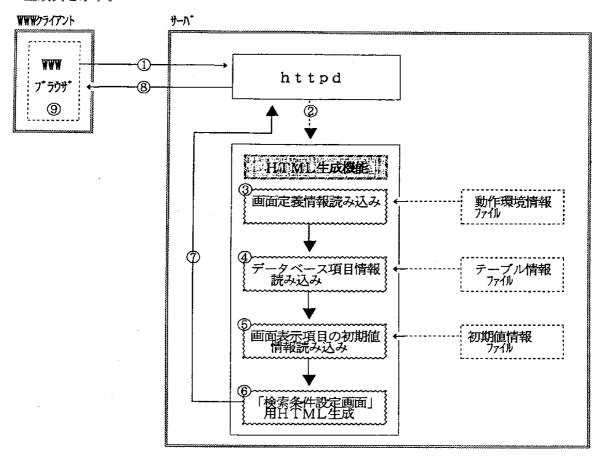


図9-4 「検索条件設定画面」用HTML生成時の処理の流れ

- ① ホームページより「検索条件設定画面」が選択されると、WWWブラウザよりhttpdに対しURL (Uniform Resource Locator)で「検索条件設定画面」用のHTMLを生成する「HTML生成機能」(プロセス)を起動する要求が出される。
- ② httpdより「HTML生成機能」(プロセス)が起動される。
- ③ 起動した「HTML生成機能」は、動作環境情報ファイルより本機能が動作するのに必要な情報や、画面の表示色等の情報を読み込む。
- ④ 画面に表示するデータベース項目をテーブル情報ファイルより読み込む。
- ⑤ 初期値情報ファイルよりプルダウンメニューの初期値に関する情報を読み込む。
- ⑥ ③~⑤の取得した情報を基に、「検索条件設計画面」用HTMLを自動生成する。
- ⑦ 「HTML生成機能」より標準出力でhttpdに生成したHTMLを通知する。
- ⑧ httpdよりWWWブラウザに「検索条件設定画面」用HTMLを通知する。
- ⑨ WWWブラウザにより「検索条件設定画面」が表示される。

<HTML生成イメージ>

<pre><html> <head><title> MultiMedia DB For WWW </title></head></html></pre> <pre></pre> <pre></pre> <pre></pre> <pre>ACKGROUND="/mmdb/env/images/paper gif" TEXT="#000000"> </pre> <pre></pre> <pre></pre> <pre> </pre> <pre> </pre> <pre> </pre> <pre>TEXT="#000000"></pre> <pre> </pre>	
〈H2〉 〈IMG SRC="/mmdb/env/images/logo.gif" ALIGN="bottom">〈BR〉 検索条件の設定 〈HR〉 〈HR〉	
<form action="/cgi-bin/i_umdb_answer" method="POST"> <table border="5" cellpadding="0"></table></form>	
<td><input src="/mmdb/env/images/search.gif" type="image"/></td>	<input src="/mmdb/env/images/search.gif" type="image"/>

通常検索
```  (   (   ((
《TD> 《OPTION VALUE="imidb_type">ファイル種別 《OPTION VALUE="imidb_date">登録音 《OPTION VALUE="imidb_date">登録音 《OPTION VALUE="imidb_no">登録番号 《/SELECT>
『midb_type"や"midb_date"等のデー外、- 2項目名
および日本語名は、「デーブル情報ファル」より情報
を取得して展開する。
を含む(~X~) で始まる(X~) で終わる(~X) である(X=)
より小さい (X<)
---
```  SELECT NAME="ol"> 表示しない
ファイル種別k/TH>

9.2.3 SQL生成機能の開発

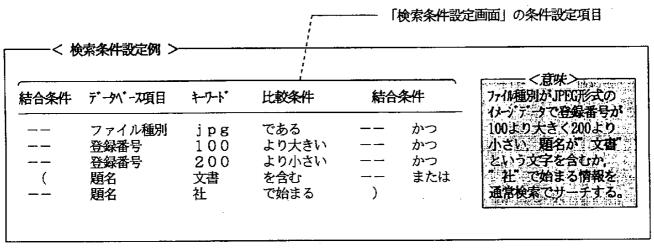
(1) 機能概要

画面より設定された検索条件をhttpd経由で受け取り解析し、SQL文を自動生成してデータベースの検索を行い、データベース検索結果をWWWブラウザに返却するまでの一連の処理を実現するのが「SQL生成機能」である。

なお、データベースエンジンは、ORACLEおよびinformixを対象としているため、データベースエンジンに合わせたSQL文を生成する。

(2) SQL生成例

図9-6にSQL文を自動生成した場合の例を示す。





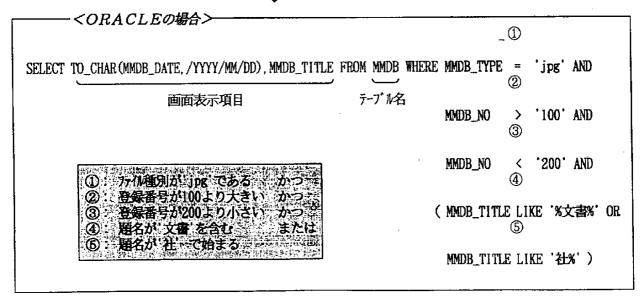


図9-6 SQL生成例

9.2.4 マルチメディア情報管理機能の開発

「マルチメディア情報管理機能」は、マルチメディア情報のデータベース登録機能とミドルウェアの機能として、大容量の資源を必要とするマルチメディア情報が格納された集合光磁気ディスク装置の制御を行う。このように、データベースとマルチメディア情報とをリンクしマルチメディア情報の統合管理を行う機能である。

以降に、今回構築したデータベースの構造等を「マルチメディア情報管理機能」の解説を行う。

(1) データベース構造

今回構築したデータベースの構造は、ORACLE、Informix共にキャビネット、フォルダー、ドキュメントという階層構造となっており、各種マルチメディア情報(実体:ファイル)とリンクする。

(2) データベース項目

今回構築したマルチメディアデータベースのデータベース項目を表9-1に示す。なお、ORACLE、informix共に同じ構造である。

表 9 - 1	- -	タベー	プロロ	1—警
75 J - 1	<i>-</i>	<i>-</i>	<u> ヘーヌ ロ</u>	1 見.

データベース項目	意 味	検索キー項目
格納場所	マルチメディア情報が格納されているファイルの ディレクトリ (パス) 名、ドキュメントファイル名 を格納する。	_
キャビネット名	マルチメディア情報が属するキャビネット名称(日 本語名)を格納する。	_
フォルダー名	マルチメディア情報が属するフォルダー名称(日本語名)を格納する。	<u></u>
ドキュメント名	当該マルチメディア情報のドキュメント名称(日本語名)を格納する。	_
ファイル種別	当該マルチメディア情報がどのような種類のファイル形式で格納されているかを識別する情報を格納する。マルチメディア情報ファイルの拡張子として使用する。	0
	<ファイル種別の例>	
	aif:AIFF(Audio Interchange File Format) 形式音声情報	
	mpg:MPEG(Moving Picture Experts Group) 形式動画情報 等	
登 録 日	当該情報を登録した際の登録日を格納する。	0
登録番号	当該情報を一意に識別するための登録番号を格納する。	0
題 名	当該情報がどのような情報なのかを示す"題名"を 格納する。	0
備考	登録時に任意に入力された備考を格納する。	0

(3) データベース登録/変更機能

今回構築したシステムでは、マルチメディア情報の取り込み部分とは連動していないため、 スキャナー等で取り込んだマルチメディア情報ファイルが存在していることを前提に各種マ ルチメディア情報ファイルとデータベースとをリンクしながら登録を行う方法で実現している。なお、データベースの登録/変更操作は、サーバマシン(UNIXワークステーション)のディスプレイで実施する。

(4) データベース化したマルチメディア情報について

今回データベース化したマルチメディア情報は、インターネットで取り込んだ静止画、動画、音声や、イメージスキャナ、35mmフィルムスキャナ、フラットベットスキャナ等の機器を利用して読み込んだ情報、市販のCD-ROM等を利用してサーバにファイル化した。

以下にデータベース化の対象としたマルチメディア情報を示す。なお、情報量については、 WWWブラウザを使用した検索システムを構築することが主眼であったため、各々数件程度 の情報である。

- (a) A I F F (Audio Interchange File Format) 形式の音声情報
- (b) GIF (Graphics Interchange Format) 形式のイメージ情報
- (c) JPEG (Joint Photographic Experts Group) 形式のイメージ情報
- (d) TIFF (Tagged Image File Format) 形式のイメージ情報
- (e) MPEG (Moving Picture Experts Group) 形式の音声無しの動画情報
- (f) QuickTime動画情報
- (g) テキスト情報

等

(5) 集合光磁気ディスク上のマルチメディア情報管理

集合光磁気ディスク上のマルチメディア情報をどのようなメカニズムで取得し、WWWクライアント返却して表示するかを、図9-7に示し解説する。

(a) 「マルチメディア情報管理機能」の起動

h t t p d より本機能が起動される場合、光磁気ディスク媒体のボリューム名とデータファイル名を環境変数で受け取る。

(b) 光磁気ディスク媒体のマウント

「マルチメディア情報管理機能」は、環境変数で受け取った当該マルチメディア情報の ボリューム名の光磁気ディスク媒体をマウントする。

(c) WWWブラウザへマルチメディア情報ダウンロード

光磁気ディスク媒体のマウントに成功した「マルチメディア情報管理機能」は、httpdに対し、当該マルチメディア情報のデータファイル名を標準出力にてhttpdに通知する。

通知を受けたhttpdは、マルチメディア情報ファイルを読み込み、WWWブラウザ

にダウンロードし、WWWブラウザは各種マルチメディア情報を表示するための外部ビューアを起動して、当該マルチメディア情報をWWWクライアント画面に表示する。

(d) 光磁気ディスク媒体のアンマウント

処理の完了したマルチメディア情報が格納された光磁気ディスク媒体をアンマウントし 処理を終了する。

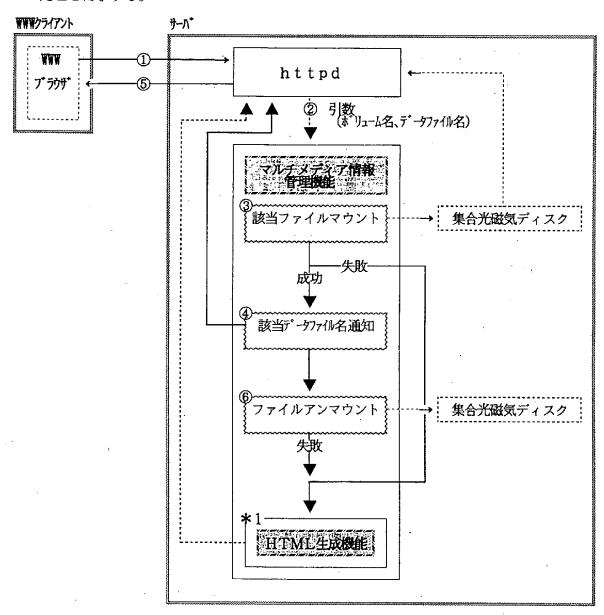


図9-7 集合光磁気ディスク上の情報取得時のデータフロー

9.2.5 予想される効果

今回開発した「マルチメディアデータベース検索システム」の核となるミドルウェアは、インターネットの利用者にデータベースの情報を提供するに当たり、サービスの対象となるマシンを限定しないため、サーバに蓄積された情報資源を多くの利用者に提供することが可能になる。このミドルウェアが普及した場合に予想される効果を以下に示す。

