

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 7 年 3 月

財団法人 データベース振興センター

KEIRIN

00

この事業は、競輪の補助金を受けて実施したものです。

序

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、データベースの拡大を図る必要がある。

このような要請に応えるため、(財)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に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 6 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 6 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 7 年 3 月

財団法人 データベース振興センター

平成6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
社会	1 報道ニュースのマルチメディアデータベース構築 2 携帯型電子新聞プロトタイプの開発 3 SPORTS POWER INDEXのパイロット版データベースの構築 4 画像付き地図資料データベースのプロトタイプの作成 5 パソコンを利用した副作用症例データベースの構築調査	日本電子計算(株)名古屋支店 (株)日本経済新聞社／(株)日経データ (株)ビデオ・リサーチ (財)地図情報センター 協立医師協同組合
中小企業振興 地域活性化	6 関西イベント&プロジェクトデータベース構築 7 ハイビジョン大型映像ソフト流通促進用データベース構築 8 分散協調型データベースのための高度運用システムの開発 と学習支援環境への応用	(株)京都新聞社 (財)大阪科学技術センター (株)エマーズ
海外	9 電子デバイス情報の電子化ドキュメントのプロトタイプ作成 10 海外ユーザ向けの国産データベース検索支援用端末ソフトウェアの開発	電子デバイス情報サービス(株) カテナ(株)
技術	11 連想検索写真データベースのプロトタイプ作成 12 レーザー研究データベースの構築 13 砂漠化防止技術データベース 14 メディア変換型感性データベースの構築方法の研究 15 フォールトトレランスな多データベースサーバシステムに関する調査研究 16 創出キーワードの自動付与に関する調査研究	シャープ(株) (財)応用光学研究所 (株)新産業創造センター (財)イメージ情報科学研究所 (株)シネジャーナルプロダクション (株)エレクトロニック・ライブラリー

目 次

1. 報道ニュースのマルチメディアデータベース構築	1
1.1 概 要	1
1.1.1 目 的	1
1.1.2 実施内容	1
1.2 実施結果	2
1.2.1 プロトタイプ作成	2
1.2.2 プロトタイプ作成上の問題	4
1.2.3 プロトタイプの評価	6
1.3 今後の課題	7
1.3.1 データベース作成上の課題	7
1.3.2 データベースサービス提供上の課題	7
2. 携帯型電子新聞プロトタイプの開発	9
2.1 開発の背景と目的	9
2.2 実施内容	9
2.3 実施結果	11
2.3.1 携帯型電子新聞プロトタイプのコンテンツ作成	11
2.3.2 携帯型電子新聞端末の基本構成	16
2.4 今後の課題と展望	18
2.4.1 はじめに	18
2.4.2 電子新聞のコンテンツ	18
2.4.3 サービス形態	19
2.4.4 おわりに	20
3. SPORTS POWER INDEXのパイロット版データベースの構築	23
3.1 はじめに	23
3.1.1 SPORTS POWER INDEX(スポーツパワーインデクス)とは	23
3.1.2 スポーツパワーインデクスのデータベースの構築について	24
3.2 仕様の検討	24
3.2.1 パイロット版データベースの目的	24

3.3	検索・集計システムの基本構成	25
3.4	パイロット版データベースの機能	25
3.4.1	検索・集計プログラムの概要	25
3.4.2	検索・集計プログラムの処理手順	26
3.5	今後の課題	28
3.5.1	機能の拡張	28
3.5.2	データベース内容の拡張（意識調査のデータベース化）	28
3.5.3	新たなサービス提供形態の検討	28
3.6	まとめ	29
4	画像付き地図資料データベースのプロトタイプの作成	31
4.1	画像付き地図資料データベースとは	31
4.1.1	本データベースの基本的考え方	31
4.1.2	本データベースの意義と特徴	31
4.2	テキストデータベースの構成	33
4.2.1	キーワードの検討	33
4.2.2	テキストデータベースの設計・記入	34
4.3	画像データベースシステムの構成	35
4.3.1	画像入力システム	35
4.3.2	基本的なコンピュータハードウェア	35
4.3.3	画像入力	36
4.3.4	画像データベースエンジン	40
4.4	今後の課題	40
5	パソコンを利用した副作用症例データベースの構築調査	43
5.1	はじめに	43
5.2	目的	43
5.3	調査方法	45
5.4	調査結果	45
5.4.1	医療現場が求める副作用情報	45
5.4.2	データベース登録項目の選定	46
5.4.3	用語とコードの標準化の試行	47

5. 4. 4	プロトタイプ作成と配布	48
5. 4. 5	データ集約と登録データへのアクセスの効率化	50
5. 5	到達点と今後の展望	51
5. 5. 1	副作用症例データの迅速な収集と提供	51
5. 5. 2	処方設計などの臨床現場での医師への情報サービス向上	51
5. 5. 3	薬剤評価への発展	51
5. 5. 4	患者への情報公開とルールの検討	51
5. 5. 5	更なる標準化の追求と提言	52
6.	関西イベント&プロジェクトデータベース構築	53
6. 1	概 要	53
6. 1. 1	目 的	53
6. 1. 2	実施内容	54
6. 2	実施結果	55
6. 2. 1	運用処理概要	55
6. 2. 2	ラインモード検索処理概要	55
6. 2. 3	近未来データの移行処理概要	56
6. 2. 4	出力処理の概要	56
6. 2. 5	実施体制	56
6. 2. 6	予想される効果	56
6. 2. 7	現システムの問題点	57
6. 3	今後の課題と将来展望	58
6. 3. 1	データベースの再利用	58
6. 3. 2	FAX配信サービスの拡大や音声サービスの展開	58
6. 3. 3	他社とのゲートウェイ接続	58
6. 3. 4	国際都市京都向けに翻訳システムの検討	58
6. 3. 5	マルチメディア対応のバージョンアップを検討	58
6. 3. 6	アクセスポイントの検討	58
7.	ハイビジョン大型映像ソフト流通促進用データベース構築	59
7. 1	概 要	59
7. 2	データベースの特徴	59

7.3	格納されたデータ	60
7.4	更新・追加データとその情報収集について	60
7.5	データの構成	61
7.6	ユーザシステム機器構成	63
7.7	データベースシステムの評価	64
7.8	データベース構築における課題	66
7.9	まとめ	66
8.	分散協調型データベースのための高度運用システムの開発と学習支援環境への応用 ...	69
8.1	目的と概要	69
8.1.1	データベース利用の現状	69
8.1.2	地方からの情報発信	69
8.1.3	今年度の開発 (MineS の強化と学習環境への応用)	69
8.2	MineS アーキテクチャとMineS プロトタイプ版	70
8.2.1	インターネット上の情報サーバの調査	70
8.2.2	MineS アーキテクチャ	71
8.2.3	MineS の実装 (MineS プロトタイプ版開発)	76
8.3	MineS 運用版	77
8.3.1	MineS 運用版の仕様	77
8.3.2	MineS の記述言語	77
8.3.3	MineS の運用評価	77
8.4	学習環境への応用	79
8.4.1	情報環境下における学習環境	79
8.4.2	MineS による学習環境の構築 (MineS2)	80
8.4.3	MineS における教材の扱い	81
8.4.4	MineS によるコミュニケーションの支援	82
8.5	まとめ	84
9.	電子デバイス情報の電子化ドキュメントのプロトタイプ作成	85
9.1	E D S C (Electronic Document Study Committee)の活動	85
9.1.1	背景	85
9.1.2	インターネットでの情報提供	85

9.1.3	研究会の構成と活動	86
9.2	ドキュメント電子化の現状と課題	87
9.2.1	電子化の現状	87
9.2.2	ドキュメント電子化の課題	88
9.3	電子化ドキュメントのモデル	88
9.3.1	メーカ側からの電子化ドキュメントモデルの検討	88
9.3.2	ユーザ側からの電子化ドキュメントモデルの検討	90
9.3.3	E D A ツールに必要な情報	93
9.4	ドキュメント電子化ツールの調査	94
9.4.1	一貫した作業ツールの適合性調査	94
9.4.2	文書作成の工程要素	95
9.4.3	S G M L ツールの調査	95
9.4.4	ツール間の接続性の調査	96
9.4.5	まとめ	96
9.5	検 証	97
9.5.1	S G M L データの作成から印刷まで	97
9.5.2	WWWによる情報提供実験	98
9.5.3	電子図書館システムによる半導体情報の提供	99
9.6	今後の課題	99
9.6.1	標準化	99
9.6.2	文書の構造化	100
9.6.3	情報管理技術としての S G M L	100
9.6.4	使用ツールの整備	100
9.6.5	I S O 9000, P L	101
9.6.6	国際的な標準システムの拡張	101
9.6.7	関係者への教育指導	102
9.6.8	ユーザインタフェース	102
10.	海外ユーザ向けの国産データベース検索支援用端末ソフトウェアの開発	103
10.1	概 要	103
10.1.1	目 的	103
10.1.2	検索サポートシステムの必要性	104

10.1.3	開発方針	105
10.1.4	システムの構成	108
11.	連想検索写真データベースのプロトタイプ作成	109
11.1	研究の目的	109
11.1.1	写真データベース構築の目的	109
11.1.2	画像検索技術の課題とキーアイディア	109
11.1.3	実施内容	110
11.2	意味ベクトル技術	111
11.3	意味ベクトルによる画像検索	112
11.4	写真検索ソフトウェアとユーザインタフェース	113
11.5	画像検索評価用ベンチマークの作成	116
11.6	今後の課題	116
12.	レーザー研究データベースの構築	119
12.1	概要	119
12.1.1	目的	120
12.1.2	実施体制及び実施内容	121
12.2	実施結果	123
12.2.1	作成データベース	123
12.2.2	データベースの利用	126
12.3	今後の課題	127
13.	砂漠化防止技術データベース	129
13.1	概要	129
13.1.1	目的	129
13.1.2	期待される効果	130
13.1.3	実施内容	130
13.1.4	実施項目	131
13.2	実施結果	132
13.2.1	入力情報調査・収集	132
13.2.2	システム開発	133

13.2.3 専門委員会の開催	133
13.3 今後の課題	134
14. メディア変換型感性データベースの構築方法の研究	135
14.1 ネットワーク・マルチメディア・メディア変換	135
14.1.1 インターネットとマルチメディアデータベース	135
14.1.2 マルチメディアとメディア変換	135
14.2 産業・研究分野でのニーズと従来の関連研究	136
14.2.1 印象語による相互変換	136
14.2.2 知覚情報による相互変換	136
14.2.3 信号による相互変換	136
14.2.4 画像と音の統合	136
14.2.5 具体的なニーズ事例・研究事例の紹介	137
14.3 メディア変換の要素技術の現状	137
14.3.1 画像から音響への直接変換	137
14.3.2 物理モデルからの映像と音の同時生成	138
14.4 メディア変換型感性データベースのプロトタイプの概要	139
14.4.1 イメージに基づくイラスト画の検索システム	139
14.4.2 感性情報に基づくイラスト画の自動生成システム	140
14.4.3 音楽からの感性情報の抽出	141
14.4.4 感性パラメータを利用した作編曲支援システム	142
14.5 将来に向けての課題と展望	143
14.5.1 システム提案	143
14.5.2 これからの課題	144
15. フォールトトレランスな多データベースサーバシステムに関する調査研究	145
15.1 概 要	145
15.1.1 目 的	145
15.1.2 実施内容	145
15.2 実施成果の要旨	146
15.2.1 データベースシステムの動向	146
15.2.2 フォールト	147

15.2.3	トランザクション	147
15.2.4	多重化方式	148
16.	創出キーワードの自動付与に関する調査研究	149
16.1	概 要	149
16.1.1	背景と目的	149
16.1.2	実施内容と結果の概略	149
16.2	実施結果	150
16.2.1	実施方法の概略	150
16.2.2	「辞書」と論理式	152
16.2.3	自動付与の実験と結果	153
16.3	実験結果の評価と展望	155
16.3.1	評 価	155
16.3.2	展 望	156

1 報道ニュースのマルチメディアデータベース構築

日本電子計算株式会社 名古屋支店

1. 報道ニュースのマルチメディアデータベース構築

1.1 概 要

1.1.1 目 的

現在、報道ニューを捕えるには、新聞、テレビ、ラジオ、雑誌などの従来からある別個のメディアで捕えている。今後、情報通信技術、コンピュータハードウェア技術、ソフトウェア技術が発達するにつれて、報道ニュース提供のあり方も大きく変わることが予想される。特に、昨今は米国の情報スーパーハイウェイ構想が引き金となって、わが国においても情報通信基盤の整備を始めとして、マルチメディアへの取り組みが活発化してきた。われわれの日常においてもマルチメディア、インターネット、CD-ROM、CATVなどのキーワードを耳にしない日がないほどである。

今回の委託研究では、これらの変化を予想して、異なるメディアを統合した今までにないニュース提供サービスの可能性をプロトタイプ作成を通じて見いだすことを目的とした。

CD-ROM 1枚に新聞記事とテレビニュースをそれぞれ約20項目登録した。新聞からは紙面をめくる感覚で、テレビからはヘッドラインを選ぶ方法で、記事を選択し、新聞記事とテレビニュースを同一画面に表現した。更に、新聞記事をアナウンサー又は機械音声変換によって読み上げることにより、目の不自由な人も新聞記事に接することができる。

このシステムはマウスで手軽に操作できるので、将来、パソコンが一般家庭に普及した場合、誰でも、見たい、聞きたいニュースにアクセスすることが可能となる。

1.1.2 実施内容

本研究では、以下の内容を実施した。

(1) ニュースデータベースの現状把握

まず、テレビニュース及び新聞記事ニュースのデータベース化の現状がどうなっているのか、また、現状でどのような問題点があるのかを各業界の専門家よりヒアリング及び説明を受け把握した。

(2) 将来の報道ニュース提供のあり方を討議

将来、報道ニュースがどのような形態で提供されるか、また、その望ましい姿はどのようなものかを討議した。

(3) 報道ニュース提供の完成イメージ作成

(2)の内容をベースにプロトタイプに組み込むべき機能及び、音声、動画、写真などをど

のように配置するか、画面構成をどうするかを検討し、プロトタイプ在完成イメージを作成した。

(4) データの収集

期間を限定して新聞記事ニュース、テレビニュースデータを収集した。また、同時に報道写真、効果音、用語解説等のデータを収集した。

(5) プロトタイプシステム作成

(3)の完成イメージを具体化するためにシステム設計書を作成し、プログラム化を行い、プロトタイプシステムを作成した。

(6) システム評価

完成したプロトタイプシステムのデモンストレーションを行い、アンケートを実施した。回収したアンケートを集計分析し、システムの評価を行った。

(7) 今後の課題と計画

プロトタイプ作成を通じて得た課題とシステム評価を参考に、今後、どのように本研究を発展させるかを検討した。

1.2 実施結果

将来の報道ニュース提供のあり方を論じ、テレビニュースと新聞記事ニュースを融合させた新しい形態のニュースメディアを想定したプロトタイプシステムを作成した。プロトタイプの作成過程及び完成後のアンケートによる評価を通して問題点を明らかにした。

1.2.1 プロトタイプの作成

プロトタイプシステムの作成に当たり、システムの機能を次の五つに分割した。

- (1) ホットニュース検索
- (2) 分野別ニュース検索
- (3) キーワード検索
- (4) 株式情報検索
- (5) 天気情報提供

システムに登録するデータは、CD-ROMの記憶容量の限界を考慮して、対象ニュースを1994年8月15日(月)～19日(金)の5日間から20件を選択した。また、一部のニュースに関しては、新聞を英語及び機械音声変換で読み上げた。更に、新聞記事の用語解説もシステムに加えた。

プロトタイプシステムの全体構成図を図1-1に、各機能における内容を(1)～(5)に示す。

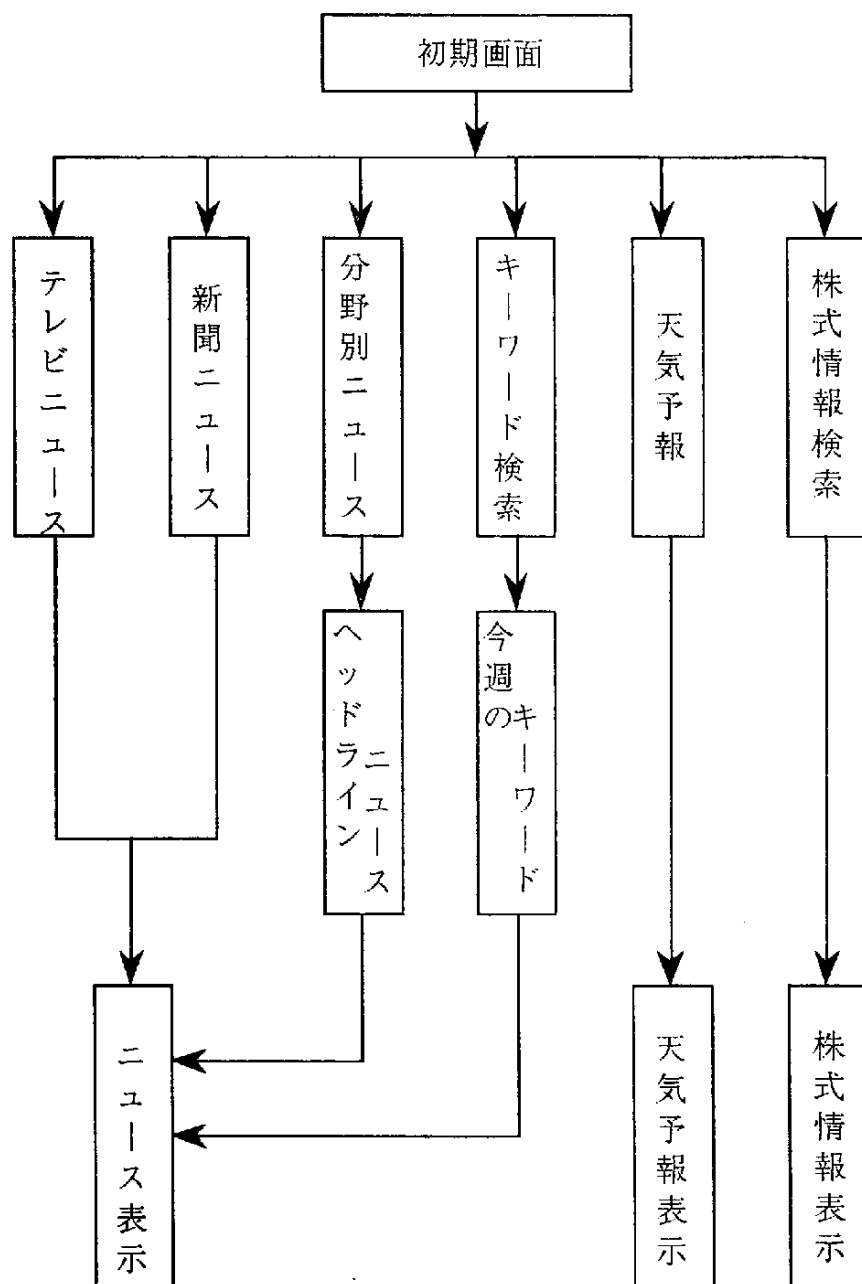


図1-1 プロトタイプシステム全体構成図

(1) ホットニュース検索

テレビと新聞の融合というねらいからテレビ、新聞の両方から検索できる内容とした。
また、1994年8月19日を、現在という設定で、8月19日のニュースをホットニュースとした。

(2) 分野別ニュース検索

8月19日以前のニュースを、「政治・経済」、「社会」、「国際」、「スポーツ」の4分野に分け、分野別に検索可能とした。

(3) キーワード検索

ニュースの中からキーワードを選択し、そのキーワードが含まれるニュースの検索を可能とした。キーワードに関する資料などはアニメーションで表示した。ただし、20件の対象ニュースの中で複数ニュースに存在するキーワードが、今回は、「渇水」だけであったため、キーワード検索は「渇水」のみとした。

(4) 株式情報検索

新聞に掲載されている株式情報を業種分野別に検索できるようにした。株式情報に関するシステムは、今日、ホストコンピュータなどで開発されているため、ここでは複雑な検索方式をとらないことにした。

(5) 天気情報提供

今日（8月19日）の天気、及び週間天気予報を表示できるようにした。

1.2.2 プロトタイプ作成上の問題

プロトタイプを作成する上で次のような問題が発生した。

(1) テレビニュースと新聞記事の不一致

今回の研究では、テレビと新聞の融合を目的としているため、ニュースの内容を映像と文字の両方で表現する必要がある。そのため、テレビと新聞の両方で取り上げられているニュースを題材として選択した。

全国的なニュースの場合、テレビでも新聞でも同じように取り上げられているのでほとんど問題はなかったが、ローカルニュースになると、両者の扱う項目には大きな違いがみられた。また、同じニュースであっても、その切り口（視点）が違い、時には全く違った内容になっていることもあった。一例を上げると、同じ渇水に関するニュースであっても、一方は産業界の対応について扱い、他方は一般家庭の対応について扱っているといったことがあった。そのため、両者の内容を確認する必要があった。

(2) 書き言葉と読み言葉の違い

プロトタイプシステムでは、視覚障害者が新聞に接する一つの方法として新聞記事の音読を行っている。ここでは書き言葉と読み言葉の違いが問題となった。略語や当て字、記事の見出しなどは、そのまま読んだのでは意味が通じないことが多いため、分かりやすく直す必要があった。表1-1に問題となった書き言葉と読み言葉の違いの主だったものを示す。

表 1-1 書き言葉と読み言葉の違い

種 類	新聞での表現	読む際の表現
略 語	四球	フォアボール
	中前打	センター前ヒット
	五輪	オリンピック
	米	アメリカ
	午後四時～同十時	午後四時から同十時まで
() 表現	「(木戸が)担架で～」	「木戸が担架で～」
	ビートたけしさん(47)	ビートたけしさん, 47歳
見 出 し	靖国参拝 自・さ7閣僚	自民・さきがけ7閣僚, 靖国参拝

(3) 著作権（二次使用の問題）

著作権は、マルチメディア時代における大きな問題の一つである。今回の研究は実験を目的としていたため、各方面の理解が得られたが、外電発のニュースや、音楽を使用しているニュースなど、著作権処理が難しい項目については使用を見送らざるを得なかった。

コマーシャル（CM）については、制作会社・広告主・放送局の三者があいまいな関係のまま権利を所有しているため、著作権のクリアには煩雑さを極めた。更に、CMに使用されている音楽は、この三者とは別に、JASRAC（日本音楽著作権協会）などが権利を所有しており、CMの二次使用の問題を複雑にしている。このため、今回のプロトタイプではこれらの処理が容易にできるように、東海テレビ放送の関連会社のCMで、なおかつ使用されている音楽についても、JASRACなどに登録されていないオリジナルの音楽を、作曲者の了解を得た上で使用している。

(4) 作業負担について

(a) 新聞記事データの収集作業

新聞記事単位の文字情報については新聞記事データベースからキーワード、日付などから検索すれば特定できる。該当記事をホストマシンデータベースよりダウンロードすることにより比較的簡単に文字情報は得られた。

新聞一面あるいは記事単位の入力については、今回はスキャナーによる入力を行ったが、新聞製作から直接デジタル情報として得ることは、先行各社では、既に完成している社もあり、システム的な課題は解決しているものといえる。

(b) 報道写真データの収集作業

写真については、写真部、資料部、通信社などから写真を入手した。部署によって、

扱う写真、保存期間、管理方法が異なり、一元化されておらず、システムの取り組みが求められている。

今回は、この事情が反映し、資料集めは、システムのなものは一部しか機能せず、ほとんど手作業で収集した。

(c) 映像データの収集作業

プロトタイプに使用したニュース映像は、テレビ放送内容を録画したビデオテープから、一つ一つ手作業で内容を確認してコンピュータに取り込む作業を行っている。この作業は非常に手間と時間がかかるため、短時間に大量のニュース項目を処理することが難しい。このため、将来的には新聞記事のオンラインデータベースのように、適当なキーワードから映像を自動的に抽出できるようなシステムの構築が待たれる。

1.2.3 プロトタイプの評価

プロトタイプシステムの評価方法としてアンケートによる評価方式をとった。

情報関係に携わる会社員36名を対象に、1995年1月13日（対象14名）と1月18日（対象22名）に分けて、プロトタイプシステムのデモを行い、アンケートを実施した。

アンケートによる評価結果をまとめると以下ようになる。

- (1) 新しい形態のニュースメディアとしての使用意向が高い。
(プロトタイプのようなシステムを使ってみたいと回答した人の割合が63%)
- (2) 「自分の仕事のため」、「教養を高めるため」に利用できる。
(30%の人が「自分の仕事のため」、「教養を高めるため」として利用できると回答)
- (3) 現行の新聞、テレビとの競合は少なく、全く新しい形態のサービスとなる可能性が高い。
(新聞は52%の人が、テレビは38%の人が今まで通り、新聞、テレビは見ると回答)
- (4) システムの月間使用料金は5,000円以下が望ましい。
(91%の人が1ヵ月当たりの使用料金は5,000円以下が適当と回答)
- (5) システムの中に広告(CM)を入れることは、使用料を安くするためには、ある程度必要である。

現在の技術及び制度上では、異なるメディアを統合データベース化し、即時に、新しい形態のニュースをサービスするには、多くの課題が残されている。しかし、今回の実験は、将来、新しい情報提供手段の可能性を秘めたものであったといえる。

1.3 今後の課題

報道ニュースのマルチメディアデータベース（異なる報道ニュースメディアを統合した新しいメディア）のプロトタイプを作成した。今後、実用化を目指すには次のような多くの課題がある。テレビ映像データベースの課題、新聞記事データベースの課題、データベース作成上の課題、データベースサービス提供上の課題などである。

1.3.1 データベース作成上の課題

(1) データベース自動作成の開発

日本におけるテレビ映像、新聞記事データベースは、それぞれの放送局、新聞社によって作成されつつあるが、まだ完成されたものはない。ニュースを特定するための手法が完成されていない。したがって自動的に放送ニュースと新聞記事ニュースをドッキングできない。データベースの自動作成の技術開発が実用化の前提となる。それまでは人手による作業が多くあり採算の合わないものとなっている。

(2) 圧縮技術、電子媒体の研究開発

今回のプロトタイプของデータベースは、CD-ROM上に製作したが、文字、音声、静止画、動画を取り込むための容量が不足している。約600MBで約20項目しか構築できなかった。圧縮技術の開発が必要である。また別な電子媒体の開発応用が求められる。

(3) 機械音声変換、機械翻訳の研究開発

文字情報から音声情報への機械音声変換の採用には品質改善が必要である。専門のアナウンサーの読みと比較してまだ技術は低く実用化にはほど遠い。また、現在は日本文の英文機械翻訳には相当の人手をかけないと実用にはならない。

(4) 検索方式の開発

検索は、ごく一般的なメニュー方式であり、検索の容易性の面から更に研究開発が求められる。どのような検索方式がよいか今後の開発が待たれる。

1.3.2 データベースサービス提供上の課題

(1) 通信回線整備

「ニュース・オン・デマンド」（必要に応じて、誰にでも、安価にニュースを提供する）にするためには、CD-ROMだけではその一部の目的を達したにすぎない。通信回線を通してオンラインでの提供が求められる。オンライン提供のためには通信回線の整備が必要である。

(2) 著作権の課題

新聞記事にしても、放送映像にしても、単独メディアとしての著作権はクリアしている。それぞれのニュースの二次利用についての著作権問題の解決がサービス提供の前提である。

(3) 採算の課題

採算の合うデータベース構築・提供を目指したコスト低減が課題である。提供サービス料金は、アンケート調査によると月 3,000円程度でないと利用されない。

(4) 視聴の明確さ

一つのディスプレイに、常時二つ以上のメディアを強調表示することは検索のとき以外好ましくない。選ばれたメディアである文字、音声、映像の媒体を強調し、表示することよりメディアの判読が明確になる。いったん選ばれたテレビ放送、新聞記事、音声ニュースをより強調した形態で利用できるようにし、他のものの扱いは小さくてよい。

(5) 鮮明度の向上

今回のディスプレイは17インチであったが、新聞紙面の場合文字の判読が難しかった。より大きなディスプレイ又は鮮明度向上が求められる。

(6) 利用端末の普及

新しい電子情報メディアの成否は利用端末の普及にかかっている。いくら適当な料金でサービスしても初期投資（利用端末投資）が高くては成功しない。多くの家庭での総合端末（現在のテレビ受像機、ファクシミリ端末、ワープロ機器、パソコン機器、電話機器などの総合機器）の普及が課題である。

最後に、最近のコンピュータの発達は、マルチメディア化の方向に発展しようとしている。報道ニュースについてみると、一つの情報メディアだけでなく、多くの情報メディアを同時に捕えることも可能になるだろう。必要なときに、最新のニュースから過去のニュースまで遡って、新聞ニュース、ラジオニュース、テレビニュースを容易に手に入れることが可能になるであろう。これを「ニュース・オン・デマンド」ということができるであろう。

21世紀には「1.3 今後の課題」は解決されることになるだろう。ここで挙げているのは技術的課題である。そのほか社会的課題（著作権、社会的ニーズなど）がある。「1.2.3 プロトタイプの評価」によると、このような「報道ニュース」サービスがオン・デマンドで提供された場合、あなたはどうしますか？ との問に対し、「ぜひ使ってみたい」、「使ってみない」の2回答で63.9%を占めた。これは注目すべき結果である。今後、実現化への課題をクリアするための研究開発と市場への実験が重要となろう。

2 携帯型電子新聞プロトタイプの開発

株式会社 日本経済新聞社
株式会社 日経データ

2. 携帯型電子新聞プロトタイプの開発

2.1 開発の背景と目的

昨年度実施した委託事業「新しい電子情報サービスに関する調査研究」結果から、新しい電子情報サービスの形態別、アプリケーション別の21世紀における将来市場の展望及びそれらが社会に普及するために必要な要件が明らかになった。特に、“電子新聞”に対する新しいメディアとしての期待は大きい。しかし、マルチメディアをベースにしたインタラクティブな電子新聞の具体的なイメージが分からない。購読料金を安価に抑え、消費者市場に対応させるための広告掲載のあり方などの問題点が指摘された。このような前年度委託事業の結果を踏まえて、近い将来実現可能となる電子新聞の具体像を描くため、特に消費者市場をターゲットとした携帯型電子新聞のプロトタイプを作成し、操作・実演することを通じて電子新聞としての情報コンテンツ、利用形態等を詳細に検証する。

また、携帯型電子新聞の事業化を進めるに当たっては、コンテンツとともに、消費者への配信(Delivery)をどのようにするかも大きな問題である。2010年ごろまでを視野に入れて、マルチメディア関連情報の伝達方法、携帯型端末機やソフトウェアなど技術動向、デジタル化された情報の著作権問題、及びマルチメディア取材・編集体制のあり方を探り、携帯型電子新聞実現へのタイムテーブルを作成する。

2.2 実施内容

開発実施に当たっては委員会設置とともに、携帯型電子新聞のコンテンツと携帯型電子新聞のサービス形態とに分けてワーキンググループを組織し、プロトタイプ開発及び携帯型電子新聞を実現するための技術的、社会的要件の考察を行ってきた。

- (1) 日本経済新聞社並びに関連会社のデータを中心に、記事、写真、映像、図表、音声などマルチメディア新聞情報の素材を収集・編集し、携帯型電子新聞として魅力ある情報コンテンツを作成した。愛称をNIKKEI NEWS BOY(日経携帯型電子新聞)とした。記事、広告などの検索は、アイコン表示をペンタッチで簡単に操作できる。記事をクリッピングしてフロッピー等にダウンロードすることも可能である。また、慶応義塾大学湘南藤沢キャンパスのMMM(MultiMedia Modeling)プロジェクトの学生からも、携帯型電子新聞という枠を外して新しいメディアの創造という大きな視点から斬新なアイディアを出してもらい、スペシャルメニューとして情報コンテンツを作成した。

(2) 携帯型電子新聞を新しいメディアとして実現するための技術的、社会的要件を検討した。

- ① 携帯型電子新聞端末、ハードウェアの展望 —— CPU, 通信用IC, DRAM, フラッシュメモリー, ディスプレイ, バッテリー, 電源電圧等
- ② ソフトウェアの展望 —— 圧縮技術, オーサリング技術, マルチメディア通信プロトコル等
- ③ データ伝送手段とコストの見通し —— BS/CS, 文字放送, CATV, B-ISDN, 移動体通信・PHS等
- ④ マルチメディアデータベースの著作権
- ⑤ マスメディアのマルチメディア取材・編集体制のあり方

なお、われわれがイメージする携帯型電子新聞は、図2-1のように描ける。即ち、フラッシュメモリーカードなどの外部記憶媒体に、夜間に電話回線又は電波放送などを経由して、翌日分の新聞記事・関連映像等を電子宅配し、朝には携帯端末で電車の中で読むことができる。

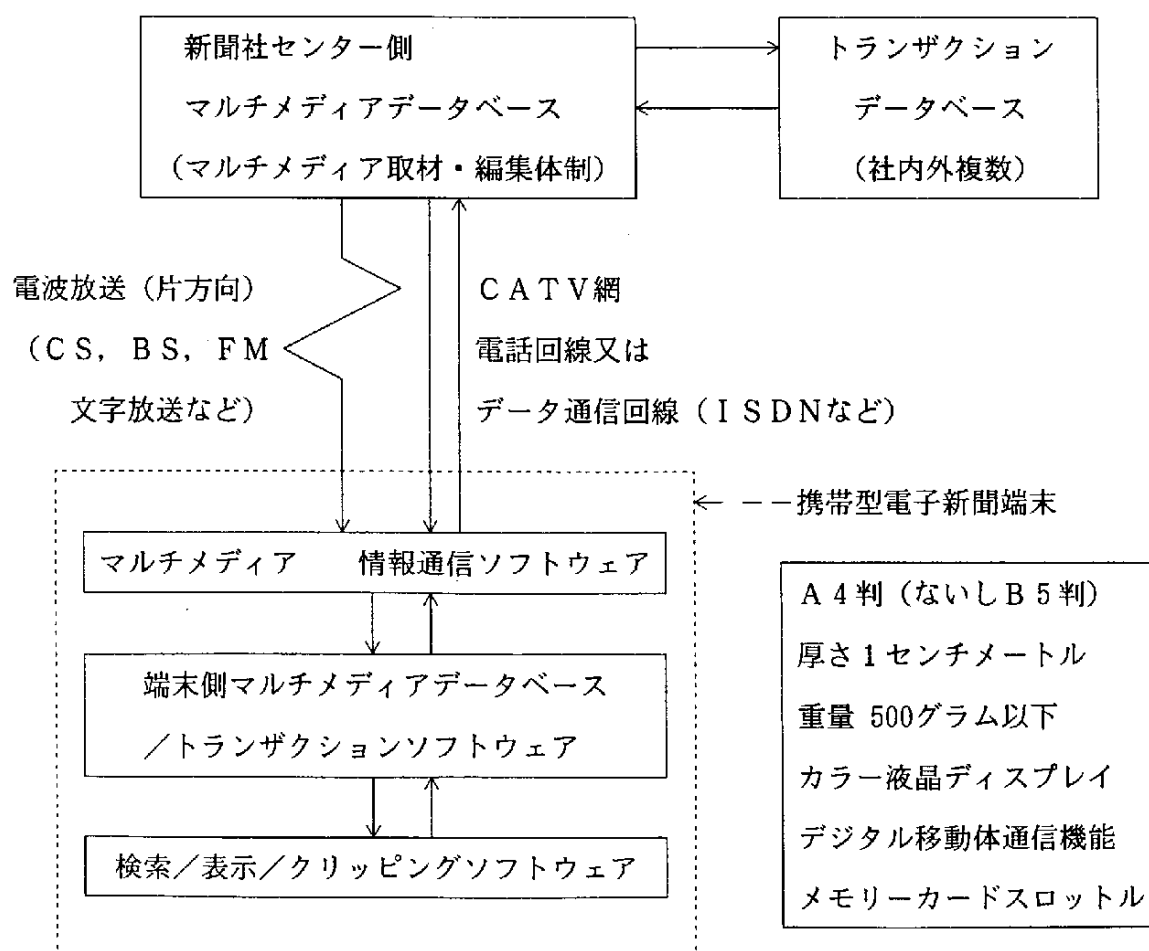


図2-1 携帯型電子新聞のサービス概念図

すべての操作はアイコンをペンタッチするだけで、誰でも手軽に読みたいジャンルの記事、写真、グラフや音声を引き出せる。満員電車の中でも新聞を小さく折る必要もない。イヤホンで読みたい記事の図表を見ながら、聞くこともできる。自分の行きたいイベント情報を発見したときは、即座にトランザクション機能を使って、商品を注文したり、チケットを予約・購入できるといった消費者市場を意識したものである。