(1) インターネットへの情報公開促進

今回開発したミドルウェアを核としたシステムを構築すれば、各研究所等の既存のデータベースをインターネットから容易に検索することが可能になり、情報公開の飛躍的な促進が期待できる。

(2) マルチメディア情報の流通の加速

音声、画像等のマルチメディア情報を容易にデータベース化できることで、各サーバには 豊富なマルチメディア情報を蓄積することができる。この情報をインターネットから容易に アクセスすることが可能となり、マルチメディア情報の流通の加速が期待できる。

(3) 既存のデータベースの有効利用

既存のデータベースには、様々な分野の様々な種類の情報が蓄積されている。このような情報を今回開発したミドルウェアを活用し、インターネットから容易に検索して参照することが可能になれば、既存のデータベースの有効利用が可能となる。

(4) 企業内での知的生産性の向上

インターネットの技術を活用した社内LANでのOA化等、イントラネットの概念にも今回開発したミドルウェアを活用することができるため、企業内での知的生産性の向上に役立てることが可能である。

9.2.6 マルチメディアデータベース検索システムの問題点

マルチメディアデータベース検索システムは、多くの利用者のニーズに応えるにはユーザインターフェースという点で不十分な面があるため、ユーザインターフェースをより充実させ、 どのような利用者に対しても容易にデータベースの管理が行え、情報を引き出すことが可能な システムへと展開させていく必要がある。また、ネットワーク負荷の問題等、十分に評価でき なかった課題についても今後、評価を行い改善させていく必要がある。

9.3 今後の課題

9.3.1 機能の拡張

(1) マルチメディアデータベース管理機能の充実

WWWクライアントからインタラクティブにイメージスキャナ等で情報を取り込み、連動してデータベースに登録できるような機能が必要である。また、データベースの登録/更新操作についても、WWWクライアントから行えるようになれば、システム運用上の操作性を向上させることができる。

(2) 検索機能の充実

階層検索機能を提供することで知識のないエンドユーザでも容易に情報を参照することができようにする必要がある。また、集合光磁気ディスク媒体に格納された使用頻度の高い情報の磁気ディスクへの動的再配置機能の実現によりレスポンスの向上を図る等の工夫を行う必要がある。

(3) HTML生成機能, SQL生成機能の充実

HTMLの自動生成やSQLの自動生成等で使用している各種管理ファイルは、エディタでの登録、変更となるため、知識のないエンドユーザには操作できない。エンドユーザでも容易にメンテナンスができるようなマンマシンインターフェースを確立する必要がある。今後、GUIベースの対話型操作でHTMLフォームとCGIプログラムを同時に作成できるようなオーサリングツールとして確立させる予定である。

(4) 市販パッケージソフトとの連動

市販の英文翻訳ソフトや和文英訳ソフト, イメージ情報をスキュナ等で読み込ませデータ ベースに自動登録させるための文字認識ソフト等を本システムに読み込み, 連動させること により, 利用範囲を拡大させる。

9.3.2 APIの提供

様々なニーズを持つ利用者に対応できるように、今回開発したミドルウェアを部品化してAPIを整備し、ライブラリとして提供する予定である。

9.3.3 マルチプラットフォーム化

今回開発したシステムのサーバマシンは、処理能力を考慮してUNIXワークステーションを対象としたが、今後、需要の広がりをみせているWindows NTサーバ対応等も検討中である。

9.3.4 将来の展望

VRML(Virtual Reality Modeling Language)の出現により立体画像も扱えるようになった。今後,ユーザインターフェースを仮想空間で実現しようとするシステムの利用範囲が広がると予想される。また,JAVA言語やリアルオーディオ機能などを使用したインターネット上で展開されるニュービジネスの可能性も無限である。このような最新技術を順次取り入れてより良いシステムへと展開できることを願っている。

10. 大規模データベースにおける構造化情報抽出方式の調査



10. 大規模データベースにおける 構造化情報抽出方式の調査

10.1 問題と目的

インターネットや高性能なデータベース処理機の普及に伴い,ユーザが利用可能な情報資源は 大幅に拡大している。また、データベースの利用目的も多様化し、画像データを中心としたマル チメディアデータ処理に対する要請も高まってきている。

このような情報資源の拡大は以下の3つの観点で捉えられる。

- 量的拡大
- ・範囲の拡大
- 表現方式の拡大

本報告ではこれらの大規模化の各要因に対応する技術として、それぞれ以下の技術をとりあげた。

- 1. KDD (データベースからの知識発見)を用いてデータ群から情報構造を抽出する。
- 2. 分散人工知能技術を用いて、ネットワーク上の複数の情報ソースから構造化情報を抽出する。
- 3. マルチメディアデータにおける意味的な並列性に焦点をあてたマルチモダリティデータ処理。

10.2 情報検索システムの現状

これまでの情報検索システムで、最も基本となるのはテキスト処理に関する技術である。しかし、フラットテキストの検索で良く用いられるフリーターム検索は、検索されるデータ量の増大につれてパフォーマンスが非線形的に低下するといわれている。さらに検索精度に関する問題も 指摘されており、他の検索処理と組み合わせた方式が開発されている。

近年、このようなフラットテキストと並んでハイパーテキスト処理に関心が集まっている。ハイパーテキストはテキストのある部分と他の部分、あるいは他のテキストをリンクと呼ばれるポインターで結合したテキストである。ユーザはテキストをシーケンシャルに読むだけではなく、リンクを辿って非線形的なアクセスを行う。ハイパーテキストはあるい概念間の関係を"埋め込んだ"テキストといえる。

文献データベース利用の観点からみれば、そのデータはそれらが持つコンテンツによって管理

されるべきであろう。インデックスやシソーラスのような意味的構造化情報は、知識量が少ない ユーザによる検索を支援する強力な情報であり、同時に文献を管理する有力な情報である。

構造化情報の作成やそれらに基づいたドキュメントの構造化は、各ドキュメントをデータベースに登録する際、専門家によって処理される。現在、この作業の大半は人力に頼っており大規模化に対応していない。また、通常の情報管理ツールの大半がこの問題をブレークスルーするような有力な技術を提供していない。

10.3 大規模化の克服

大規模化した情報の利用や管理の円滑化そして容易化を図るためには、いかに各ドキュメントの意味構造や文書構造を抽出するかが問題となる。このような構造化情報抽出プロセスの自動化を図るために知識工学的手法を応用したシステムの研究開発はこれまでにも多数行われてきた。それらのシステムでは、"深い"構造化すなわち精緻な意味解析を行うことに焦点があてられている。情報は本質的には状況依存性を持つものであり、各状況に対応可能な意味処理は非常に困難である。

近年、このアプローチとは異なったアプローチをとるシステムの研究開発が行われている。このアプローチは意味処理の精緻さよりも、処理速度や処理量の大きさを優先したシステムである。これにより意味処理の正確性は低下するが、これは検索→確認のサイクルを早くすることによってカバー可能である。

このアプローチによるシステムは、ゼロからの構造化を行うのではなく文書の持つ構造を積極 的に利用して自己組織化を行う、意味構造の抽出よりも情報縮約を優先する、などを特徴として いる。本報告ではこのアプローチに立った、以下のようなシステムをその例として紹介している。

- 1. 自己組織化型情報ベース〔1〕における帰納的推理機能〔3, 4〕。このシステムにおける帰納的推理では、シソーラスを背景知識として用いている。
- 2.Webに対するインデックス,要約,そしてリンク作成を自動的に行うシステム。
- 3. インターネット上の情報の収集,抽出,構成,検索,キャッシュなどを行うツールである。

10.4 KDDによる構造化情報発見

データベースからの知識発見(KDD: Knowledge Discovery in Databases)の大量のデータ 群からの知識発見を行う手法や理論である。KDDはエキスパートシステム、知的データベース、 知識獲得、機械学習、事例ベース推論、統計学などの分野と関連があり、伝統的な機械学習の研 究の成果を継承しながら新たな理論展開を模索している。

10.4.1 GLS発見方法論

GLS発見方法論〔8〕はKDDの特性や問題などを考慮して開発された方法論であり、大 須賀らによる一般問題解決モデルをKDDへ応用したものである。GLSの核はデータベース からのグローバルな学習スキーマ(Global Learning Scheme)であり、図1のように表される。

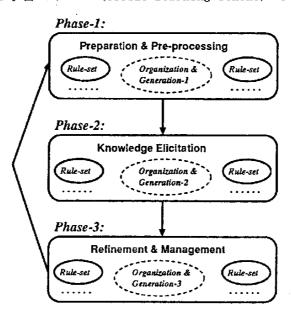


図1:Global Learning Scheme

GLSにおける知識発見はプリプロセス、知識抽出、知識の精緻化/管理という3フェーズから構成されている。

- 1. プリプロセスは知識抽出の準備フェーズである。このフェーズの主要なタスクはユーザ との対話によってユーザ要求を収集し、使用したい学習方法(知識指向統計推論法や分割 に基づく帰納法など)、データベース名と属性名のなどを確定し、データを収集と整理な どを行う。
- 2. 知識抽出のフェーズではプリプロセスフェーズの結果を基に知識抽出を行う。
 - 3. 知識の精緻化/管理フェーズでは前フェーズで発見された初期概念や仮説をルールとして表現して知識ベースに追加し、さらに精緻化や管理を行う。すなわち、データベースから初期概念や仮説を発見し、それを知識ベース管理システムによって管理すると共にメタ推論と多重世界の機能を用いた知識の精錬手法と組み合わせて、それを精緻化する。

GLSの三つの学習フェーズによって、知識ベースとデータベースとを統合した環境で多面的データ分析および多段階学習や概念抽象化などを行う知識発見が可能になる。また、GLSでの発見/学習プロセスはユーザの要求によるKDDだけではなく、動的な発見プロセスの組

織化や発見プロセスの制御と性能改善なども含んでいる。知識ベースを用いて処理が行われた GLSシステムでは、すべての知識が多重世界のメカニズムによってマルチレベルに分けられ て管理される。

10.4.2 テキストデータベースの構造化プロセス

GLS発見方法論に基づいて考えると、テキストデータベースからの構造化の情報を抽出するためのプロセスは以下の3つの学習フェーズに分けられる。

1. プリプロセス

このフェーズでの重要な処理はユーザとの対話に基づいてユーザの要求を収集し、使用 する構造化情報抽出法を確定し、テキストデータの収集と整理などを行う。

2. 構造抽出

構造抽出フェーズではプリプロセスフェーズの処理に基づいて構造抽出を行う。たとえば、このフェーズでは統計および確率的な手法、一階帰納学習、ニューラルネット技術、遺伝的アルゴリズムなど方法を利用し、概念や概念クラスタを生成する。

3. 管理/精緻化

前フェーズで抽出した初期概念や仮説を適当な表現形式に変換して知識ベースに追加して精緻化や管理を行う。すなわち、テキストデータベースから初期概念や仮説を発見し、それを知識ベース管理システムで管理すると共にメタ推論と多重世界の機能を用いた知識の精錬手法と組み合わせて、それを精緻化する。また、テキストデータベース上の変化や新しいデータを収集した時、これに対応して精緻化や管理を行う。

10.5 エージェント技術の応用

WWWロボットはWeb上のドキュメントを再帰的に検索することによってWeb上のハイパーテキストを自動的にトラバースするソフトウェアである。WWWロボットはサーチエンジンと密接な関連がある。サーチエンジンとは、自ノード内のデータベースに蓄積されたデータ使用して検索を行うプログラムであり、WWWロボットが収集したHTMLドキュメントのデータベースを用いて検索を行うシステムである。WWWロボットでは、あるテーマとそのテーマに関連するリソースのアドレスが主な処理対象となる。