2.3 実 施 結 果

2.3.1 携帯型電子新聞プロトタイプの内容作成

(1) コンテンツの基本的フレームワーク

今回開発するプロトタイプの内容のファイル構造は、図2-2に示したように①定時情報、②速報情報、③遡及（クリッピング）情報、④用語解説、⑤サウンド（BGM）から成り立っている。

① 定時情報エリアには、ニュース及びニュースのバック情報、広告などが収録される。

紙メディアの新聞が毎朝宅配されてくるように、携帯電子新聞も文字、写真、CG、映像、音声などからなるマルチメディア編集情報が、衛星通信あるいはB-I SDNなどによって夜間のうちに電子宅配される。ユーザは文字を読む代わりに図表や写真を見ながら音声で聞くこともできる。今回は音声情報としてアナウンサーが朗読した情報を別途収録し、オプションによりテキスト情報と音声情報を選択できるようにした。将来は携帯型電子新聞端末機に内蔵されるソフトウェアによって、テキストから音声への自動変換が考えられる。

② 速報情報エリアには定時情報を受信した後、新しく発生したニュースが収録される。

現在発生中のニュースは、自宅やオフィスであれば衛星通信や電話回線で、屋外であれば移動体通信などでリアルタイムに見ることができる。

③ 遡及（クリッピング）情報エリアには、新聞購読者が必要な記事を切り抜いてファイルするように、電子新聞の定時・速報情報をクリッピングして収録できる。テキスト情報であれば3ヵ月間は蓄積可能であり、いつでも必要に応じて検索できる。長期間にわたる遡及記事は、別途用意されるフラッシュメモリーカードやCD-ROMなどのパッケージ情報を利用することもできる。あるいは新聞社のデータベースから直接検索することもできる。

④ 用語解説エリアには、日経の経済新語字典や人名・地名などを収録し、定時情報や速報情報の記事の用語解説を参照できる。

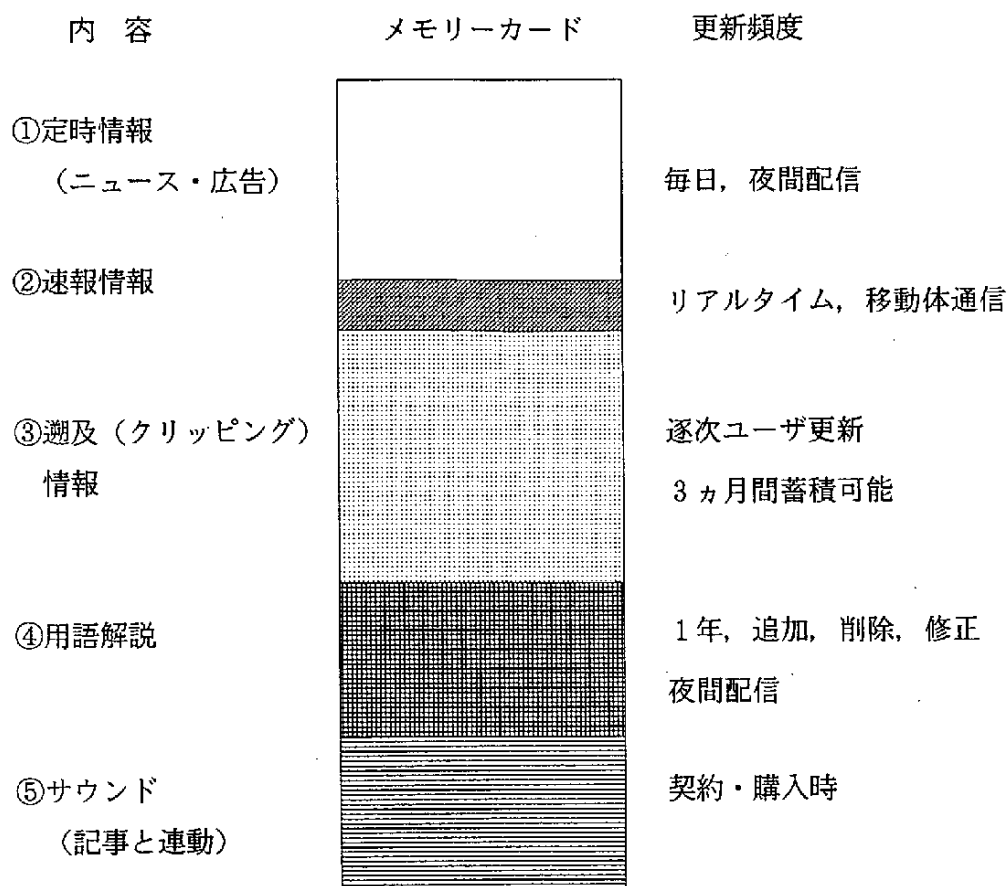


図 2 - 2 プロトタイプデータファイルの構造

⑤ サウンド (BGM) はテキスト情報と連動しており、ニュース画面のバックグラウンドミュージック (BGM) として利用される。BGMの切替えはオプション指定により、ユーザが自由に操作できる。

(2) 素材データ及びファイル容量

携帯型電子新聞プロトタイプのコンテンツ作成に当たっての素材は、紙メディア新聞との比較を可能にし、その特徴を明確にするため、次の特定日付の日本経済新聞を使用した。

- ① 1994年 8月20日 (土曜日) 付、日本経済新聞朝刊の一部
- ② 各記事に関係するニュース映像 (テレビ東京による提供協力)、CG等
- ③ 各記事に関する遡及記事データ
- ④ その他 (広告、1995年 1月のテレビ番組欄等)

記事本数はフルテキストで約 140本、遡及記事約 160本、写真及びCGは約 100枚、動画は11記事で約12分34秒、音声情報はヘッドラインが62本で約13分、ナレーション (記事の朗読) が25本で約25分である。これらを合わせてファイル容量は約 223メガバイトである。

(3) 広告の扱い方

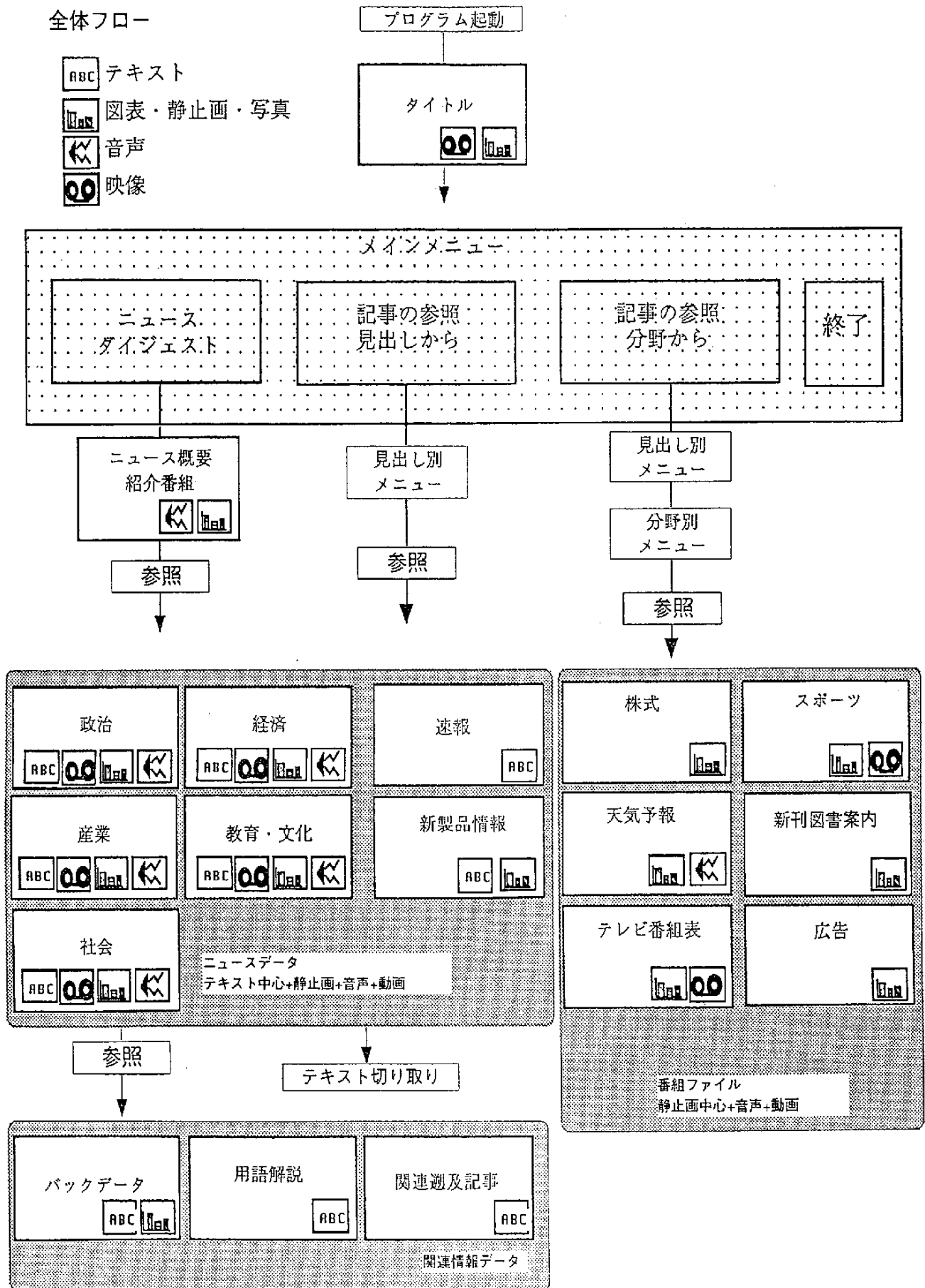
現在の新聞が家庭への配達料金を含めて月額 4,300円（日本経済新聞）と低価格で提供できるのは、掲載される広告に負うところが大きい。携帯型電子新聞が、現在の紙メディア新聞と同じように消費者市場に浸透していくためには、できるだけ低廉な価格設定が必要である。そのためには、広告の扱い方が大きなポイントとなる。

電子新聞に収録される広告は、紙メディア新聞のようにめくっていくことによって「思いがけない出会い」といった広告効果は期待できない。ユーザが広告を参照するという行動は、情報としての目的意識をもった検索であろう。その一つのサービス形態としてトランザクションサービスが考えられる。また、4th Wave, Inc. のJohn Latta氏の調査によると、米国では最近、CD-ROMを起動した際に製品のビジュアル広告を3～4秒間画面表示する、いわゆる雑誌の裏表紙広告のようなカメオ広告が登場しているとのことである。この広告に対する消費者の反感はないとのことである。本プロトタイプでは、ウィンドウ画面を設けて絶えず広告を流し、クリックすると広告の詳細が表示される方法と、広告データベースとしての新製品情報をメニューに加える方式をとっている。後者は、今回実験できなかったが、新製品情報メニューからトランザクションサービスへの連動を考慮したものである。

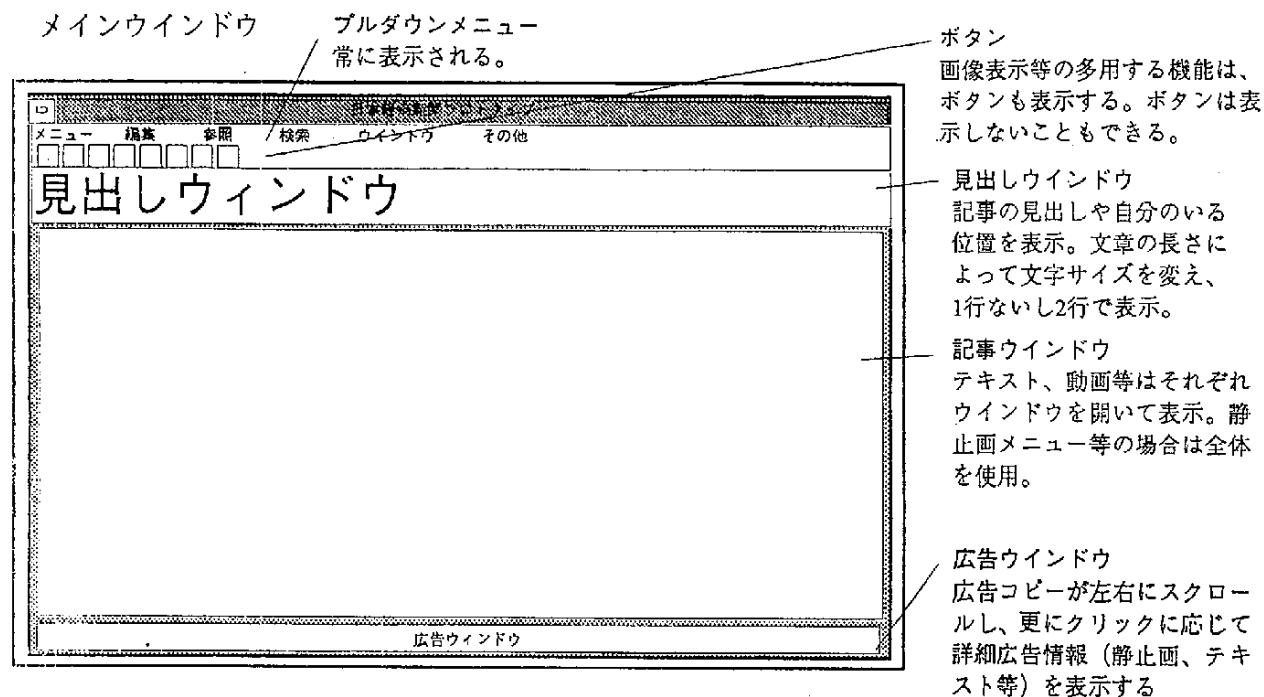
(4) 全体のフローチャート

全体フロー

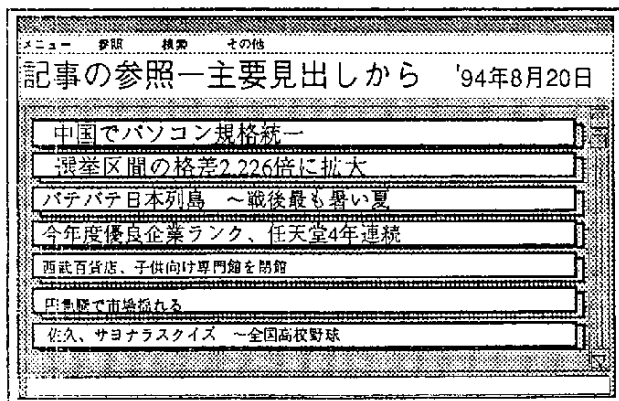
-  テキスト
-  図表・静止画・写真
-  音声
-  映像



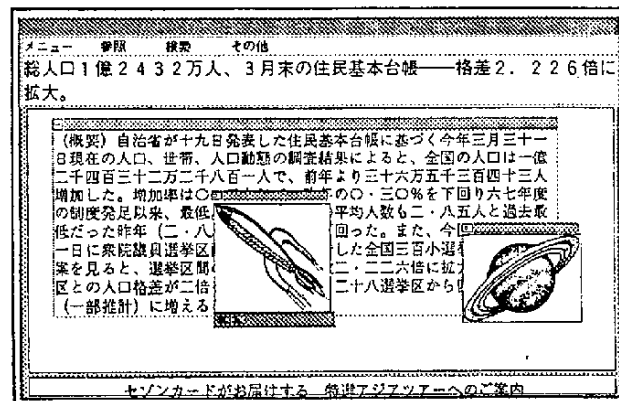
(5) 基本的な画面構成



メニュー表示例

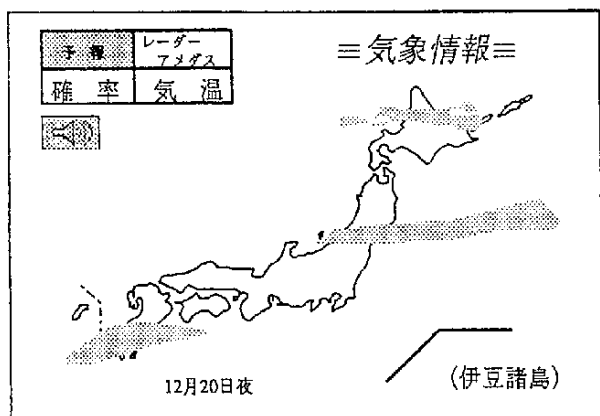


本文（記事）表示例

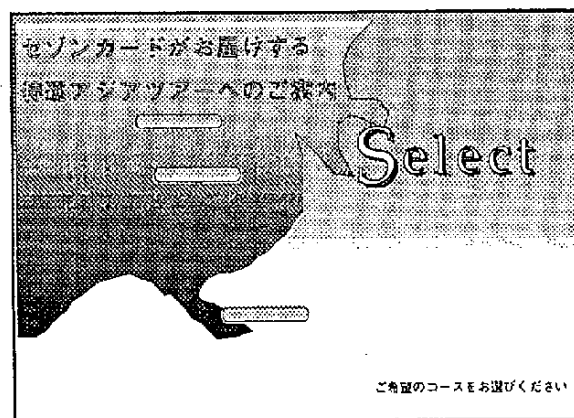


番組呼出し時

本文（静止画）表示例



広告表示例



2.3.2 携帯型電子新聞端末の基本構成

電子新聞端末の基本構成を模式図にまとめたのが図2-3である。携帯型電子新聞プロトタイプ開発ワーキング（サービス形態検討）委員会で詳細に検討した情報伝送手段（片方向・双方向）のほか、ハードウェアのカギとなる表示装置、並びに外部記憶装置について、考えられる選択肢の組み合わせを示した。

おおまかにいって、電子新聞端末ハードウェアは次の2形態に分けて考えることができる。一つは基本的に家庭やオフィスの決まった場所に据え置いて、有線での通信が可能なシステム。もう一つが、今回メインに考えているかばんの中に入れて簡単に持ち運べる薄型軽量の電子新聞端末で、無線方式での情報収集が可能なタイプのものである。

われわれの構想する電子新聞の基本的な利用形態は、この据え置きの端末に夜間に翌日の新聞の大量のデジタル情報を電子宅配してメモリーカードなどに情報を蓄積し、翌朝メモリーカードを脱着して携帯型端末（ビューア）でマルチメディア情報を検索・表示させるというものである。もちろん、この二つの端末は一つのハードに統合することができる。

まず、据え置き型端末では、利用可能な通信手段として、通信衛星（BS/CS）、CATV網、B-ISDN、そしてもちろん公衆電話網などが考えられる。衛星通信は無線方式であるが、BS/CSアンテナは指向性が強くサイズの持ち運びは困難なので、実質的には有線通信と同じく、据え置き型端末での利用となる。BS/CSを利用すれば、リアルタイムで、しかも規模の経済性がはたらきユーザ数が増えるほど通信料金は安く、大量のマルチメディア情報を伝送することが可能である。持ち運ぶことを考えないので、表示装置は従来のCRTや、又はTVモニターでもかまわない。外部記憶装置はメモリーカードの他、ハードディスクドライブ（HDD）も利用できるだろう。もし、家でしか新聞は読まない、しかも決まった場所で読めればいい、というユーザがいるならば、今回の携帯型電子新聞の基本コンセプトからは外れるが、この据え置き型端末だけでも用が足りることになる。

次に、携帯型電子新聞端末では、利用可能な通信手段として、FMデータ伝送、PHSなどの携帯電話システムが有望である。表示装置は携帯性を考えると、2000年頃までの技術動向からみて現実的なデバイスは液晶しかないだろう。外部記憶装置としてはハードディスクと同様に書き換え可能で、電源を切っても記憶内容が失われないフラッシュメモリーカードが軽量小型で理想的だが、ハードディスクよりかなり高価になるので、採用にはコストとの兼ね合いが必要と思われる。

先に述べたように模式図ではこのように2形態に分けて考えたが、実際の端末は携帯型の一つのハードであり、有線又は衛星とアダプターを介して接続可能なインタフェースを備えていけばよい。携帯中にもリアルタイムで情報を取れるFM、PHSなどはオプションと考えることができる。

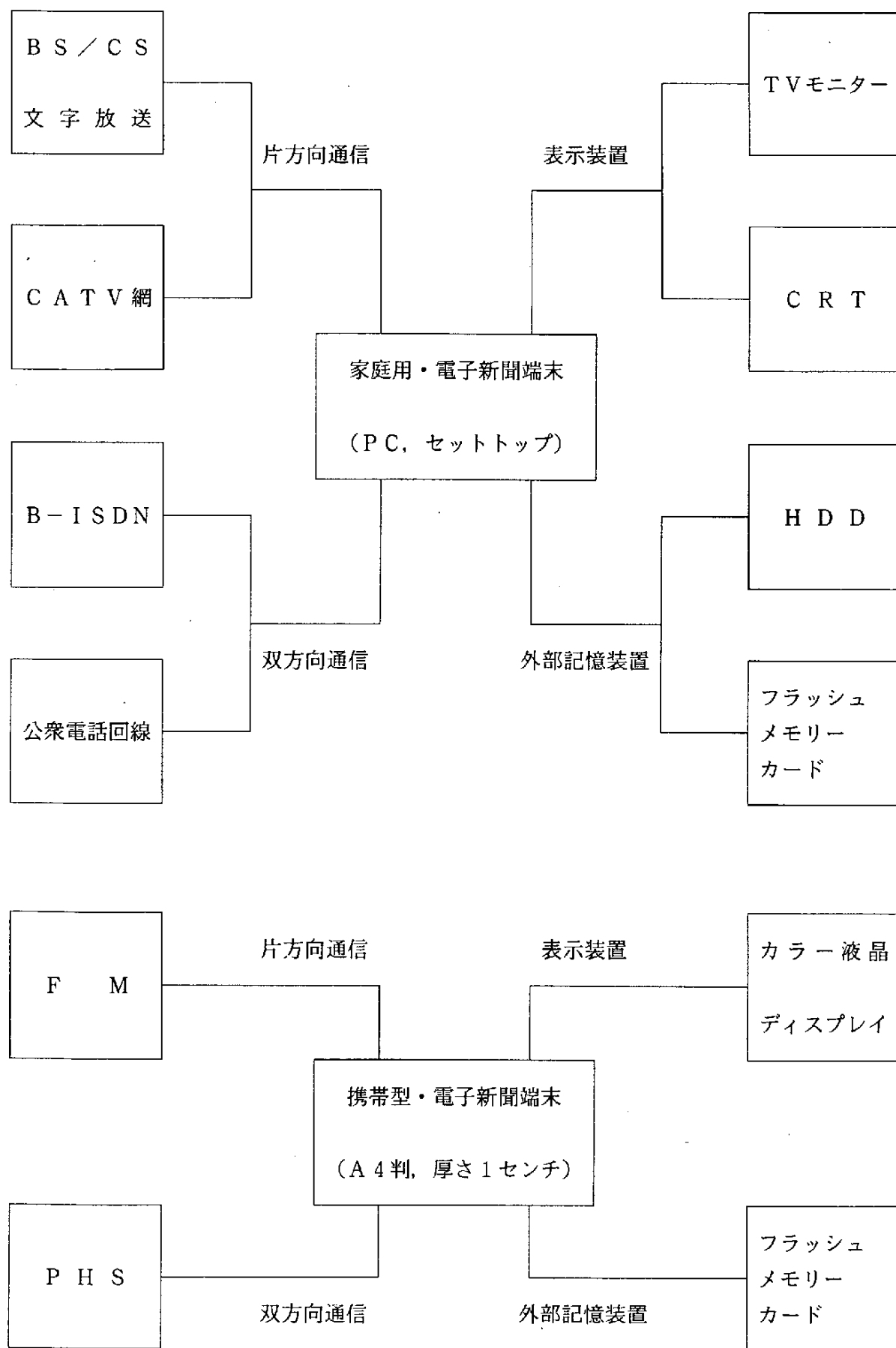


図 2 - 3 電子新聞端末の基本構成

2.4 今後の課題と展望

2.4.1 はじめに

昨年度の「新しい電子情報サービスに関する調査研究」の結果を受けて、今年度は携帯型電子新聞のプロトタイプを開発することになった。コンピュータの出現とともに進むかと思われたペーパーレス化は、むしろ反対に紙使用の増加という結果を生じてしまった。しかし近未来には、遠からず紙メディアの新聞から電子メディアの新聞に移行するなど、本格的なペーパーレス化を指向している。特に、ノートパソコンやパームトップコンピュータやPDAなどの普及によって、携帯型の電子新聞が現在の紙メディアの新聞の役割の一部に置き代わる可能性は小さくない。

本委員会では、電子新聞のコンテンツと電子新聞のサービス形態とに分けてワーキンググループを組織し、プロトタイプ開発を行ってきた。特にコンテンツについては、携帯型という枠を外して、近未来の電子新聞のイメージ作りという視点から、慶應義塾大学湘南藤沢キャンパス（SFC）のMMM（MultiMedia Modeling）プロジェクトの学生メンバー達からもアイディアを出してもらう形を取った。

2.4.2 電子新聞のコンテンツ

最初に電子新聞の位置付けを明確にしておく必要がある。電子メディアの新聞は紙メディアの新聞と共存すべきものか、それとも置き代わるべきものか。あるいはまた、電子メディアの新聞とテレビとはどのような棲み分けをすべきか、われわれがこれまでに議論した範囲では、電子メディアの新聞はネットワーク社会での新しいジャーナリズムの一つであり、従来の紙メディアの新聞やテレビなどの果たしている役割と一定の部分は重複するけれども、それらに置き代わるものではない。即ち、テレビのもつニュースの速報性や、新聞のもつ質の高い報道や教養文化の啓蒙といった機能を部分的に含む。また、ネットワーク上でアクセスできるデータベースへのインタフェースの役割も果たす。

電子新聞は、次のようなキーワードで特徴付けすべきである。まず第一に、楽しくわくわくするようなメディアであること。新しい技術が可能にする電子新聞が、社会に定着し生活に浸透するには、相当の魅力をもったものでなければならない。意外と保守的なジャーナリズムの世界に新しい切り口をもち込むには、わくわくする楽しさという性格を欠くことはできない。

二番目に、インタラクティブなメディアであること。情報の供給側が一方的に利用者に情報を流すだけのメディアなら、ネットワーク環境を利用する意味が少ない。マルチメディア

のもつインタラクティブな性格を、電子新聞でもフルに生かすべきである。

第三に、前項とも関連しているが、利用者がカスタマイズできるメディアであること。インタラクティブであれば、利用者がそれぞれの好みに応じたサービスを受けることができるはずである。そうしたカスタマイズの仕掛けを用意しておくことで、これまでの紙メディアの新聞とは異なったメディアになる。既に1980年代半ばに、MITのメディアラボのニューズピークの研究グループが、そのようなコンセプトを取り込んだプロトタイプを制作している。

第四として、視覚情報をより重視し、動く新聞という形であること。紙メディアであってもコンピュータディスプレイであっても、二次元の情報表現であるということは同じである。しかも新聞紙に比べて、コンピュータディスプレイの方はずっと画面が狭い。そこで、画面に表示する情報に動きをもたせることによって、三番目の次元を導入し、限られた画面の大きさを有効に使うべきである。また、文字中心の情報表現では、まだ紙メディアの方が電子メディアよりもずっと優位にある。むしろ映像などの視覚的な情報表現を多用することで、電子メディアの優位性を確保できる。

ここで動く新聞とは、単にビデオ映像を多用するというだけではない。むしろ、従来の新聞記事の一次元的な書き方でなく、電子新聞向きの平面プラス動きという三次元的な書き方を、新しく確立していく方向を考えるべきである。この点に関しては、慶応義塾大学SFC学生の明田守正君が、最初のアイデアとプロトタイプを出している。ここを出発点として、電子ジャーナリズムのあり方を新しい視点から見直してみることは、極めて大きな意義がある。

電子新聞が視覚情報を重視するとはいっても、音声音響を利用することもまた大切である。音声音響もまた、もう一つの次元を与えてくれるからである。単にBGMを流すだけといった使い方だけでなく、部分的に記事内容の朗読を伴うなど、積極的な利用方法を講じるべきである。一般に耳で聞いた内容は意識のなかに停留することが多い。そうした残留した情報の断片を画面上の情報表示と重ねることによって、人間の体験記憶を強め、記事を立体的に伝達することができる可能性がある。

2.4.3 サービス形態

電子新聞の普及で最も大きな問題は、情報伝達方法と、財政的な経費回収方法である。情報伝達に関しては、テレビのような無線方式、電話のような有線方式、本やCD-ROMのようなパッケージ方式をすべて採り入れた多面的なものにすべきであろう。新聞の縮刷版や用語辞典のような内容は、CD-ROMのパッケージと、インターネットのような回線とを

使い分ける。定時情報については、やはり回線から夜間にダウンロードする。そして速報性のある情報は、移動体通信・PHSのように無線通信による。それらを合わせた複合体として、一つの電子新聞を形作ることになる。サービス形態のワーキンググループでは、それぞれの方法ごとに、現在から近未来にかけての技術水準を見積り、それぞれの役割分担を考察している。

経費回収のしくみとしては、紙メディアの新聞では講読料と広告料がある。テレビの場合、NHKは視聴者から一律料金を徴収しているが、民放は広告収入ですべてまかなっている。また有料テレビについては、使用頻度に応じた料金徴収をすることができる。それでは電子新聞はどうすべきか。カスタマイズ機能が強力であれば、読者が特に興味をもたない広告を強制的に見せることは困難になる。もちろん見たくなくなるような広告を工夫すべきではある。しかし紙メディアの新聞の特徴の一つである、思いがけない出会いといった性質は、電子新聞ではあまり大きな期待はできない。むしろ、特定の情報を求めて検索する型での広告の方が、電子新聞では重みを増すことであろう。

それでは講読料の方はどうか。紙の新聞という「もの」と比べて、電子新聞という手で触れることができない「ソフト」に対して、講読者は本当に適正な金額を快く支払ってくれるか。日本の社会の慣習では、まだまだソフトに対する経済価値を低めにみる傾向が残っている。それでも最近では、家庭用コンピュータのソフトなどを始め、ソフトが有料の情報であるという概念が、ようやく一般に浸透してきた。二十年前とは大きな違いだといえる。電子新聞の講読料についても、それを認知する風土に落ち着くことは、時間が解決してくれることを期待しい。

2.4.4 おわりに

電子新聞はニュースを伝えるためのものか、それともデータベースから情報を検索することが主な役割か。昨年度の調査結果では、遡及情報の検索の方に大きな期待をもっている人が多いという結果が出ていた。一般に何か新しいものが、二つの異なる性格のどちらであるかを論じる場合、ちょうど現代物理学で光が粒子か波動かを論じたのと同じ結論になるように思う。即ち、光は粒子でもあり、波動でもあった。同様に、電子新聞はニュースを伝えるためのものでもあり、データベースから情報を検索するためのものでもある。この両面性を生かしたデザインをすることで、紙メディアの新聞やテレビにはない、電子新聞独自の性質を確立することができるのであろう。

今回のプロトタイプ開発結果、今後5～6年の間にコスト的にも実用化可能と思われる、一般家庭向け携帯型電子新聞のコンテンツ、また、情報伝送手段を柱とするサービス形態が

徐々に明らかになってきた。しかし、携帯型電子新聞のプロトタイプ開発は、まだその端緒についたばかりである。まだまだ多くのアイディアが出てくる余地があり、その可能性も高い。今後もう何年かの試行錯誤を経過することによって、本当の電子新聞とは何か、何であるべきかが見えてくることであろう。

3 SPORTS POWER INDEXの パイロット版データベースの構築

株式会社 ビデオ・リサーチ

3. SPORTS POWER INDEXの パイロット版データベースの構築

3.1 はじめに

3.1.1 SPORTS POWER INDEX（スポーツパワーインデクス）とは

「スポーツパワーインデクス」は、当社（株）ビデオ・リサーチ）がスポーツビジネスをサポートする情報サービスとして、平成5年度より調査、データ収集、また一部販売を実施しているものである。

Jリーグの発足とその爆発的な人気の獲得は、スポーツが社会現象の一つの流れを生み出すことを改めて示し、スポーツが新たな脚光を浴びたことは記憶に新しいところである。現在国内では、毎日何らかのスポーツイベントが開催されており、その件数は年間5,000件以上にも及んでいる。また、新聞・テレビ等のマスメディアを通じてスポーツについてのさまざまな情報が届けられており、その数は、TVのスポーツ中継番組では年間4,000本以上、TVニュースでは年間20,000件以上、また、新聞では一般紙1紙当たり年間10,000件以上といった膨大な情報量に及んでいる。これらの情報がスポーツの人気や普及に大きな影響をもたらしていると考えられるが、それを総合的に裏付けるデータはこれまでに存在しておらず、特にTVニュースに関してはその内容・パブリシティ量を把握することは困難であった。

一方、スポーツにはさまざまな企業や団体が支援を行っており、資金援助から商品提供、人的な支援までその幅は広く、一説にその市場規模は全体で4兆円を超えともいわれており、今後とも成長が予想されている。更に地方自治体でも「地域興し」の施策としてスポーツを活用するケースが増えており、スポーツに関する総合的な動向や、スポーツイベントの効果測定が求められている。

「スポーツパワーインデクス」はこれらのニーズに対応するもので、スポーツイベントの開催情報を核として、TV中継やニュースの紹介状況や、スポーツイベントへの企業の協賛動向をイベントごとに体系付けてデータの収集を行っており、観客動員、TVの視聴率を始めとしてTVニュースでの各スポーツイベントの紹介状況、協賛企業別の集計など、さまざまなマーケティング指数の提供を可能とするものである。また、一般生活者を対象としたスポーツ意識調査も実施しており、競技別の支持層のプロフィールや、各々のスポーツイベントの認知、スポンサー企業のイメージといった側面からの情報の把握も行える。

更に「スポーツパワーインデクス」では、こうした各種データの総合的な分析と検証を通して、スポーツイベントについての統一的な指標の提供を目指している。これにより、スポ

ーツイベントの相対的な位置付けや、イベントへの協賛の効果の測定を行うツールとしてだけでなく、統一的な指標をベースにした「スポーツイベントビジネスの活性化」及び「競技の振興」に貢献するサービスの提供を意図している。

3.1.2 スポーツパワーインデックスのデータベースの構築について

前述のようなサービスの提供に当たっては、多方面、かつ膨大な情報の一元的な管理がその前提であり、また、実際の活用シーンとしては、それらの情報から必要とする部分を抽出し、相互の比較・検討といった分析となるであろう。これらのためには、情報のデータベース化、及びそれらのデータベースを効率的に管理するシステムの存在が不可欠であり、また、実際のユーザとして想定されるマーケティングの担当者が自ら活用が可能な操作性をもつものでなければならない。

したがって、今回のパイロット版のデータベース構築においては、開発が比較的容易で仕様の修正・変更においても柔軟、かつ低コストでの対応が可能なスタンドアローンのパソコン環境で動作するシステムとして開発を行った。このシステムの実際の運用を通じて、顧客のニーズの把握や、スポーツイベントに関する指標の整備や検証、また、今後のより本格的なデータベース構築へのステップとなるべく作業を開始させた。

3.2 仕 様 の 検 討

3.2.1 パイロット版データベースの目的

「スポーツパワーインデックス」を事業とし確立し成功させるためには、各々の情報を効率的に管理し・運用することが可能なデータベース管理システムの開発が不可欠であることは、企画の当初の段階から確認されていたことではあるが、どのような情報をデータベース化するのか、また、それを定期的かつ効率的に収集する方法など、全くの新規事業ということもあり多くの面で課題があった。

そこで今回のパイロット版データベースの構築に当たっては、こうした課題を解決するとともに、実際の社内での運用を通じて、業界の統一的な指標としてのあり方や、情報の顧客のニーズに対応した分析方法等を検討するためのツールとしての開発が開始された。ここで提示されたデータベースシステムの要求仕様を集約すると下記のようなものとなる。

- ① 任意のスポーツイベントを検索条件として、抽出されたデータを「試合別」「競技カテゴリ別」「競技種目別」など任意の単位での集計が可能なこと
- ② 任意の企業を指定して、その企業が協賛したスポーツイベントが検索できること

- ③ データベースの知識のあまりない一般のユーザで操作が可能なこと
- ④ 今後予想される仕様の変更・追加に対応が容易かつ低コストであること
- ⑤ 必要に応じて他のアプリケーションでのデータの活用が可能なこと
- ⑥ 将来的に、「検索システム+データ」を組み合わせたパッケージ販売を想定した開発を行うこと

3.3 検索・集計システムの基本構成

- ① 「スポーツパワーインデックスのパイロット版データベース」の動作環境
 - ・本体：日本語MS-Windows 3.1以上が動作する機種（I486以上の機種を推奨）
 - ・メモリ：8M以上
 - ・OS：日本語MS-Windows 3.1以上
- ② データベース開発環境
 - ・検索プログラム開発言語：Microsoft Visual Basic Version 2.0
 - ・データベースフォーマット：Btrieve 形式

3.4 パイロット版データベースの機能

3.4.1 検索・集計プログラムの概要

「スポーツパワーインデックスのパイロット版データベース」の検索・集計プログラムの機能は、「検索対象とするスポーツイベントの指定」と「抽出されたスポーツイベントの表示・再集計」に大別される。

まず、スポーツイベントの検索指定では、

- ・競技の開催情報（競技名・日付・チーム名等）
- ・スポーツイベントに協賛した企業

を組み合わせ指定し、検索結果を画面に表示、又はファイル保存することができる。また、検索結果に対して、

- ・イベントのパワーの指数（観客数・メディアでの露出など）によるランキング
- ・競技の単位（種目／競技カテゴリ／シリーズ・大会など）による各指数の再集計などの集計処理が可能である。

3.4.2 検索・集計プログラムの処理手順

(1) 検索条件の指定

① 企業を検索条件とする場合

- ・「業種」、又は「企業名（漢字）」を指定し、検索候補となる企業を絞り込む
- ・候補企業から検索の対象となる企業を指定、検索を実行（画面出力をクリック）

② 競技（スポーツイベント）を検索条件とする場合

- ・検索対象とするスポーツイベントの開催期間
- ・検索対象とする「競技名」「競技カテゴリ」「大会名」「チーム名」等を指定
- ・最大5件までの検索条件の指定が可能（それぞれ「OR条件」となる）

図 3 - 1 検索条件指定画面（企業）

図 3 - 2 検索条件指定画面（競技）

③ 検索結果の一覧表示

- ・ 指定した検索条件に該当するスポーツイベントが表示される
- ・ 抽出された項目のうち、各データにおいて共通するものは画面上部に表示

SPORTS Power Index					
ファイル(F) 検索(S) 表示(O) ヘルプ(H) 実行(E)					
集計: (シリーズ・大会) (期間: 1993/9/1~1993/12/31)					
競技	ゴルフ				
カテゴリ	JPGA ツアー				
	シリーズ・大会	観客数	延接触 世帯数	紹介 回数	紹介秒数
25	日本プロゴルフマッチプレー選手権ユニシ	11,019	4,257	28	1,490
26	サントリーオープン	22,982	3,687	28	1,320
27	全日空オープン	32,518	3,745	26	1,430
28	シー・ササゼン・ジュンクラシック	17,024	4,028	21	1,010
29	東海クラシック	47,180	4,832	94	972
30	第58回日本オープンゴルフ選手権	18,141	8,330	38	1,750
31	アサヒビールゴルフダイジェスト	14,357	5,036	29	1,284
32	ブリヂストンオープン	21,385	3,025	29	994
33	ABC主催LARK CUP GOLF	43,838	6,053	33	1,421
34	ダイワインターナショナル	9,975	3,078	27	1,324
35	住友VISA太平洋マスターズ	23,249	5,378	34	2,362
36	ダンロップフェニックス	42,175	4,970	36	2,048
37	カシオワールドオープン	25,805	3,911	40	2,440
38	第30回ゴルフ日本シリーズ日立カップ	45,193	7,259	41	3,903
39	大京オープン	27,394	1,990	33	1,839

図 3 - 3 検索結果表示

④ 集計単位の切り替え

- ・ 検索結果に対して表示する単位（競技種目／競技カテゴリ／シリーズ・大会別／試合別）で再集計をすることが可能
- ・ 集計結果は各画面ごとにファイルに保存することが可能

SPORTS Power Index					
ファイル(F) 検索(S) 表示(O) ヘルプ(H) 実行(E)					
表示項目の選択(I) ソート(O)					
集計単位(U)	1-29回/2/31				
詳細情報(I)	カテゴリ	TV 放映回数	延接触 世帯数	紹介秒数	
	シリーズ・大会				
1	サッカー	209	341,702	56,597	
2	野球	956	24,781,097	5251,062,481	
3	ゴルフ	489	1,746,493	210 258,235	
4	モータースポーツ	273	444,600	0 8,482	
5	バレーボール	72	0	0 5,846	
6	バスケットボール	83	0	0 3,324	
7	マラソン・駅伝	63	0	0 12,850	
8	陸上(トラック競技)	28	0	0 3,029	

図 3 - 4 集計単位指定

3.5 今後の課題

今回の「スポーツパワーインデックスのパイロット版データベースの構築」により、従来の「記録」を中心としたスポーツのデータベースとは異なる、スポーツに関連する「ビジネス」をサポートするデータベースという新たな情報提供サービスの第一歩を踏み出すという当初の目的には達することができたものと考えている。

しかし、仕様・機能の双方ともにまだ不十分な点もあり、今後とも改良を続けていきたいと考えている。具体的な開発の課題としては下記のような点を検討している。

3.5.1 機能の拡張

今回のパイロット版の開発では、スポーツという未知の領域をテーマとしたこともあり、扱うデータの領域、各機能の仕様の決定などにかなりの時間を費やしてしまい、システム開発の進行に大きな影響を及ぼす結果となってしまった。以下の点は早急に改良したいポイントである。

- ・ 検索結果のプリントアウト
- ・ 検索結果のグラフ表示
- ・ 検索速度の向上
- ・ ユーザインタフェースの見直し

3.5.2 データベース内容の拡張（意識調査のデータベース化）

パイロット版においては「スポーツパワーインデックス」で調査・収集している情報のうち、スポーツイベントの「開催情報」「メディア露出」「企業協賛」についてデータベース化の作業を行う一方、「スポーツへの関与度」「協賛企業の認知・イメージ」など意識調査の情報については、調査項目が多岐にわたること、また、日数の不足などの理由からデータベース化を見送らざるをえなかった。意識調査の情報をデータベース化することで、既存のデータベースと連携することにより、「メディアへの露出状況と競技への関与度との関係」「企業の協賛状況と企業認知度・イメージとの関係」等、より多角的視点からの分析・検証が可能となる。

3.5.3 新たなサービス提供形態の検討

現在の「スポーツパワーインデックス」のデータベースは、単体のパソコン環境でのみ動作するものであり、今後、より有効かつ広範囲な情報活用を目指して、次に挙げるような形態

でのデータ提供サービスの実現性を構想にしている。

- ・検索システムとデータを組み合わせたCD-ROM等によるパッケージ販売
- ・パソコン通信やインターネットでのオンラインデータ提供サービス
- ・「TVニュースの内容データベース」の構築（スポーツに限定せずニュース全般を対象）
- ・TV中継やニュースにおいて画面に登場した「ロゴマーク」の露出量の測定サービス

3.6 ま と め

（株）ビデオ・リサーチでは、テレビ視聴率調査を始めとして、各種のメディア、及びマーケティング分野でのリサーチ事業を中心としてさまざまな形態の情報サービスに携わってきた。

今回の「スポーツパワーインデクス」はこうした中で、イベントなど新たな事業展開への第一歩と位置付けられ、また、パイロット版データベースでの経験が新たな情報サービスの開発につながるものと期待している。

4 画像付き地図資料データベースの プロトタイプの作成

財団法人 地図情報センター

4. 画像付き地図資料データベースのプロトタイプの作成

4.1 画像付き地図資料データベースとは

4.1.1 本データベースの基本的考え方

地図資料のデータベース化は、地図資料の内容・形式があまりにも多岐にわたるため、一般図書と比べるとはるかに遅れている。

従来、(財)地図情報センターを始め、さまざまな機関が地図資料のデータベース化に対してアイデアを提供し、あるいは技術開発を行ってきた。しかし、そのほとんどは文字情報によるものである。もちろん、文字情報の有用性は極めて高く、更に研究を進める必要があろう。

しかし、地図という極めてビジュアルな資料を、いかに詳細に文字で表現しても限度がある。この点、絵画や写真などの画像情報のデータベース化技術が大いに参考になる。本研究では、さまざまな形の地図資料に合わせて、文字・画像両情報の長所を生かしたデータベース構築を研究しようとするものである。

現実問題としては、画像を主とするデータベース化ではなく、文字情報を主、画像を従とするデータベース化が有効であると考えられた。通常のディスプレイで検索することを想定した場合、画像を主とするものでは現在の技術では不十分であると考えられるからである。こうして画像付きのデータベース化を想定し、研究を進めることとした。

4.1.2 本データベースの意義と特徴

(1) 地図のもつ特異性

地図は、例えば、国土地理院発行の1万分1・2.5万分1・5万分1地形図、20万分1地勢図などの、いわゆるシリーズ地図（ここでは、統一された縮尺、図式規定に従って作成された地図を指す）では、例えば日本国際地図学会30周年記念『近代日本地図集成』における分類項目（一部加筆修正）をみると、

0：整理番号	1：発行機関	2：図の番号
3：シリーズ名、図葉名	4：縮尺	5：紙の外寸
6：色数	7：測量年・改測・補測など	8：発行年・印刷年など
9：購入・閲覧場所	10：備考	

のようになっている。

これは、シリーズ地図には十分な項目であるが、主題図の多くは、シリーズ化されてい

ない。このシリーズ化されていない地図類では、地図1枚に1枚ごとのシートの項目が必要になってしまうことがある。

明治時代以降に作成された地図類では、比較的統一のとれた傾向をもつ規格が多いので、何とか分類項目を1～2増やすことにより解決されるが、明治時代以前のいわゆる古地図（以降、明治時代以前に測量・印刷・発行された地図を古地図と称する）では、その規格がまちまちで統一性に乏しく、統一化されたシートでは、備考、又は、その他項目に多くの情報が入力され、逆に、分類項目未記入が増えてしまう。

(2) コーディングシートの作成作業

（勘）地図情報センターでは、シート作成に当たり、上述の地図の特異性を考慮に入れたシート、特に、入力すべき項目の選定に、項目策定のための委員会を設置し、多くの時間を費やした。

多くの時間が必要であった理由は、上述の問題点を踏まえながら、従来から、多くのバラエティに富む主題図、一般図などが印刷され、学校、図書館、博物館、資料館、文書館などの諸施設で地図が所蔵されているにもかかわらず、統一された地図データベースのためのシートの項目が定まっていないため、標準的な基準作りをも考慮に入れたシート並びに、項目作りを試みたからである。

(3) 画像情報の必要性

地図情報はデジタル的・アナログ的視覚資料であるということから、文字情報ばかりではなく、画像の情報に関するデータベースも必要になってくる。

地図には異なった種類の地図であっても、ほとんど似たような情報をもつ地図があり、見なければ区別がつかないという地図もある。このような場合には、シートを見ながら、地図見本をパーソナルコンピュータディスプレイで見ることができれば、学問研究・教育の一助となることは間違いない。現在のディスプレイ上での色彩の豊富さ、色再現の正確さはかなりのレベルにある。

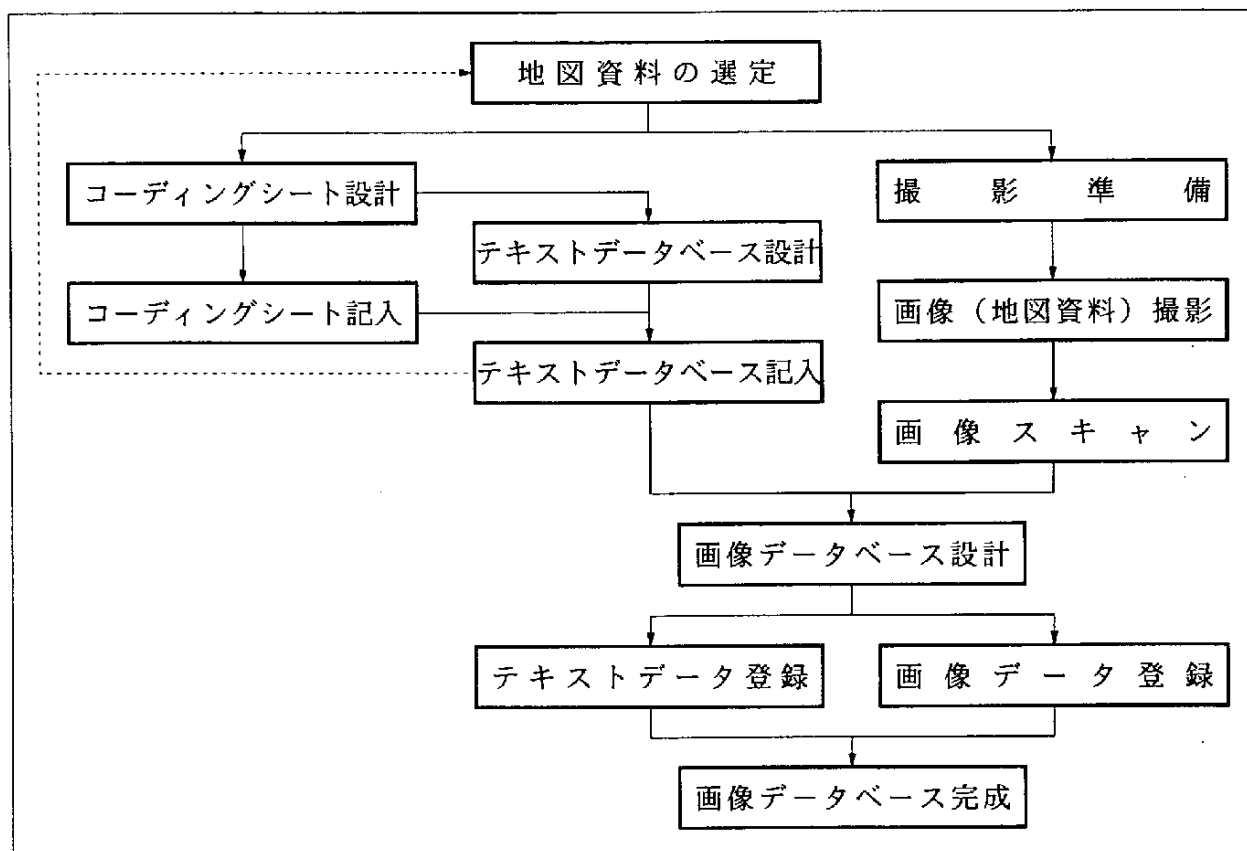


図4-1 作業の流れ

4.2 テキストデータベースの構成

4.2.1 キーワードの検討

一般に地図には、欠かすことのできない要素ともいえるものがある。それらは図名、1枚の地図に収められている範囲（地域）、縮尺、発行年、測量年、発行所、色数、形態その他製作者、印刷所等々である。しかし、本研究の対象となる地図資料＝主題図及び外国の地図について、その特性を十分に勘案してシートへの記入項目の取り決めを行い、図名・地域・縮尺・発行所・発行年・測量年・形態・色数とすることとした。これら以外については備考欄を広く設け、すべてこの欄に記入することとした。更に、地図資料自体の項目以外にも整理番号・保管場所・地図資料受入年月日・寄贈者等地図資料所蔵機関として必要な項目を設けることとした。

整理番号		保管場所		
図 名	(シリーズ名)			
	地域名			
	縮 尺 1 :			
	寸 法 × × ———			
発 行 所			寄 贈 者	
発行年 年 月 日 (年)				
測量年 年 月 日 (年)			受入年月日 年 月 日 (年)	
形 態	一枚もの	彩色 単・多		
備 考				

図 4 - 2 コーディングシート

4.2.2 テキストデータベースの設計・記入

地図データはまず前述したコーディングシートに記入された後、後述するような方法で写真撮影される。したがって、書誌データの記入時期と画像データの記入時期の間にはタイムラグが存在する。したがって、最終的な画像データベースを作成する以前に、書誌情報のみを記入するテキストデータベースを作成する必要があった。

テキストデータベースの設計としては、使用アプリケーションが文字・数値形式や固定長といった概念をもっておらず、レコードセルの中で体裁を整えるために、改行を許すという特異なスペックがあり、最終的なハードコピー出力においては整った印刷ができるものの、DOSアプリケーションとの連携においては多少のこつを有する。

表 4-1 テキストデータベースのフィールド

フィールド名	データ形式	データ内容	備考
番号	数字(可変長)	5桁シリアル	空白値不可
保管場所	文字・数字(可変長)	地図情報センター定義	空白値不可
シリーズ名	文字・数字(可変長)		空白値, 複数値容認
図名	文字・数字(可変長)	原本に準拠	空白値, 複数値容認
地域名	文字・数字(可変長)	地図範囲を含む地域名, 地図情報センター定義	空白値, 複数値容認
縮尺	文字・数字(可変長)		空白値, 複数値容認
寸法	文字・数字(可変長)	縦×横×N U L (実地計測)	空白値, 複数値容認
発行所	文字・数字(可変長)	原本に準拠	空白値, 複数値容認
発行年	文字・数字(可変長)	西暦(年号併記)	空白値, 複数値容認
測量年	文字・数字(可変長)	西暦(年号併記)	空白値, 複数値容認
形態	文字・数字(可変長) 文字・数字(可変長)	94年度は一枚もののみを対象とした	すべて同一, 空白値不可
彩色数	文字・数字(可変長)	多色・単色の2値	空白値不可, 2値選択
寄贈者	文字・数字(可変長)	地図情報センター定義	空白値, 複数値容認
受入年月日	文字・数字(可変長)	寄贈者に連動	空白値, 複数値容認
備考	文字・数字(可変長)	図名等が旧字体の場合, こちらに記入	空白値, 複数値容認

4.3 画像データベースシステムの構成

4.3.1 画像入力システム

画像入力システムとして今回最大の目標としたのは,

- ・システム全体が低価格で設置が容易である。
- ・誰でも簡単に撮影やデータベース化ができ, 特別な技術を要しない。
- ・専門のスペースを用いない。
- ・汎用性のある市販アプリケーションを連携させる。

ということである。大規模博物館や図書館などは所蔵物の撮影を行う部署やスペースがあり, 専門の技術者や学芸員がその知識と技術により高度な撮影を行っている。しかし, 中小の博物館などでは, そのような資材も人材も技術も乏しく, しかも外部に委託する場合, 多額の撮影料が必要となる場合が多いと考えられる。

4.3.2 基本的なコンピュータハードウェア

ハードウェアとして通常のパーソナルコンピュータを使用することにした。これは最近の R I S C 技術による C P U のスピード, メモリ空間の向上, メモリチップの急激な価格低下によるユーザメモリの拡大など, 比較的低価格のパーソナルコンピュータが数年前の大型汎用機や十数年前のスーパーコンピュータに匹敵する能力を有するようになってきたためである。

4.3.3 画像入力

(1) 撮影装置

地図資料には、切手やハガキ（切手・ハガキに印刷されている地図）のような小さなものから、国絵図のように大広間でないと広げられないような大きさのものまである。しかし当センターに所蔵されている地図資料は、A4サイズから四六サイズ（約78×109cm）のものが多く、したがって、今回は四六のサイズに余裕をもたせ、90×120cmの撮影装置を設計した。

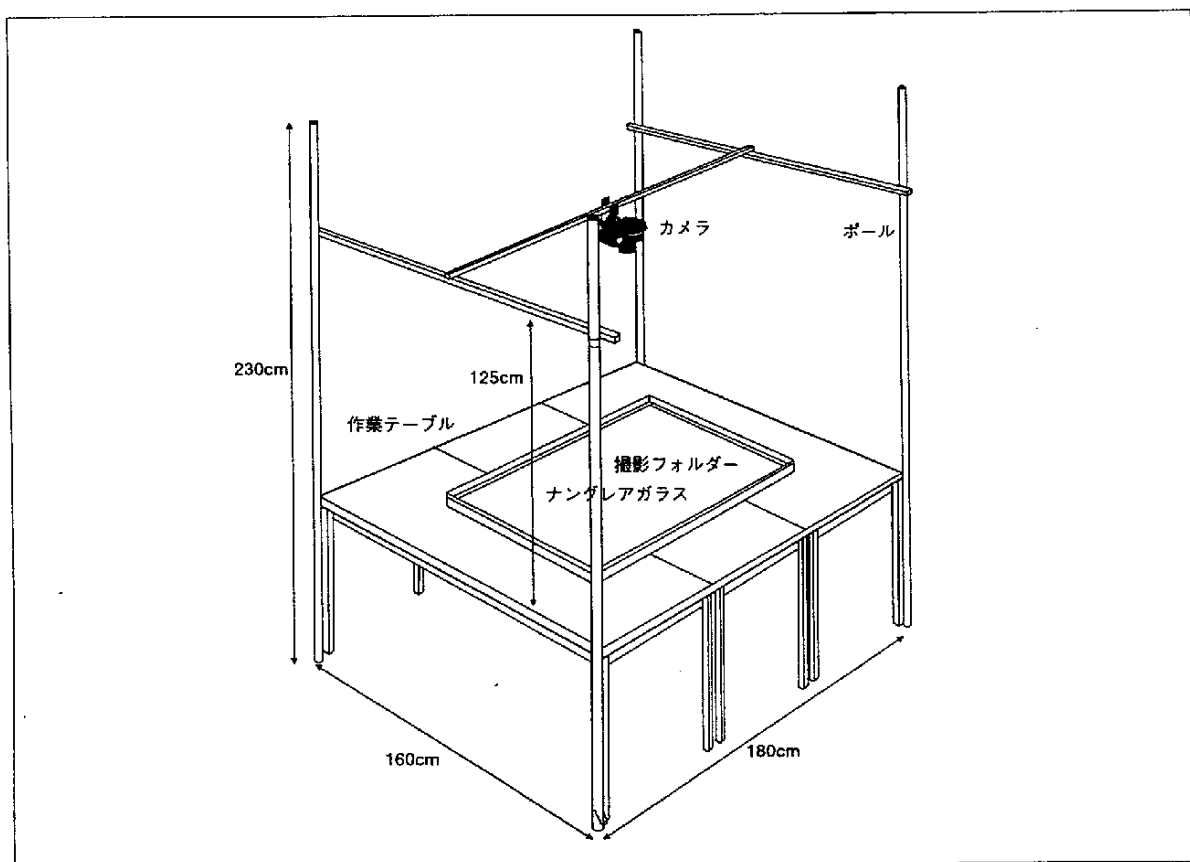


図4-3 撮影台の概要

(2) スキャニング

スキャニングについては撮影方法も含め幾つかの方法が考えられた。

表 4-2 スキャニングの手法

データ入力 方 法	最大原稿 サ イ ズ	利用解像度	最 小 判 読 文 字	入 力 時 間 (セッティングは 含まず)	デ ー タ 容 量	鮮 明 度	色 表 現	特 記 事 項
一般フラット スキャナ ダイレクト	普通 A 4 以 下	300dpi程度	A 4 原稿 1mm角	5分程度	1～40M	◎	◎	A 4 以上の原稿の読み取りは困難。リトライは容易。部分入力に適する。
電子カメラ	制 限 な し	40万画素程度	四六原稿 50mm角	2分程度	0.5 ～ 1 M	△	○	大きさに制限なし。解像度は最も低い。リトライは容易。コンピュータと接続の必要あり。
Photo CD	制 限 な し (光学カメラ使用)	800dpi程度 (フィルム 換算)	四六原稿 20mm角	写真撮影 + CD焼き 込み時間	20M以上	◎	○	出来は撮影技術に依存。 データの再編集は困難。 最大解像度ファイルは印刷レベル。詳細は後述。
写真・印刷 紙スキャナ	制 限 な し (光学カメラ使用)	読みと読み 取り範囲に 依存(普通 300dpi)	四六原稿 20mm角	写真撮影 + 5 分	1～20M	△	△	解像度は印画紙サイズに依存。整理が容易。ランニングコストは高い。間に介在するデータが多い。
写真・一般 フィルムス キャナ	制 限 な し (光学カメラ使用)	1200dpi 程度 (フィルム 換算)	四六原稿 20mm角	同 上	1～20M	○	○	印画紙スキャンより鮮明度は高い。撮影条件は最もシビア。ランニングコストは高い。

(3) 入力解像度

地図資料のスキャニングにおいては、入力解像度は大きな問題である。ここでは35ミリフィルムを基準として考える。35ミリフィルムは有効画像面積は長辺35ミリ短辺24ミリである、したがって、短辺は約1インチである。入力解像度を1200dpi とすると24ミリについて約1200ピクセルの解像度ということになる。これを全面に地図を撮影したとすると、地図サイズによって判読できる文字サイズが決まってくる。例えば主要なサイズである90×70cmの地図の場合、短辺70cmが1200分割されるわけであるから、1ピクセル当たり地図上で約0.6mmの解像度ということになる。実際は有効面積上の誤差などもあり、1ピクセル当たり0.7mmの解像度であろう。それでは実際の地図上の1cm角程度の文字が1200dpiで判読可能かという一概にそうとは限らない。それはコントラストによるハレーションや文字の形態等によるからである。一方、解像度を上げていけば、より細かい文字が判読できるかというこれも問題である。まず、カメラレンズの解像度の問題がある。更にフィルムの粒状性、フィルムのラチチェードなど、阻害要因は数限りなく存在する。例えば、

白地にグレーのコンター（等高線）などはフィルムの表現範囲を超えている。また、使用したスキャナー（NIKON COOLSCAN）の最高解像度ではISO 100程度のボジフィルムの粒状性まで拾ってしまう。

表 4 - 3 P I C T画素数とファイル容量

PICT画素数	24bit (PICT/JPEG)	16bit (PICT/JPEG)	8 bit (PICT/JPEG)
840×1200dpi	2.9/0.2MB	1.7/-	0.9/-
1260×1800dpi	6.49/0.3	4.07/-	2.12/-
2520×3600dpi	26.0/1.2	16.4/-	8.4/-

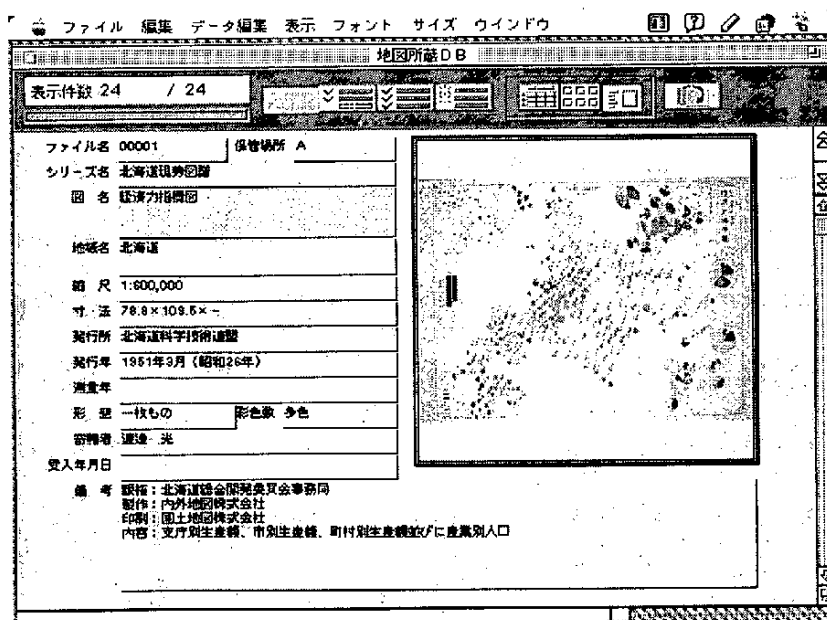
また、ファイル容量も大きな障害になる。P I C Tでは1260×1800dpi できさえも約 6.5 MBという大きなサイズとなってしまう、フロッピーディスクでは分割しなければ入らない。

(4) 画像フォーマット

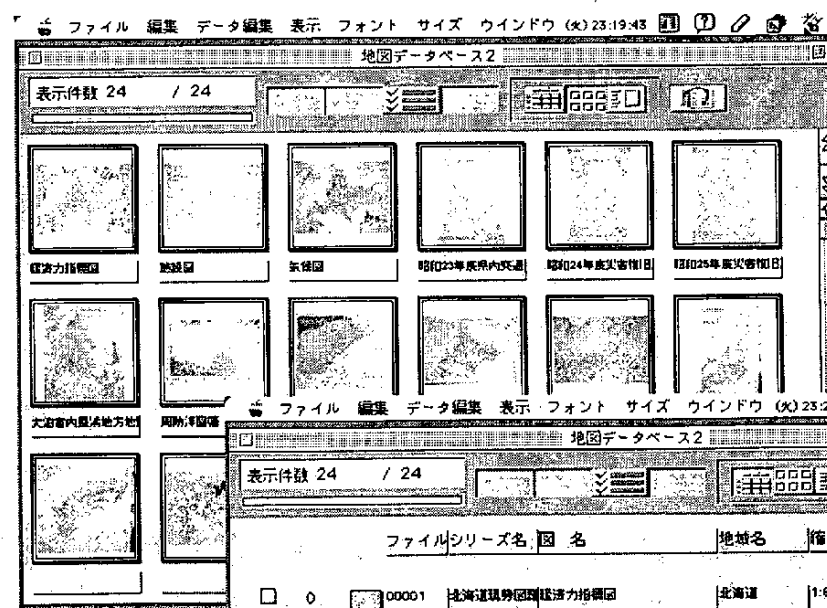
ファイル、特に画像系のファイルはその性質上、巨大データ量を要求する。J P E G (Joint Photographic Expert Group)はフルカラーを非常に効果的に圧縮できる技術であり、最高でソースデータの1/100程度まで圧縮が可能である。J P E Gが他の圧縮アルゴリズムと最も大きく違うのは非可逆的な圧縮方式であるという点である。つまりJ P E Gアルゴリズムはソースデータを劣化させることによって高い圧縮率を保つのである。ただ、劣化といっても人間の視覚的特徴を十分に研究しているため、1/10 や1/20 の圧縮度ではディスプレイ上での劣化は確認ができないほどの優秀なアルゴリズムであり、印刷する場合でも1001pi程度の解像度ではほとんど劣化は確認できない。

結論として、今回は 840×1200pixer(約800dpi) のP I C TファイルをJ P E Gで1/20 程度に圧縮し、1ファイル当たり約 200KB程度、500画像で約 100MB程度とした。

画像ファイル名として、地図名などを使用することはファイル単体として使用する場合は便利であるが、今回はそれぞれ別個のシーケンシャルな番号を付与しているため、ファイル名称は5桁の番号とした。またこれはMS-DOSのようにファイル名に制限の多いシステムでも表示が可能である。



ブラウズモード



リストモード

ファイル 編集 データ編集 表示 フォント サイズ ウィンドウ (x) 23:20:30

表示件数 24 / 24

ファイル	シリーズ名	図名	地域名	縮尺	寸法	備考
☐ 0	00001	北海道現勢図経済力指標図	北海道	1:600,000	78.8×109.5×	原図: 北海道開発委員会事務局
☐ 1	00002	北海道現勢図気候図	北海道	1:600,000	78.8×109.5×	原図: 内外地図株式会社
☐ 2	00003	北海道現勢図気候図	北海道	1:600,000	78.8×109.5×	原図: 内外地図株式会社
☐ 3	00004	昭和23年度県内交通	鹿児島県	1:200,000	104.5×71.5×	原図: 鹿児島県庁
☐ 4	00005	昭和24年度災害復旧工事箇所図	鹿児島県	1:200,000	104.5×71.5×	原図: 鹿児島県庁
☐ 5	00006	昭和25年度災害復旧工事箇所図	鹿児島県	1:200,000	104.5×71.5×	原図: 鹿児島県庁
☐ 6	00007	大泊管内風害地方誌	サハリン南部	1:100,000	86.2×72.7×	原図: 建設省
☐ 7	00008	(地質図) 周防・宇留	周防・宇留	1:200,000	37.3×50.5×	原図: 建設省
☐ 8	00009	(地質図) 日和佐図	日和佐	1:200,000	36.2×49.0×	原図: 建設省
☐ 9	00010	北海道農村分布図	北海道	1:1,000,000	54.5×78.7×	原図: 建設省
☐ 10	00011	最新日独大地図(東京版)	日本	1:1,400,000	79.0×108.7×	原図: 建設省
☐ 11	00012	人口密度図第1図(1930年度)	日本	1:1,000,000	91.7×59.0×	原図: ライフ
☐ 12	00013					
☐ 13	00014					

0件

図4-4 画像データベースの表示画面

4.3.4 画像データベースエンジン

「ビレットEX」(株クボタ)は低価格ながら複数検索を実行できる画像データベースである。このアプリケーションの特徴はPower Macintoshに対応したパワーアプリケーションであるという点である。これによりパワーマックで高速なラインタイムが行える。今回はこのアプリケーションを用いて画像データベースを構築した。特に、ランタイムが作成できることにより、アプリケーションをもたない他のコンピュータでのブラウジングや検索が行えることから、CD-ROMによるデータベースの配布の道が開けることになる。