WWWロボットの問題点は自律性が不足していることである。ネットワーク上の環境は頻繁に変化しており、静的な情報だけで行動することには限界がある。トランスポータブル・エージェットはこのような特徴をもつネットワーク上での自律性を実現するために提案された〔2〕。ト

ランスポータブル・エージェントはヘテロなネットワーク上をナビゲートするソフトウェアモジュールであり、ネットワーク上の任意のサイトへマイグレートできるソフトウェアである。このタイプのエージェントはその実行を任意の点で一時停止し、他のサイトへ移動した後に実行を再開できる。

トランスポータブル・エージェントはファイル、データベース、ネットワークトラフィックそして他のエージェントの状態の検知を行う仮想センサー、そして、マイグレーションによって異なった物理的位置に移動する仮想エフェクターのネットワークから構成されている。トランスポータブル・エージェントは以下のような点でネットワーク上での処理に適している。

- ・エージェントは全てのデータを要求されたサイトに送らず、小さなエージェントをデータソースに送るのでネットワークトラフィックを減らせる。
 - ・データをサイトから移動させる必要がないのでデータインテグリティを向上できる。

ネットワーク上で情報検索やリソース情報を作成するためにはトランスポートエージェント機構は有効な機構であろう。ここまでの議論から、ネットワーク上のエージェントが協調分散処理を行うためには、以下の機構が必要である。

- 共通知識
- ナビゲーション情報

これらの要求を統合する機構として、自己組織型情報カタログ(図2参照)が提案されている。

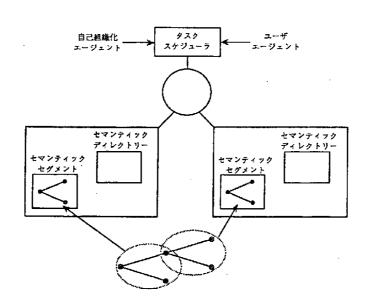


図2:情報カタログ

情報カタログはネットワークワイドのシソーラス機構であり、またリソースマップである。情報カタログがシソーラスであることにより、各エージェント間で構造化された知識を共通知識として使える。このシステムの特徴は次の各点である。

- ・ネットワーク上のリソースを探査する場合、一定の知識体系に基づいてネットワーク上のト ラバースが可能になり、タスクプランニング管理が容易になる。
- ・課題の分担、並列探査、エージェントのランデブー点の指定などが容易になる。

10.6 マルチモダリティデータ処理

現在製品化されているマルチメディア管理ツールは各ドキュメントが内包する意味処理という 点ではこれまでのツールよりも前進がある。しかし、これらのツールにおける意味情報とは、異 なったメディアをその生成時の同期関係、各メディアを提示する時の同時関係などメディア管理 に関連するものである。

一方,いくつかのWebサイトで非テキストデータ主体の情報提供が行われている。このようなサイトでは、蓄積されている各データには、それぞれデータリソースを代表するインデックスが付けられ、そのインデックスによって分類されている。これらのサイトにおける検索は、テキストベースによる検索が主体となっている。このような検索方式に加え、さらにイメージの内容を表すアイコンを検索の補助手段として用いているサイトもある。この方式の問題点として以下の各点が挙げられよう。

- ・インデックスが表す内容からそのコンテンツを特定しづらい。
- ・専門知識が必要となる。
- ・複数の話題を同一リソースを持つ場合その内容を特定しづらい。
- データ更新のタイムラグが大きい。

検索や検索支援技術そしてデータソースの自動インデックス作成技術の成熟度は低い。その一方で既存のパターン認識技術や知識処理技術で、これらの課題をブレークスルーする技術に関する明確な展望が得られていない。この問題の根源は非テキスト情報を表象するデータと、意味を表象するシンボルとのセマンティックギャップを解消する技術が未熟であることである〔6〕。

10.7 マルチモダリティデータの構造化

たとえば、動画像は静止画像とは違って一定のストーリーを持っている。このストーリーはさらにいくつかのサブストーリーに分割でき、さらにこのサブストーリーもいくつかの構成要素に分割できる〔7〕。これはテキストで表現された情報には文書→段落→文→単語といった階層が存在するのと似ている。非言語的データにおいても適切な単位に分節化することが、より深い処理を行うための出発点となる。

この方式のポイントは、状況依存性を無視すれば情報の根源的意味は一つであるが、表象形態 すなわちモダリティが異なったデータ集合としてマルチメディアを位置付けられる点である。つ まり、マルチメディアとはモダリティが並行したデータ集合といえる。

このようなモダリティの並行性を生かした処理を行うためには、それと適合した情報ソースを見つけることと並んで、各モダリティ間のレベル合わせが必要となる。これは、各モダリティ内の他段階処理とモダリティ間の同期化が必要であることも意味している〔5〕。モダリティの特殊性を考慮した多段階処理と、かつ意味的な等価なモダリティ間の同期処理のモデル化のためにもGLSのスキームが適用できる。

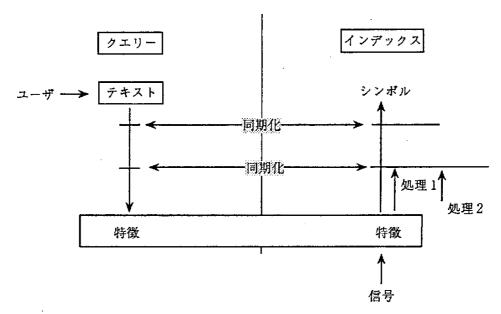


図3:多段階モダリティ処理

1. プリプロセス

このフェーズでの重要な処理はユーザとの対話に基づいてユーザの要求を収集し、使用したい構造化情報抽出法を確定し、テキストデータの収集と整理等を行う。

2. 構造抽出

構造抽出フェーズではプリプロセスフェーズの処理に基づいて構造抽出を行う。このフェーズではエッジ検出、プリミティブの処理、統計的、および確率的な手法、ニューロネット技術などの方法を利用しクラスタを生成する。

3. 管理/精緻化

前フェーズで抽出した基本的プリミティブやプリミティブ間の類似性と知識ベース中の知識とを照合し、さらに大きな図形単位への合成などを行う。この画像上のプリミティブ合成過程は、そのプリミティブに対応する概念の合成となる。ゆえに、あるプリミティブ同士の矛盾性のチェックを、それに対応する概念の組合せにおける矛盾のチェックとできる。

この三つのフェーズ処理によって、モダリティ間の並行性を生かした多面的データ分析、多段階の概念抽象化そして意味情報の抽出が可能になる。さらに、ユーザの要求による構造化だけでなく、動的な発見プロセスの組織化、発見プロセスの制御と性能改善なども行う。このプロセスから、マルチモダリティデータ処理に関するメタ知識も抽出できよう。

参考文献

- (1) Y. Fujiwara, et al. The information base system for materials research. CODATA Bulletin, Vol. 24, No. 1, pp. 1-7, 1990.
- [2] R. Gray, et al. Transportable information agents. Technical Report PCS-TR96-278, Dartmouth University, 1996.
- [3] Y. Kakemoto. Inductive reasoning using thesaurus. In Proc. of JAPAN-CIS Conference on Knowledge-base Software Engineering., 1994.
- [4] Y. Kakemoto. Inductive reasoning using thesaurus. In Proc. of Advece in Data-Bace and Information System (in Lecture Note in Computer Science), 1995.
- [5] Y. Kakemoto and N. Zhong. A study of self-organized image database, 1996. (submitted).
- [6] 美濃導彦. 知的映像メディア検索技術の動向. 人工知能学会誌, Vol. 11, No. 1, pp. 3-9, 1996.
- [7] 柴田正啓. 映像の内容記述モデルとその映像構造化への応用. 電子情報通信学会論文誌, Vol. 78-D-II, No. 5, pp. 62-72, 1995.
- [8] N. Zhong and S. Ohsuga, GLS a methodology for discovering knowledge from databases. In P.S. Glaeser and M.T.L. Millward, editors, New Data Challenges in Our Information Age, pp. A20-A30, 1992.

11. モーバイルデータベースシステムに関する 調査研究

(株)イフ・アドバタイジング

-		

11. モーバイルデータベースシステムに関する調査研究

11.1 概要

11.1.1 目的

情報通信技術の進展によって、今日、情報通信ネットワークの活用が急速に拡大しつつあるが、その多くは有線による位置固定のコンピュータ間に結合するものである。確かに、無線LANの利用も多くの場で見られるようにはなったが、それでもコンピュータおよび関連機器は、通常、設置場所が固定されている。

こうした中で、最近、無線等を利用した移動型の情報通信システムが大きな脚光を浴びており、特にパーソナル・ハンディホン・システム(PHS)等の進展によって急速に普及しようとしている。

膨大な情報を効率的に活用しなければならないユーザにとって、移動型ステーションを通じて必要な場所から、必要な特に即座に様々なデータベースにアクセスできれば、さらにサーバ内のデータの更新もできれば、ユーザの利便性が飛躍的に向上する。しかし、こうした場合、位置が固定されない移動型のシステム間での通信処理(モーバイルコンピューティング)が重要となる。また移動型ステーションは小型・軽量であるだけに、サーバとクライアント間の通信量および通信時間の軽減が必要とされる。

このため、固定位置のデータベースサーバを移動型ステーションのクライアントが利用する問題について様々な側面から考察する。さらに、こうしたシステムを利用した有効データベースサービスについて検討を行うとともに、併せて既存の大規模ネットワーク、たとえばインターネット等の融和性についても考察する。

以上の調査・研究によって、わが国のデータベース振興に資することを目的としている。

11.1.2 実施内容

本調査・研究における主な実施内容は、以下の項目に示す通りである。

1)移動型検索方式

移動型ステーションを通じて、データベースサーバ内のデータを検索する方式について 検討する。ステーションとサーバ間の通信量を減少するためにサーバ内のデータをステー ション内にキャッシングすることにより、通信量を減少する方式を検討する。

2)移動型トランザクション管理方式

次に、複数のデータベースサーバ内のデータを更新する方式を検討する。特に、利用者

が移動しながらサーバを利用するときに、サーバ内のデータの一致性を保つための同期方 式を検討する。

3) データの配置方式

移動型ステーションから、データベースサーバを利用するために、データを複数のサーバにどのように配置するかについて検討する。

4)移動型ステーションによるデータベースサービス。移動型ステーションを利用した有効なデータベースサービスについて検討を行う。

11.2 実施成果の要旨

11.2.1 システムモデル

現在のデータベースシステムは、DBMS機能のうち利用者インターフェースの部分をクライアントに、データベースに対する基本操作部分をサーバに機能分散させる形態をとっている。こうしたデータベースシステムは、通常、複数のクライアントとサーバがLAN等の通信ネットワークによって相互接続されている。

一方、これまでの通信ネットワークは、有線ネットワークでありクライアントもサーバもネットワーク上の固定の位置に接続されていた。これに対して、現在、携帯電話、自動車電話、PHS等の無線を用いたモーバイル通信技術が進展し普及しつつある。また、コンピュータもきわめて小型で高性能化が進みつつある。したがって、こうした超小型コンピュータを携帯型ステーションとした移動型クライアントとサーバ、つまりモーバイルクライアントとサーバに関するシステムモデルについて提案を行った。

11.2.2 検索方式

まずクライアントとサーバが移動している場合、クライアントから特定のサーバに通信を 行う方法について検討を行った。モーバイルデータベースシステムでは、同様のデータを保 有する複数のサーバが存在するため、最も有効なサーバを特定し利用する必要がある。した がって、データベース内のデータをオブジェクトとして、オブジェクトがデータベースサー バ内に分散されている場合の通信方法について、伝送速度、伝送品質の変化等を加味しつつ 検討した。その上で、クライアントがオブジェクトをベースとしてサーバにおけるデータを 効率的に検索する手順について提案を行った。

11.2.3 トランザクション管理方式

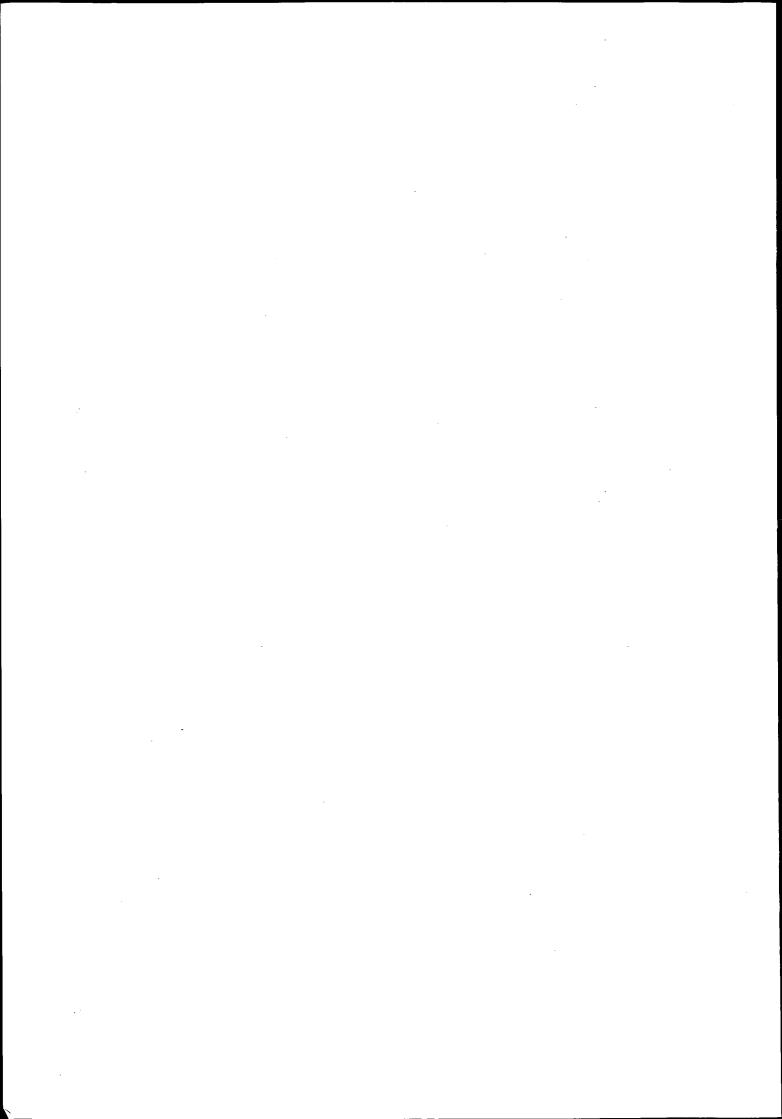
モーバイルデータベースシステムでは、各ユーザが所有する移動型ステーションが長時間の通信を行うための十分な容量のバッテリを持てない。このため移動型ステーションはサーバとの通信を最小限に止める必要がある。また、たとえ移動型ステーション内のクライアントがサーバとの通信状態を切断されても、サーバ内のデータを利用したプログラムを継続させる必要がある。また、利用者も移動しながらサーバを利用する場合に、サーバ内にデータの一致性を保たねばならない。こうした視点にたって、複数のデータベースサーバ内のデータを更新する方式、つまりトラザクション管理方式について研究を行った。

11.2.4 データの配置方法

複数のデータベースサーバから構成される分散データベースシステムの目的の一つは、利用者に対し分散されているデータがそれぞれ独立した形をとっていることを示すことにある。この場合、利用者から見える分散データベースシステムのデータベースは、テーブルの集合と言える。したがって、ここでは移動型ステーションからデータベースサーバを利用するため、データを複数のサーバにどのように配置するかについての検討を行った。

11.2.5 モーバイルステーションによるデータベースシステム

これまでのデータベースシステムは、固定的なコンピュータ内に構成され、データベース管理者の下で集中的に管理されてきた。こうした統合的データベースシステムは、情報提供型大規模データベースシステムとして今後も必要とされよう。しかし、一方ではコンピュータの小型化、高機能化に伴いこれらを携帯型端末としたデータベースの検索はもとより、これら自身が個人的データベースを保有する移動型分散データベースシステムが急速に進展していくものと考えられる。さらに、今日のインターネット等のネットワークの急速な進展にともない、多くの個人が保有するデータベースを有機的に相互接続するような超分散データベースシステムが構成されていることになろう。ここではこうした場合の各データベースシステムの関連付け、データベースの検索、構成されるモーバイルデータベースシステム等の問題点等について考察した。



12. 情報収集ロボットによるInternetでの WWW所在検索データベースの構築

	•	
	•	

12. 情報収集ロボットによるInternetでのWWW所在検索 データベースの構築

12.1 概要

近年、インターネット(Internet)という世界的規模のネットワークが広く知られるようになった。インターネットは全世界の少くとも81か国以上、200万台以上のコンピュータが接続されている $11\cdot 21$ 。今では学術研究分野で主に使われてきたが、最近は科学技術分野だけではなく、人文科学を始めとする大学教育全般を始め、小・中・高等学校をも巻き込んだ普及・展開がされてきている。更に、企業・団体の参加の増加社数は群を抜いており、これからの通信情報伝達手段として、大きな期待が持たれている。

特に最近注目を浴びているのは、WWW(World Wide Web)と呼ばれる、マルチメディア対応の情報蓄積・検索システムである。これは、インターネットに接続されているホストマシン(多くはUINXが利用されている)上に、HTMLと呼ばれるSGMLベースの言語で記述されているデータを、PCを始めとする多くのクライアントシステムから呼び出して見ることができるものである。HTMLでは、リンクと呼ばれる仕掛けによって、世界中どこのホストにあるデータも直接参照することができ、これが新しいデータベースの基本構造として認識され始めている。WWWでは、膨大な量のデータに直接アクセスできるメリットがあるが、一方、求める情報を

捜し当てることがなかなか難しいといわれている。このような状況から、希望する情報を素早く 捜し当てる手立てを講じる必要がある。 本報告は、上記のような要求に答え、インターネット上のWWWサーバ(HTMLによる情報

を蓄積し提供するホストマシン)を自動的にソフトウェアによるロボットを使って網羅的に探索し、URL (Uniform Resource Locator)と呼ばれるWWWの所在情報と、その中に含まれる情報を抽出・整理して検索用のキーワード情報を作り出すための基礎調査を行い、ならびにその結果を利用したWWWページ検索用データベースのプロトタイプを構築するものである。

12.1.1 目 的

WWWページを探索する既存のロボット(ソフトウェア)は、学術分野では既にいくつかの 試行が行われている^{8, 7, 3)}。まずこれらの調査・分析を行い、利用できる技術やシステムの確 認と調査を行う。

既存のロボットでライセンス関係の問題が無く、ベースとして使えそうなものがあれば改良 試作を行い、日本語での運用動作について確認するとともに、今後の拡張についても評価する。 特に日本語による世界中の主要WWWページの探索情報を作成することを最終目標とし、その ための基礎データを収集することを目的とする。

ロボットによって収集したHTMLデータの評価・分析を行い、今回の実行結果をWWWブラウザから検索できるようにして見る。このプロトタイプデータによる情報検索サービスを試行してみる。これは将来、有料サービスへの移行を検証するものである。

12.1.2 実施内容

12. 1. 2. 1 ロボットのベースシステム

現在インターネット上でロボットと称されているシステムはかなり多数にのぼる。 Martijn Koster氏の「List of Robots」というWWWのページ $^{3.47}$ には世界中のロボット関係のシステムの一覧表がある。また、NTTの「Search Engines in Japan」ページ 67 には、日本の検索システムやリンクページの情報がある。これらの内、ロボットとして認識できるものを表1にまとめた。

日本で研究されている、WWWページを検索する既存のロボット(ソフトウェア)では、早稲田大学の千里眼などが知られているが、今回はアメリカのコロラド大学で開発・公開されているハーベスト(Harvest) $^{8, \ 8, \ 10}$ に注目した。Harvestの最新版はV1.4であるが、論文 8 にもあるように、さまざまな基本的な機能を実現している。今回はこのHarvestをベースシステムとして採用することとする。

12.1.2.2 日本語情報に対応するための拡張

HarvestはWWW上で提供されている情報を収集し、それをデータベース化するための、 ソフトウェアロボットである。 Harvestは欧米語とくに英語で書かれたもののみを対象とし ている。このため、日本国内で提供されている情報を収集し、データベース化するために、 日本語を扱えるように拡張を行った。

Harvestでは、収集した各ページから、データベースに収録したい文書構造のみを抽出し、その特定の文書構造に含まれるテキストから、データベースを作ることが可能である 81 。たとえば、個々のHTML文書中に含まれる、〈 1 H1〉、〈 1 H2〉、、〈 1 H6〉のヘッダー情報と、〈 1 TITLE〉をデータベースに収録し、その他は収録しないというような指定ができる。これらの情報の抽出には、 1 SGMLのパーサを使っており、まずは、これの日本語化を行った。