4.4 今後の課題

(1) コーディングシートの再構築

これで満足せず、より汎用性の高いシートの完成を目指すべく、多種多様な地図資料をシートへ記入し、問題点を摘出していききたい。

本研究の対象とした一枚ものの地図以外にも、地球儀、地図帳、立体模型地図、掛け図など地図資料はさまざまであるが、このような状況にも耐えられるシートの追加構築・改変も考慮する必要がある。

しかし、枚数を無限に広げないことである。

(2) 画像入力

画像情報を入力することは現代の技術では簡単であり、これを加工することも簡単にできてしまう。ここで生じる問題点は、著作権の問題である。パーソナルコンピュータにデジタル情報として入力できれば、デジタル情報として、また、アナログ情報として加工・編修し出力できることはあたりまえである。この加工品を個人の利用範囲から逸脱し、販売してしまうこともできるのである。

このようなことを考えた場合、地図の画像情報はどこまで入手が許されるのかを検討する必要がある。個人で利用する範囲であるのであるならば、引用した原図名・作成機関などを発表する論文などの末尾に記載すれば済むことなのか、また、加工し発表する前に作成機関などへ届け出をするべきなのかなどが未だ不明である。

考えられる状況を整理し一つ一つ判断していかななくてはなるまい。

(3) パーソナルコンピュータプラットフォーム

近年のパーソナルコンピュータは、機種を限定しないアプリケーションソフトやデータの互換が可能となってきたが、データベース構築のために利用したパーソナルコンピュータシステムと他の機種との互換性をどのように整合性をもたせるかも、地図データベ

ースの普及の鍵となるところである。

特定の機種・アプリケーションソフトに偏らない、普及を図るデータベースの構築を考えなければならない。最近ハイブリッドタイプのCD-ROMなどが存在し、これらの採用も考えるべきであろう。

データベースは情報として売買される対象となる。データベース作成者が個人使用の範囲で利用している場合には問題はないが、これに関与しない第三者によるデータベース閲覧の申し出があった場合、ディスプレイ上・プリントアウト・ディスクなどを有料・無料、パーソナルコンピュータ通信に乗せるのかなど第三者の閲覧に関する取り決めが必要になるであろう。

などが、今後の課題として残されているものと考えている。一つ一つが重要な内容であるだけに、じっくりと考察を重ね吟味していきたい。

5 パソコンを利用した副作用症例 データベースの構築調査

協立医師協同組会

5. パソコンを利用した副作用症例 データベースの構築調査

5.1 はじめに

ソリブジン薬禍のように、医薬品情報の不適切な伝達により、不幸な事件の発生が続くことは、日本の医療・医薬品業界にとって第一に改善せねばならないことである。現在、日本で使用される医療用医薬品は14,000品目を超えている。薬事法によってメーカーに対して義務付けられている医薬品添付文書による情報提供は、こまぎれで画一的な内容のため、情報の洪水に拍車をかける状態である。また、有効性を宣伝する情報は繰り返し大量に伝えられても、製品の販売にとって不利益となるとみなされる安全性情報は、メーカーから積極的に提供されることは少ない。

協立医師協同組合は、医薬品の適正使用につながる情報整備に努めてきた。そしてこれまでも、医薬品情報室を整備して実践活動をしてきた。特に、副作用モニター情報の収集では、厚生省のモニター活動とは別に独自の活動を続けてきた。情報活動の結果は、ニュースと医薬品情報誌の発行と、電話・FAXでの質疑応答活動として、組合員に返してきた。

近年、パソコンを始めとする情報機器の医療機関への普及に伴って、医薬品情報を印刷物ではなくコンピュータで利用できる形態で供給することを求められるようになってきた。また情報交換と分析の効率化を図る目的で、副作用モニター情報をパソコンのデータベースで整理することが、医療機関からの自発的活動として開始された。このような状況の中、情報活動の統一的な整備が求められていた。

副作用症例データベースの構築と運用を協同組合内の事業として実現し、他の医療関係者へも普及することをめざしてきた。

今回、機会を得て副作用データベース構築調査の活動を行うことができたので、その成果を報告する。

5.2 目的

データベース構築の過程で、今回は以下の3点について明らかにすることをめざして、調査研究を行った。

(1) 医療現場で実用的に使える副作用データベースの要件

副作用情報を入手するためのデータベースとしては、現時点でも複数のものがある。しか

し、医療現場では文書の形態での情報収集と交換がいまだに主流のままである。普及の遅れの原因を探り、診察室や薬局で求められている実用的なデータベースのあり方を検討した。

(2) 登録データの水準を保ちながら、ユーザからデータ提供を受ける方法

現在、当組合に加盟している医療機関のうち副作用モニター契約を行っている 124施設から集中する副作用の症例件数は、年間2,000件を超えている。

信頼できるデータベースを作るために、情報の質を保ちながら更に量の拡大を図る必要がある。副作用情報が発生する各医療機関での的確な情報収集、十分な訓練を受けた専門家グループによる情報評価、確実な情報の選択と登録をするために要求される業務規範の作成を検討した。

(3) 低コストで運用するデータベースを作成する方法

将来にわたりデータベースを維持するために、機器やコード体系などを保守し続ける必要がある。その費用をいかに少なくするかは、民間医療機関での情報活動にとって重要な課題である。また、組合員が分散している中で通信コストの節減も重要である。今回は特に、データ内容の整備だけでなく、低コストで運用できるデータベースの設計を試みた。

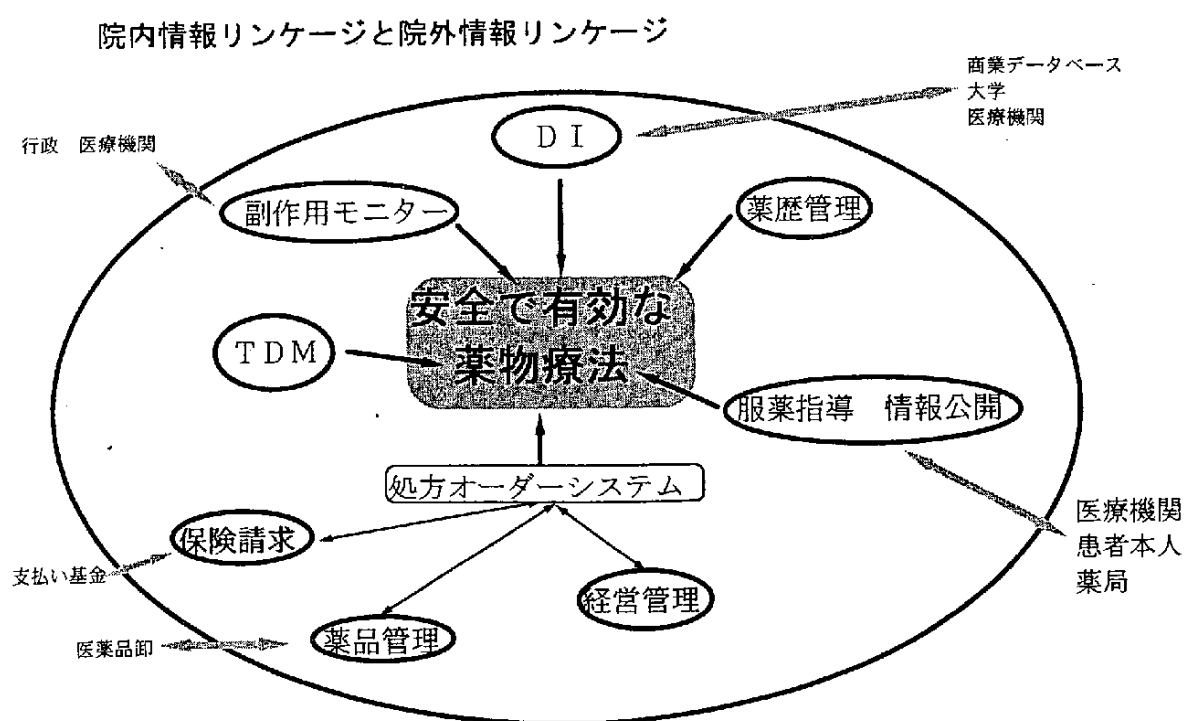


図 5 - 1 医療と情報の結合

5.3 調査方法

調査は、薬剤師のワーキンググループを集めて現状の問題点を明らかにすることから始めた。

医薬品業界や医療関係で広く使われているコード表や、調査票と報告書の類を収集し分析した。厚生省の副作用モニター報告用紙の改善を主な題材としながら、データベース収載項目の設計を進めていった。

更に調査結果を実務の中で検証し、実用システムに仕上げるために、プロトタイプを作りながら設計を進めていった。

5.4 調査結果

5.4.1 医療現場が求める副作用情報

製薬企業からの緊急安全性情報や、学会誌の情報は散発的で、重要な情報を予め伝えられていても、臨床判断が必要なときにはなかなか見つけることができない。

また、索引が整備された文献は発行間隔が長く、速報性がない。

一方、医療関係で利用できるデータベースとしては、MEDLINEやEMBASEのように古くからある文献データベースが代表的なものである。また、近年になって医薬品卸が中心になって整備されてきているARISのようなオンラインの副作用情報や、これから公開される厚生省の公的情報など選択肢は増えていっている。

われわれは、既存の文献データベースに不足する速報性と事実伝達の中立性を重視した、症例ファクトデータベースの構築をめざすことにした。文献データベースと症例ファクトデータベースは、それぞれ補足しあって情報の完全性を保障するものとする。

表 5-1 文献データベースと症例ファクトデータベースとの比較

	文 献 デ ー タ ベ ース	症例ファクトデータベース
情 報 の 質	公表されるまでに十分な検討が されるため正確。	検討不十分なものも途中経過で 登録される。
情 報 の 偏 り	著者の関心ある範囲で情報が取 捨選択され、加工されてしまう。	比較的偏り無く広く集まるが、 調査計画を事前に設計しない限 り調査対象の実態を反映しない。
情 報 数	比較的少ない。	多い。
タ イ ム ラ グ	投稿されてから審査編集を受け 出版され、なおかつその後に登 録されるので遅い。	出版の過程がなく、事実経過が 簡単なチェックだけでそのまま 登録されるため早い。
公 開 の 有 無	種々の媒体において公開中。	現在は組織内で公開され、対外 的には非公開。
利 用 方 法	教訓や対応、処置方法をまとめ た資料とする。	速報的に副作用発現のシグナル を検出する。報告者と違った視 点で分析する。

5.4.2 データベース登録項目の選定

(1) 項目の選定

アメリカFDA、スウェーデン、イギリスの副作用報告用紙及び日本の幾つかの施設で使用している用紙を厚生省の用紙と比較検討した。

医療現場で利用に必要と要望された事項の中から検討の上、副作用報告用紙に記載するもの、その中での必須項目、データベースに登録する項目、データベースの運用上で事務局が使用する項目の4種類に整理した。94項目中、報告書に入れたものが64件、データベースに登録するものが58件、入力必須項目が30件であった。

原則として、同じ内容を厚生省薬務局安全課にも報告するようにしたため、厚生省の報告用紙と同等の項目内容を本データベースにも収集することとした。

(2) 登録データのレベルを引き上げる入力フォームの作成

決定した項目を入れて、副作用報告用紙とそれに対応したパソコンでの入力フォームの画面を作成した。その中で、以下のような設計方針をとった。

- ① 副作用を報告する者が記入しやすい「項目配列」とした。

- ② 分析に必要な情報を洩らさないように、記入を誘導する「項目名」の設定。
- ③ 既知で軽微な副作用は略して記載ができるように、「必須項目」を明らかにした。
- ④ 詳細症例の報告用にコンピュータ用とは別に薬歴形式の記入票を作成した。
- ⑤ この用紙のままで厚生省副作用モニターにも報告できる、「標準化」の遵守。

図 5 - 2 パソコン入力画面ハードコピー

5. 4. 3 用語とコードの標準化の試行

(1) コード化の基本的考え方

今回、データベース構築に当たってコードの利用を検討したのは、医薬品（商品名・一般名）、薬効分類、副作用名、病名についてである。

WHO・厚生省・学会で標準化がされている用語・コードを使用する。コード化、標準化がすぐにできないものは規格化を急がず、まず自由文で入力・蓄積して、ある程度の情報が蓄積された段階で基準化を行うこととする。

コード表が、どこの医療機関でも安価に入手でき、しかも簡単な訓練でコード引きができるような分かりやすいものを求める。

この観点で医療現場で利用できるコード体系のあり方を調査した。

(2) 現状と問題点

一般に使用されているコード表に関して、以下のような問題点が存在した。

① 標準コードがない

容易に入手できるコードは、厚生省関係でも薬務局経済課・安全課・保険局医療課が管掌する別々のコード表が使用している。

オンラインデータベースには日本語のシソーラスも整備されているが、安価に実務の現場で入手できない。

② 継続性の欠如

長期にわたって一貫して保守されている医薬品の I D コードがない。薬価コードも再

審査コード、レセプト電算化の薬品コードもすべて薬効コードを内部にもっており、分類コードとして設計されているため、薬効分類が改訂されるとコードに変更が発生する。

③ ID機能の欠如

薬品コードは薬品の識別ができるIDとしての登録番号だけにすべきである。前記の欠点のため、IDとして使えるものが標準では存在しない。

④ 階層構造の欠如

薬品コードでは、成分名（一般名）－薬品銘柄名（メーカ別のブランド）－商品コード（規格・容量・包装別）のような一貫した階層構造がほしい。

薬効コードでは、対象疾病別－治療メカニズム別と化学構造的分類の二つの分類を混同せずに作ってほしい。

(3) 改善へ向けての対応策

① 医薬品コード

医薬品コードとしては独自コードを使う。業界標準の他のコードとは、自社の薬品マスター内部で変換表をもつことにする。

② 薬効コード

薬効コードは薬業時報社の保険薬事典のコードを使う。現行の87分類が改定されたときは、過去のデータも更新する。

③ 副作用名

副作用名とコードは小田島のHYTARDを使用する。

④ 病 名

病名の標準化は今回は見送る。現時点では、日本語で入手できる標準コードがない。

5.4.4 プロトタイプの実成と配布

データベース関係はリレーショナル型データベースソフトウェアである桐Ver. 5を使用し作成した。可変長の文字データを標準で扱えるため、データ設計は簡単であった。今回は、データベースの基本となる表の設計、データ送付と回収の一括処理プログラム、データ入力用のフォーム画面、最低限のプルダウンリストの作成を行った。

通信関係はコストの関係で、クライアント側は件数が多いためフリーソフトウェアを中心にシステムを構築した。通信ソフトウェアのオートパイロットのマクロプログラムと、桐のチャイルドプロセスで起動する通信のバッチプログラムの作成を行い、モニター施設に配布した。

バイナリーテキスト変換ツール「ISH」(石塚匡哉)

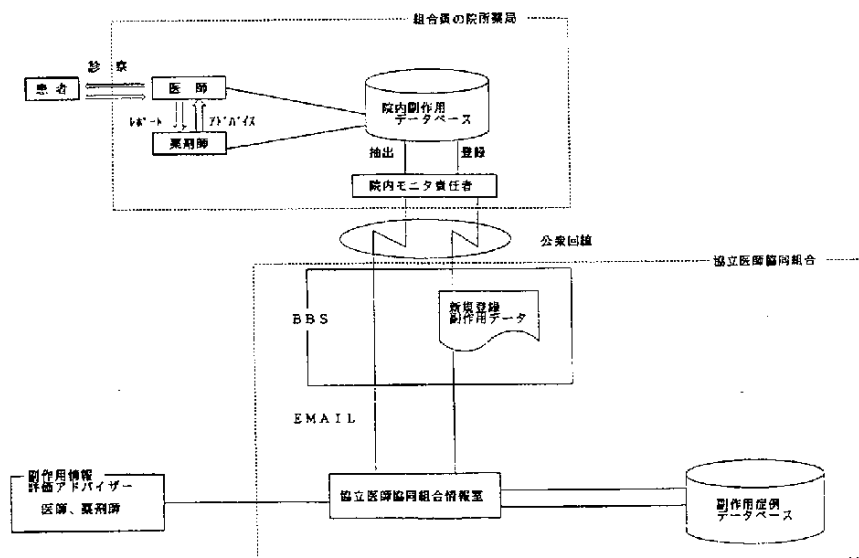


図5-4 組合員と協立医師協同組合との間のデータフロー

5.4.5 データ集約と登録データへのアクセスの効率化

(1) パソコン通信の利用

運用コストの低廉さと、守秘義務の面で安心できる独自の電子掲示板システム（BBS）を開設することとした。一次情報を電子メールにすることで、加工済みのデータしか一般のユーザはアクセスできないようにした。

データのダウンロードなどは、通信ソフトウェアの自動実行機能でユーザに複雑な操作を要求しないで実行できるようにした。

更にこのBBSの掲示板機能を用いて、医薬品情報の質疑応答結果や各施設の薬事委員会で行った新薬の評価活動の成果などの交流も可能となった。

(2) オフラインでのデータ検索

オンラインで検索まで行うシステムの構築は、通信費用が高いことと検索ソフトウェアの開発難易度が高いため、今回は見送った。新規登録のデータはユーザの責任でダウンロードして、ローカルデータベースに追加して利用することとした。

これにより、通信費用を気にせず十分なデータ検索ができることが本システムのメリットとして特徴付けられた。

なお、今回開発したものは桐の会話型処理で検索するところまでで、検索語の入力から自動検索する一括処理プログラムの開発は、来年度の課題とした。

5.5 到達点と今後の展望

本事業課題の中で実用にたえるプロトタイプを作成することができた。設計上の不備は、試験的運用の中から洗いだして改善していきたい。

5.5.1 副作用症例データの迅速な収集と提供

データをパソコン通信で集約しすみやかにフィードバックできることから、副作用情報を必要なときに、必要な場所で利用できるようになる。

5.5.2 処方設計などの臨床現場での医師への情報サービス向上

コンピュータで利用できるデータを、診察室や薬局の現場に返送することにより、重篤な副作用発生の防止のための処方設計の改善にも活用できる。

これから蓄積するデータを、だれでも利用できる、分かりやすい資料に加工する方法をこれから検討していきたい。

5.5.3 薬剤評価への発展

医薬品の製造承認前の評価の限界から市販後調査の充実が必要とされている。また薬物療法を適正に進めていくための薬剤使用評価（DUE）が注目されている。本データベースが、そのための重要な資料の供給源となりえる。

しかし、疫学調査に使えるデータの要件を満たす内容にするためには、更に検討と整備が必要である。

5.5.4 患者への情報公開とルールの検討

患者の「知る権利」・カルテの公開など、医療をめぐる情報公開の要求が急速にたかまってきた。それに備えて、医療機関の情報整備が必要となっている。今回コード体系の整備を始めたことで、いままで医療機関の中だけにあった薬歴を共同して利用する道を開くことができた。

患者はさまざまな疾病をもち、複数の医療機関で受診する。その場合でも患者自身が自己の投薬歴情報をもつことによって、医薬品の相互作用による副作用などを未然に防ぐこともできるようになる。

患者本人へのデータ提供、地域に広がる医療連携の中で医療機関同士の情報交換などで、プライバシーを守りながらどこまで情報を公開できるか、議論を重ねていく必要がある。

今回の調査研究によって、医薬品の安全な使用を進める活動の促進とともに、患者の人権に係わる重要な課題に取り組むきっかけを与えられたことを感謝する。

5.5.5 更なる標準化の追求と提言

日本の保険医療分野で情報業務を充実するためには、コードの整備がボトルネックとなっていることを、今回の調査研究で強く感じた。そこで、関係各機関に以下のことを要望したい。

(1) 標準化機構の設立に政府の指導と財政補助を

コードの基本設計と保守には経済的保障が必要である。現在、各省庁で個別に行っているコード化を一元化して、国民共有の財産として公開できるようにしてほしい。

(2) 薬品のIDコードでのデファクトスタンダードの作成

薬品のIDコードの標準化を、大至急実現してほしい。

このことを実現すれば、医療情報の面における日本の重大な遅れを取り戻すことができるだろう。

6 関西イベント&プロジェクト データベース構築

株式会社 京 都 新 聞 社

6. 関西イベント＆プロジェクトデータベース構築

6.1 概 要

阪神大震災による被災地の復興や、京都・滋賀を中心に、地域経済の活性化情報となる、地域に関係する近未来（現在から5年先）に予定されるイベントやプロジェクトなど、各種プログラム情報のみを蓄積するデータベースを構築し、地域社会の近未来情報の共有化を進める。

これまでのデータベースは、過去の情報をベースとしており、これらの事例、事象情報を利用する形態が一般化している。全国的に公開されている記事データベースは、マクロ的には十分その目的を果たしていると思われるが、地域情報をみた場合、拾いきれず十分とはいいがたく、ミクロ的な弱点をみせている。

これを補完する意味からも、地域独自の情報、特に近未来情報を集約、記事のみならず関連資料を付加した各種プログラム情報をデータベース化することによって、現在の記事データベースの拡充強化を図る。

これらの情報を活用することによって、更に確度の高い市場開発の可能性が見込まれ、地域の経済活性化を支援することができる。これを現代社会に浸透しつつあるパソコン通信とファックスを介して提供していくことを本データベース構築の狙いとした。

6.1.1 目 的

新聞社はこれまで「新聞」を第一義とし、紙面化されてしまえば、その情報は省みられることはほとんどなかった。また紙面化されるのは、集められた情報のごく一部であり、大多数は日の目を見ないことが多かった。さすがにこれではもったいない、とデータベース化して既存情報の保管を始めた。同時に、情報の多重活用、情報提供の多角化が研究されるようになり、その観点から生まれたのが記事データベースであり、今回の「関西イベント＆プロジェクトデータベース」計画である。

また、平成元年（1989）から蓄積を開始した記事データベースは、昨年4月から一般に公開しており、そのデータは本社の汎用コンピュータシステムで集中管理している。本社や支局は、取材上の必要から、近未来の事業計画、イベント情報を求めることが多く、タイムリーな情報検索ができない場合は、電話でこれらの情報に関する資料を請求する、という方法をとっていたが、一般ユーザからはリアルタイムに情報検索する方法がなかった。これをなんとか解決したい、という気持ちをかねてからもち続けていた。

更に、これら近未来の情報を新聞社内だけで利用するだけでなく、広く一般に公開することに

よって、地域経済の活性化に寄与できないか、過去情報の蓄積である記事データベースと併用することによって、新しいデータベース活用の道が開けるのではないか、こうした考えのもとに「関西イベント&プロジェクトデータベース」の構築計画をスタートさせた。

6.1.2 実施内容

(1) システム要件

前章の目的から次の基本的なシステムの考え方をまとめた。

- ① 新聞事業を補完する情報発信とする。
- ② 記事データベースと同様に商用化を図る。
- ③ システムはシンプルな構成とし、ワークステーションと汎用コンピュータで稼働するパッケージソフトウェアを採用し、費用負担を抑える。
- ④ 提供情報の入力は一ライトの負担を軽減するため、既存の記事データベースの活用や出稿後にバックアップでフロッピーディスクに収録された情報を活用する。
- ⑤ 情報提供は京都発信の情報に加え、地域情報も充実を図る。
- ⑥ 情報提供は有償サービスを前提に会員制とする。
- ⑦ 他のパソコン通信ホスト局とのネットワーク化を図る。
- ⑧ 情報提供はパソコン、ワープロ、ファックス、電話など既存の機器利用を前提とする。

(2) 機器システム構成

基本的な考え方を踏まえて、経済的なシステム構成でサービスを可能とする、現行の記事データベースと同様のシステム形態を採用することになった。

これは従来の汎用コンピュータシステムをゲートウェイで接続した、ワークステーションをフロントマシンに、光ファイリングシステムやFAX配信システムをLAN (Local Area Network) で接続し、課金システムはパソコンで行うなど、既存システムと連携を進め、軽量化と経済性を図ったシステムとした。

(3) ソフトウェア構成

ソフトウェアプロダクトは既設のパッケージシステムを採用。汎用コンピュータによる関西イベント&プロジェクトデータベースの構築、及び検索システムには、(株)平和情報センターの「Super/Bride」を採用、ワークステーションのパソコン通信ホストシステムには、インターリンク(株)の「SCOPE21」をカスタマイズによりシステムの構築を図った。

(4) データエントリー方法

登録対象の近未来関連情報は、汎用コンピュータを使った新聞組版システムから、端末

側のパソコンの操作で自動切り出した記事や、本社、支社、支局の出先機関から入手した情報のほか、行政情報や地元企業情報などを加えたものとした。また、関西イベント&プロジェクトデータベース構築までの近未来情報は、社内でパソコン単体に蓄積してきた。これらの情報についても、データ変換して取り込む。

量的には確度の高いデータベースとの観点から、現在以降5年後迄を条件として、1ヵ月200件から500件程度を予定した。

(5) 出力方法

関西イベント&プロジェクトデータベースは、社内的にはオンラインでフルスクリーンモードによる検索が可能。社外向けには京都新聞パソコン通信「情報BOX都大路」を介してラインモードにより検索サービスを行う。画像情報があれば、配信リクエストによりファックス出力が可能である。

6.2 実 施 結 果

関西イベント&プロジェクトデータベースのシステム構成は、各レベルのコンピュータシステムの混合システムとなった。ダウンサイジング化とか、マルチベンダーとかいわれるが、汎用コンピュータには、既に記事データベースが蓄積されており、これと新データベースの類似点を生かした結果となった。

汎用コンピュータをデータベースマシンとして、既設の光ファイリングシステムにイメージを取り込み、新データベースと連携をとった。また、既設FAX配信システムと、既設の課金システムも同様に連携を図った。汎用コンピュータ、光ファイリングシステム、FAX配信システムの中継は、パソコン通信ホストマシンのワークステーションで行うシステム構成となった。

6.2.1 運用処理概要

汎用コンピュータ系の運用処理の概要は、稼働中の京都新聞記事データベースから、定期的に近未来関連のデータを自動抽出し、合わせて校正リストを出力している。このリストをもとに近未来情報のキーワードとなるデータを、校正処理と同時に付加して修訂正処理を行い、関西イベント&プロジェクトデータベースを作成する。

6.2.2 ラインモード検索処理概要

パソコン通信を介しての検索処理形態。エンドユーザは、京都新聞パソコン通信「情報B

OX都大路」で関西イベント&プロジェクトデータベースにアクセスし、1行単位に画面表示されるメニューや指示に従って検索する。

サンマイクロシステムズ(株)のワークステーション(SUN-S/4-LX)に富士通(株)のF6680エミュレータを搭載、社外からのアクセスに備えている。

平成7年3月現在、アクセス可能な回線数は10回線設けており、このうち、一般ユーザには一度に7ユーザまで、関西イベント&プロジェクトデータベースのアクセスが可能である。

6.2.3 近未来データの移行処理概要

社内で独自にパソコンなどで処理される既存の近未来関連情報のデータ変換処理を行う。社内的には(株)アスキー社の「The CARD ver.5」カード型パソコン用データベースソフトなどが使われている。これらのデータを吸収するため変換処理を行い、これまで蓄積してきたデータの取り込みを図った。

6.2.4 出力処理の概要

関西イベント&プロジェクトデータベースを構築した後の、社内向けの出力処理として、従来から各局で実施されていた近未来情報のリスト作成を行うことにした。

6.2.5 実施体制

総合メディアセンター、制作局(電算部)で、近未来データベース作業部会を設置し、必要の都度、情報提供者として編集局が参加する。技術的な支援として* 京都電子計算(株)が参加する形態の実施体制をとった。

* 京都電子計算(株)=京都新聞社全額出資の関連会社でシステムサポートを行う。

6.2.6 予想される効果

平成7年4月より社内的な使用テストを繰り返しながら、リリース時期の設定を急ぎ、本来の以下の効果を図ろうと考えている。その予想される効果としては、次の4点に集約される。

- (1) 近未来(現在から以降5年先まで)を捕らえたデータベースで新たな市場開拓の情報源となる。
- (2) 地域の近未来情報を共有化できる。
- (3) 近未来情報の利用による地域の産業の活性化効果が期待できる。更には阪神大震災以降

の経済活動の復興を補う情報源となり得る。

- (4) 自治体を限定しない事業計画などの広範で公共性の高い情報源となる。

利用対象者については地域住民や地元産業界を想定している。利用用途は、地域産業界における将来の事業展開などへの活用を始め、地域住民には地元自治体や、近隣地域の事業計画情報などの利用が想定される。利用の程度については、1日50件以上のデータベースアクセスを期待している。

6.2.7 現システムの問題点

- (1) 分散型システムの脆弱性

汎用コンピュータ、ワークステーション、光ファイリングシステム、パソコンなどをLAN接続した構成は、時代の混乱を物語っている。集中型システムと比べて、分散型のシステムは、全体のシステムが同時に壊れることはないが、全体の障害発生率は上がることになる。また、汎用コンピュータ、ワークステーション、光ファイリングシステムなど機器管理が社内で分散しており、機器管理面で改善課題が残る。

- (2) 情報管理の分散化

近未来のような情報は各部局で分散管理されており、将来的なリンケージを想定すれば、情報の一元化などによる情報管理が必要である。

- (3) 将来性や拡張性

- ① オープンシステム化やダウンサイジング化

当面、汎用コンピュータをデータベースマシンの位置付けとしたが、近い将来には更に経済効率を上げるため、汎用機からの脱皮を具体化しなければならない時期がくる。その日のために、今回のワークステーションの採用を位置付けている。

- ② 汎用コンピュータによる経済性の問題

汎用コンピュータの維持管理や、新規システムの開発の経済性は、時代の要求とは遊離した方向に向かっている。

- ③ マルチメディア対応の問題

最近のキーワードの一つであるが、まだ先が見えない状態であり、これらの対応は今後の課題とした。

- ④ 高速通信やアクセスポイントの問題

画像の取り扱いや京都という知名度を想定した場合、遠隔地からのアクセスユーザ向けに高速の通信網とアクセスポイントの設置が不可欠である。

6.3 今後の課題と将来展望

6.3.1 データベースの再利用

普及しつつあるCD-ROM化に向けて調査研究中である。既に先進社の全国紙は縮刷版との併用で展開している。弊社もこの度の近未来情報や地域の色彩を出したCD-ROMの制作を目指したい。また、出版物の版下出力、連載ものの出版について、データベースからの再利用などは進めている。

6.3.2 FAX配信サービスの拡大や音声サービスの展開

パソコン通信によるデータベースサービスなど情報提供サービスに加えて、インタラクティブ（双方向）で、人に優しい地域への情報サービスとして、音声による情報サービスや、ファックスによる情報サービスも計画中である。

6.3.3 他社とのゲートウェイ接続

現在のサービス状況を踏まえて、より充実を心掛け、これらの期待に応えるためにはゲートウェイ接続やインターネットとの接続などを計画中である。

6.3.4 国際都市京都向けに翻訳システムの検討

関西国際空港の開港で、京都はこれまでに増して国際観光都市として注目を集めている。また、京都市街地では、大学が府内に分散するなど、学生離れが目立つ。これらの歯止めとはならないが、「大学のまち」や国際都市に相応しく、インターネットサービスや翻訳システムの採用を課題としている。

6.3.5 マルチメディア対応のバージョンアップを検討

現在の文字情報サービスに加えて、画像や音声を含んだサービスの展開を検討中である。

6.3.6 アクセスポイントの検討

京都新聞社は京都・滋賀をテリトリーとすることから、京都府内や滋賀県向けを始め、支社を置く東京や大阪にアクセスポイントの設置を検討している。しかし、会員数の増加や他社とのゲートウェイ接続を考慮にいれながら、ユーザの負担を抑え、より良い情報提供を念頭に置き具体化を図りたい。

7 ハイビジョン大型映像ソフト 流通促進用データベース構築

財団法人 大阪科学技術センター

7. ハイビジョン大型映像ソフト流通促進用データベース構築

7.1 概 要

ハイビジョンを中心とした大型映像システムは、各市町村の公共施設に急速に導入されてきたが、同時にそのソフトの不足が顕在化してきた。そのために、大型映像ソフトの内容をデータベース化し、大型映像ソフトの流通促進を図り、地方自治体等のソフト不足の解決とシステムの有効利用を行う。そのためには、既存のハイビジョンに使用可能なソフトをユーザフレンドリィでかつコンパクトな形のデータベースシステムの構築を行い、それをCD-ROM化することとする。

本データベース構築計画では、具体的に、

- ① 試用データベースの構築
- ② 配布データベースの構築
- ③ 上記の運用のためのユーザフレンドリィな周辺ソフトウェアの生成

が必要とされ、それらの構築を行った。

ここで開発するデータベースは、「ハイビジョン大型映像ソフト流通促進」を第一義とする。使用するハードウェアは開発用として高性能ワークステーション、ユーザが使用する機器としては高性能パーソナルコンピュータ、又はパーソナルペンコンピュータを想定し、ソフトウェアはそれらの機器で汎用的に使用可能な関係データベースシステム(RDBMS)とした。

7.2 データベースの特徴

本データベースを構築するに際し以下に示すような特徴をもたせることとした。

(1) 操作の容易なシステム

- ・コンピュータシステムを使用したことのない人でも操作マニュアルを見ることなく操作できるやさしいシステムにする
- ・汎用性を考え、ウインドウズ上で動作するシステムとする
- ・マウスで操作できるようにする

(2) 豊富で柔軟性のある検索を可能にする

- ・ソフトの検索に必要なと思われる項目を網羅する
- ・中間一致方式の検索(項目のどの位置に検索文字列があっても検索できる方式)ができる

(3) 内容やイメージで検索する手段を設ける

- ・作品の内容及びイメージのキーワードの項目を設け、これらによる検索も可能にする
- (4) 静止画像を表示できるようにする
 - ・映像ソフトを代表する写真（静止画）を取り込み、検索結果の画面で表示できる
- (5) 作成したデータベースのデータをCD-ROM化する

7.3 格納されたデータ

本データベースに格納されるデータは、「貸出・販売可能ハイビジョンソフト台帳」に基づいている。これは、近畿地方における実践的なハイビジョンの普及を目的に設立された「近畿ハイビジョン・マルチユース協議会」（1994年10月1日より「近畿新映像マルチユース協議会」と改称）が、1994年1月にハイビジョンソフトを所有している団体を対象に実施した「ハイビジョン・ソフトプライス調査」及び「ハイビジョンユーザソフト調査」の結果をまとめたものである。

この台帳には、以下に示すような件数のハイビジョンソフトの情報が収録されている。

(1) ハードメーカー	7機関	123件
(2) 放送局	4機関	100件
(3) ソフト制作企業	22機関	111件
(4) 公的団体	4機関	47件
(5) ハイビジョンユーザ	43機関	173件
総計	80機関	554件

また、それらのソフトを検索するために、台帳には既に

- (1) 業種別・団体別の一覧表,
 - (2) ジャンル別の一覧表,
 - (3) タイトル50音別索引,
- が用意されている。

貸出・販売可能ソフトの個票には、2種類の仕様があり、仕様1は公共団体が保有するソフトについての個票、仕様2はその他の機関が保有する個票である。

7.4 更新・追加データとその情報収集について

今回のデータベース構築のために、従来からある台帳の映像情報のほかに、作品のジャンル、作品の内容及び作品のイメージに関する情報を検索に使用できるように配慮した。特に、作品

内容や作品イメージはハイビジョンソフトをその内容に関連したキーワードやイメージを表す形容詞で検索しようとする試みである。また、ハイビジョンソフトを代表する写真（静止画像情報）を検索結果の文字情報と合わせて表示できれば、検索結果の表示の質が格段に向上すると考え、写真の取り込みも行った。