表1:Internet Robot一覧表

システム名	作者または管理者	記述言語
JumpStation	Jonathon Fletcher < j.fletcher@stirling.ac.uk>	Perl 4, C, C++
RBSE Spider	David Eichmann <eichmann@rbse.jsc.nasa.gov></eichmann@rbse.jsc.nasa.gov>	C, Oracle, WAIS
WebCrawler	Brian Pinkerton < bp@webcrawler.com>	C
The NorthStar Robot	Fred Barrie <barrie@unr.edu></barrie@unr.edu>	
W4 (the World Wide Web Wanderer)	Matthew Gray <mkgray@mit.edu></mkgray@mit.edu>	<u> </u>
fish search	Paul De Bra <debra@win.tue.nl></debra@win.tue.nl>	C
The Python Robot	Guido van Rossum < Guido.van.Rossum@cwi.nt>	_
html_analyzer-0.02	James E. Pitkow <pitkow@aries.colorado.edu></pitkow@aries.colorado.edu>	<u> </u>
MOMspider	Roy T. Fielding < fielding@ics.uci.edu>	Perl 4
HTMLgobble	Andreas Ley <ley@rz.uni-karlsruhe.de></ley@rz.uni-karlsruhe.de>	
WWWW - the WORLD WIDE WEB WORM	Oliver McBryan <mcbryan@piper.cs.colorado.edu></mcbryan@piper.cs.colorado.edu>	_
W3M2	Christophe Tronche <tronche@lri.fr></tronche@lri.fr>	Perl 4, Perl 5, C++
Websnarf	Charlie Stross < charles@fma.com>	
The Webfoot Robot	Lee McLoughlin < L.McLoughlin@doc.ic.ac.uk>	
Lycos	Dr. Michael L. Mauldin < fuzzy@cmu.edu>	(finite memory model)
ASpider (Associative Spider)	Fred Johansen < fred@nvg.unit.no>	CGI script
SG-Scout	Peter Beebee <ptbb@ai.mit.edu, beebee@parc.xerox.com=""></ptbb@ai.mit.edu,>	· —
EIT Link Verifier Robot	Jim McGuire < mcguire@eit.COM>	HTML, CGI script
NHSE Web Forager	Robert Olson <olson@mcs.anl.gov></olson@mcs.anl.gov>	Perl 5
WebLinker	James Casey <jcasey@maths.tcd.ie></jcasey@maths.tcd.ie>	<u> </u>
Emacs-w3 Search Engine	William M. Perry < wmperry@spry.com>	Lisp
Arachnophilia	Vince Taluskie <taluskie@utpapa.ph.utexas.edu></taluskie@utpapa.ph.utexas.edu>	<u> </u>
Mac WWWWorm	Sebastien Lemieux < lemieuse@ERE.UMontreal.CA>	HyperCard
churl	Justin Yunke <yunke@umich.edu></yunke@umich.edu>	<u> </u>
tarspider	Olaf Schreck < chakl@fu-berlin.de>	_
The Peregrinator	Jim Richardson <jimr@maths.su.oz.au></jimr@maths.su.oz.au>	l' —
Checkbot	Hans de Graaff < j.j.degraaff@twi.tudelft.nl>	Perl 5
webwaik	Rich Testardi at HP	С
Harvest	hardy@bruno.cs.colorado.edu	C, Perl 4, Perl 5
Katipo	Michael Newbery < Michael Newbery@vuw.ac.nz>	c
InfoSeek Robot 1.0	Steve Kirsch <stk@infoseek.com></stk@infoseek.com>	Python
GetURL	James Burton <james@snark.apana.org.au></james@snark.apana.org.au>	ARexx
Open Text Corporation Robot	Tim Bray < tbray@opentext.com>	_
The TkWWW Robot	Scott Spetka < scott@cs.sunyit.edu>	Tck/Tk
A Tcl W3 Robot	Laurent Demailly < di@hplyot.obspm.fr>	Tcl
CS-HKUST WWW Index Server	Budi Yuwono < yuwono-b@cs.ust.hk>	С
Spry Wizard Robot	Spry <info@spry.com></info@spry.com>	<u> </u>
weblayers	Loic Dachary <loic@afp.com></loic@afp.com>	Perl 5
WebCopy	Victor Parada < vparada@inf.utfsm.cl>	Perl 4 or 5
Scooter	Louis Monier < monier@pa.dec.com>	С
Aretha	Dave Weiner davew@well.com>	Macs
WebWatch	Joseph Janos < janos@specter.com>	C++
ArchitextSpider	Architext Software < spider@atext.com>	Perl 5 and C
HI (HTML Index) Search	Razzakul Haider Chowdhury <a94385@cs.ait.ac.th></a94385@cs.ait.ac.th>	Perl 5
explorer	Paul Bourke < pd.bourke@auckland.ac.nz>	C++
Senrigan	TAMURA Kent < kent@muraoka.info.waseda.ac.jp>	Ċ
FunnelWeb	David Eagles <eaglesd@pc.com.au></eaglesd@pc.com.au>	C, C++
The Jubii Indexing Robot	Jakob Faarvang < jakob@jubii.dk >	Visual Basic 4.0
Jobot	Adam Jack <ajack@corp.micrognosis.com></ajack@corp.micrognosis.com>	Perl 4
DeWeb(c) Katalog/Index	Marc Mielke <dewebmaster@orbit.de></dewebmaster@orbit.de>	Perl 4
Web Core/Roots	Carlos Baquero <wc@di.uminho.pt></wc@di.uminho.pt>	Perl 5
Robot Francoroute	Marc-Antoine Parent < maparent@crim.ca>	Perl 5, Sql plus
	Larry Burke < lburke@aktiv.com>	
Duppies IncuWines	Simon Stobart <simon.stobart@sunderland.ac.uk></simon.stobart@sunderland.ac.uk>	C++
IncyWincy IBM Planetwide	Ed Costello <epc@www.ibm.com></epc@www.ibm.com>	Perl 5
IBM_Planetwide	Richard Sonnen < sonnen@cs.colostat.edu>	Perl 4
Nomad UCSD Crawl	Adam Tilghman <atilghma@mib.org></atilghma@mib.org>	Perl 4
	onty pittsburgh, l.p. <webmaster@onty.com></webmaster@onty.com>	C++
webfetcher	• •	C++
libertech-rover	Anil Peres-da-Silva <adasilva@libertech.com></adasilva@libertech.com>	C++
HTDig	Andrew Scherpbier < andrew@sdsu.edu>	C, C++
TM1-557 -1	Kevin Hoogheem <khooghee@marys.smumn.edu></khooghee@marys.smumn.edu>	
BlackWidow	Migab A Williams / migab@convent unofer adu	l C
Pioneer	Micah A. Williams < micah@sequent.uncfsu.edu>	С _
	Micah A. Williams < micah@sequent.uncfsu.edu> 原田 昌紀 < harada@graco.c.u-tokyo.ac.jp> titan-admin@isserv.tas.ntt.jp	C

表 2: 開発実行環境

Machine	IBM Risc System 6000 model 43P-133
OS	AIX Risc System Version 4.1.4
Memory	64MB
HDD	total 9.4GB (including internal 1GB HDD)

また、開発・テスト実行環境として、表2のような環境を用意した。このため、AIX環境で再コンパイルする際に必要になった、AIX固有のローカルな修正も行った。修正箇所の主な部分を下記に列挙する。

- ・SGMLパーサである"sgmls"を日本語化した。
- ・xfree() 関数内で呼び出しているfreeシステムコールが正しく動作していないことを対 処した。
- ・実行時の仮想記憶容量を1.3GBとした。

12.1.2.3 得られるデータのフォーマット

日本語化された Harvestにより収集されたHTMLデータは、表 3 によるHTMLのタグから Harvestで採用されているデータ記述方式であるSummary Object Interchange Format (SOIF)のフィールドに変換される。具体的には、表 4 のようなフォーマットで検索データが得られることになる。

データは、@FILE $\{\dots\}$ の形式をしている。このデータをDBMS化して検索に利用することになる。形式は単純なので、さまざまなDBMSやGUIに対応することができる。

表3:HTMLタグとSOIFフィールドの対応表

HTML のタグ	タグの意味	SOIF フィールド
A HREF	リンク先	url-references
ADDRESS	アドレス/著作権表示	address
В	ボールド体	keywords
EM	強調 (イタリック体)	keywords
H1	ヘッダ行(レベル1)	headings
H2	ヘッダ行 (レベル 2)	headings
H3	ヘッダ行(レベル3)	headings
H4	ヘッダ行(レベル4)	headings
H5	ヘッダ行(レベル5)	headings
H6	ヘッダ行(レベル6)	headings
I	イタリック体	keywords
META:CONTENT	メタ情報	\$NAME
STRONG	強調 (イタリック体)	keywords
TITLE	タイトル	title
ТТ	タイプライター体	keywords

表4:SOIFデータフォーマットの雛型

```
@FILE { http://harvest.cs.colorado.edu/harvest/user-manual/node99.html
update-time{9}: 793962520
Time-to-Live{7}: 9676800
Last-Modification-Time{1}:
Refresh-Rate(7):
                       2419200
                      Netwarked Information Discovery and Retrieval
Gatherer-Name{57}:
Gatherer-Host{21}:
                       bruno.cs.colorado.edu
Gatherer-Version{3}: 1.0
Update-Time{9}: 781478043
Type { 4 }:
               HTML
File-Size{4}:
               2551
               c2fa35fd44a47634f39086652e879170
MD5{32}:
               Darren Hardy
author{42}:
Wed Feb 15 13:27:56 MST 1995
keywords [68]: about document drakos harvest html index latex manual
nikos this user
url-references{274}:
                      user-manual.html
node98.html
node3.html
                       About this document ......
Partial-Text { 601 }:
Title{84}:
               IRTF Research Group on Resource Discovery
IRTF Research Group on Resource Discovery
Keywords {121}: advanced agency bowman danzig darren hardy harvest manber mic
michael peter problems projects research schwartz talk udi
```

12.1.2.4 データベースの日本語化

抽出されたSOIFフォーマット情報を、データベース化する際に、日本語による検索が出来るようにしなければならない。Harvestでは、glimpse、waisなど、様々なindexingメカニズムをサポートしている 9 。日本語で検索できるようにするために、実験段階では $MWAIS^{11-12}$ を利用したが、最終的には任意のWWWサーバ(httpd)でも検索処理ができるように、評価試験用の検索システムを作成した。

12.1.2.5 WWW検索用インターフェース

WWWのブラウザを利用して検索するためのインターフェースを作成した。検索システムの概要を図1に示す。また、検索を行う際のブラウザによるGUIの例を図2に示す。

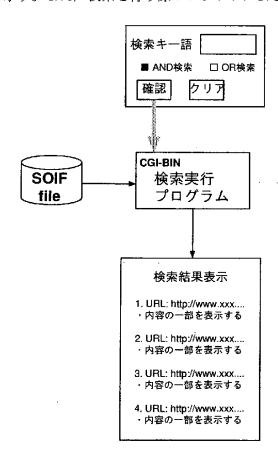


図1:検索インターフェース概要

	w Go Bookmarks Options Directory Window	Help L
日外で	アソシエーツ Web ページ検索システム	
「種様素キーを)	治野のWWWサーバーにあるページを検索します。 く力してください。 弦切って複数のキー語を指定してAND/OR検索ができます。	
機業十一語:	[* AND検索(絞り込み) * OR検索(全で検索) * 8 * **	
	検索 。 クリア	2.20
7/sij		

図2:WWWブラウザによる検索インターフェース

12. 2 検討結果

12.2.1 ロボットによるデータの収集

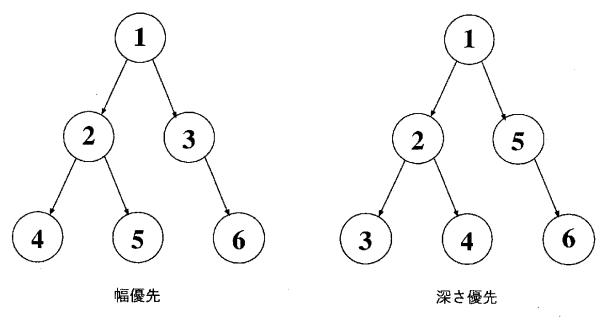
まず、ロボットをどのURL地点から走らせるかを検討する必要がある。学術研究用の URLを効率的にたどるために、あらかじめ下準備として、早稲田大学の村岡研で収集した URLデータを利用した。

データの収集は、学術系の Webサーバが集中しているつくば市地域を対象として行った。 表 5 にWWW上の情報を収集した機関名とドメイン名の一覧表を示す。

ロボットによる検索は、長時間に渡って大量のトラフィックを発生させるので、実際に実行する際には十分に注意しなければならない。今回は、 $Power\ P\ C\ +\ A\ I\ X$ のハードウェア環境を用意し、 $1.5\ Mbps$ の高速回線を利用することで、毎秒1回の割合で検索要求を出したが、より低速の回線では過大な負荷となるので注意すべきであろう。試行で収集したデータのサイズを表6に示す。

各サーバの探索方法には、表3のような方法がある。それぞれに特徴があるが、今回のよ

うな探索では幅優先の探索が適している。幅優先とすることによって、指定した複数のスタートポイントから、均等に深さを徐々に深くして検索するので、各スタートポイントから満 遍なく情報を取得することができる。



幅優先探索	各スタートポイントから同じ深さで徐々に降りてくる
深さ優先探索	最初のスタートポイントを見終わってから徐々に他を検索する

図3:各サーバの探索方法

表 5:機関名とドメイン名の対応一覧表

機関名	ドメイン名			
国立研究所および政府関連機関				
通産省工業技術院	aist.go.jp			
科学技術庁防災科学技術研究所	bosai.go.jp			
建設省国土地理院	gsi-mc.go.jp			
文部省高エネルギー研究所	kek.jp			
気象庁気象研究所	mri-jma.go.jp			
宇宙開発事業団	nasda.go.jp			
環境庁国立環境研究所	nies.go.jp			
科学技術庁無機材質研究所	nirim.go.jp			
科学技術庁金属材質技術研究所	nrim.go.jp			
建設省土木研究所	pwri.go.jp			
科学技術庁研究交流センター	statci.go.jp			
農林水産省農業生物資源研究所	abr.affrc.go.jp			
農林水産省農林水産技術会議事務局	cc.affrc.go.jp			
農林水産省果樹試験場	fruit.affrc.go.jp			
農林水産省森林総合研究所	jircas.affrc.go.jp			
農林水産省国際農林水産業研究センター	jircas.affrc.go.jp			
農林水産省農業研究センター	narc.affrc.go.jp			
農林水産省食品総合研究所	nfri.affrc.go.jp			
農林水產省農業環境技術研究所	niaes.affrc.go.jp			
農林水産省家畜衛生試験場	niah.affrc.go.jp			
農林水産省蚕糸・昆虫農業技術研究所	nises.affrc.go.jp			
農林水産省農業工学研究所	nkk.affrc.go.jp			
教育機関				
東京家政学院筑波短期大学	kasei.ac.jp			
筑波大学	tsukuba.ac.jp			
筑波技術短期大学	tsukuba-tech.ac.jp			
図書館情報大学	ulis.ac.jp			
つくば市立桜南小学校	ounan-es.tsukuba.ibaraki.jp			
非営利機関				
新情報処理開発機構	rwcp.or.jp			
社団法人 農林水産先端技術産業振興センター	staff.or.jp			
一般企業				
筑波研究コンソーシアム	trc-net.co.jp			

表 6:機関名とWWWで提供している情報量

機関名	II	提供している情報量			
1級関位	ページ数	バイト数	サーバ数		
通産省工業技術院	2863	6611059	1		
科学技術庁防災科学技術研究所	106	140002	3		
建設省国土地理院	796	2888405	2		
文部省高エネルギー研究所	+1797	+9338589	+33		
気象庁気象研究所	156	315963	1		
宇宙開発事業団	1237	2289395	8		
環境庁国立環境研究所	1670	4495317	1		
科学技術庁無機材質研究所	30	236777	1		
科学技術庁金属材質技術研究所	674	1619592	2		
建設省土木研究所	141	153147	1		
科学技術庁研究交流センター	73	173076	1		
農林水產省農業生物資源研究所	12	26051	1		
農林水產省農林水產技術会議事務局	197	393463	5		
農林水産省果樹試験場		_			
農林水産省森林総合研究所	197	689171	1		
農林水産省国際農林水産業研究センター	137	363434	1		
農林水産省農業研究センター	2	2859	1		
農林水産省食品総合研究所	9	10610	1		
農林水產省農業環境技術研究所	259	654323	1		
農林水産省家畜衛生試験場	849	4876288	2		
農林水産省蚕糸・昆虫農業技術研究所	35	70587	1		
農林水産省農業工学研究所	_				
東京家政学院筑波短期大学	13	110128	1		
筑波大学	+2477	3803491	+53		
筑波技術短期大学	107	150243	4		
図書館情報大学	1652	4615636	7		
つくば市立桜南小学校	95	77273	1		
新情報処理開発機構	738	15819352	1		
社団法人 農林水産先端技術産業振興センター	100	868384	1		
筑波研究コンソーシアム	17	32590	1		
合 計	+16439	+60825205	+137		

12.3 今後の課題

12.3.1 Harvestの問題点

Harvestはある特定の話題や、本来ある限られた範囲の情報を収集することを目的としたソフトウェアロボットである。このため、広い範囲の情報を収集し、データベース化するのには若干工夫する必要がありそうである。今回の計画では、「日本国内で提供されている情報すべて」という広い範囲の情報をこのソフトウェアロボットを利用して収集しようとしたが、この点については再度検討を行って実際の運用方法を考えなければならない。

12.3.2 漢字コードの文字化けについて

SOIF形式のデータには若干ではあるが、漢字コードの一部が壊れて解読不可能になっている部分がある。これは、該当URLのファイル自身が壊れていた可能性を否定できないが、転送後のファイル変換での問題であるかも知れない。

Harvestは基本的には8ビットスルーでデータを収集する。このため、元ファイルがJIS、EUC、SJISなどで記述されていると、そのコードのままひとつのファイルにまとめて集積を行う。このファイルを入力コードを仮定して処理すると、誤変換が多く発生してしまうため、今回は行毎に漢字コードを判定して変換を行った。

これで普通であれば正しく変換できるはずであるが、誤変換している部分は更に詳細な追試を行う必要がありそうである。WWWページに日本語を掲載する場合は、ISO2022に基づく言語環境シフトコードを伴ったJISコードを使うことが推奨されているが、まだ直接SJISコードで書かれているページも多く見受けられる。このような問題は、webmaster と呼ばれるWWW管理者の教育や、専門職としての認知と処遇がまだまだ浸透していないことによる面があるように思われる。

12.3.3 動作環境の増強・

今回のつくば地域での試行で、約6 MB弱のデータを採集した。全国的な規模で行うと、1,000倍以上のデータ量となることが予想されるので、更に大規模なサーバと大容量回線を用意する必要がありそうである。

12.3.4 今後の検索データベースに求められるポイント

Harvestを用いて広い範囲の情報を収集しデータベース化するには、主題別の各々の複数の データベースを作成し、それらを一つのフロントエンドで検索するような形で行うのが良さそ うである。

但し、今後WWW上で提供されている情報がますます多くなっていること考えると、あらか じめ決められた特定の情報のみを収集し、データベース化するソフトウェアロボットが必要と なってくると予測する。

現在は主に人間によるチェックによってインデックス別に集められたURLを羅列するサービスが行われており、日本ではNTTのものが良く知られているが、今後は、主題別に整理された情報をロボットでも集められるようにする必要がありそうである。

参考文献

- (1) Matthew Gray. Measuring the growth of web. http://www.netgen.com/info/growth.
- [2] Netree Internet Statistics Estimated. http://www.netree.com/netbin/internetstats
- [3] Martijn Koster. World Wide Web Robots, Wanderers, and Spiders. http://info.webcrawler.com/mak/projects/robots/robots.html
- (4) Martijn Koster. List of Robots. http://info.webcrawler.com/mak/projects/robots/active.html
- [5] Search Engines in Japan. http://rodem.slab.ntt.jp:8080/w3conf-bof/search.html
- [6] 中川 格. IRA homepage. http://voyager.ulis.ac.jp/
- [7] Kento Tamura. Senrigan search. http://www.info.waseda.ac.jp/search.html
- [8] Peter B. Danzig, etc. Harvest: A Scalable, Customizable Discovery and Acceass System. CU-CS-732-94, Department of Computer Science University of Colorado, Boulder. March 1995.
- (9) Darren R. Hardy, etc. Harvest User's Manual. CU-CS-743-94. Department of Computer Science University of Colorado, Boulder, Version 1.4. January 31, 1996.
- (10) The Harvest Information Discovery and Access System. http://harvest.cs. colorado.
 edu/
- [11] 中本 賢一,山本 毅雄,長谷部紀元.複数ハッシュふるい分け法の日本語情報システムへの応用.情報システムVol.48, No.7.1994.3.15.
- [12] 中川 格. Query Interface to the ULIS Pages. http://voyager.ulis.ac.jp/Harvest/brokers/ULIS/
- [13] 安齋宏幸. JAPAN/MARC提供実験. 図書館情報大学修士課程における研究に伴う試験運用. http://www-student.ulis.ac.jp/contrib/anzai/JMarc3/japanese home.html

13. インハウスデータベース用CGI作成の調査研究

13. インハウスデータベース用CGI作成の調査研究

13.1 目的

インターネットの普及に伴いWWWサーバによる情報提供は急速に進んでおり、企業のみならず行政機関を含む公的機関も、従来の紙による情報提供からインターネットでの情報提供を積極的に行っている。

ネットワークを使用することで、多様で最新の情報を迅速にかつ低コストで多数のユーザに対してサービスが可能なことから、新しい情報提供手段の一つとしてあげられ、さらに静止画像、動画や音声等のネットワークでは高コストでしかも困難であったマルチメディア対応も可能となっている。

またこの機能を利用して、最近ではイントラネットと呼ばれる社内LANでの Webサーバを利用した組織内向け情報サービスが盛んとなり、一層の情報提供機能の拡充に拍車をかけている。

これらの Webサーバを利用したサービスでは、提供されるデータはHTMLエディタ等により 容易に作成でき、また階層的な構造を持つファイルのリンクによる他の情報のアクセスやURL による他の情報源への到達が可能となり、さらに、これの検索・情報を表示するソフトウェアで あるブラウザは、操作が簡単で標準化がなされていることから、初心者でも容易に利用することができる。

しかしながら、Webサーバでの情報提供が進むにつれ、従来の表層的なHTML形式によるデータのみならず、大量のデータの提供や、これらのデータをユーザが指定したキーワードで検索可能とする機能が必要となってきた。また、組織内では共有するデータをデータベース化していることが多く、これらのデータを容易に、他の部門や外部へ提供する傾向が強くなってきた。

そこで、本調査研究では、既に構築済みのデータベースのデータを利用し、これらのデータファイルと Webサーバ間に C G I (Common Gateway Interface)を組み込み、容易にインハウスデータベースの内容をサーバで可能とする機能を調査研究し、その実現を図るとともに、この成果を直ぐに応用できることを実証するため、(助データベース振興センターで作成している「データベース台帳総覧」のデータを使用して、インハウスデータベースの利用普及を促進することとした。

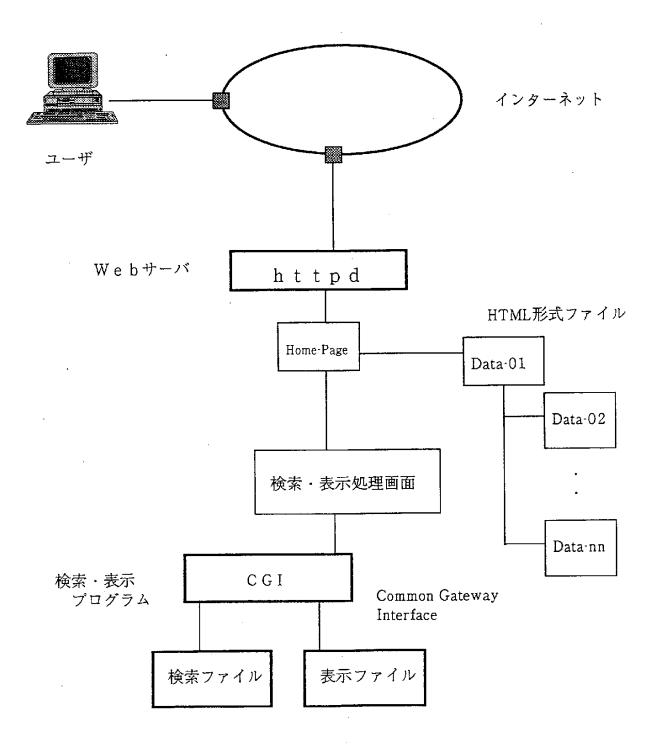


図13-1 システム概念図

13.2 実験システムの概要

13. 2. 1 設計方針

調査研究の実施にあたり、作成される実験システムの設計方針を以下のように定めた。

一般的な検索機能を有する Webサーバでは、サーバ・ソフトウェアに付属して、CGI、データベース・ソフトウェアとCGIとの連携を図るゲートウェイ・ソフトウェアがあり、さらにそのゲートウェイ・ソフトウェアにデータベース・ソフトウェアが接続されている。

これらのゲートウェイ・ソフトウェアとデータベース・ソフトウェアは高価であるとともに、 それぞれのゲートウェイやデータベース・ソフトウェアに依存したCGIを作成することが必要であることから、今回の調査研究では、これらのソフトウェアを使用せずに、パソコンに構築されたデータベースに蓄積されたデータを提供対象とした。