これらの情報の収集のために、本研究所では「近畿ハイビジョン・マルチユース協議会」と検討の上、ハイビジョンソフト所有団体にソフト情報の提供を依頼した。

依頼した情報の内容は以下のとおりである。

(1) 作品のジャンル

作品を芸術、自治体紹介など19のジャンルに分ける

(2) 作品内容のキーワード

作品内容を表すキーワードを複数付けてもらう

(3) 作品イメージのキーワード

作品のイメージを表す形容詞を複数付けてもらう

(4) 写真（静止画）

データベースに搭載すべき作品を代表する写真を添付してもらう

この写真をイメージスキャナーで取り込んで、データベースに組み込む

(5) 写真の表題

データベースに登録

7.5 データの構成

データベースのデータ構成は、現在のところ最も市場化が進んでいるリレーショナルモデルのデータベース管理システムを用い、データの格納効率及びデータの利用時の分かりやすさを考慮して決定した。当然ながら、データの重複にも配慮し、正規化を行っている。

大型映像データベースのテーブル構成は、以下に示す作品1テーブル、作品2テーブル、団体テーブル、写真テーブルの4テーブル構成とした。

作品1テーブルはデータベースの中心となるテーブルで、検索に必要な項目をすべて網羅している。

ジャンルは重要な検索項目であり、予め決められた19のジャンル名を文字（漢字）で蓄えることとした。この項目は予め決められた固定的な項目であるので、コードを作成してもよいが、検索時のデータの可読性を重視して文字のままとした。

内容のキーワード及びイメージのキーワードの項目についても、文字（漢字）のままの情報

を入力し、中間一致（検索時に文字列のどこかに検索したい文字列があれば一致したとみなす）方式の検索方法をとることにした。

・作品1テーブル

項 目 名	型	幅	備 考
作 品 コ ー ド	文 字	10	半角文字, インデックス
作 品 名	文 字	93	全角文字
制 作 年	文 字	8	YYYY-MM
再 生 時 間	数 値		単位・分
画 像 種 類	文 字	11	全角
団 体 コ ー ド	文 字	6	半角文字, インデックス
ジ ャ ン ル	文 字	46	半固定
内容のキーワード	文 字	86	複数可
イメージのキーワード	文 字	52	複数可
媒 体 テ ー プ	文 字	30	
媒 体 デ ィ ス ク	文 字	21	
貸 出 可 否	文 字	2	可又は否
貸 出 条 件	文 字	34	公的機関, 企業など
運 営 形 態 の 制 限	文 字	40	有料も可, 無料行事のみなど

・作品2テーブル

項 目 名	型	幅	備 考
作 品 コ ー ド	文 字	10	半角文字, インデックス
作 品 内 容	文 字	255	
媒体記録機器メーカー	文 字	24	
媒体フォーマット	文 字	25	
記 録 方 式	文 字	20	
マ ス タ ー 媒 体	文 字	22	
特 記 事 項	文 字	54	
費 用	文 字	96	
複 写 サ ー ビ ス	文 字	16	
備 考	文 字	100	
台 帳 ペ ー ジ	整 数		

・団体テーブル

項 目 名	型	幅	備 考
団 体 コ ー ド	文 字	6	半角文字, インデックス
団 体 名	文 字	28	
部 門 名	文 字	31	
T E L	文 字	29	

・写真テーブル

項 目 名	型	幅	備 考
作 品 コ ー ド	文 字	10	半角文字, インデックス
写 真	グラフィックス		BMP形式
写 真 の 題 名	文 字	40	

7.6 ユーザシステム機器構成

データベースシステムを組み込むコンピュータシステムのハードウェアは、融通性、互換性及びコストを考慮して、次に示される汎用パーソナルコンピュータとした。更に、現在のウィンドウズ環境の市場を鑑み、最も汎用性のある「WINDOWS 3.1」が動作すればよいことにした。

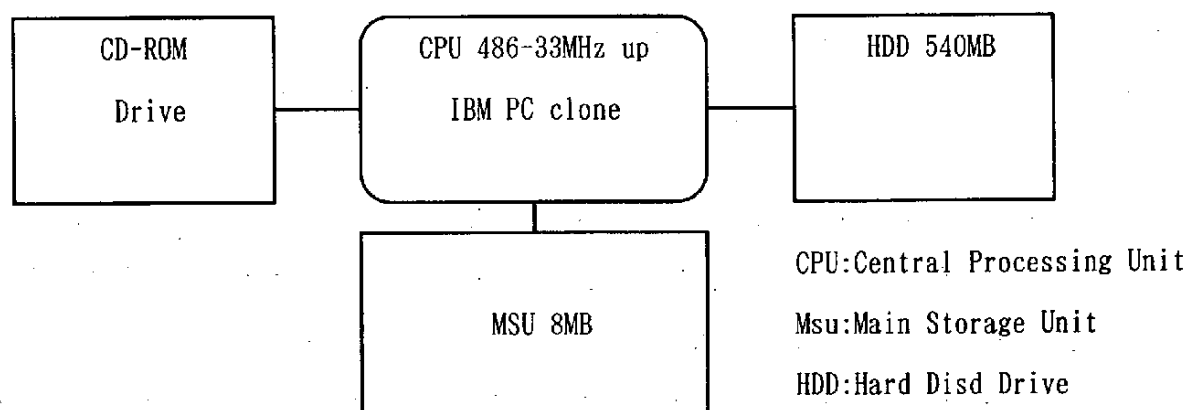


図7-1 ハードウェア構成

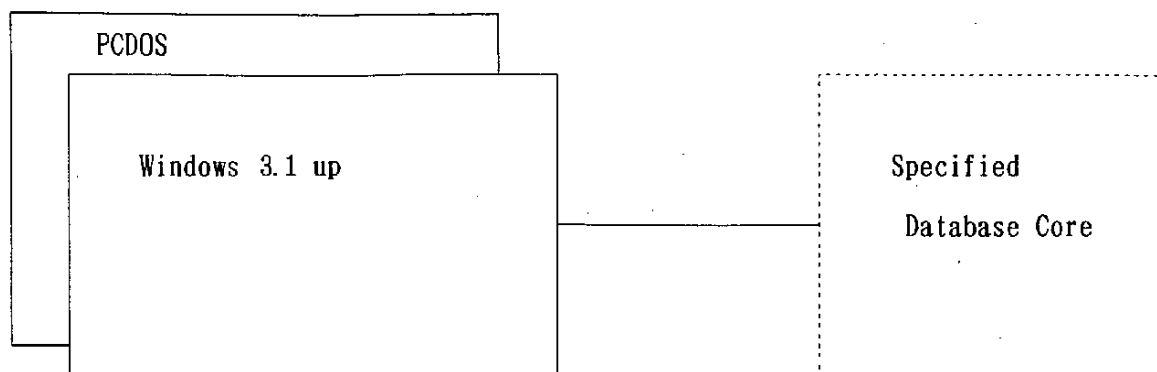


図7-2 ソフトウェア構成

7.7 データベースシステムの評価

本データベースの評価テストを行ったシステムの仕様を示す。

表7-1 評価用コンピュータの仕様

CPU (クロック数)	80486(66MHz)
内蔵RAM	16MB
ハードディスク	500MB
OS	DOS J 5.0/V
ウインドウ環境	WINDOSWS 3.1

作成した映像ソフト検索のデータベースシステムの評価として、条件検索に要する時間を計測した結果と、内容やイメージのキーワードを用いた検索結果は次のようになった。

(1) 検索時間

評価用パーソナルコンピュータ上で今回構築したデータベースシステムを本番のシステムと同一条件で動作させ、検索時間を計測した。なお、検索はユーザメニュー上に作成した検索用のフォームから起動され、検索されるデータの件数は562件であった。また、データベースのすべてのファイルはハードディスク上に置かれている。表7-2に検索時間の測定結果を示す。

表 7 - 2 検索時間

検 索 実 験 の 場 合 設 定	検索時間
数 値 項 目 に よ る 検 索 (1 条件)	1 秒以下
文 字 項 目 に よ る 検 索 (1 条件)	1 秒以下
文 字 項 目 の 中 間 一 致 検 索 (1 条件)	2 秒以下
文字・数値の組み合わせによる検索 (2 条件)	2 秒以下
文字・数値の組み合わせによる検索 (多条件)	2 秒以下

検索結果をみると、あらゆる条件における検索におけるレスポンスタイムが満足すべき状態で終了していることが分かる。検索時間は、ハードウェア構成、データベース管理システムが確保するワーク領域の大きさなどによって影響を受けるが、この結果は他で行われたベンチマークテストの結果とともに合致し、信頼できる数値ではないかと考えた。この結果、「Paradox」をデータベース管理システムとして使用したハイビジョンソフト検索システムは、実用上問題ない検索速度を有しているとの見解を得た。

また、文字と数値の組み合わせによる検索時においても、文字項目の検索は中間一致方式を採用しており、設計時に懸念していた中間一致方式の検索速度も十分使用に耐えうることが判明した。

(2) 内容やイメージのキーワードによる検索

本検索システムでは、ジャンルやソフトの再生時間などの検索項目のほかに、内容とイメージに関するキーワードを取り入れ、ジャンルが「スポーツ」で内容が「サッカー」、ジャンルが「音楽」でイメージが「勇壮」など、ソフトの内容やイメージについての情報でも検索できるように企画した。

検索実験を繰り返し行った結果、内容のキーワードを用いた検索については、ほぼ期待通りの検索ができているものの、イメージのキーワードを用いた検索については、ほとんどのケースでうまく機能しないという結果になった。

内容のキーワードについては、キーワードそのものも付けやすく、また、内容の種類もある程度の数に収束すると考えられる。

一方、イメージのキーワードについては、自由に付けることができる自由語方式をとったこともあり、作者によってキーワードの付け方がまちまちになったり、イメージのキーワードそのものが付けにくい性質のものがあつたなど、検索に有効に利用できるまでにはいかなかった。今後は、これらの情報を得るに当たり、ある程度キーワードの種類を制限することなどの工夫が必要と思われる。

7.8 データベース構築における課題

本データベースの構築において、文書データについては既にその本文からのキーワード自動抽出を実現している。画像や音響などのメディアについても、その内容を理解し、画像あるいは音響データからキーワードが出力されれば、

- (1) データ蓄積におけるキーワード付与の作業時間が短縮され、
- (2) データ蓄積のコストが軽減でき、
- (3) キーワードの使い方のばらつきがなくなって検索が容易になる、

などのメリットが得られる。

現在の画像処理の技術水準では

- ① 静止画の色彩分析から暖、寒など感性的な判定をする、
- ② 画像理解により自然の風景か、CG画像か、あるいは人が登場するか、
などの判定をする、

程度しか行えない。キーワード自動付与を実現するには画像や音楽についてこの分野の研究を本格的に実行する必要がある。

7.9 ま と め

データベースの情報量は以下のとおりとなり、システム評価と併せて、汎用のパーソナルコンピュータシステムで十分に使用可能な状態となった。しかし、このデータベースを稼働するためにデータベースコアの費用が数万円程度要求されるので、ただ単に「ハイビジョン大型映像ソフト」を検索するという目的には高価なものとなろう。したがって、汎用でかつ安価な標準型のデータベースコアの普及が渴望される。今後、CD-ROMの使用可能なハードウェアシステム及びそのソフトが普及することは疑いがないので、いわゆるデータとしてのソフトを管理する意味でのデータベースコアの出現が待たれる。

ファイルの種類	容 量 (K B)
作品 1 テーブル	260 6
作品 2 テーブル	387 6
団体テーブル	10 6
写真テーブル	18
注意 格納にはファイル が分割されている	27,885 6 1
トップフォーム	10
トップヘルプフォーム	9
検索フォーム	40
検索ヘルプフォーム	16

詳細表示フォーム	25
詳細表示ヘルプフォーム	9
一覧表示フォーム	22
メンテトップフォーム	12
作品 1 メンテフォーム	16
作品 2 メンテフォーム	14
団体メンテフォーム	9
写真メンテフォーム	8
印刷レポートフォーム	20
総 計	28,795

(注) 作品テーブル 1, 作品テーブル 2, 団体テーブル, 写真テーブルは複数のファイルに分割されているため容量欄は複数行記述にしている。

8 分散協調型データベースのための高度運用システムの 開発と学習支援環境への応用

株式会社 エ マ ー ズ

8. 分散協調型データベースのための 高度運用システムの開発と学習支援環境への応用

8.1 目的と概要

8.1.1 データベース利用の現状

現在、わが国では、各方面でさまざまな情報化作業への取り組みが積極的に展開されている。情報化作業では、インターネットを用いたコンピュータネットワークの形成と、その上で行われる今まで蓄積されたデータ（地域情報、個人情報等）のデータベース化、またデータベース化された情報の有効利用が成功の鍵を握っている。

インターネットについては、今まで学術ネットワークが中心となり形成されてきたが、日本でも商用インターネットプロバイダの登場に伴い、個人でもインターネットへの接続が比較的手軽に行えるようになり、まさに破竹の勢いで成長している。

しかしながら今まで蓄積された既存データのデータベース化、また既にデータベース化された情報の有効利用については、なかなか進展しないのが実情である。その原因として挙げられることは、データベースを積極的に利用する環境が整っていない、またデータの収集、蓄積等に多くの時間と労力を費やす必要がある、データベースの構築が実現したとしても、その後の管理維持がなかなか困難である、といったことである。

8.1.2 地方からの情報発信

データベース化された情報の情報発信については、これほどコンピュータネットワークの充実に力が注がれているにもかかわらず、首都圏からの情報発信がほとんどであり、地方にも存在する、有用で存在価値の高いデータベースを使った情報発信は無いに等しいものであった。

まさしくこれは中央集権型モデルとして捉えることができる。その大きな要因としては、コンピュータネットワークを用いて、データベースの構築とその利用が容易に行えるソフトウェア、又は環境が今まで存在しない状況であったからである。

8.1.3 今年度の開発（MineS の強化と学習環境への応用）

われわれは今まで述べてきたデータベース利用をとりまく現状を踏まえて、昨年度(財)データベース振興センターの委託を受け「地域情報を対象にした分散協調型データベースシステムの開発」を行い、その中でプロトタイプシステムを作成した（以降このシステムをMineS

プロトタイプ版と呼ぶ)。昨年度の開発では、データベース利用に関する問題を解決するための情報環境（MineS が目指す情報環境）の提案を行い、その環境構築の一つの実現手段としてMineS プロトタイプ版を作成し、実現の可能性を実証した。

今年度は、MineS プロトタイプ版を実運用に耐え得るように強化、拡張をし（以降このシステムをMineS 運用版と呼ぶ）、運用評価を行った。

また、MineS 応用として、現在コンピュータ利用が急速に広まりつつある学習支援環境を選び、そのために必要な機能をMineS に追加し（以降このシステムをMineS2と呼ぶ）、MineS の学習環境への応用を試みた（図8-1）。

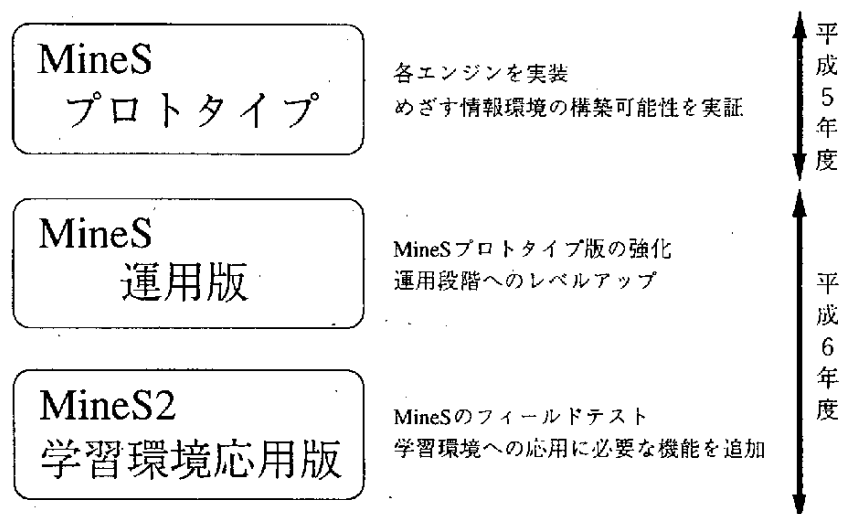


図8-1 昨年度、今年度のMineS の開発

8.2 MineS アーキテクチャとMineS プロトタイプ版

8.2.1 インターネット上の情報サーバの調査

昨年度のMineS プロトタイプ版の開発と情報提供システムの調査をもとに、今年度は、MineS と他のインターネット上の情報提供システムとのより詳細な比較を行った。ここで比較対象としたシステムは情報提供者と情報利用者が存在し、両社が非同期に利用可能なシステムである。表8-1がその比較を表している。

表 8 - 1 MineS と他の情報提供システム

名 前	情報の提供方法	情報の存在場所	情報の構造化法	アーキテクチャ
WWW	ハイパーテキスト の記述	分散配置	ハイパーテキスト	サーバ/クライアント モデル
gopher	インデックスを記 述	分散配置	インデックス&デ ィレクトリ	サーバ/クライアント モデル
WAIS	インデックスを記 述	分散配置	インデックス&デ ィレクトリ	サーバ/クライアント モデル
fip	ファイルを配置	ローカル	ディレクトリ構造	サーバ/クライアント モデル
archie	fip サイトへの問 い合わせ (自動)	ローカル	インデックス	サーバ/クライアント モデル
finger	ファイルの作成	ローカル (重複)	な し	サーバ/クライアント モデル
NetNews	テキストの送信	ローカル (重複)	ディレクトリ構造	サーバ/クライアント モデル&サイトリレー
E-mail	テキストの送信	ローカル (移動)	な し	サーバ/クライアント モデル&サイトリレー
MineS	メタ記述可能なU I 記述言語の送信	分散配置	DBを含みプログ ラマブル	すべてプログラム可能 なパートナモデル

8.2.2 MineS アーキテクチャ

(1) MineS アーキテクチャ

MineS が目指す情報環境の実現を達成するために、システム構築の際の仮定に相当する事柄、メディアの取り扱い、ネットワークの利用形態、及び具体的なシステム構成について述べ、全体の動作の概要を説明する。MineS が目指す情報環境を図 8 - 2 にまとめた。

① MineS の目標

MineS は、目指す情報環境を実現するために以下の三つの視点から、システムを構築する。即ち、(a)パーソナルデータを対象にする、(b)それぞれのコンピュータの機能を活かしながら複数の端末に対応するユーザインタフェースをもつ、(c)インターネットを中心としたネットワークでありながら個人が利用できる範囲の広いネットワークを利用す

る。

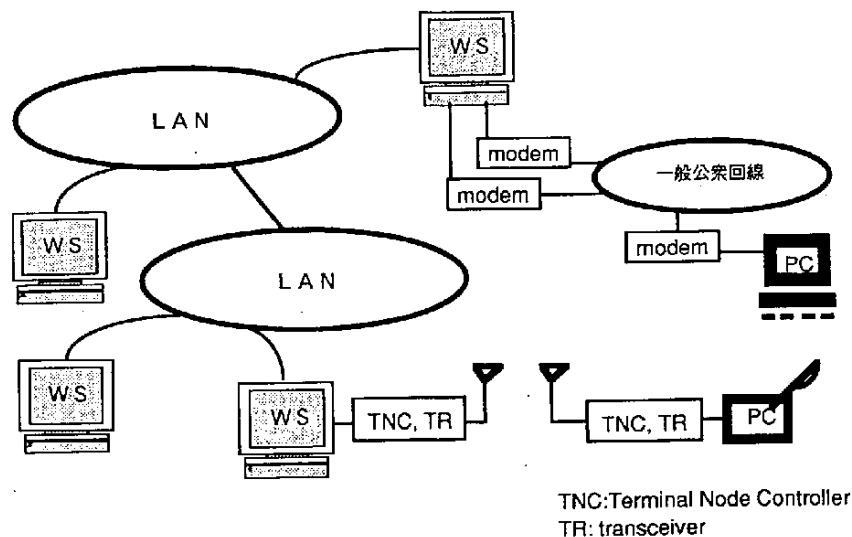


図 8 - 2 MineS のアーキテクチャ

② メディアアーキテクチャ

MineS で扱うメディアは 2 種類である。即ち、(a)記録媒体の抽象化としてのメディア（いわゆるマルチメディア）、(b)記録メディアと転送メディアを統合し抽象化したメディア（ユーザメディア）、である。MineS ではユーザに対して環境として情報を提供するために、ユーザメディアも統一的な方法で複数扱う（マルチユーザメディア）ことを目的としている。

③ ネットワークアーキテクチャ

MineS では、利用者がネットワークや端末を意識しないでシステムを利用できるように、図 8 - 2 で示したような、インターネット、公衆回線、無線などのあらゆるネットワーク利用形態を想定した。

④ システムアーキテクチャ

MineS のもつ三つの機能と、それらが実装されるエンジンの対応を表 8 - 2 に示す。また MineS のシステムアーキテクチャを図 8 - 3 に示す。各機能は上記①及び②を満たすために、機種、(②で述べた)メディア、ネットワークを抽象化し、そのインタフェースを定めることを目的としている。また、各機能に（特にユーザインタフェース）はシステム稼働中にデータに合わせ柔軟に変更可能なようになっている。以下では各機能について概説する。

表 8 - 2 MineS の各機能

MineS の三つの機能	機 能 詳 細	実装されるエンジン
MineS/D B 機能	・データ格納 ・データ検索 ・他のDBシステムの利用	D B E
MineS/U I 機能	・ユーザとのインタラクション ・ユーザインタフェースの記述言語によるアプリケーションの作成 ・ユーザインタフェースによるデータのカプセル化	U I E
MineS/C S 機能	・ネットワークの抽象化 ・MineS/C S 間での通信	C S E

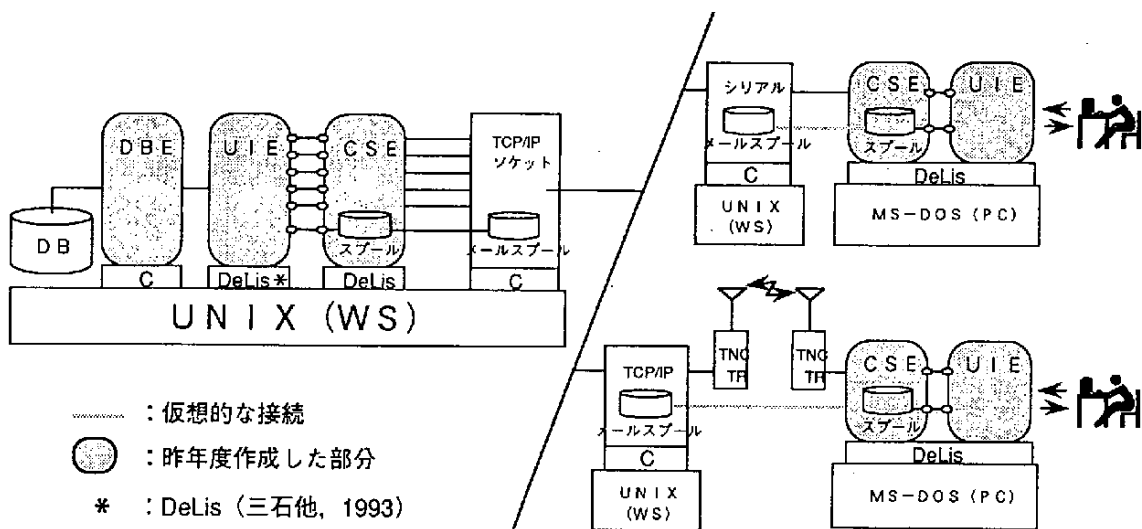


図 8 - 3 MineS システムアーキテクチャ

(a) MineS/D B (データベース) 機能

・データ格納:

MineS/D Bはデータの格納機能を提供し、データの作成を支援する。データはすべてテキストで表現されているので、MineS で提供するデータメンテナンスツールを使用しなくともデータの作成が可能であるとともに、変更も容易である。

・データ検索:

MineS/D Bのデータ検索機能は、複雑な検索機能をもたず、データの変更、

追加の容易性からフルテキスト検索方式を用いている。またマルチメディアデータの検索については、独自の管理機能を設け実現している。

- 他のDBシステムの利用：

MineS/DBでは、MineS/UI（後述）からの検索要求を受け取り、その検索要求を満たすデータを返すインタフェースが規定されている。即ちこのインタフェースを既に存在するDBMSに対して準備することで、そのDBMSがMineS上で利用可能となる。

(b) MineS/UI（ユーザインタフェース）機能

- ユーザとのインタラクション：

MineS/UIはコンピュータの機種を意識することなく、使い易いユーザインタフェースを提供する。このためMineS/UI機能を担うユーザインタフェースエンジン（UIE）ではUIの構造を記述するための言語（UI記述言語）を定義し、これによって書かれたUIをインタプリタにより実行するインタプリタ方式をとる。インタプリタは各機種ごとに作成し、多機種に対応可能にする。

- ユーザインタフェース記述言語によるアプリケーションの作成：

MineSではユーザにマルチメディアデータをいかに有用に提供するかを考え、ユーザ固有のメディアの形（ユーザメディア）を構築できる機能をもつ。このためUIEでは、ユーザインタフェース構築ツールを提供する。ユーザインタフェース構築ツールはUI記述言語により提供される。

- ユーザインタフェースによるデータのカプセル化：

検索や蓄積の対象となるデータそのものに、それぞれのデータの種別に適合したユーザインタフェース情報を付加することによって、データベース構築時の負荷を減らし、利用者にとっては使い易いユーザインタフェースの提供を実現する。

(c) MineS/CS（コミュニケーションサポート）機能

- ネットワークの抽象化：

具体的なネットワークを抽象化したMineS特有のネットワーク機能をMineS/DB、MineS/UIに対して提供する。そのためにMineS/CSは、（1A）複数の同期チャンネル（リアルタイム利用のための通信ポート。個数は実装される機械の能力による）と、（1B）一つの非同期チャンネル（電子メール転送のように蓄積されながらデータの転送を行うためのポート）、を提供する。

- MineS/CS間での通信：

それぞれのネットワークのデータ転送のための通信プロトコルをそのまま利用す

る。即ち、インターネットであればTCP/IPで直接通信を行う。また、パソコン通信ホストでは無手順でデータ通信を行う。これらは（既に現存する）プロトコルの変換器を通して互いに通信可能である。

(2) 全体動作

MineS 全体の動作の流れを図8-4に示す。DBE、UIE、CSEそしてDBEが管理するDB（データベース）がすべて実装されたシステムをMineS フルシステムと呼び、データの蓄積・検索要求を受け付けるサーバとなり得る。また、UIE、CSEのみが実装されたシステムは、クライアントとしての機能のみをもつMineS サブシステムである。MineS フルシステムは、自らのUIE、CSEをMineS サブシステムのように用いて、他のMineS フルシステムと通信することも可能である。

この図ではユーザは左側にあるMineS サブシステムのUIEと直接インタラクションをしている。MineS サブシステムのUIEはユーザの入力をUI記述に変換する。UI記述とはMineS のUI（ユーザインタフェース）の構造を記述した言語である。MineS サブシステムのUIEはCSEのネットワーク抽象化機能により、ネットワークを意識せずに、UI記述を用いてMineS フルシステムのUIEと通信することができる。UI記述を受け取ったMineS フルシステムのUIEは、UI記述をインタプリトし、DBEへ処理要求を出す。DBEはUIEから処理要求を受けて、DBに対するデータ検索、データ蓄積などの処理をする。

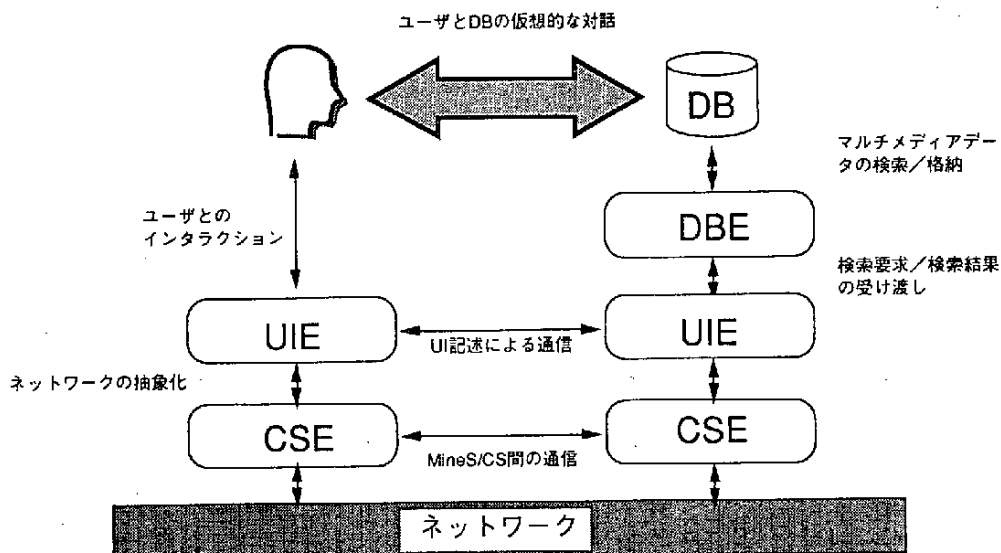


図8-4 全体の流れ

DBEは、処理結果をUIEに返す。結果を受け取ったUIEは、MineS サブシステムのUIEに検索結果を渡すため、データのカプセル化を行う。カプセル化されたデータは

CSEのネットワーク抽象化機能によりMineS サブシステムのUIEに渡され、UIEはそのデータをインタプリトしてユーザに表示する。

このように、実際はUIE、CSE、DBEを通じてユーザの要求が処理されるが、ユーザにはあたかもDBと直接対話をしているようにみえることになる。

8.2.3 MineSの実装 (MineS プロトタイプ版開発)

今まで述べてきたMineS のアーキテクチャを実現するために、前述した各エンジンの機能をMineS プロトタイプ版に実装した (表8-3)。

表8-3 MineS プロトタイプ版の実装

	各種OSとMineS の通信方法		現在の各エンジンの実装			代表的機種
	OS	プロトコル等	DBE	UIE	CSE	
1	UNIX	TCP/IPによる同期通信	○	○	○	UNIX ワークステーション UNIXを載せたパソコン
2		無手順での同期通信	×	×	×	
3		uucpなどによる非同期通信	×	×	×	
4	MS-DOS	TCP/IPによる同期通信	×	○	○	PC-9800パソコン
5		無手順での同期通信	×	×	×	
6		uucpなどによる非同期通信	×	×	×	
7	MS-Windows	TCP/IPによる同期通信	×	○	○	PC-9800パソコン
8		無手順での同期通信	×	×	×	
9		uucpなどによる非同期通信	×	×	×	
10	MacOS	TCP/IPによる同期通信	×	×	×	Macintosh
11		無手順での同期通信	×	○	○	
12		uucpなどによる非同期通信	×	○	○	

(注) 「各種OSとMineS の通信方法」の「無手順での同期通信」とは、一般のパソコン通信のようにシリアル回線を制御するソフトを用いた場合を指す。

「現在の各エンジンの実装」の○印は実装済を示す。

8.3 MineS 運用版

8.3.1 MineS 運用版の仕様

(1) MineS 運用版とは

われわれが考えるMineS 運用版とは「多くの人が違和感なく利用でき、ユーザの要求するデータの取得が高い信頼性のもとで実現できる水準に達している状態のもの」である。即ち昨年度のMineS プロトタイプ版はわれわれの考えることの実現ができるか否かを実証する場にすぎなかった。そして今回の開発目的の一つはそれら実証されたことをベースにして、実運用に耐え得るMineS 運用版を構築することである。

(2) MineS 運用版の仕様

MineS 運用版は大きく分けて二つのシステムから成り立っている。一つはDBE, UIE, CSEを実装したMineS フルシステムで、複数のMineS サブシステムからの検索要求に対応することができる。またもう一つはUIE, CSEを実装したMineS サブシステムで、MineS フルシステムに対してデータの検索要求等を発行する。

本年度の開発では、MineS フルシステムとMineS サブシステムをわれわれが考える実運用レベルに値するものへと強化するために各仕様書を作成し（約 100頁）、それに従って開発を進めた。具体的にはネットワーク機能の充実、ユーザインタフェースの改良、検索され得るデータの構造の変更等を行った。そしてこの拡張を施したことにより、MineS 運用版は仕様書とともにネットワークを通じて広範に配布することができるとともに、実運用を行うことが可能となった。

8.3.2 MineS の記述言語

MineS 運用版は、表 8-4 で示す関数群から構成されている。これらの関数を用いることにより、MineS フルシステム、MineS サブシステムを構築することができ、またユーザ自身も自由に変更が可能となっている。また、関数の仕様がこのように明確になったことにより機種、又はネットワークを意識せずにシステムを構築することができた。更に、完全な仕様かでき上がったことにより、広範な機種への移植が可能となった。

8.3.3 MineS の運用評価

(1) 運用形態

今回の運用評価はネットワークには東北インターネットを用い、素材としては昨年度同様、東北地域情報サービス企業データ（約 1,000件）を使用し、利用者のワークステーション

ョン上にMineS フルシステム, MineS サブシステムを構築し運用評価を行った。

利用者としてはわれわれ開発者以外の第三者に依頼し, 客観的立場で利用してもらい, アンケート調査及び直接の聞き取り調査により, 利用の感想, 意見等を集約した。

(2) 運用結果

表 8 - 4 MineS の記述言語

	MineS フルシステム	MineS サブシステム	共 通
C S E	—	—	send receive connect-channel
U I E	open-db close-db search-data search-caped-template giveme-caped-template giveme-data data-expand Set-connect-id make-template-id keep-template encapsulate	query-db-connect query-db-disconnect query-search-data query-search-caped-template query-caped-template sync make-outdb-name close-win save-win connect-and-make-port-list expand-sexp brows-first-rcdno brows-last-rcdno next-brows-cnt prev-brows-cnt brows-rcd disp-rcd gethost-by-db	set-encap-data mines-template stack-init stack-append stack-push
D B E	DbSearch	—	—
その他	server daemon getpeername squeeze	open-mines	

運用結果について、MineS 運用版はわれわれが考える運用レベルに達したものであり、利用者からも良い反応が得られた。

しかしながら、まだまだMineS 運用版は開発者の立場で構築された部分があり、実際にユーザの使いやすいユーザインタフェース、操作性等を考えなければならないことが明確な形でリストアップされた。これは使用者の意見等をもとにして改良していく必要性がある。

総括として、MineS 運用版はまだ改良の余地ありといった結果を残したが、しかしながら十分実運用に耐え得るものであった。

8.4 学習環境への応用

8.4.1 情報環境下における学習環境

近年、コンピュータやネットワークの発展に伴い、小中学校の授業といった学習環境（授業を行うクラスルーム等）におけるコンピュータ利用が増加しているが、その利用形態はスタンドアローンで教育用ソフトウェアを利用するものや、単にネットワークを用いて教育用ソフトウェアを共有するだけのものが主流である。

また、その教育用ソフトウェアは、その多くが特定教科向けで市販されているものであるため、さまざまな外部のデータを取り込む等の拡張性に乏しく、授業での利用方法が限定されている。このような学習教材はコンピュータネットワーク環境を有効に利用しているとはいえない。