(1) 対象とするデータの選定

組織内で利用されるデータベースは、顧客管理、経理処理等の数値データを主体にした事務処理で多く利用されているが、最近ではワープロの普及に伴い文書情報(文字情報が主体)の電子化が進展し、これらを収集しデータベース化している。

そこで、本調査研究ではこれらの文書情報で構成されるデータベースをソースとし、データとしては、一般的な文書形式であり、本調査研究の普及啓蒙に貢献可能な「データベース台帳総覧」ファイルを対象とする。

(2) CGIの機能選定

① 検索機能

検索条件については、論理和、論理積及び否定の指定が可能とし、検索語は、入力の切替えの煩雑さを無くすため全て2バイトコードとした。アルファベットや記号についても同様に2バイト系を採用する。

② 表示機能

- ・表示は、検索結果の概要を示す画面と、この概要の一部をクリックすることで、当該の 詳細データが表示される画面の二つとする。
- ・HTML形式への変換は、それぞれの処理に従いその都度変換することで、事前のHT ML化作業の煩雑化を避ける。

(3) Webサーバの選定

インターネットでは各種サーバソフトウェアが使用されているが、本調査研究では最も一般的であり無償で利用できるCERN httpdを採用する。

(4) ブラウザの選定

Mosaic、Netscape等の一般的なブラウザを対象とし、日本語の入力、出力が可能なバーションものを想定する。

13.2.2 システムの概要

13. 2. 2. 1 文書情報の形式

(1) 「データベース台帳総覧」のデータを市販のデータベース・パッケージの「ファイル・メーカ II」でデータベース化し、これを原データとした。

(レコード件数 4,576件,総バイト数 4,515Kバイト)

- (2) Webサーバ用のデータ作成のために、このデータベースから必要なデータを抽出し Visnal Basicで、加工編集行った。
 - ① 検索部分と表示部分の区分けとそのファイル作成
 - ・検索対象及び表示対象になるデータ項目の選定
 - ② 表示ファイルの編集
 - ・データ項目の分割(項目の先頭に"#"を付与)
 - ③ 検ファイルの編集
 - ・データ項目の分割(項目の先端に"#"を付与)
 - ・英文字の統一(小文字に統一)

13. 2. 2. 2 ハードウェア・システムの構成

以下のハードウェアをLAN上で使用した。

・インハウス・データベース用 Macintosh Quadra610 32MB

(漢字TALK 7.1)

データ加工用

HP-600C

(Windows 3, 1)

• Webサーバ用

SPARC station 5

- C P U mico SPARC-II (85MHZ)
- MAIN MEMORY 32MB
- HARD DISK 2.1GB
 (Solaris 2.4)

ネットワーク

LAN Ethernet 10B

インターネット接続 64k

サーバマシン Sun Sparc 5 POS/Vマシン HP-600C Macintosh Quadra-650

図13-2 ハードウェア構成図

13, 2, 2, 3 ソフトウェア・システムの構成

利用及び作成したソフトウェアは以下通りである。

インハウス・データベース用

File-Maker Pro-II

Mac-Ms-dos変換ソフトウェア

データ加工用

Visual Basic (新規作成 作成プログラム2本)

・Webサーバ用

WWW server CERN httpd 3.0

CGI用スクリプト言語 jperl version 4.019+1.3(SJIS)

perl CGI ライブラリ cgi-lib.pl 1.8

- データベース検索(新規作成)
- ・データ表示 (新規作成)

13.3 実験システムの詳細

13.3.1 データ整備

今回の調査研究に使用したデータは「データベース台帳総覧」であり、データは以下の構成となっている。

(1) 原データの構成

インハウス・データベースから表示・検索に必要な項目を以下のとおり抽出した。

- (01) レコード番号:
- (02) データベース名:
- (03) ブロデューサ名:
- (04) 国名:
- (05) 内容:
- (06) タイプ:
- (07) 言語:
- (08) サービスシステム名:
- (09) サービス企業名:
- (10) 申告企業名:
- (11) 住所:
- (12) 担当部署:
- (13) キーワード群:
- •(01)のレコード番号はファイル内ではユニークである。
- ・レコード番号を除いて、他のデータ項目は全て2バイト系で表記されている。
- ・データ項目は全てTabコードで区分されている。
- (2) 表示ファイルの処理と例

上記のデータ構成から必要なデータ項目を抽出し、データ項目の先頭に" #"を付与した。

#4001

#The Palm Beach Post

#'The Palm Beach Post

#米国#「Palm Beach Post」は首都圏地域のスポーツ、ビジネス・産業を完全に収録する一般日刊紙です。本紙で扱われる主な分野は、航空・航空宇宙、コンピュータ・通信技術、農家および金融です。この地域に所在している大企業にはPratt&

Whitney社およびAtlantic Sugar Millがあります。全コラムの全文、投書、論説(地域人の訃報を含む)、葬儀告知、地域・ニュースの通信社記事が本ファイルに収録されます。(PAPERS参照)

#1 一般, (新聞)

#報道

#文書(全文).

#英語.

#DIALOG

#Knight-Ridder Information, Inc.

O N.

#株式会社 紀伊國屋書店

1 5 6 東京都世田谷区桜丘 5 - 3 8 - 1

#電子情報部

#4109

#半導体(一般電子部品含む)データ・ベース

#電子デバイス情報サービス(株)

#日本

#半導体及び一般電子部品を含む電子デバイスに関わる情報をデータベース化して提供する。 国内・海外の電子デバイスメーカより提供される、製品に関するデータシート・データブッ クの抄録・原文情報を網羅的に蓄積するほか、新製品ニュース、開発ニュース、保守・廃止 品情報、イベント案内、メーカ・代理店ガイド等、広範な情報を蓄積し、提供する。

#2 自然科学・技術. (電気/電子/情報)

#電気·電子産業

#文書(全文). 文書(抄録). 図表.

#日本語. 英語.

#ELISNET

#電子デバイス情報サービス(株)

#ON. OFF. (MT. CD. COPY.)

#電子デバイス情報サービス株式会社

#102東京都千代田区平河町2-4-14 平河町KSビル

#企画部企画担当

(3) 検索ファイルの処理と例

画面ファイルと同様の処理を行うとともに、英文字についてはすべてを小文字化にした。 #4001

#the palm beach post #the palm beach post

#米国

#「palm beach post」は首都圏地域のスポーツ,ビジネス・産業を完全に収録する一般日刊紙です。本紙で扱われる主な分野は,航空・航空宇宙,コンピュータ・通信技術,農家および金融です。この地域に所在している大企業にはpratt & whitney社およびatlantic sugar millがあります。全コラムの全文,投書,論説(地域人の訃報を含む),葬儀告知,地域・ニュースの通信社記事が本ファイルに収録されます。(papers参照)

#1 一般. (新聞)

#報道

#文書(全文).

#英語.

#dialog

#knight-ridder information, inc.

#on.

#株式会社 紀伊國屋書店

#4109

#半導体(一般電子部品含む)データ・ベース

#電子デバイス情報サービス(株)

#日本

#半導体及び一般電子部品を含む電子デバイスに関わる情報をデータベース化して提供する。 国内・海外の電子デバイスメーカより提供される、製品に関するデータシート・データブッ クの抄録・原文情報を網羅的に蓄積するほか、新製品ニュース、開発ニュース、保守・廃止 品情報、イベント案内、メーカ・代理店ガイド等、広範な情報を蓄積し、提供する。

#2 自然科学·技術. (電気/電子/情報)

#電気・電子産業

#文書(全文). 文書(抄録). 図表.

#日本語. 英語.

#elisnet

#電子デバイス情報サービス(株)

#on. off. (mt. cd. copy.)

#電子デバイス情報サービス株式会社

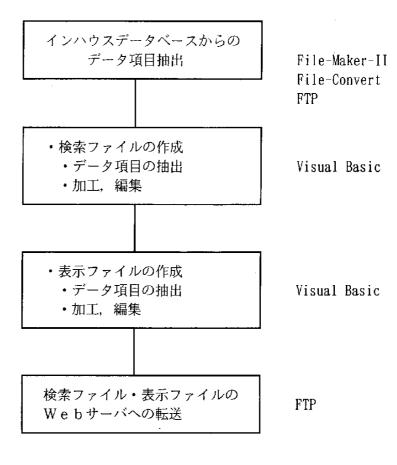


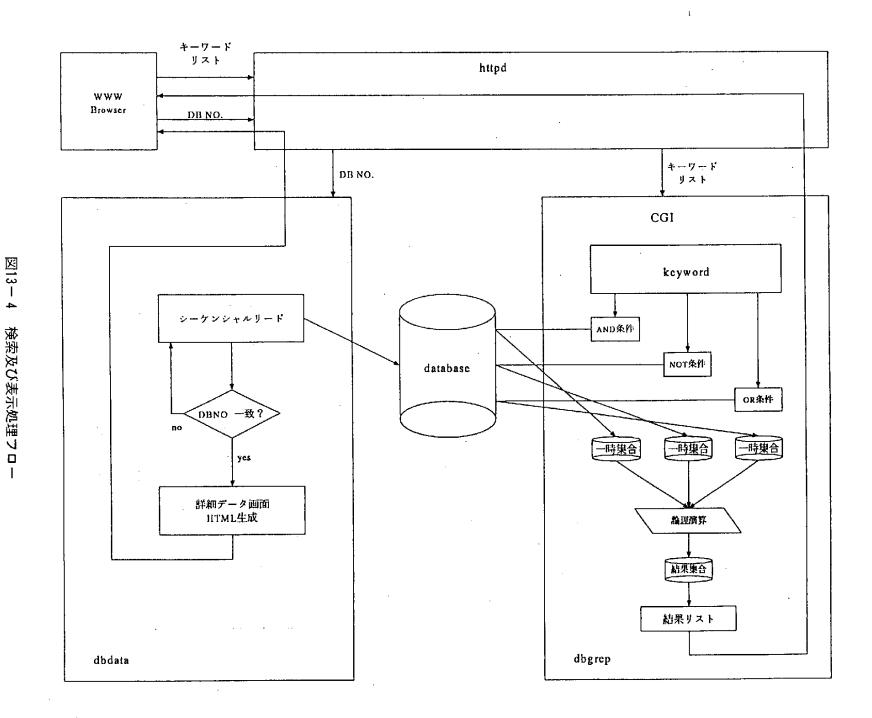
図13-3 データ処理の流れ

13.3.2 検索システム

13, 3, 2, 1 検索ロジック

- (1) 本システムはインハウス向けという事もあり、検索対象となるデータは比較的小規模 (2~3 MB) なテキストのみのものを前提とする。全データを主記憶にロードする為、データ規模は主記憶容量に依存するが、読み込みのオーバーヘッドを踏まえると実用的には 9~10MB程度が限界と考えられる。
- (2) 実際の検索は主記憶にロードされた全件データ (検索ファイル) を対象にperlの関数 grepを用いてユーザによって指定された検索キーワードそれぞれについて行い, 得られた 集合間の論理演算によって検索結果を導出する。

- (3) 検索対象となるテキストデータベース1件1レコードで構成され、項目は任意のセパレータ("#")により区切られていたものとする。
- (4) 検索の実質的な意味は1レコードを1つのオブジェクトとしてみなし、1つのキーワードに対応する(ヒットした)レコードオブジェクトからなる部分集合を生成することである。
- (5) 最終的に得られた検索結果から、必要なデータ項目を抽出し、HTML形式に変換の後、 クライアントに返す。詳細画面の検索に必要なレコード番号も出力対象とする。
- (6) 詳細画面の生成は、クライアントのクリックにより与えられたレコード番号により、表示ファイルを検索し、必要な項目を抽出・編集したのち、クライアントに返す。