学習教材には学習の効果を最大に引き出すという目的があるため、教師が授業に向けて学習教材を準備する段階では、教師の意図を反映させやすいことが要求され、その利用段階では、教材をもとにした議論の発展や、生徒の想像力、表現力の養成が期待される。このような学習教材は学習環境において重要な位置を占めている。

そこでわれわれは、教師と生徒が各自、ネットワークで接続されたコンピュータを利用できるような情報環境下において、次のような学習教材を中心とした学習環境の構築を試みる。

(1) 学習教材の準備段階

教師による学習教材の準備段階では、教師自らが、ネットワーク上に存在するさまざまなマルチメディアデータなどを自由に組み合わせて表現できる学習教材を作成できる環境を提供する。これにより教師は、授業で伝えなければならないことを素直に学習教材へ反映させることができる。また、一度作成した学習教材を、ネットワークを通じて再利用することもできるような環境を提供する。

(2) 学習教材の利用段階

授業などの学習教材の利用段階では、教師の作成した学習教材が、生徒個人に割り当てられたコンピュータ上で、ネットワークを介して利用できるような環境を提供する。更に、生徒が提示された教材をもとに、自分なりの視点によって自由にデータを入れ替えたり、データの並べ替え等を行うことができるような環境を提供する。

また、実際の授業では、学習教材をもとに教師と生徒の間に次のようなコミュニケーションが存在する。

- ・教師が学習教材を用いて進める授業の中の、教師や生徒の間の自由な意見、質問の交換。
- ・準備された学習教材をもとに、生徒の発見した新たな視点の生徒自らによる、教師、他の生徒への表現。

このような学習者間のコミュニケーションを教材によって支援することができれば、教室に一同に介するといった物理的・時間的制約に縛られることのない、ネットワーク上の仮想的な教室（バーチャルクラスルーム：図8-5）を実現できる。

バーチャルクラスルーム内部では、そのクラスルーム内にいる人々間のコミュニケーションの相互作用という一種の共同作業により、次々と新たな視点が生み出される。また、他のバーチャルクラスルームとの間でコミュニケーションが行われれば、その共同作業の規模、効果は留まるところを知らない。

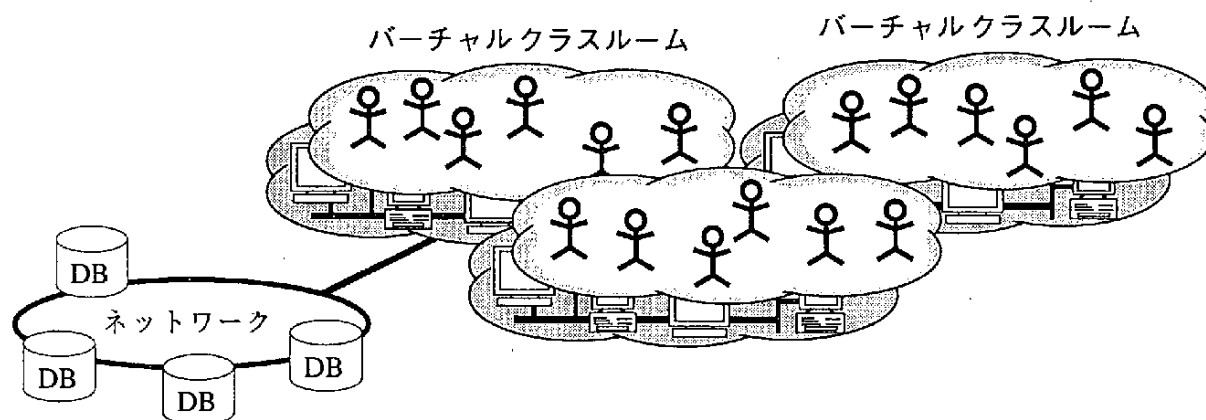


図8-5 バーチャルクラスルーム

8.4.2 MineS による学習環境の構築 (MineS2)

上記のような学習環境をMineS を用いて実現するためには、MineS に幾つかの機能を追加する必要がある。われわれはMineS を学習環境に応用するためMineS に必要な機能を追加し

たものをMineS2と呼ぶ。必要な機能は以下のとおりである。

(1) 素材データの取り込み機能

教材を有効に利用するためには、目的に合ったデータがネットワーク上に豊富に存在することが不可欠である。そのためには誰もがデータベースを容易に構築できることが望ましい。素材データの取り込み機能とは、このデータベースの構築という作業を自動化し、有効なデータを簡単にデータベース化できるということであり、これは、MineS システムの外部に存在するデータをMineS のアーキテクチャの内部にどのように取り込んで行くかといった課題である。

(2) 教材の作成機能

教師や生徒が、自由に教材作成、変更できるということは、即ち、ネットワークやコンピュータ、データベースの専門家でなくとも、簡単にこれらを制御できるということである。そのためMineS2では、GUIを用いた教材構築ツールを提供し、誰でも容易に教材を構築できる環境を提供する。

また、このようにして作成した教材を個人のデータベースに格納することにより、ネットワークを通じた教材の再利用も可能とする。

(3) 学習者間コミュニケーションの支援

学習者間の複雑なコミュニケーションを支援するためには、意見交換のための音声や動画像による会話や、自らの意見を皆にプレゼンテーションできるような機能が必要である。そのためMineS2では、教材に学習者間のコミュニケーションを支援する機能を追加する。

MineS に以上の機能を追加することにより、情報環境下における学習環境が容易に構築できる。

8.4.3 MineS における教材の扱い

学習環境の実現に際し、MineS におけるユーザインタフェース及びデータは以下のように対応する。

(1) メタ教材

メタ教材は教材作成ツール等によって、利用者が作成したデータのビューワである。UI 記述言語で記述されており、MineS のユーザインタフェースに当たる。メタ教材に対して素材データを入れ替えることにより、異なる教材にすることが可能である。メタ教材には数値データのグラフ化機能、ボタン等のユーザインタラクション機能等がある。またメタ教材をデータベースに納めることで、ネットワークを通じた再利用が可能である。

(2) 素材データ

MineS がデータベースに管理しているデータである。MineS2では、データベース構築の自動化によって外部から取り込まれた、気温や湿度のような時間軸に依存して変化する情報を定期的に取り込んだデータなども、他のデータと同様に扱うことができる。

(3) 教材

素材データをメタ教材に埋めこみ表示したものである。MineS2は主にプレゼンテーションを目的として教材を扱う。MineS2の教材はさまざまなデータベースから収集される素材データと、その素材データに応じたメタ教材によって構成されている。

教材を使用した例を図 8-6 に示す。これは、作成したメタ教材に、仙台と札幌の気温という素材データをグラフ化して表示し、また同時に当地の写真を一緒に表示した教材の例である。

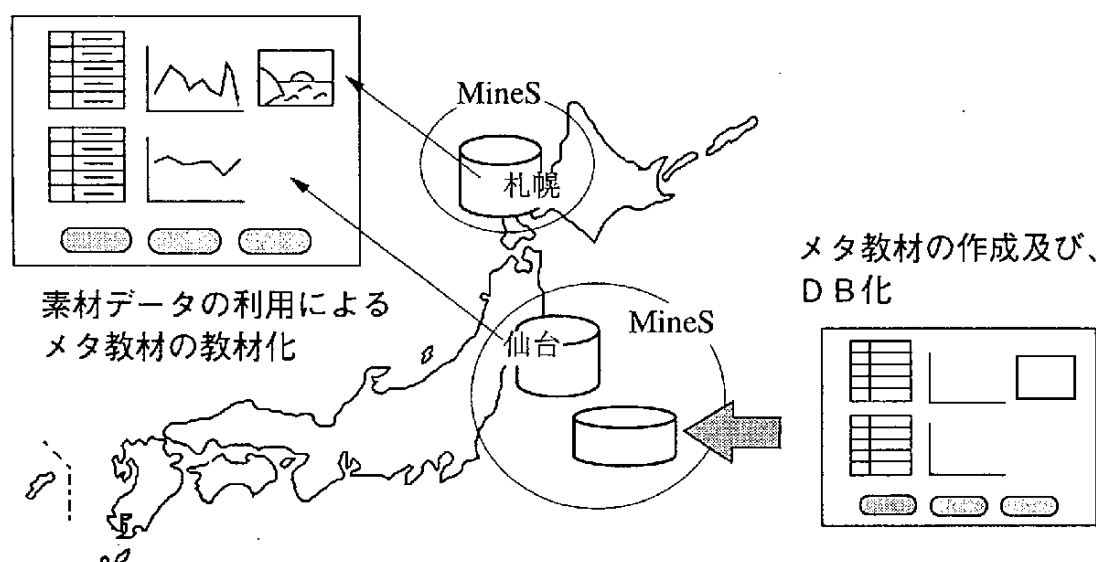


図 8-6 メタ教材，素材データ，教材

8.4.4 MineS によるコミュニケーションの支援

学習者間でなされるコミュニケーションをMineS で実現するためには、以下のような機能が必要となる。

(1) 会話機能

音声、文字、動画等を用いて利用者同士で会話をする機能である。同期的会話、非同期的会話、ブロードキャストによる会話などが選択可能である。これにより、利用者同士の意見交換などが可能となる（図 8-7）。

(2) プレゼンテーション機能

表示している教材をもとに、その教材に対する操作や、マウスや簡単な描画による教材への指示等を、そのまま他の利用者の教材に表示させる機能である（図 8-7）。

また、会話機能やプレゼンテーション機能を学習者間で有機的に行えるようにするため、教材を表示している際に動的にメタ教材を作成・変更できる機能も必要である。それが動的メタ教材作成機能である。

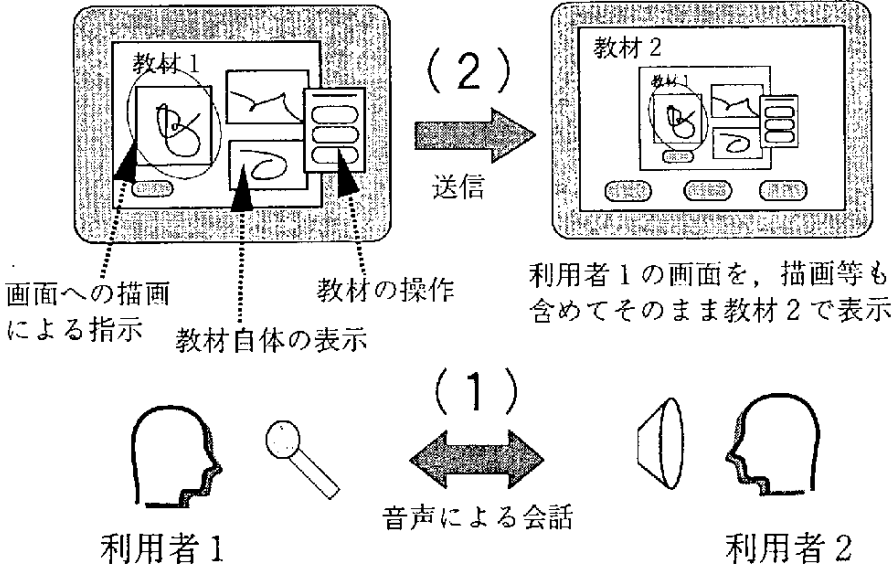


図 8 - 7 会話機能とプレゼンテーション機能

(3) 動的メタ教材作成機能

現在利用者が表示している教材に対して、GUIによる教材構築ツールを用い、表示中の教材を変更したり、新たなメタ教材を生成する機能である。これにより、表示中の教材を、他の利用者と通信できる教材に変更できる（図 8 - 8）。

以上の機能を組み合わせることで、学習空間内の学習者間のコミュニケーションの実現が可能となる。

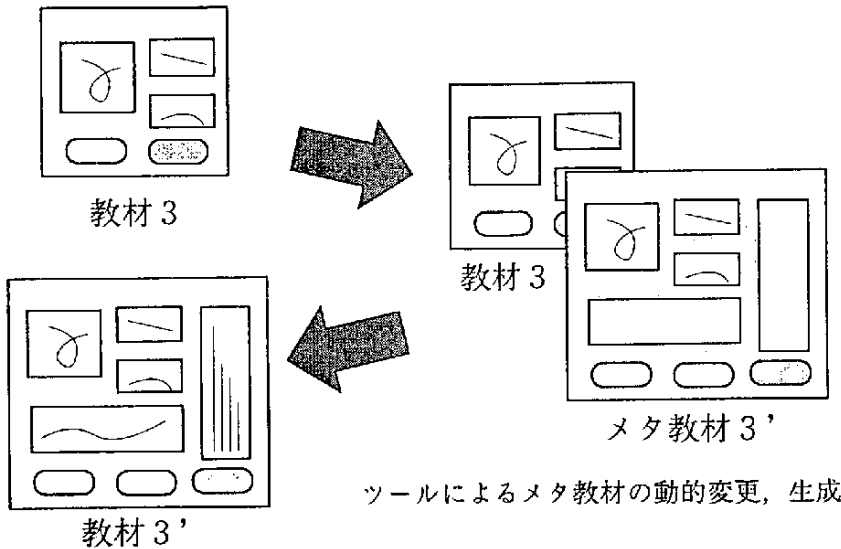


図 8 - 8 動的メタ教材作成機能

8.5 ま と め

本年度の開発ではMineS が目指す情報環境が運用レベルで実証できたこと、依然問題は残っているが多くの人に使用していただき貴重な意見を頂いたことはわれわれにとって大きな収穫であった。MineS 運用版を基礎として、今まであまり進展していなかったコンピュータネットワーク上のデータベース利用を発展させるよう、あらゆる角度からアプローチを図るとともに、更なるMineS 運用版の発展をさせたいと考えている。

また、学習環境でのデータベース利用の増加も見込まれる。そこで今回提案したMineS の学習環境への応用の提案をもととし、学習環境でのコンピュータ利用の発展に貢献できることを確信している。

参考文献

(財)データベース振興センター, 1994 ; データベース構築促進及び技術開発に関する報告書, 地域情報を対象にした分散協調型データベースシステムの開発。

三石他, 1993 ; 分散環境のための言語系 DeLis, 情報処理学会研究報告93-PRG-10, pp. 57-64.

9 電子デバイス情報の電子化ドキュメントの プロトタイプ作成

電子デバイス情報サービス株式会社

9. 電子デバイス情報の電子化ドキュメントのプロトタイプ作成

9.1 E D S C (Electronic Document Study Committee)の活動

9.1.1 背景

システム的なICのドキュメントは、必ず長大なものになる。またそれは、単に回路設計者のみならず、装置ソフトのプログラマ、製造技術者、検査担当者、保守責任者、装置ユーザなど、多種多様な人々に、装置寿命が続くかぎり、いろんな局面で使用されるものになる。システムICのドキュメントの特色をまとめるとき、これらを紙媒体に印刷し、データブックとして綴じている現在の手法の困難が実感される。最先端の装置を最適に設計しようと志す技術者は、新規に追加される情報と、また一方では生産中止になった製品の情報を持ち、また追跡しようと懸命である。また、半導体メーカーも、新製品を広報しようと懸命である。

ところが、両者の関係を困難にしている要因も多い。それを列挙すれば、

- (1) ごく少数の情報消費者に、的確に情報をとどけること。
- (2) 消費者の必要とする情報を、精度高くとどけること。
- (3) 大量の情報ストックを可能とすること。
- (4) 情報の追加・削除・修正のメンテナンスを迅速にできること。
- (5) 数値・画像情報などを含められること。
- (6) 一覧性の高い、横断的な検索が可能なこと。
- (7) 設計、製造、検査、保守など各作業ステップに適した情報を提供できること。

などである。

9.1.2 インターネットでの情報提供

研究・商用ネットワークとしてインターネットはその地盤を固めつつあり、またMosaicに代表されるように、インターネットならではの通信応用も出てきた。

インターネットに各種のデータベース技術を付加したものが、半導体部品情報の提供メディアとして可能性が高いと判断される。

インターネットでの部品情報提供を考えると、初めて紙媒体を脱却する電子ドキュメントを構想すべき必然性が生じる。従来の紙メディアを指向する情報の作成・流通を越えて、予めインターネットでの情報提供を主流にすえた、ドキュメント構築法、流通法を開発しなければならない。

9.1.3 研究会の構成と活動

半導体技術情報を提供する上で、部品情報の消費者である設計者が、迅速に情報を入手できるシステムであることがまず必要で、次に部品情報の提供者であるメーカーが迅速に情報を発信するためのシステムであることが必要である。

米国において、半導体ドキュメントをSGML (Standard Generalized Mark-up Language) で作成、蓄積、共有しようという計画が進行しており、Pinnacles と呼ばれている。EDSCは、この活動と互いに相補うことで、協調するものである。

EDSCとPinnacles 両者に共通の要素として、電子部品（半導体、その他素子）の技術ドキュメントを電子媒体に移そうという、具体的な活動自体を挙げることができる。両者とも、CAD情報など、これまでのドキュメントに含められなかったコンピュータデータも電子データブックとして統合することを考慮しており、そのための技術開発を指向するなど、同様の方向性をもつ。

一方、異なる要素もある。Pinnacles の目的とするところは、主として参加者（メーカー間）相互の部品情報共有であり、それをエンドユーザのために検索したり配布したりすることについては、一応スコープの外にしている。

また、設計担当者などの技術情報消費者へ情報を届けるためには、その消費者のかかえる問題に即した検索方法や、特に紙媒体による提供が不可欠である。そのため、EDSCでは文書内容に即した検索方法の開発とあいまって、最初からレイアウト化された部品情報を蓄積し、消費者にはレイアウト後の画面（紙に印刷できるもの）を届けられるように配慮している。一方Pinnacles においては、紙媒体への配慮は検討の対象外で、その設計は行わないことになる。

Pinnacles とEDSCの目的の相違は、その達成のためにとる手段の相違となって現れる。即ち前者では機械処理による文書の共有を主たる目的とするため、メーカー間での情報記述形式の統一が主たる興味となる。そこで、具体的な技術としてSGMLをとりあげ、それを中心とする作業を展開している。

一方、EDSCの目的は、完成された文書の配布である。したがって、その元文書の記述方法を統一することは第一義の目的ではない。むしろコンピュータを利用していかに合理的に文書入力、整形、蓄積、配布を行うかの観点から実態把握、配布に適した情報フォーマット (Postscriptなど) の検討、検索システム (Mosaicなど) の検討を行う。以上のようにEDSCとPinnacles は相補的な性格が強いため、将来にわたって連携を保っていく必要がある。

EDSCでは、その提案プロトタイプ設計と評価を、下記の三つのグループにより進めた。

- ① WG0 全体の調整、情報提供メディアの選定・実験
- ② WG1 情報提供者からみたEDSCシステムの検討
- ③ WG2 情報の消費者からみたEDSCシステムの検討

まず、WG2において、現在の利用者の望んでいる半導体情報提供形態を検討した。

WG1においては、逆に情報生産者のメーカーの立場から、現在の情報生産工程を前提に、将来の電子化ドキュメントのあり方を解析した。

WG0では、近未来の提供者システムを検討するとともに、特にネットワークによる流通に向けて、そのプロトタイプシステムを構想し、実験した。SGMLに基づく情報記述法の検証と、ネットワークによる流通実験を行っている。

9.2 ドキュメント電子化の現状と課題

9.2.1 電子化の現状

(1) 半導体メーカー各社に対して、電子化ドキュメントの取り組み状況を調査し、以下の点が分かった。

- ・現状では、DTPを使用してドキュメントを作成し、紙でユーザに提供している。
- ・紙情報での製作、配布が難しくなっており、また製本コストも年々増している。カスタマーも保管場所や検索、またCADデータとしての再利用のための電子化を望んでいると判断している。
- ・今後は、オンライン、CD-ROMなどで提供することが必要と認識しており、ドキュメント情報をデータベース化し、異機種システムとの情報交換を可能（再利用可能データ）にしたい。
- ・将来の技術予測に基づく、デバイスメーカー及びユーザの要求を満たす標準化が必要である。
- ・インターネット等を利用し、World-Wideに情報提供できることが必要である。

(2) 半導体メーカー各社に対して、情報内容ごとに提供する媒体を調査し、以下の点が分かった。

- ・現状では、ドキュメントに関しては紙情報のみの提供が圧倒的である。
- ・将来的には、インターネットなどオンラインでの提供が大きく伸びることを予測している。ただし、紙情報での提供は10%程度減少するに止まる。
- ・CD-ROMはデータブックやマニュアルに、オンラインはデータシートや技術情報誌やCADデータの提供に利用しようという判断があることが読みとれる。

9.2.2 ドキュメント電子化の課題

紙への表示（印刷）を前提としていた従来の情報の作成／提供形態では、情報の作成、加工といった技術と情報作成後の出力（印刷）や流通・配布といった技術が、それぞれ独立して開発／運用されてきた。

しかし、「情報の管理」「共通化」「再利用」「多岐にわたる電子媒体での表示／提供」を中心とした今後の電子化ドキュメントに対する社会的要請に応えるには、「情報の作成／加工」から「情報の管理」、「再利用」、「表示／出力」、「流通／配布のネットワーク」といった電子文書交換（EDI：Electric Documentation Interchange）に係わるすべての工程を念頭においた総合的「情報技術の確立」と、「国際的な標準化」、関係者への啓蒙、教育、指導を含んだ「運用環境の整備」が必須である。

9.3 電子化ドキュメントのモデル

9.3.1 メーカー側からの電子化ドキュメントモデルの検討

(1) データシートの要素分析

半導体デバイスのデータシートの電子化を進めるに当たり、データの相互交換を可能とするためSGML文書化を図ることとした。その第一歩としてまず各社のデータシートを持ちより、データシートのモデル化を行った。

モデル化の作業として、一から構成要素を洗い出す作業はPinnaclesにおいて詳細な分析がなされており、アメリカと日本で同一作業を繰り返すような非効率を避けるため、PCISを参考にして作業を進めることとなった。委員会が標準化作業を開始した時点（1994年10月）では、まだPCIS 1.1が入手できなかったため、当面の作業として各社のデータシートの大まかな論理構造に該当する要素を列挙しクロスリファレンスを作成することになった。

この作業を行うに当たっては、電子デバイスと一口にいても製品によって記載項目がかなり異なるため、すべてを一つの表にまとめるのには作業がやりづらいという意見が多かった。そのため、製品群ごとの表にまとめた。

表における各項目の統一名称はPCISのタグ名に当てはまるものを日本語にして用いた。検討の結果、後々のデータ交換やツールのサポートを考えるとタグ名はできるだけPCISに準拠させたほうがよいという結論が得られた。既存のSGMLエディタ／コンバータにおいても日本語のタグ名に対応できないものもあるので、タグ名には英語を使用す

ることが一般的である。したがって日本語の統一名称は形式的なものであまり意味をもたないので特に議論もなかった。

今回の調査における共通要素はおおざっぱなもので、P C I Sのタグ名をそのまま用いることができるが、P C I Sのタグライブラリ及びD T Dを詳しく検討する時間的な余裕がなかったため、本委員会で挙げた要素とP C I Sの要素の内容が完全に一致しているか検証する必要がある。

また、W G 1では、図表の取り扱いをどうするかを検討する際に必要となるので、使用しているツールや出力可能なデータ形式の調査も行った。ただし、Pinnacles と同様に本委員会においてもツールやデータ形式の統一や推奨は考えていないため、これは参考として行ったものである。グラフィックデータや表の取り扱いについてはこの段階で標準を設けることは困難なため、できるだけP C I Sに準拠する形で各社がD T Dのサブセットを用いて対応する方向である。

(2) 構造化

上記のデータシートの構成要素分析により、各社のデータシートにおける記述項目の違いを把握した。この分析をもとに、最低限メーカ側として提供すべき情報をツリー構造で表現し、各項目のデータフォーマットや属性について考察した。

① 各社の提供する情報の項目に共通部分は多くあるものの、細部は微妙に異なっている。

そのため、データシートの構造化は、データシートの情報を提供する上で、最低限必要な情報に対してのみ行う。各社が独自に提供したい情報については、本活動では議論せず、各社が自由に項目を加えることを許容することにした。

② S G M Lのタグ名を決定する際のルールを次のように定めた。

- ・同一の内容を記述している項目については、（項目の名称が各社で異なるかどうかに関わらず）統一されたS G M Lのタグ名を定義する。
- ・一部の会社が独自に提供している項目については、S G M Lのタグ名称を定義しない。
- ・各社が独自に提供したい情報については、各社ごとにS G M Lのタグ名称を定義する。
独自に定義したタグを用いてS G M L文書を自由に定義してかまわないが、各社間で統一されたタグを別の意味で用いてはいけなない。
- ・タグ名には日本語（2バイトコード）を用いず、英語で表現する。
- ・支障のない範囲で、P C I Sで定義されているタグ名称（又はタグ名称定義のルール）に合わせる。

③ 上記の方針に則り、データシートの構成要素の中から、最低限メーカとして提供すべき情報の項目をピックアップした。次にピックアップされた各項目について、Pinnac-

esのP C I Sを参考にしながら、タグの名称を定義した。なお、作業の時点ではP C I Sの正確な仕様を入手していなかったため、P C I Sに則って作成した日本T I(株)のD T Dのタグ名に合わせることにし、P C I Sの詳細が判明した後にタグ名を再考することにした。

- ④ 情報の各項目について、どのようなデータ種別によってユーザに提供するかを決定した。データの種別は、次の四つ（及びそれらの組み合わせ）から選択することにした。

・テキスト ・図 ・表 ・写真（イメージ）

なお、それにより詳細なレベルのデータフォーマット（例えば、図の場合C G M, I G E S, P I C T等）は、ここでは決定しないこととした。これは次の二つの理由による。

- ・ユーザが提供されたデータを二次加工するような状況を想定した場合、データフォーマットはユーザが望むものを提供すべきである。
- ・ユーザが二次加工しないデータに関しては、各社が現在用いているデータフォーマットをそのまま用いても問題がない。また各メーカは、それぞれ異なるデータ作成ツール／データフォーマットを用いており、データフォーマットを統一することは混乱を招くのみで、メリットがない。

- ⑤ 次の情報の項目について構造化を行った。構造化は、情報を大分類／中分類／小分類の三つのレベルに分け、ツリー構造で表現することにした。

(3) 簡易D T Dの作成

上記(2)の⑤で行った構造化の結果をもとに、簡易的なD T Dを作成した。

9.3.2 ユーザ側からの電子化ドキュメントモデルの検討

(1) ユーザからみた電子化の狙い

製品の設計を始めようとする設計者にとっては、部品のカタログはなくてはならないものである。部品カタログから部品の仕様、機能、性能などを読み取り、これから設計しようとする製品に使えるかどうか判断する必要があるからである。

しかも、部品カタログは、設計の初期からなくてはならない。まず製品仕様を検討する段階では、部品の仕様、性能、機能を確認するために部品カタログを活用する。部品価格は、当然部品カタログに記載がないが、部品コストに直接影響があるので、メーカから価格情報を入手している。次いで実際の設計が始まると、論理シンボルを作成したり、電気解析のために電気的特性を読み取ったり、基板設計のためには部品外形寸法や、ピンピッチを読み取ったりするなどさまざまな情報を部品カタログから読み取っている。

最近では、C A Dによる設計の進展が著しく、部品カタログから読み取った情報をコンピュータへ入力して電子データとして活用するケースも多くなっている。

ユーザからみた電子化の狙いは次のとおりである。

- ・新製品情報の早期入手
- ・最新情報の入手
- ・検索の容易性
- ・データダウンロードによる情報の活用

(2) 電子化ニーズの調査

電子機器開発メーカーのエンジニア（15社21名）に部品カタログの活用について調査を実施した。設計の工程で、部品カタログ記載内容及び非記載項目でも必要と推定されるデータ項目を、紙で欲しいか、電子化して欲しいか記入をお願いした。

(3) 調査結果

① 設計のどの段階で部品のカタログのどこを参照しているのか

- ・製品仕様検討（設計の初期）

製品企画の段階から部品の調査が始まっている。参照するデータも部品カタログ全般にわたっている。価格、品質情報など部品カタログには記載のない情報も入手している。

次いで、市販部品を自社内で使うための部品選択／評価、認定を行う段階では、部品のカタログスペック以外に評価、認定に必要な情報を入手している。認定業務のための資料作成時に、部品カタログのデータをテキストで流用したいというニーズがある。

- ・設計

設計段階では、工程ごとに必要とするデータが異なっている。C A Dツールにより設計が専門化しており、必要とする情報が異なるからである。回路図入力では、シンボルの形やピンの属性情報（ピン番号、入出力など）が必要となる。電気解析では、電気的特性が必要である。特にアナログ的な解析にはS P I C Eを使っている事業所が多く、S P I C Eモデルの提供が重要となってくる。熱解析では、消費電力や熱抵抗を必要としている。最近では、C A Dを使って設計している事業所が多く、何らかの形でデータブックの電子化を望んでいる。部品外形寸法を入力したり、電気特性のV－R曲線から座標値を読み取ったりしている設計者もいる。

- ・手配／購買

部品調達段階である。この段階は発注業務主体であり、部品カタログの参照は少ない。価格や納期の交渉が主体である。自社内で部品情報システムを構築している事業所で

は、これらの情報を部品データベースへ登録して活用している。

・組立／製造

電子機器では、プリント基板の組立、筐体の製造が作業の主体である。基板組立では、自動挿入機が普及しており、データ生成のために部品の機械的情報を必要としている。部品の外形寸法や組立時のパッケージ耐熱温度などがある。

② 設計者が必要としている電子化データ

設計の各段階で部品カタログを参照していることが上記で確認できた。次に参照しているデータはどのような形式で必要なのかを分析した。つまり、データを「イメージで参照できればよいのか」、「テキストや表で流用したいのか」を次のとおり、明確にした。

- ・参照したい情報：目次／概要、機能説明、技術情報、開発情報
- ・流用したい情報：管理情報、パッケージ情報、関連情報、C A D関連、その他（資材、品質）

電子化の順序として、第一段階として機能説明、技術情報、パッケージ情報、第二段階として資材／品質関連、管理情報、C A D関連情報と続けるとよい。

③ 電子化データのフォーマット

アンケートの中で、利用者が使いたい電子化フォーマットとして上げたものは、(a)テキスト、(b)イメージデータ、(c)プリンタ出力フォーマット、(d)専用ベクトルデータ、(e)その他の五つに分類される。

利用者のレベルがさまざまであり、使っているツールもさまざまな状況では、実際の処理系では一つにしぼるのは難しい、複数を扱えるようにする必要があるだろう。

今後の活動の中で、標準化について討議する必要がある。特にC A Dツールと密接に関係したフォーマット（例えば、H D L、S P I C E、I B I S）ではC A Dベンダーを含めた標準化の促進が必要であるとともに、部品メーカーの情報提供の体制整備が不可欠である。

④ 電子データの入手方法

ユーザアンケートの分析結果では、「速報性」と「検索容易性」に力点がおかれているので、F A X送信若しくはインターネット経由での情報提供が必要となる。自社内で部品情報データベースを構築している利用者若しくはC A Dの利用を促進している利用者からは、外部の情報データベースから電子データをダウンロードして自社内で使うことを希望している。電子データの活用からみれば、インターネット経由が理想的である。いずれにしても検索の容易なツールの準備が前提である。

⑤ 電子化ドキュメントの提供方法

部品カタログを電子化して提供する場合、イメージで提供するか、テキストで提供するかなどの選択が必要である。イメージでの提供が一番容易であるが、利用者の望んでいる「流用」ができない欠点がある。一方、流用できる形での電子化は図や表の電子化形態をどうするという点が解決されていない。今すぐ始めるためにはサプライヤ側、ユーザ側とも準備不足である。今後のツールの整備状況を待つ必要がある。

部品カタログの情報の電子化は記載項目の特性によって、文字部分、図形部分、表部分に分割できる。それぞれの部分をどのように電子化するかについて検討した。

それぞれを電子化する場合は、イメージにする方法と、流用できる形でコード化する方法とがある。コード化する場合は、それぞれの特長に応じてフォーマットや形式を決める必要がある。

ユーザの要求からみれば、今後、文字データ、表データ、図データの順で電子化する必要がある。表データについては、最低限CSV (Comma Separated Variable) 形式を支援すれば可能であろう。図については、電子的に流用しようとする場合、標準フォーマットが定まっていないので、今後の標準化の検討を先に進める必要がある。

9.3.3 EDAツールに必要な情報

近年の集積回路技術の進展に伴って、エレクトロニクス製品は、ますます高機能に、より高密度になってきている。より複雑化している電子機器の設計現場では、設計作業そのものは最新機能を持つEDAツールの導入によって効率化されてきても、ツールに使用する部品情報は、データブックやカタログから必要な数値データやグラフを読みとり、多大な工数をかけて入力しているのが現状である。

そうした状況を背景に、商品カタログから読みとったデータを電子化したり、EDAライブラリに加工して専門的にユーザに提供するVAP (Value Added Information Provider) と呼ばれるサービス業種も登場しているが、同じく必要な情報の収集と加工・蓄積には多大な費用を必要とするため、現状ではVAPからの情報提供は限られた範囲に止まっている。

① EDAツールを提供する立場と利用する立場のそれぞれより、EDAツールに必要な情報の内容調査を実施した。EDAツールに必要な情報の割り出しには、EDAツールによる設計の各工程ごとに、ツールの必要としている情報を抽出する作業を行った。

② 調査分析

EDAツールが利用する情報のほとんどは、テキストあるいはイメージとして、コンピュータ上で流用できる形式が望まれている。したがって、今後は提供する情報のフォーマ

ットの標準化について検討が必要である。

E D A ツールに必要な情報の種類は、調査に参加した各社で違いはあるものの、調査事項の具体的な受けとめ方に若干の相違もあり、この点を加味すれば、ほぼ同じ内容のものを必要としていることが分かった。これらの要望には次のような共通性がある。

- (a) 直流特性、交流特性等に対して、各個別の特性値の電子化に対する要求が多かった。
- (b) 各特性値では、標準値及び最悪値（最大値若しくは最小値）が必要とされている。
- (c) 掲載されている特性グラフは、人手若しくはツールによる読み取りが行われるため、正確さが要求されている。

回路シミュレーションと熱解析においては、現状のデータブックでは情報があまり網羅されていないこともあり、必須項目のバラツキが目立った。メーカ及びE D Aベンダーにて検討の必要がある。

9.4 ドキュメント電子化ツールの調査

9.4.1 一貫した作業ツールの適合性調査

現在利用可能なドキュメントの電子化ツールとして利用されていて、データブック制作に利用されている主なD T Pシステムを調査した。

- ・ PageMaker
- ・ QuarkXpress
- ・ Interleaf
- ・ FrameMaker
- ・ M S - Word
- ・ J - Star
- ・ DynaText
- ・ Interleaf < S G M L > Tool

上記のシステムの主な機能と対応機種を調査し、各システムの作業環境ディスクトップイメージや出力サンプルを添付した。

いずれのD T Pでも印刷された段階で、せっかく用意されたC A Dデータやテキストデータなどの電子化データの交換／流用が阻まれている問題点があった。

9.4.2 文書作成の工程要素

S G M L化された文書を作成する場合は、作業工程ごとに専用のツールを組み合わせる使用することが一般的である。次の作業工程に整理し分析した。

(1) 文書記述

D T Pツールやワードプロセッサを用いて非S G M L文書を作成することである。文書のS G M L化を行う以前の文書作成工程に相当する。

(2) 構造化文書の記述

D T Dで定義されたS G M L文書の構造に則って文書を記述することである。具体的にはS G M Lのタグが付いた文書を作成する作業である。

(3) 構造化の検証

作成したS G M L文書がD T Dに則って正しくタグ付けされているかどうかを検証する作業である。この作業で使用する検証ツールは、パーサと呼ばれる。

(4) 文書のデータベース化

作成したS G M L文書をデータベースなどの情報管理システムに蓄積する作業である。

(5) レイアウト記述

S G M L文書自体は、文書の論理的構造を定義しただけであり、文書がどのように表示されるのかといったレイアウト情報についてはいっさい定義されていない。S G M L文書の各タグをどのように表示するのかを定義する作業である。

(6) S G M L文書の出力

スタイル定義に従ってS G M L文書をレイアウトし、紙や画面等の物理媒体に出力する作業である。この作業で使用するツールのうち、紙で印刷するものがフォーマット、画面に表示するものをビューア又はブラウザと呼ぶ。

9.4.3 S G M Lツールの調査

S G M Lの支援ツールはまだ広く普及しているわけではなく、各ツールがどれだけの機能を提供しているのかを知るのは容易ではない。新聞／雑誌などのメディアにおいてもS G M Lツールについて議論されることは、ほとんど皆無である。S G M Lツールを導入している企業も幾つか存在するが、製品を網羅的に検討したわけではなく、自分たちの使用しているツールの他は、どのような機能を提供しているのか分からないのが現状である。アメリカ合衆国では政府主導で公的文書のS G M L化が進められたこともあり、多くのツールが市販されているが、それらのうち、日本語の扱えるツールがどれくらいあるのかも不明である。

したがって、このような状況を改善するため、どのようなS G M Lツールがあるのかをま

ず網羅的に調査した。

9.4.4 ツール間の接続性の調査

前節にてSGMLの網羅的な調査を実施したが、SGML文書を作成するまでのすべての作業工程を、一つのツールで支援するものは存在せず、複数のSGMLツールを組み合わせで利用しなければならないことが確認された。

しかし、ツール同士を組み合わせた場合に問題なく動作するかどうかは不明である。また、カタログスペックだけに頼った情報も多いため、実際に利用した際に問題点が生じることも考えられる。

このため、SGMLの接続性について調査を行った。特にカタログスペックよりも実際にSGMLツールを使用している人の意見をなるべく取り入れるように配慮した。

9.4.5 まとめ

(1) SGMLツールの現状

- ① 日本語に対応したSGMLツールは少ない。またそのツールも試験的に使われているのが現状であり、実際に業務に導入した事例は更に少ない。
- ② 任意のDTDに対応した製品が少なく、それがツール間の接続性を阻む一因となっている。またSGML宣言まで自由に変更できる製品は皆無に等しく、SGML宣言を変更できるツールでも宣言の記述に何らかの制限がある。また、任意のDTDへの対応を唱ったツールにおいても、完全に動作を確認したDTDの種類は少ないのが現状であり、今後新たに作成するDTDが問題なく動作するかどうかは不明である。
- ③ SGMLエディタなどの入力ツールの多くは、次に入力可能なSGMLのタグをガイダンスしてくれる形態のものである。しかし、このようなツールを使うには、ある程度SGML全般の知識やタグに関する知識が必要であり、誰でも入力できるというものではない。
- ④ 今回調査したSGML文書用のデータベースは、主にバージョン管理や複数人で文書を執筆する際の排他制御管理など、文書単位での管理を指向した製品である。つまり、SGMLのタグレベルの構造をデータベースの中にマッピングして管理するものではない。一方で、SGML文書をさまざまな文書へと変換し再利用することを主目的としたデータベースは、ほとんど製品化されていないのが現状である。
- ⑤ 日本語の表示ができるフォーマットやビューアは幾つかあるものの、行頭に、“)”などが来ないようにする各種禁則処理や、縦書き表示を備えたものは皆無に等しい。

(2) 今後の課題

- ① 日本においては大量のSGML文書を扱ったシステムの事例自体が少ない。半導体分野のように多種多様な文書を統合し、大量にデータを処理する事例は、国内では皆無に等しい。これらを実現しようとした場合、市販ツールの組み合わせだけではシステムを実現するには無理がある。そのため、市販ツール間の接続を支援するツールや市販ツールのカスタマイズを行うツール、若しくは専用のシステムを開発することを考慮すべきである。
- ② 各メーカは絶えず半導体情報の文書を作成しているため、ある日突然に現在の文書作成システムからSGML文書作成システムへ移行することは不可能である。したがって、現状のDTPツール主体の文書作成システムからSGMLツールへのスムーズな移行を行うにはどうすべきかを議論する必要がある。
- ③ 現在インターネット上の資源を有効に利用するマルチメディアツールとしてWWW (World Wide Web), Mosaicなどが普及している。これらは、インターネット上の資源をハイパーテキスト的に検索／取得できるソフトウェアである。

MosaicなどWWWクライアントで表示可能なデータは、HTML (Hyper Text Markup Language) というSGMLに則った言語で記述されている。WWWの爆発的な普及に伴いHTMLに関連したツールが多く出回っており、HTMLの普及が今後のSGML普及の推進役となる可能性も大きい。

WWWクライアントはHTML専用の固定DTDしか処理できないため、任意のDTDを扱うSGMLの世界とは個別に議論がなされてきたが、近年SGMLへの橋渡しとなるツールも開発され始めている。インターネットはユーザへの情報提供を実現する上で、今後有効なインフラとなることは確実であり、これらWWW関連ツールとSGML関連ツールとの共存を図ることが重要である。

9.5 検 証

9.5.1 SGMLデータの作成から印刷まで

これまでの調査により、各種DTPツールやSGML関連ツールが存在することが分かったが、実際にこれらのツールを利用してSGMLドキュメントデータを作成し、印刷やCD-ROM出版などの可能性を検証した。

- ・テキスト／グラフィックスの素材データを用意
- ・データブックの構造を分析して、ある構造を想定したSGML宣言とDTDを作成

- ・そのDTDに基づき、タグを付け、文書インスタンスを作成し、グラフィックスデータは外部参照ファイルとして用意し、パーサ「MARK-IT」で検証
- ・そのSGMLドキュメントを利用して、次の3通りの出力実験を行った。
 - ① GRACEによるフォーマッティングと出力
 - ② RTF変換によりMS-WORDでレイアウトして出力
 - ③ LATEX変換により、DVIファイルを用意し、PS変換して出力

9.5.2 WWWによる情報提供実験

(1) HTMLの特性を考慮した記述

半導体データシートをWWWに載せるには、その内容をHTMLで記述する必要がある。これはEDSCで提唱する一般的なデータシートの記述形式から、自動的に生成されるべきものである。

しかし、HTMLないしMosaicに内包されている下記の特徴・問題点から、数ページ以上にわたる半導体のデータシート／データブック全体をHTMLに変換すべきでない。

- ① Mosaicは移殖性の観点から、フォントの出力機能をユーザのウインドウ表示機能で標準的に使用できるものに限定している。そのため、英字について文字種の制限、文字の大きさの制限、文字飾りの制限がきつく、上付き、下付きの添字のある語彙の表示、数字の表示、表組の表示などは不可能である。
- ② 日本語については更に条件がきつく、数通りの大きさの明朝体を使用できるにすぎない。
- ③ HTMLは文書をコンピュータ画面で表示することを基本に設計されているが、長文の文章について、その判読性や使い勝手は紙媒体より良いとはいえない。そのため、数ページを超える文書では、Mosaicにより画面表示で読ませること自体実用的ではない。
- ④ HTMLではハイパーテキストの思想から、複数の文書をリンク付け、渡り歩いて参照できるようになっている。しかし、そうした文書の利用法自体、人間が心地よいと思う原則にかなっているかどうか検証できておらず、無条件には採用できない。

既存の半導体データシートをHTMLで記述することは表示機能上の制約から困難でまたたとえそれが可能であっても、コンピュータ画面の使い勝手の問題から、有用性は低いのである。したがってデータシートのHTML化は、WWWの特性に合わせて紙媒体とは別に設計する必要がある。

(2) HTMLによる半導体情報記述

Mosaicの特性に合致させるためには、数10行の短い広告と同じ形式で、文書を構成する

ことが必要である。図表や画像はG I F形式の画像として貼り込んだ。

(3) Mosaicを活かした情報提供

Mosaicの特長の生きているのは、むしろ紙媒体にとらわれない情報提供形態を指向した場合である。携帯無線機の各部の半導体情報を提供するのに、まず無線機のブロック図を表示する。そしてブロック図の一部をクリックすると、対応する個別半導体のデータシートが表示されるのである。

実験では、最終的な情報表示にPostScriptを使用した例を作成した。この場合、紙の上にレイアウトされたデータシートと同じものが再現できる。ただ、次のページを表示するのに少し時間がかかるなど一覧性が悪く、また画面では解像度の点でも問題がある。画面上ではあくまでも内容のチェック程度が可能で、実際に文書を使用する局面では、おそらく紙にプリントアウトされることになるだろう。

9.5.3 電子図書館システムによる半導体情報の提供

電子図書館システムは、学術情報センターが1994年から実験的に実施している電子図書館サービスのためのドキュメント・デリバリー・システム（文書配布システム）である。電子図書館では、主に紙面情報を電子的に配布することを考えており、紙面を静止画像として蓄積、検索し、それを高速に配布するシステムである。

まず、原ファイルが必要で、大別して画像情報ファイルと、検索するための二次情報（インデックス情報）からなる。

ソフトウェアは、上記の情報管理用・検索用のデータベース管理ソフトウェア、及びそれを通信でユーザサイトに届ける電子図書館サーバソフトウェア、ユーザサイトでは、電子図書館クライアントソフトウェアを使用する。

半導体データブックを搭載して試験的にサービスを構成したところ、データシートをめくる感覚で次々にページ表示ができる。その速度はネットワークに依存するが、LANのような高速なネットワークを使用する環境で次ページ参照は約 0.5秒である。比較的に遅い広域ネットワークを通した場合でも、4秒程度である。

9.6 今後の課題

9.6.1 標準化

ドキュメントを従来の紙で表していたテキスト、図表だけでなく、CADデータ、映像、あるいは音声といった紙では表現できない情報を伝達するには、国際的に共通する標準仕様

の採用が必須であり、これら情報を電子化された「情報の集合」として捉え、これらの情報をネットワークやCD-ROMなどの新しいメディアを活用して情報交換・伝達を図らなければならない。

半導体業界におけるドキュメントは知的生産物であり、その過程において知的活動を増幅してきたといっても過言ではない。その制作過程では最も身近なコンピュータ応用にもかかわらず、その共有性や再利用性では多くの問題が内在していたことが電子化ドキュメントに対する半導体メーカ、ユーザ側アンケートで判明した。データフォーマットの互換性、蓄積情報の抽出／検索性、異機種間利用等ユーザが自由に活用できないことなどである。

短い期間ではあったが、SGMLを中心とする構造化言語を使用し、ユーザが検索容易なDTDとは何かを検証することができたと思われる。しかし、これまでの活動はまだ初期段階である。構造化言語については多くの場所で検討されているが、半導体業界DTDについては半導体関係者で検討する必要がある。

9.6.2 文書の構造化

データシートとして必要な構成要素の設定と、各社のデータシートの項目／内容を分析し、品種間並びにメーカ間のクロスリファレンスを作成した。また共通項目を抽出し、文書構造を規定し、タグセットを設計した。その際、各項目のデータフォーマットや属性についても考察した。

9.6.3 情報管理技術としてのSGML

今回の一連の活動において半導体各社の電子化ドキュメントへの取り組みはようやく走り始めたところであり、今後大きなテーマとして位置付けされることになると思われる。更にこれを利用することにより、メンテナンスが非常に容易となる。

9.6.4 使用ツールの整備

今回の調査・研究によって、半導体デバイスのドキュメントは、ユーザ側のさまざまな作業工程で使用されていることが明らかになった。

例えば、設計部門では企画設計、回路設計、PCB設計の各工程で、また品質管理部門では受入れ試験や品質管理に、更に資材・購買部門ではデバイス選定に当たって使用している。

また、入手した情報に社内の情報を付加し、電子部品情報管理システムに、更にEDAツール用のライブラリを独自に構築していることも明らかになった。もちろんデバイスメーカ側では、さまざまな部門がドキュメント作成に加わり、さまざまな部門が半導体デバイスド

キュメントのユーザとなっている。

以上のことから、電子化ドキュメントの作成／利用において、いかにハードウェアやソフトウェアといったツールの整備が重要であるかということが理解できる。

作成／蓄積された情報をユーザに利用してもらうためには、情報流通／情報配送の仕組みが安価で手軽に利用できるようなしなければならない。最近利用者が増え特に注目を集めているインターネットもこのような仕組みの一つの有力な候補である。各サプライヤが作成／蓄積した情報を各サプライヤ自らがインターネットに向かって情報発信することが比較的容易に行えるようになると考えられる。

このような環境になればユーザは各サプライヤの情報サーバを渡り歩きながら情報収集をすることができるが、利用する上において次のような課題が残っている。

- (1) 各サプライヤの電子デバイス情報を横断的に検索することができない。
- (2) ユーザは情報サーバごとの異なる情報検索のためのユーザインタフェースを使わなければならない。
- (3) ユーザは情報サーバの出現など欲しい情報に対するメタな情報に注意を払わなければならない。
- (4) もし提供情報が有償であるとするユ－ザ／サプライヤともに課金の管理をそれぞれの対応で行わなければならない。
- (5) ユーザは各情報サーバごとに自分の欲しい情報を指定しなければならない。

電子デバイスに関する情報については、各サプライヤの情報を横断的に検索できる有償サービスをインターネットの上で段階的にビジネスとして成立させる可能性がある。

9.6.5 ISO9000, PL

ISO9000の中でレビジョン管理があること、及び7月から施行予定のPL問題に対応できるようにするにはSGMLを情報管理の核に据える必要がある。

9.6.6 国際的な標準システムの拡張

半導体デバイスのユーザは、国内を始め、米国、欧州、アジアと国際的な広がりをもっている。また、半導体デバイスの開発部門も国際的に広がり、情報の作成部門もそれに従い国際的に広がっている。そのため、国内を念頭においたシステムではもはや機能しなくなっている。今後は、Pinnaclesを始めとして各国の関係機関との連携を深め、国際的な標準化と啓蒙活動を進めることが必須である。

9.6.7 関係者への教育指導

電子化技術の開発は、社会的要求であり、電子化ドキュメントの作成／利用環境は間違いなく、早晩に開発されるはずである。しかし、その一方で、それらのシステムが普及するかどうかという問題がある。つまり、電子化ドキュメントの作成や加工に従事する作業者や作成された電子化ドキュメントを利用する利用者らがそれらの環境に慣れ、使いこなせるかという問題である。

電子化ドキュメントのシステムが普及し、関係者に広く使われるためには、啓蒙、教育、指導といった課題が残っている。研修センターの設置や啓蒙活動の推進などにより、ドキュメント電子化のための専門家の育成や、利用促進を図る必要があろう。

9.6.8 ユーザインタフェース

ユーザに提供する電子化ドキュメントにおいて、そのフォーマットは多種多様であるが、ユーザのメリットを第一義に考えた提言を行う必要がある。利用者の情報利用環境がさまざまなレベルにあり、複数のフォーマットを利用可能とし、利用ツールも含めた情報提供の配慮が必要である。特に、CADツールと密接に関係したフォーマット（例えば、HDL、SPICE、IBISなど）ではCADベンダを含めた標準化の促進が必要であるとともに、部品メーカーの情報提供の体制整備が不可欠である。

10 海外ユーザ向けの国産データベース検索
支援用端末ソフトウェアの開発

カナテ株式会社

10. 海外ユーザ向けの国産データベース 検索支援用端末ソフトウェアの開発

10.1 概 要

10.1.1 目 的

日本は、バブル景気の終焉とともに、近来、稀にみる不況期を向かえている。世界的にみても景気はアメリカが若干の回復の兆しがみえ、ベトナム、香港、台湾、マレーシア、シンガポールといった東アジア諸国が好調であるほかは、冷戦構造の崩壊とともに民族や宗教に起因する地域紛争が多発しており明るい兆しは少ない。しかしながら、日本GNP（国民総生産）が第2位、国際連合への負担も第2位などからでも、日、米、欧と三極化した世界経済の一翼を担う重要な存在となっている。戦後50年を経過し、国際的な影響力の強くなった日本でも、国際的に活躍する舞台が多くなるに従って、いろいろな問題点が出てきている。例えば、莫大な対米貿易黒字に起因する日米貿易摩擦などといった形で現れたり、マスコミなどの報道をみても世界が日本を正しく理解しているとはいえないし、その反対に日本も世界を正しく理解していない部分がある。こうした誤解を生んでいる原因の一部として次のものが挙げられる。

- ・言葉の違い
- ・歴史的、文化的な土壌の違い

これらを完全に克服することは不可能であるが、可能な限りの努力は必要である。その一つの方法としてコンピュータを利用した情報の交換がある。コンピュータを利用した情報の交換を具体化したものの一つがオンラインデータベースサービスであり、成長を続けている産業である。日本でもアメリカには普及面で遅れをとっているものの法令情報、学術文献情報、記事情報、特許情報などの多くのデータベースが登場している。しかし、これらのデータベースは、日本語のキーワードでしか検索できない。これでは、1990年代ぐらいから出ている海外のデータベースユーザから日本のデータベースを検索したいという要望を満たすことはできない。

そこで、海外のデータベースユーザが日本国産のデータベースを検索したいという要望を満たすために、本課題は、端末側のソフトウェアを開発する試みであり、昨年度の課題「英日キーワード変換機能をもつデータベース検索システムの開発」の機能を拡張したものである。

10.1.2 検索サポートシステムの必要性

一般企業が外部から情報を得る方法の一つとしてオンラインデータベースの情報を利用することがある。しかし、オンラインデータベースの情報を利用しているすべての企業が直接的なユーザではなく、代行検索業者や調査会社に委託する場合もある。このような検索を依頼される業者にはサーチャと呼ばれる特殊技能をもつ技術者達がいる。その目的は、依頼者の必要な情報を一番コストのかからない方法で検索することにある。即ち、サーチャとは、最も効率的な検索をするためのノウハウを駆使する職種である。このように他社に委託してまでもデータベースの情報を得たいという事実は、情報の需要があるということだけでなく、裏を返せば、自社でデータベースを検索してもコスト的に釣り合わないということともいえる。「何故、サーチャという職業が必要なのか?」、「何故、自社でデータベースを検索するとコスト的に釣り合わないのか?」といった問題がオンラインデータベース産業の現状ではないかと思われる。

現在のオンラインデータベースの利点からいうと、

- (1) 大量のデータ資源が存在する。
- (2) データの加工の仕方によっては、有効な資料として使用できる。
- (3) 検索の仕方によっては、データが速く手に入る。
- (4) 必要なデータを検索させるために検索方法を工夫している。
- (5) データが分野別に体系的な分類がなされている。

しかし、上に示した利点は、条件付きのものが多く、利用の仕方によっては活きてくる潜在的資源であることには間違い無い。さて、ユーザがこうした有効資源を効率的に取得するためには、何をなすべきか?、障害は何なのか?、ということについて整理する。

(1) 普及状況の立ち遅れ

データベースの数、ユーザ数ともにわが国は、欧米先進国より立ち遅れている傾向にある。データベースの売上高こそアメリカにつぐものがあるが人口比や為替相場という経済状態を考えれば、伸びている傾向にはあるもののまだ普及面が立ち遅れていることは否めない。

(2) 検索手続の難しさ

一般の検索手続は、検索項目(検索キーワード、書誌事項、etc)を入力して、大きな概念の検索集合から小さな概念の検索集合へとデータを絞り込むものである。最もコストのかからなく最も良い結果を出したいという目的を実現するためには、適切なタイミングで検索項目や検索式を与えなければならない。これは、経験のないユーザが操作すると時間がかかる。時間がかかると金がかかるといった悪循環を引き起こし、ユーザのデータ

ベース離れの原因となってくる。

(3) 標準化されていない検索式

検索式の標準化は、欧米ではある程度の標準化がなされているが、わが国では、初めて、オンラインのデータベースサービスが始まったとき端末機から漢字を入力することができない時代であった。そのため検索キーワードは、半角のカタカナで入力していた。こうした背景から現在、わが国の商用データベースは、同じ意味の検索キーワードが漢字であったり半角のカタカナであったりするといった非統一の状態で共存している。また、最近では、メニュー選択形式も多くなってきたが複雑な検索を行うには、コマンド入力が必要である。

10.1.3 開発方針

端末側の開発は昨年度のシステムの評価を踏まえて次のような方針で実施した。

(1) A T 互換機端末

基本的に日本語MS-Windows V3.1 が稼働しているパーソナルコンピュータであれば稼働が可能であるが、本システムの目的である海外ユーザ向けということと導入コストの観点から、A T 互換機、いわゆる、DOS/Vマシンを端末として採用する。

(2) DOS/V, MS-Windows環境

本システムの目的の一つは、コマンド入力からの脱却であるため旧来のコマンド入力型のオペレーティングシステムとは相容れない面があるため、GUI（グラフィカルユーザインタフェース）を採用する。現在パーソナルコンピュータでGUIを実現しているオペレーティングシステムは、アップルのマッキントッシュ系のOSとMS-DOS上で稼働するMS-Windowsがある。本課題では、操作上はGUIを採用するが、扱うデータはテキストデータ（新聞記事）なので、GUIのみのオペレーティングシステムであるマッキントッシュ系よりもテキストデータの加工が容易なMS-Windowsを採用した。

(3) 検索コマンドの自動生成

検索コマンドは、MS-Windows のリストボックスによる選択や、ボタンの選択によってコマンドを生成する。

(4) 昨年度課題で作成した英日変換用のシソーラスを採用

昨年度の課題「英日キーワード変換機能をもつデータベース検索システムの開発」で使った次のような3種類のシソーラス辞書を使用する。

① アルファベット順リスト

シソーラスの用語の英訳したものを、アルファベット順に整理したもので、一つの利用語に対して、ローマ字、漢字、スコープノート（用語の使用法、注釈）、上位語、下位語、関連語、類語、同義語が掲載されている。また、英日キーワード変換にも使用する。

② 階層リスト

シソーラスの用語を英訳したもので、上位、下位の階層関係を示したリストで上位階層からの用語検索に使用する。

③ KWICリスト

用語を構成する単語別に分類しアルファベット順に並べたもので、用語がn個の単語で構成されている場合は、n回リストに登場する。主に単語から、その単語を含む用語を検索するのに使用する。

(5) 英日キーワード変換を補助するための英日単語変換辞書を内蔵

前年度課題のシステムの英日変換用シソーラスは、日経シソーラスの一部なので件数が少ない。したがって、ユーザが任意に入力する用語を日本語に訳せない場合が多い。検索対象の日経テレコンBizは、シソーラスに掲載されている用語（統制語）以外にもフリーターム（自然語）が数多く用意されているため入力した用語を一般的な日本語の名詞に訳すことができれば検索キーワードとして十分使用できる。この英日単語変換用辞書は、英日変換用シソーラスのように一対一の翻訳だけでなく一対多の翻訳ができる特長があり一つの英語の用語から複数の日本語キーワードを生成することが可能である。このことによって、検索ヒット率を高めることができる。

(6) ユーザ登録用英日単語変換辞書を内蔵

英日変換用シソーラス辞書、英日単語変換用辞書いずれにも存在しない用語を登録するための辞書で、新語のように時代とともに発生するようなものや、固有名詞を登録すれば英日単語変換辞書と同様に使用できる。

(7) キーワード入力での日本語表示

キーワードを1件入力する都度、対応する日本語を表示する。

(8) 二次検索中のキーワード検索

前年度課題のシステムでは、キーワード検索は、一次検索でしか使用できなかったが、論理演算による絞り込みや検索条件による絞り込みなどの二次検索を行った後でもキーワード検索を行い再び論理演算や検索項目で絞り込むケースもあるので、これを実現する。

(9) 関連語表示

関連語をもつキーワードを入力したときは、用語の右に「->R」を表示する。

(10) 階層検索時下位語表示

階層検索時の下位語が存在する場合はその右に「 +」を表示する。

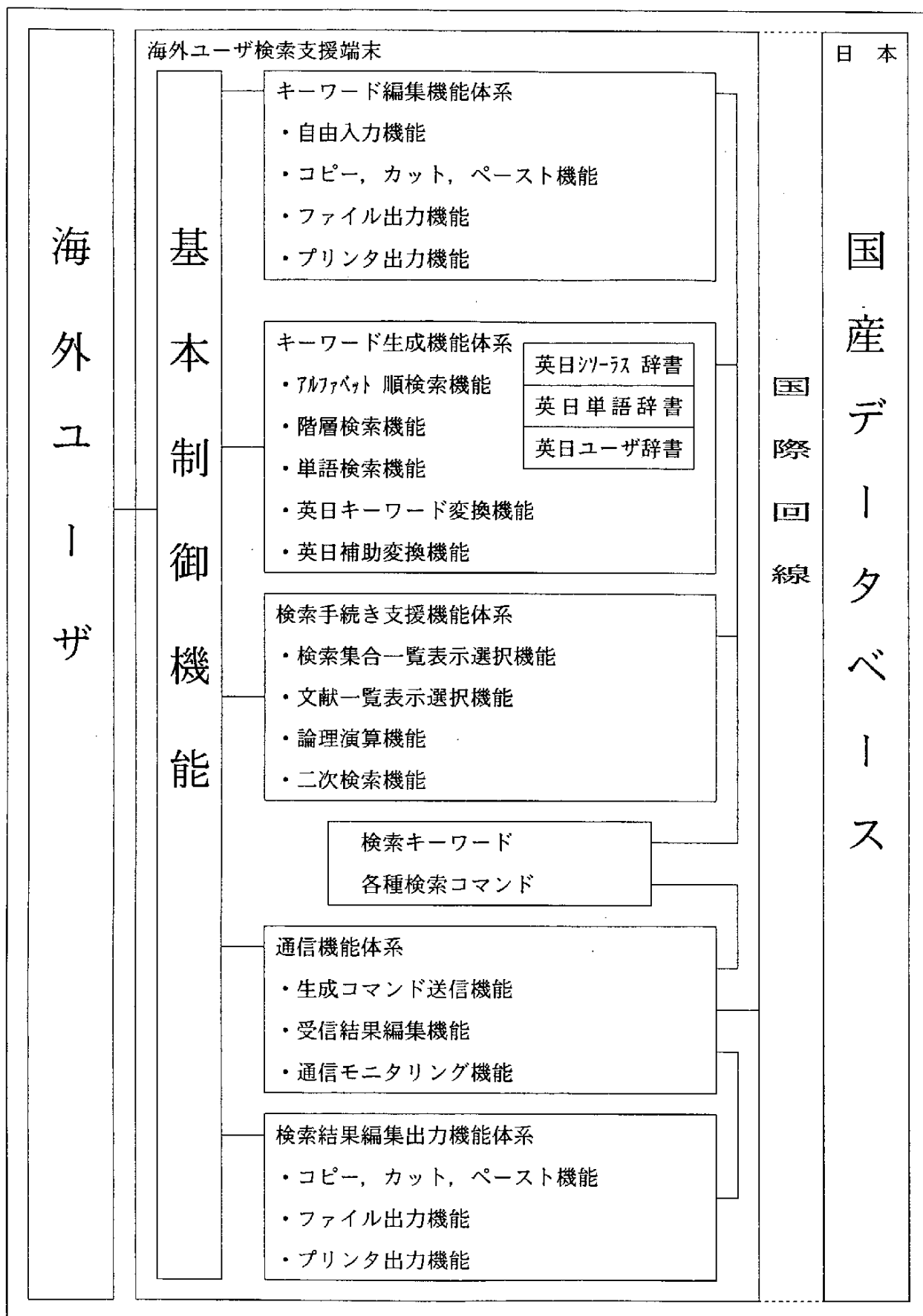
(11) 検索結果表示

昨年度課題のシステムは、リストボックスに表示していたが、この形式だと表示があまりにも固定化されすぎていてデータの出力、移動や加工編集に難点があるため、エディタ形式のウィンドウ表示をする。これによって、クリップボードを利用した別アプリケーションへのデータ転送や印刷、ファイル出力が容易になり、簡易ワープロ的な利用法ができる。

(12) 通信状況の表示

通信ソフトは、人間が通信状況を画面から受診データを見ながらコマンドや選択番号を入力する方式をとっている。本システムのように受信データを解析してWindows のメニューに表示しユーザの選択した結果で送信コマンドを生成する方式は、予測できる範囲内のことは、システムが判断しているが、予測できない通信障害が起こったときに弱い面がある。そこで、通信状況を表示することにより予測できない通信障害の原因が分かるようになり、適切な対処が可能になる。

10.1.4 システムの構成



11 連想検索写真データベースの プロトタイプ作成

シャープ株式会社

11. 連想検索写真データベースのプロトタイプ作成

11.1 研究の目的

近年、写真データベースの構築が急速に進展している。しかし、これらのデータベースは、一般に、キーワード検索あるいは全文検索技術を用いて構築されるため、写真の内容を詳細に記述する必要があり、データベース作成者の負担が大きく、十分な検索が困難だという問題がある。

本委託課題の目的は、シャープ(株)情報技術研究所において開発された「意味ベクトルによる検索技術」を用いて、通常の写真のタイトルやキャプション（説明文）程度の少ない情報をもとに、多様な検索目的に対応できる写真検索ソフトウェアのプロトタイプをパーソナルコンピュータ（PC）上に作成することである。PCには、本写真データベースのユーザになると想定されるデザイナーに広く利用されているマッキントッシュを選択した。

11.1.1 写真データベース構築の目的

写真データベース構築の目的は、次の2種類に大別できる。

(1) 大量画像の管理

フォトライブラリのように膨大な数の写真を管理し、必要に応じてストックした写真の中からユーザの要求を満たす写真を探し出すためには、写真番号、ジャンル、撮影地、写真家などで整理して蓄積しておく必要がある。

(2) 画像の検索

画像データベースの中味を知らないユーザを対象に、検索された画像の二次利用を目的にした写真データベースの利用方法がある。フォトライブラリにとっては、貯蔵写真の電子的なカタログになる。また、企画会議等における発想支援のツールとして写真データベースの利用も考えられる。

本委託課題は、膨大な数の写真が格納された写真データベースの中身を全く知らないユーザが、何等かの目的（DTP、絵葉書の作成、発想支援、図鑑としての利用等）のために、簡単に利用できる写真データベースのプロトタイプを作成することである。

11.1.2 画像検索技術の課題とキーアイデア

分類表方式やキーワード検索技術の課題を次に示す。

- ・キーワードを思い出せないと検索できない

- ・すべての人の要求を満たすようなキーワード付けや分類は困難
- ・分類表の作成困難
- ・分類表は一次元の分類のため、多角的な検索ができない

分類表方式やキーワード検索の課題を解決し、キーワードに頼ることなく人間のような連想機能をもった検索手法として、言葉に伴って想起される概念の多次元的表現（意味ベクトル）を用いた連想検索技術を開発した。基本単語 3,700語の意味ベクトルを、人手で 266種類の F E A T U R E（画像の特徴を記述する要素）との関係を付けることで作成し、この基本単語のベクトルと百科事典のテキストを用いた統計的（ブートストラップ）学習によって、百科事典に含まれる全単語の意味ベクトルを生成した。この意味ベクトル辞書を用いて、風景、ビジネスシーンなどの写真を対象に画像データベースを試作し、意味ベクトルを用いた検索を行ったところ、素人によって数語程度の説明文を付けただけでも、多様な検索ができることが分かった。これによって、例えば、キャプションに「ハワイワイキキ」という単語があるだけで、「リゾート地」や「海岸」といった「ハワイワイキキ」と連想関係の強いキーによってその写真が検索されるようになる。

11.1.3 実施内容

百科事典をもとに作成した単語の意味ベクトルを写真のキャプションに適応させることで、少ないキーワードでも連想的に写真を検索できる写真データベースのプロトタイプをマッキントッシュ上に作成した。実施内容を以下に示す。

(1) プロトタイプ詳細設計（平成6年7月～平成6年9月）

マッキントッシュ用の写真検索ソフトウェア、検索結果を視覚的に表示するための操作性の良いユーザインタフェースを設計した。

(2) プロトタイプ作成（平成6年10月～平成7年2月）

マッキントッシュ（Power MAC）上でC言語を用いて、写真検索ソフトウェアとユーザインタフェースのプロトタイプを作成した。ただし、単語の意味ベクトル辞書は、今回の委託課題の対象外とし、ワークステーション上で作成した既存のものを用いた。

(3) 評価用写真画像の選択（平成6年12月～平成7年1月）

画像検索ベンチマークを作成するためのツールをワークステーション上に開発し、画像の内容の主観的印象から14グループに分けられる 104件の写真を選択した。

(4) 連想検索写真データベースプロトタイプの検証テスト（平成7年2月）

今回作成した画像検索ベンチマークを用いて評価を行ったところ、上位3位以内に80～100%の確率で少なくとも一つの正解画像が検索されており、意味ベクトルによる検索の

有効性を確認した。

11.2 意味ベクトル技術

意味ベクトルは、多くのFEATUREとの意味的な関係をベクトル表現したものである。 n 個の概念分類をFEATUREとし、各次元が一つのFEATUREに対応した n 次元ベクトル空間上の一点で、意味を表現するものである。単語 i の意味ベクトル $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$ の各要素の値は、次のように定義される。

$x_{ij} = 0$ if 単語 i が FEATURE j と関係なし

$x_{ij} = 1$ if 単語 i が FEATURE j と関係あり

例えば、次の6個のFEATUREを選択した場合に、

[人間, 悲しい, 芸術, 科学, 興奮, 政治]

単語「パイロット」の六次元の意味ベクトル X は以下のとおりである。

$X = (1, 0, 0, 1, 1, 0)$

単語と各FEATUREとの関係を、2値で表現することで、分野に依存しない汎用的な単語の意味ベクトル辞書として蓄積できると考えられる。意味ベクトルを単語の意味表現として用いれば、ベクトル演算、例えば、各単語の意味ベクトルの和を文章の意味ベクトルとすることによって、ベクトルの向きで文脈の意味を表現することができる。

FEATUREの組み合わせによって、写真の内容を分類できるようにFEATUREを選択する必要がある。表11-1に示すように、6種類の大分類、29種類の上位概念に属する266種類のFEATUREを選択した。

百科事典中の単語すべてについて、266個のFEATUREとの関連を人手で付けるのは現実的ではない。そこで次の2種類の仮説のもとに、ブートストラップ学習と呼ぶ手法を開発した。

- (1) レコードの意味ベクトルは、一定数（例えば、5個）以上のコア単語が含まれていれば、そのコア単語の意味ベクトルの加重和によって、適切な文脈情報が表現される。
- (2) 単語の意味ベクトルは、その単語が含まれているレコードの意味ベクトルの加重和によって、その単語が使われる適切な文脈情報を獲得できる。