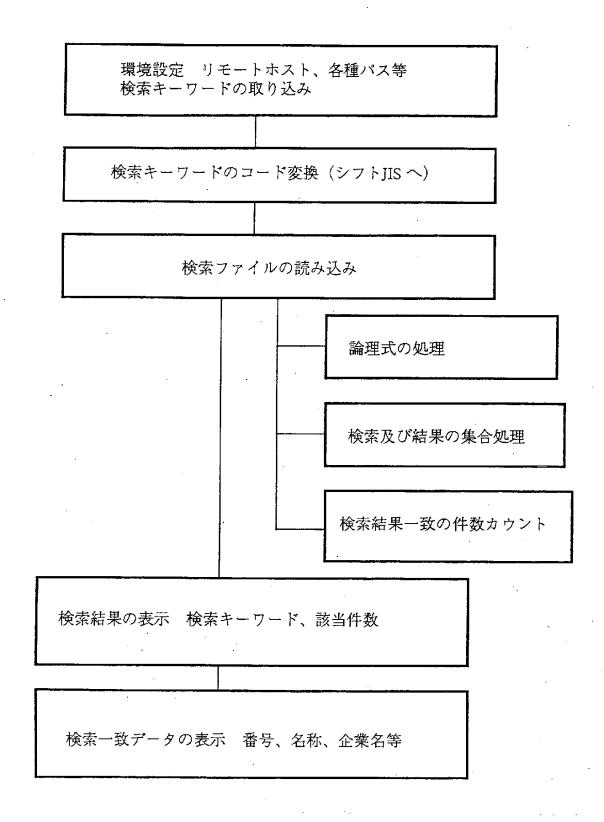


図13-5 検索処理の概要

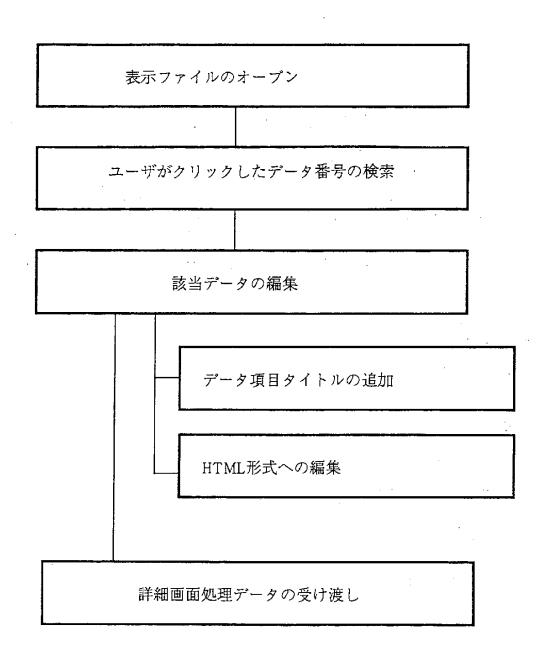


図13-6 検索結果表示処理の概要

13.3.2.2 画面の構成

以下に検索キーワード入力画面、検索結果画面、詳細画面を示す。

NCSA Mosaic: Document View (L10N+)	<u> </u>
<u>File Options Navigate Annotate News</u>	<u>Н</u> еІр
Title: DATABASE SEARCH	
URL: http://www.sci.jip.co.jp/sci/soc/dbgrep.html	
データベース総覧検索	:
ノーダンへ応見伏光	
●検索語を入力して下さい。	•
医薬 +承認	
	:
<u>検索 再入力</u>	
データベース総覧について	
検索式について	
• データ項目について	1

www-admin@sci.jip.co.jp	
Back Forward Home Reload Open Save As Clone New Window Close Window	

図13-7 検索キーワード入力画面の例

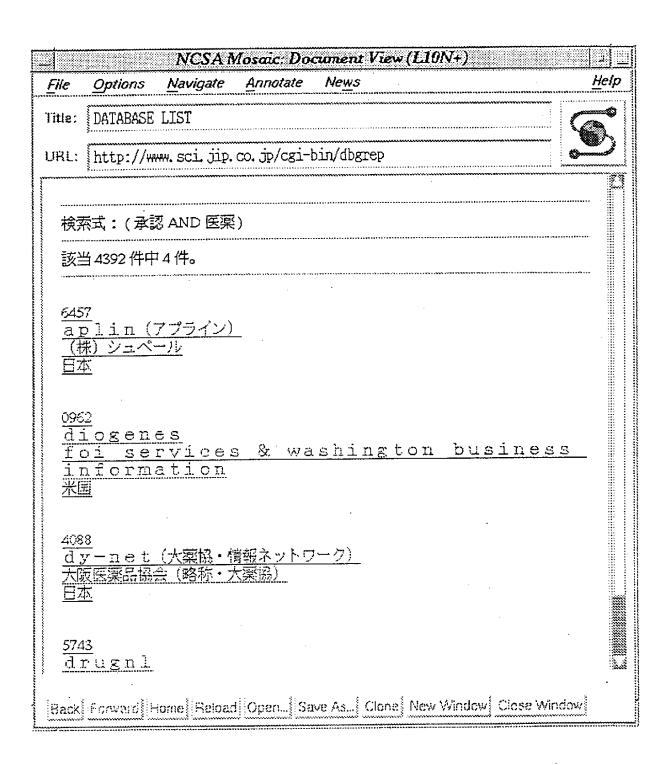


図13-8 検索結果画面の例

NCSA Mosaic: Document View (L10N+) Annotate File **Options** Navigate News Help Title: DATABASE DETAIL URL: http://www.sci.jip.co.jp/cgi-bin/dbdata?6457 データベース名: APLiN(アプライン) 内容: 医薬品等の承認・許認可手続きに必要な情報が満載。改正のもとになった通知 や廃止された通知も見ることができる。また、当該通知に関連する通知も収 録。データは厚生省薬務局審査課の監修。データのダウンロードは自由。 プロデューサ名: (株) シュペール (国名): 日本 分野: 2 自然科学・技術、(医学/薬学/生命学/生物) タイプ:文書(全文). 言語: 日本語。 サービスシステム名: APLiN (アプライン) サービス企業名: 日本電子計算株式会社 問い合わせ先: 住所: 135東京都江東区東陽2丁目4-24 キーワード: 医薬

図13-9 詳細画面の例

Back Forward Home Reload Open... Save As... Clone New Window Close Window

13.4 成果と課題

13.4.1 成果

- (1) 小規模なデータベースから、簡単に Webサーバ用のデータを作成することが可能となった。
- (2) 処理ルーチンやデータの表現方法を明確にしたことで、個人や組織内でのインターネット やイントラネットでのデータベース情報提供仕組みや理解が容易となった。
- (3) 現在の表層的な情報提供から、深みのある大量のデータを提供できることで、ユーザの情報利用が容易となった。
- (4) WSの低廉化やPCでのUNIX利用が可能になっていることから、今後 Webサーバによる情報提供が容易になるだけでなく、提供する情報量の拡大や個人レベルの情報が容易に発信することが可能となり、埋蔵していた新しい情報資源が活用できる。

13.4.2 課題

(1) ユーザ・インターフェースの向上

今夏の実験では、原データの内容は変更せずに直接その内容を検索対象としたが、内容の 表現や表記はデータ提供者に依存していることから、記述されている内容は正規化されてお らず、ユーザの検索結果への満足度は低いと思われる。

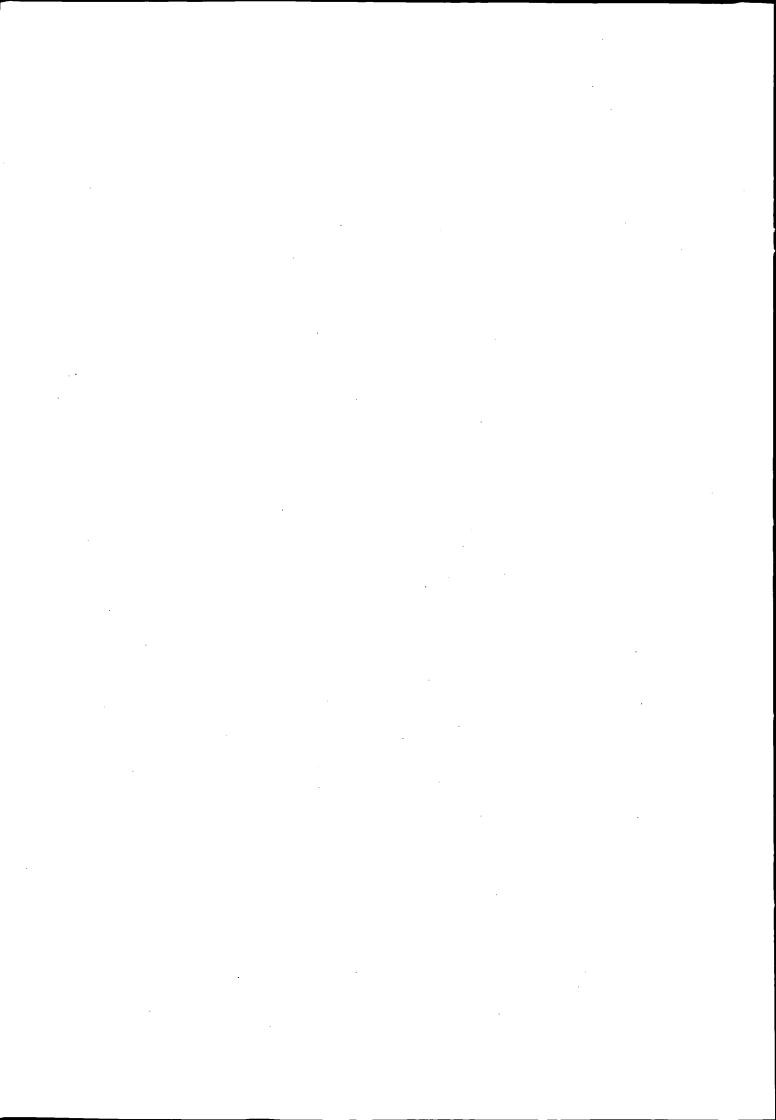
そのため、このシステムの追加機能として、データベースに蓄積されているキーワードの 表示や、同義語さらにはシソーラス等の追加よりユーザ・インターフェースを拡充する必要 がある。

しかしながら、これらの機能を充実するには、標準的なシソーラスの作成、日本語解析の ための環境やソフトウェアの進展が必要となる。とくに、英語処理の環境においては、上記 の機能はすでに整備されていることから早急な解決が望まれる。

(2) プラットフォームの整備

今後は組織内LANの整備とともに、組織や個人での小規模のデータベースの構築作成は 一層進展することが考えられるが、これらに蓄積されたデータは将来組織内のみならず、広 く提供されると予想され、今回の実施した調査研究の成果はこれらに十分に対応できると思 われる。

しかしながら、コンピュータに習熟していない人々がこの成果を容易に利用することは、 もっと簡単にしかも経費のかからない方策を検討する必要がある。このためには、オペレー ティングシステムやプログラミング言語をよりわかりやすいものにすることや、現在市販さ れているソフトウェア・パッケージ等がさらに簡単に利用できる環境の整備が望まれる。



禁無断転載
平成8年3月発行
発行財団法人データベース振興センター
東京都港区浜松町二丁目4番1号
世界貿易センタービル7階
TEL 03-3459-8581

印刷所 株式会社 正 文 社 〒113 東京都文京区本郷 3 丁目12番 2 号 TBL 03-3832-9571



		S				
	·					
	•					
			,			
					·	
			•			
	·					
•						