表11-1 FEATUREの分類

大 分 類 (上位概念数)	上 位 概 念	FEATURE数	FEATURE例
人 間 ・ 生 命 (6)	人 間	8	人間, 人名, 男性, 女性, 子供, 大人, 老人,
	生 物	6	動物, 水棲生物, 鳥類, 虫, 微生物, 植物
	生 命	7	生命, 生死, 誕生, 病気, 老い, 殺生, 性
	・	14	
人 間 環 境 (5)	人 造 物	3	道具, 機械機器, 建造物
	交通・通信	6	通信, マスメディア, 交通輸送, 自動車,
	社会制度	7	会社職業, 教育育児, 福祉年金, 施設設備,
	・	17	
自 然 環 境 (2)	地 域	14	地名, 国名, 日本, 都会, 地方, 海外,
	自 然	10	陸地, 山岳地, 天空, 海洋, 気象気候, 環境,
抽 象 概 念 (4)	精神・心理	5	感覚, 感情, 喜楽, 悲哀, 恐怖
	抽 象 概 念	8	様子様態, 変化, 関係関連, 秩序順序, 数量,
	・	36	
物 理 ・ 物 質 (8)	動 作	2	動作, 行為
	運 動	4	運動, 停止, 動的, 静的
	物理現象	7	蒸気気化, 凝固凍結, 溶解液化, 発熱発光,
	物 質	3	鉱物, 資源, 素材材料
	・	47	
文 明 ・ 知 識 (4)	人 文	20	民族人種, 知識, 言語発話, 書物書籍,
	学 術	6	数字, 物理学, 天文学, 地学, 医学薬学,
	・	36	

11.3 意味ベクトルによる画像検索

百科事典をもとに生成した10万単語以上の意味ベクトル辞書を用いて、ワークステーション上に試作した画像検索システムのファンクションについて述べる。本システムは、自然言語文による質問文を入力し、この質問文の意味ベクトルに近い特徴ベクトルをもつ画像を検索してその結果をランク付けして表示するものである。画像の特徴ベクトルには、画像に付与された

説明文から作成した意味ベクトルを使用する。

画像検索システムをブラックボックスとして、質問文に対して検索された画像とその説明文だけから判断した、実現されたファンクションを次に示す。

- (1) 単語の意味の類似性に基づく検索
- (2) 連想的検索
- (3) 百科事典の知識に基づく検索
- (4) 感性検索

このように、対象とする画像に関して専門的な知識をもたない素人が付与した簡単な説明を、画像の特徴ベクトルとしても、百科事典をもとに生成した意味ベクトルによる検索によって、妥当な画像が検索できることを確認した。

11.4 写真検索ソフトウェアとユーザインタフェース

図11-1にシステム全体のフローを示す。システムはスタート後に単語辞書の読み込み等の初期化処理を行った後、メイン画面とメニューを表示する。検索を行う場合には検索対象となるデータベースを指定し、検索方法を指定した後、質問文を入力し検索を行う。検索結果は上位9画面ずつ出力され、画像での表示と説明文での表示との切り替えができる。データベースを登録する場合には後述する登録ウインドウ上でテキストファイルとイメージファイルを指定して行う。

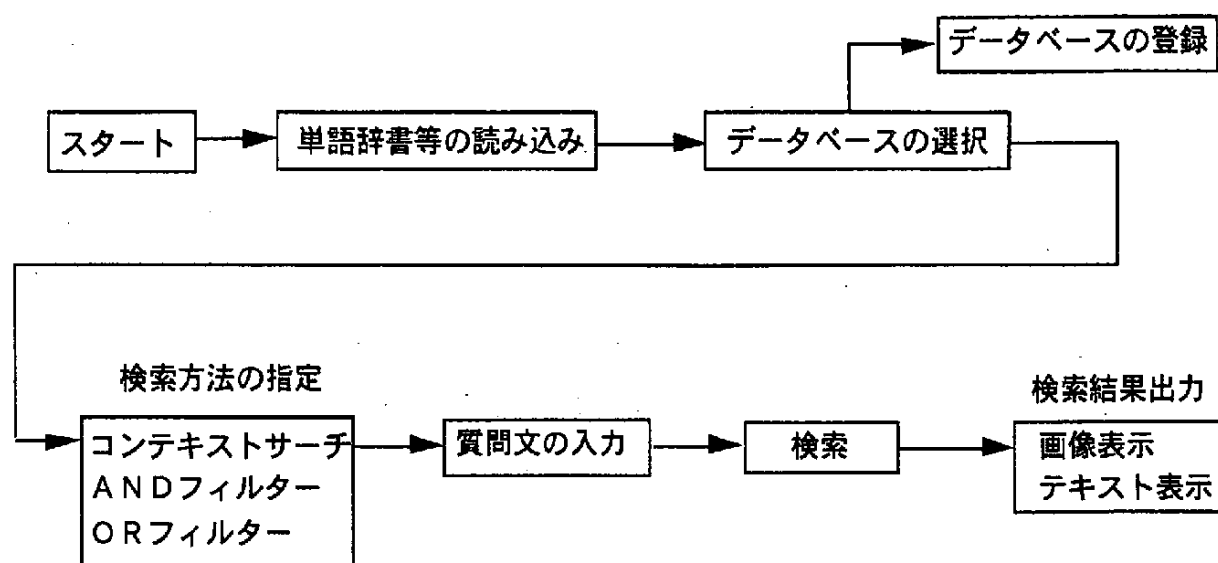


図11-1 写真検索システムのフロー

図11-2はシステムのメイン画面で、左画面上でラジオボタンによる検索方法の指定と質問文の入力を行い、続いて検索ボタンを押すことで検索が開始される。右画面には検索結果として9画像が出力される。出力はラジオボタンにより画像出力と画像の説明文の出力との切替が可能である。したがってこのシステムは説明文の出力に限定すればテキストデータの検索にもそのまま使用できるものとなっている。また検索された画像からのフィードバック検索も可能となっている。各ボタンの役割は表11-2のとおりである。更に検索された画像をクリックすることでフルサイズの画像を表示することも可能である。

表11-2 写真検索システムのメイン画面上のボタン

CONTEXT SEARCH	ベクトルのみによる検索の指定。
AND FILTER	質問文から抽出された単語の論理積を満たすレコードを対象にベクトル検索。
OR FILTER	質問文から抽出された単語の論理和を満たすレコードを対象にベクトル検索。
検 索	検索の開始。
IMAGE	検索結果を画像で表示する。
TEXT	検索結果をテキストで表示する。得点とイメージファイル名も同時に表示する。
FEEDBACK	フィードバック検索の指定。

このシステムは上で述べたように写真検索に限らず百科事典などのテキスト検索にもそのまま利用できる。また出力方法さえ変更すれば音声、動画等も含むマルチメディアデータの検索にも拡張できる。また利用者にとって使いやすいようにユーザインタフェースに工夫を凝らした。今後の実使用において更に操作性が増すようにユーザインタフェースも改良されるべきである。本システムが実際の写真検索の場で生かされ、またユーザインタフェース部分も含めて、今後の動画等も含むマルチメディアデータ検索装置に応用されるものと期待する。

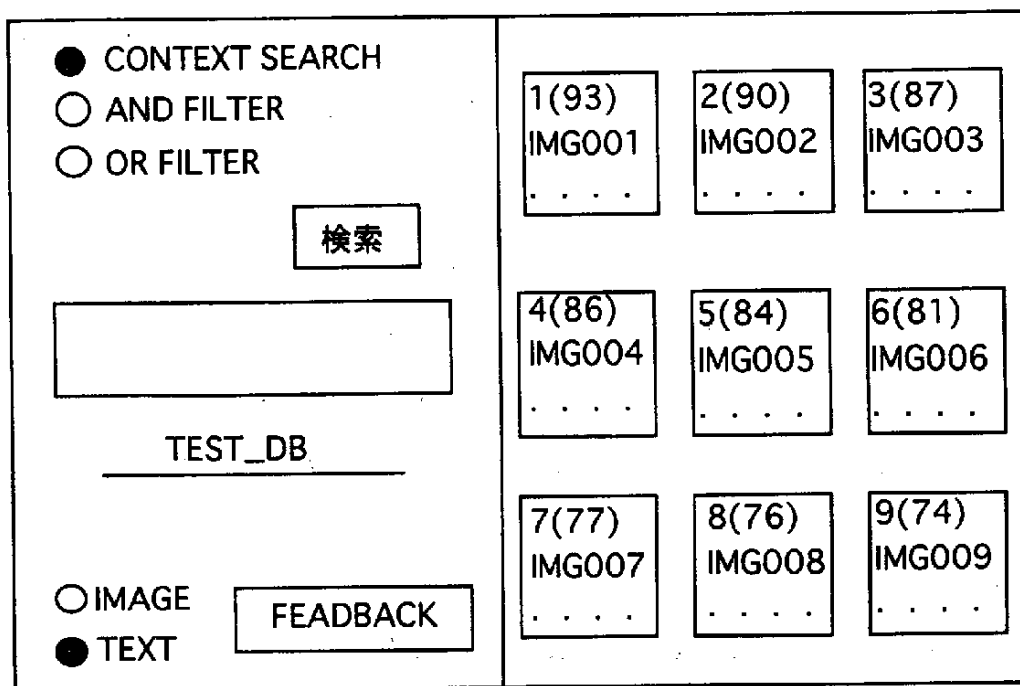
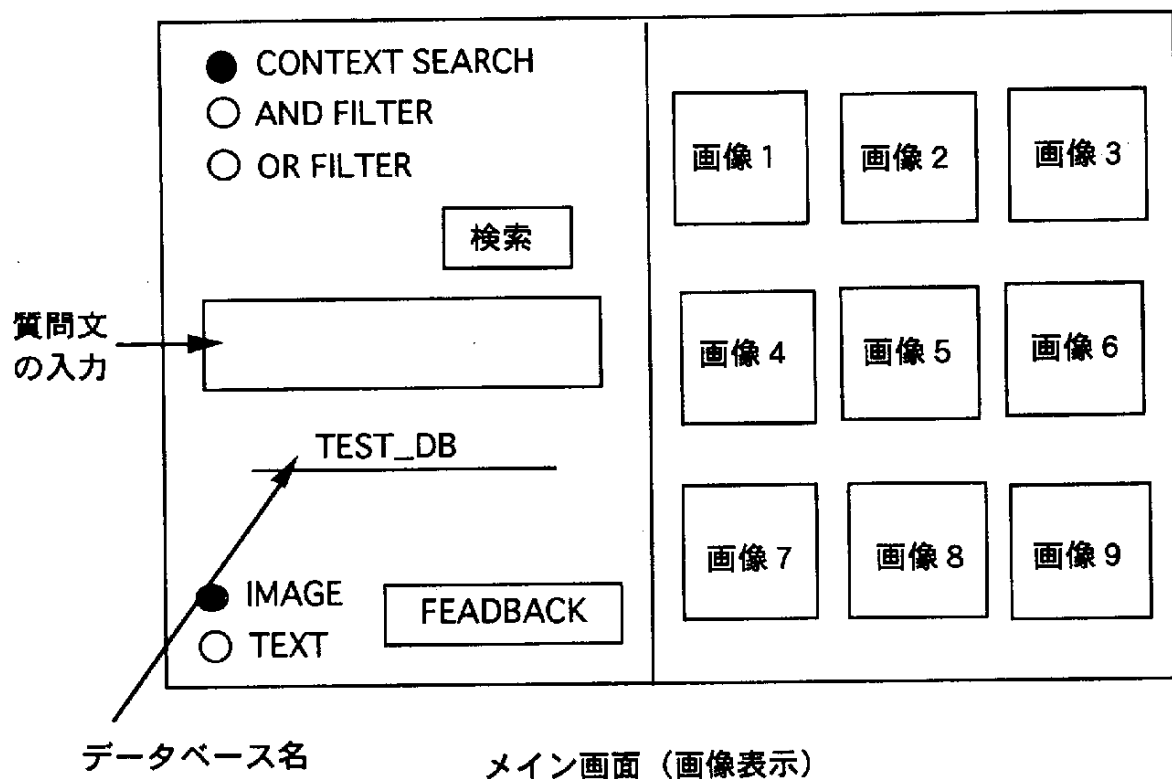


図11-2 写真検索システムメイン画面

11.5 画像検索評価用ベンチマークの作成

画像検索システムの評価方法の概略は次のとおりである。まず画像データから一定数抜き出したものを評価用のサンプルデータとし、サンプルデータに対して複数の質問文と正解のペアーを予め人手によって作成する。サンプルデータを対象に同じ質問文をシステムに入力し、得られた検索結果と人手で作成された正解データとの一致度で評価を行う。ここで予め正解データを作成しておくのは、検索の度に出力データに対し正誤判断を行った場合には、人間の判断が検索の時々で一定せず、したがって評価が一定しない、あるいは評価の自動化ができないなどの理由による。

画像のもつ意味ベクトルは画像に付加された説明文から機械的に作成されたものであるため、画像の主観的特徴を正しく表しているとは限らない。したがって画像検索システムを評価する場合、画像から得られる主観的類似度に基づいた評価が不可欠であると考ええる。そこでサンプルデータ及び質問文と正解のペアーの作成を説明文を考慮に入れず画像のみを見て行った。

システムが出力する検索結果は画像を質問文に近いものから順に得点を付けて出力するものであるため、従来の検索で用いられた含有率、適合率による評価法を直接適応することはできない。また従来の検索とは異なり曖昧性を許すものであり、また検索された画像をもとにしたフィードバック検索も可能なため、検索結果の上位に少しでも正解画像があれば良いとした評価を用いる。

以上で述べてきた評価方法は、画像検索において利用者が大まかな印象を入力した場合に妥当な画像が検索されるのかを評価する場合を想定したものである。評価結果から上位3位以内に80～100%の確率で少なくとも一つの正解画像が検索されており、意味ベクトルによる検索の有効性が確認された。しかし実際の検索では例えば絵画の検索を行う場合に作者名を明示的に指定する場合もある。このような場合も含めた評価方法を確立すべきである。またこの検索方式は入力キーワードを設定せず任意の言葉で検索可能なものであるため、本来は利用者が入力すると想定されるすべての質問文に対しての評価が必要であろう。したがって今後百科事典、広辞苑等の電子化辞書を用いた大規模な評価が必要であると考ええる。本章で提案したものは主に画像から得る主観的な印象に基づいて検索を行う場合の1評価方法である。

11.6 今後の課題

写真データベースの利用者は、デザイナーが多く、連想的、感性的に写真を検索したいとの要求が強い。現在、実用化されている写真データベースは、単純なキーワード検索方式が主流

であり、デザイナーの要求を満たしていない。本プロトタイプのように、デザイナーが利用するPC（マッキントッシュ）上で連想的に写真を検索できるツールが提供されれば、わが国の写真データベースの普及に大いに寄与すると考えられる。

今後の課題を以下に記す。

- (1) フォトライブラリ業者と協同で、膨大な数の写真を格納した写真データベースを開発し、デザイナーや出版社に提供し、ユーザの意見を集めて検索技術やユーザインタフェースに反映させる。
- (2) 本プロトタイプシステムを利用して写真データベースを構築するためには、ワークステーションでアプリケーション用の意味ベクトル辞書を作成する必要がある。汎用的な意味ベクトル辞書を対象テキストに適応させる学習機能をマッキントッシュ上のツールとして作成する必要がある。
- (3) 意味ベクトル技術は、言語への依存性が少なく、日本語以外の言語への変換も可能である。次のステップとして、英語への対応を行う。

今後このような研究を実施することによって、写真データベースによる地域紹介、旅行案内、美術鑑賞、図鑑としての利用、家庭での写真の整理など、各種の目的に応じた写真データベースの構築が促進されることが考えられる。

12 レーザー研究データベースの構築

財団法人 応用光学研究所

12. レーザー研究データベースの構築

12.1 概 要

本事業ではわが国におけるレーザー及びレーザー応用技術の研究開発に関するデータベースとして、①レーザー研究開発データベース（データ項目：タイトル，要旨，発表者，所属機関，対象レーザー種別，対象応用分野等10項目），②レーザー研究者データベース（データ項目：研究者名，所属機関，学・職歴，レーザー研究の経験，主要実績等22項目），③レーザー研究機関データベース（データ項目：研究部署名，レーザー研究内容，レーザー研究の実績等20項目）の三つのデータベースを作成した。これらはレーザー研究という一つの特定の分野に焦点を当て，研究内容，研究者，研究機関を総合的に把握できるようにしたもので，次のような特色をもっている。

- ・レーザー研究開発の研究発表を中心として，わが国の30種を越える学会誌，学会予稿集，講演集，講演課題集，研究課題集，技術雑誌等の公表メディアからレーザー研究開発に関する広範なデータ情報を収集している（1994年1年間を対象とし，約3,000件）。
- ・研究者，研究機関に関しては，公表を前提としたアンケート調査によってそれぞれの研究実績や研究の方向性を含めたより具体的なデータ情報を収録している。
- ・データベースシステムはパソコンをベースにした自立型のシステムを採用しているため，研究者等で手軽に利用でき，データの更新や改訂も容易で，レーザーという特定領域に焦点を絞ったきめ細かな検索利用ができる。
- ・使用ソフトウェアは従来型のカード型簡易ソフト（Let's アイリス），各種データベースソフトで読み込めるCSV型ファイル，及びメニュー方式により操作性を高めたACCESSの3種類で多様な利用者に対応できる。また，ACCESSでは三つのデータベースがあたかも一つのシステムで取り扱われているかのような高い操作性が実現できている。
- ・検索方法は分類コード（レーザー種別，応用分野）とワイルドカードの両方が可能であり，利用がし易くなっている。

このようなレーザー研究開発分野に特化した総合的なデータベースの利用の事例を委員会メンバーを中心に検討していただいているが，このデータベースに関しては以下のような多くの利用が想定され，レーザー及びレーザー応用技術情報の発信の中心的メディアとなることが期待される。

① 利用想定者

- ・レーザー研究機関

(国公立研究機関, 産業界, 電機, 電子, 重工業, 機械, 精密, その他)

- ・レーザー及びレーザー機器製造産業

(電機, 電子, 機械, 精密, 重工業, その他)

- ・レーザー機器利用産業

(電力, ガス, 重工業, 鉄鋼, 化学, 石油, その他)

- ・行政機関

(産業政策, 研究開発関連)

② 用 途

- ・研究情報の検索, 研究情報の入手, 研究動向の分析

- ・研究者研究機関情報の利用, 研究の共同化, 研究委託

- ・研究開発政策, 技術政策, 産業政策の立案等

- ・更に研究者や研究機関等の自己PR

また, 今後の発展方向として, 海外向けデータベースやレーザー機器技術, 産業・企業等の新たなデータベースの基礎となるものである。

12.1.1 目 的

レーザー技術は既存のあらゆる科学技術分野で使われており, 産業全体を支える基盤的技術として位置付けられている。更にこのレーザーは「光の世紀」ともいわれる21世紀における最重要技術の一つとして, わが国内外で積極的な研究開発が進められている。レーザー技術は, 「秒進分歩」と呼ばれるほど進歩が速く, また基盤的, 横断的技術であるため, その研究開発に当たっては部門間の相互交流や総合化が重要な課題となっている。

このような来世紀の中心となるレーザー及びその応用技術の開発を促進し, 日本ひいては世界におけるレーザー産業を育成するためには, レーザーに関する研究情報を的確に集積し, 関係者に広く利用され得るデータベースの整備が強く望まれる。しかし, レーザーに関する技術の進歩の速さ, また, 情報の収集や評価に専門的知識を必要とすること等から, わが国においてはまだレーザー技術に関するデータベースの整備は極めて不十分な状況にあるといわざるを得ない。

本データベース構築は, わが国において実施されているレーザー研究開発, 研究者, 研究機関等, わが国におけるレーザー研究の状況を的確に把握できるようなデータベースの構築を行い, 国内外の利用者に提供することを目指したものである。

12.1.2 実施体制及び実施内容

本事業は(財)応用光学研究所レーザー技術研究会のプロジェクトチームによって行われたが、その実施に当たっては難波進大阪大学名誉教授を委員長とし、大学、研究機関及び民間企業のレーザー関連の専門家20名の委員からなる「レーザー研究データベース構築委員会」を設置し、5回の委員会を開催し事業内容のチェック、各種データソースの提供、レーザー研究者及び研究機関アンケート票の検討、作成したデータベースの評価とデータベース情報をもとにした分析例の提供を受けた。

本データベースの作成において実施した主要な作業は次の5点である。

① 研究開発データの情報収集

わが国におけるレーザー及びレーザー応用技術の研究開発に関する情報を学会誌、学会予稿集、講演集、研究課題集、技術雑誌等の公表メディアから抽出した。対象とした公表メディアは35種で、抽出データ数は約3,000件である(表12-1)。

② アンケートによる研究者及び研究機関データの収集

公表メディアからだけでは、研究者や研究機関のバックグラウンドや現在の従事業務内容、関心分野等についての具体的な情報を得ることは難しい。また、仮に種々のデータソースからの情報の組み合わせでそれが得られたとしても、研究者や研究機関にとっては本位でないデータとなる可能性もあり、プライバシーの問題が生ずる。このため本事業では、研究者と研究機関に対しては公表を前提としたアンケート調査を行い、それぞれの属性からレーザー研究開発の対象分野(レーザー種別、応用分野)、経歴、主要実績等の情報の提供を受けた。アンケートの実施は今回は初めてのケースであったが、研究者で約400、研究機関で約300の回答を得た。

③ データベースシステムの検討

本データベースシステムはレーザー研究開発という特定分野に焦点を置きながら、研究開発情報、研究者、研究機関という、総合的なデータベースを作ろうとしたものである。この作成に当たっての基本的な考え方は、研究者を始めとするレーザー技術開発に携わる人々(ユーザ)が簡便に目的に応じたきめ細かな検索等が可能となるデータベースを提供することにある。この簡便さと目的に応じたきめ細かな検索という機能的要求は、急速な発展と普及をみせた現在のパソコンで十分に満たせるものである。

本データベースでの使用ソフトウェアとしては市販のデータベースソフトウェアをそのまま使うということでLet's アイリスを、完全なメニュー化により操作を簡易化するものとしてACCESSを、また、ユーザが既に保有している種々のデータベースソフトウェアに読み込める形のものとしてCSV型ファイルを用いることとした。

表12-1 データソースとした公表メディアと抽出データ数

公 表 メ デ ィ ア	期 間	抽出件数
1. 応用物理	94年1月～12月	35
2. 応用物理学関係連合講演会	94年3月, 9月	1,230
3. 原子力学会年会	94年3月	19
4. 高温学会誌	94年1号～6号	5
5. 光 学	94年1月～12月	47
6. 工業技術院試験研究所研究計画	93年10月	13
7. 高分子	94年3月～10月	14
8. 高分子論文集	94年2月～12月	7
9. 精密工学会誌	94年2号～12号	28
10. 精密工学会大会	94年3月, 10月	52
11. 鉄鋼協会講演論文集	94年秋	2
12. 鉄と鋼	94年1号～7号	3
13. 電気学会全国大会	94年3月	73
14. 電子情報通信学会	94年春・秋	645
15. 電子情報通信学会誌	94年1月～12月	2
16. 電子情報通信学会論文誌	94年1月～12月	42
17. 日本ME学会雑誌バイオメディカルエンジニアリング	94年2号～10号	9
18. 日本化学会合同大会	94年春・秋	135
19. 日本化学会誌	94年1月～12月	9
20. 日本機械学会誌	94年1月～12月	3
21. 日本機械学会全国大会	94年8月	26
22. 日本機械学会論文集	94年1月～11月	7
23. 日本金属学会講演概要	94年春・秋	6
24. 日本金属学会誌	94年7号～8号	2
25. 日本建築学会大会学術講演梗概集	94年9月	2
26. 日本原子力学会誌	94年1月～12月	2
27. 日本分光学会	94年春・秋	49
28. 日本レーザー医学会誌	94年3月～12月	28
29. 分光研究	94年1号～6号	2
30. 文部省科学研究費補助金採択課題	94年10月	120
31. レーザー学会	94年1月	218
32. レーザー研究	94年1月～12月	80
33. 溶接学会誌	94年2号～6号	8
34. 溶接学会全国大会講演概要	94年4月, 10月	72
35. 溶接学会論文集	94年2号～4号	5
合 計		3,000

(ただし、トピックス等を含む)

④ データ入力作業

データの入力データベースとして最も入力作業やデータチェックに馴染んだLet's アイリスで行い、そのデータを一括してCSV型ファイル及びACCESSへ変換する形を採った。

⑤ テストラン及びテスト分析

入力したデータをもとに、レーザー種別及び応用分野別のレーザー研究の動向、レーザー研究者、研究機関についての試験的分析作業を行った。

12.2 実 施 結 果

12.2.1 作成データベース

本データベース構築事業において最終的に作成したデータベースは次の3種類である。

① レーザー研究開発データベース

データソース：1994年発表35種のメディア

データ数：3,000

データ項目：①整理番号、②タイトル、③要旨、④対象レーザー、⑤応用分野、
⑥報告種別、⑦研究者、⑧研究機関、⑨機関の所在地、⑩出典

② レーザー研究者データ

データソース：アンケート調査

データ数：380

データ項目：①整理番号、②研究者名、③ふりがな、④住所、⑤所属機関名、⑥所属
部署、⑦職名、⑧機関連絡先、⑨TEL、⑩FAX、⑪学歴、⑫卒業年、
⑬職歴、⑭学会、⑮従事業務（技術）、⑯従事業務（管理）、⑰従事期
間、⑱対象レーザー種別（従来）、⑲対象レーザー種別（今後）、⑳対
象分野（従来）、㉑対象分野（今後）、㉒主要研究

③ レーザー研究機関データ

データソース：アンケート調査

データ数：254

データ項目：①整理番号、②機関名、③ふりがな、④部署名、⑤所在地、⑥部署責任
者、⑦TEL、⑧FAX、⑨連絡先、⑩TEL、⑪FAX、⑫従事業務
（技術）、⑬従事業務（管理）、⑭レーザー関連スタッフ、⑮内訳、⑯
対象レーザー種別（従来）、⑰対象レーザー種別（今後）、⑱対象分野

(従来)，⑱対象分野（今後），㉔主要研究

各種データベースに共通するレーザー種別を表12-2に、また、レーザー応用分野を表12-3に示す。（ただし、研究者及び研究機関のアンケート回答によるものは、レーザー応用分野に関しては簡易版を使用している。）

表12-2 対象レーザー種別分類表

000	レーザー一般
100	気体レーザー（一般）
110	希ガス／希ガスイオンレーザー
120	金属／金属イオンレーザー
130	CO ₂ ／COレーザー
140	化学レーザー（HF，DF，ヨウ素等）
150	エキシマレーザー
190	その他の気体レーザー
200	液体レーザー（一般）
210	色素レーザー
290	その他の液体レーザー
300	固体レーザー（一般）
310	ガラスレーザー
320	Nd：YAGレーザー
330	YAG以外のNdレーザー
340	Ndを除く希土類レーザー
350	波長可変固体レーザー
	（アレキサンドライト，Ti： サファイア等）
390	その他の固体レーザー
400	半導体レーザー（一般）
410	Ⅲ－Ⅴ族半導体レーザー
420	Ⅱ－Ⅵ族半導体レーザー
490	その他の半導体レーザー
500	自由電子レーザー
600	真空紫外，X線レーザー
700	不明

表12-3 対象応用分野分類表

0 0 0 基礎科学研究（一般）	4 2 0 センサー
0 1 0 レーザー基礎過程	4 3 0 ディスプレイ
0 2 0 共振器，モード特性	4 4 0 メモリ，記録
0 3 0 材料・光物性	4 5 0 通信
0 4 0 分光	4 9 0 その他の計測，情報
0 5 0 光伝播，伝送	5 0 0 医用・バイオ利用（一般）
0 6 0 原子核，核融合	5 1 0 外科
0 9 0 その他の基礎科学研究	5 1 1 レーザーメス
1 0 0 機械工作加工（一般）	5 1 2 ファイバー利用
1 1 0 切断，穴あけ	5 1 3 血管手術
1 2 0 溶接	5 2 0 はり及び低出力レーザー治療
1 3 0 表面処理	5 3 0 眼科
1 3 1 熱処理	5 3 1 コアギュレーション
1 3 2 組成変更	5 3 2 角膜治療
1 3 3 アブレーション（PVDを含む）	5 4 0 皮膚科
1 4 0 マーキング（手法は問わない）	5 5 0 医用計測・診断
1 9 0 その他の機械工作	5 6 0 細胞処理
2 0 0 電子工業応用（一般）	5 7 0 DNA処理
2 1 0 CVD	5 9 0 その他の医用，バイオ利用
2 2 0 PVD	6 0 0 特殊用途分野（一般）
2 3 0 エッチング	6 1 0 宇宙利用（一般）
2 4 0 ドーピング	6 1 1 ロケット推進
2 5 0 アニールリング	6 1 2 エネルギー伝送
2 6 0 リソグラフィ	6 1 3 通信
2 7 0 材料創生	6 1 9 その他の宇宙利用
2 9 0 その他の電子工業応用	6 2 0 軍事利用（一般）
3 0 0 化学応用（一般）	6 2 1 兵器
3 1 0 元素分離	6 2 2 レーザー・レーダー
3 2 0 同位体分離（除U，Pu濃縮）	6 2 3 ジャイロ
3 3 0 U，Pu濃縮	6 2 9 その他の軍事利用
3 4 0 化学反応制御	6 3 0 環境利用（一般）
（化学反応型表面処理を含む）	6 3 1 センサー
3 5 0 化学分析利用	6 3 2 レーザー・レーダー
3 9 0 その他の化学応用	6 3 3 リモートセンシング
4 0 0 計測，情報（一般）	6 3 9 その他の環境利用
4 1 0 計測機器	9 0 0 その他

これらのデータベースを使用するためのソフトウェアシステムとしては、現在のパソコン環境、利用者の便宜を考え以下の3タイプからの選択方式とした。

① Let's アイリス

比較的初期の段階から供給されているソフトウェアであること、MS-DOSベースの古いマシンでも動くこと、日本において保有シェアが圧倒的な、NEC9800シリーズに対応していることから、初期の段階からパソコンを利用してきたユーザ、古いマシンを保有するユーザなどに妥当なものとして選定した。

② ACCESS

最近のウィンドウズ環境の普及を考慮した場合ウィンドウズ対応ソフトウェアが望まれること、完全にメニュー化することが可能でありACCESSそのものの操作手順を覚える必要がないことなどから、ウィンドウズマシンユーザや、操作の容易性を望むユーザに妥当なものとして選定した。

③ CSVファイル

①、②の方式ですべての潜在ユーザのパソコン環境に適應することは困難である。一方、表計算ソフトウェアやデータベースソフトウェアの多くは、CSVテキストファイルを通じて異なるソフトウェアで作成されたデータを読むことができるようになっている。既に他のデータベースソフトウェアを利用しているなどで、上記①、②以外の方式での供給を望むユーザに対してはこのCSVテキストファイルが妥当と考える。上記①、②は、ソフトウェアを含めて供給することを想定しているが（もちろんLet's アイリス、ACCESSを保有するユーザにはデータのみを供給）、他のデータベースソフトウェアを保有していたり、慣れ親しんだユーザは、CSVファイルから自らのデータベースソフトウェアで読み利用することが可能である。

12.2.2 データベースの利用

本データベースは多くのデータの中からレーザー研究を実施する場合に参考とする情報を検索利用することが主たる目的であるが、このような特定分野に関して、ある程度のまとまった情報が整備されている場合には、単なる情報検索以上の多様な利用が可能となるが、その最も重要なものは現在のレーザーの研究開発の中心はどこにあるのか、今後どのような研究が重要となるのか、といったようなレーザー研究開発に係わる動向の分析である。このため、本データベース構築ではこれらの分析事例として、以下のものを行った。

① レーザー種別の研究動向

希ガス／希ガスイオンレーザー、金属／金属イオンレーザー

CO₂ / COレーザー

化学レーザー (HF, DF, ヨウ素等)

エキシマレーザー

色素レーザー

Nd : YAGレーザー

アイセーフ固定レーザー (Er, Tu, Ho等)

波長可変固体レーザー (アレキサンドライト, Ti : サファイア等)

III-V族, II-VI族半導体レーザー

自由電子レーザー

真空紫外, X線レーザー

② 用途別分野

機械工作加工

電子工業応用

化学応用

計測

医用・バイオ分野

また、アンケート調査で得た情報をもとにレーザー研究者及びレーザー研究機関の動向の整理分析を行った。

12.3 今後の課題

本レーザーデータベースの構築は概ね所期の目的を達成したが、今後の課題として次のものが挙げられる。

① 継続的なデータ収集体制の充実

公表メディアからの研究情報のチェック体制を充実させるとともに、研究者、研究機関情報についての定期的なアンケート等によるサーベイ調査を行うことによって、本データベースはより発展するものと考えられる。

② 本データベースの発展方向の検討

・海外向けデータベース

現在のデータベースを英文化し、海外に向けて日本のレーザー技術の研究データを発信する

・レーザー機器技術データベース

各レーザー及びレーザー利用機器メーカーの生産、販売している機器、技術をデータベース化し、ユーザの機器の選択や、研究の方向性などの基礎資料を提供する。

- ・レーザー企業データベース

研究機関データベースと同様の形で作成が可能で、レーザー関連のユーザとサプライヤの関係にとって有効なものと考えられる。

13 砂漠化防止技術データベース

株式会社 新産業創造センター

13. 砂漠化防止技術データベース

13.1 概 要

「United Nations Convention to Combat Desertification in Those Countries Experiencing Serious Drought and/or Desertification, Particularly in Africa」(深刻な干ばつ又は砂漠化に直面している国(特にアフリカの国)における砂漠化の防止のための国際連合条約:略称:国連砂漠化防止条約)が1994年6月政府間交渉会議で採択され、日本政府は1994年10月に署名した。今後、国際社会は乾燥地等の砂漠化や深刻な干ばつの与える悪影響に関し、十分な認識と具体的な対応を早急に実施していかなければならない。特に、1977年の国際連合の砂漠化防止会議において採択された砂漠化防止行動計画以来の取り組みや経験を反映し、新規にまた一層効果的で接続可能な開発を目指して進められようとしている。

わが国に対しては、保有する高度な産業技術や人的な貢献が求められていくものと考えられ、独自技術の開発や人材育成の拡充・強化が肝要と考えられる。既に、わが国の砂漠化防止関連の取り組みは、学術調査や開発国援助という形で種々の事例が有りながら、過去の成果や現在の取り組みに関する一元的な情報の集積がなく、関与する自然科学系、人文科学系の広範な分野間での連携が少ないのが実状である。

また、わが国には砂漠が存在せず、また海外の砂漠そのものの地域性が大きいことから、砂漠化防止に係わる技術・機器等の適用に関する情報は貴重である。しかしこれらの情報は広範な分野に横断的で、かつ海外に蓄積された情報が多く、日本人研究者、技術者にとっては情報収集や利用で効率的な作業が行い難い状況である。

このような背景から、本データベースは砂漠化防止技術に特化し、調査研究・計画立案業務の効率化と研究者・技術者の情報交換の活発化を図ることを目的に構築を行った。

13.1.1 目 的

近年、地球環境問題の一つである砂漠化防止に対する国際的な関心が高まっており、今後わが国に対しても従来以上にこの分野における人的、技術的貢献が求められていくと思われる。砂漠化防止は、農学、工学を始めさまざまな分野に関係するため、情報源が分散しており利用者にとって情報収集しづらい状況になっている。

本データベースは、砂漠化防止技術関連情報に特化することにより、利用者の使い勝手を向上させると同時に、砂漠化防止に係わる研究者、技術者の活動情報を収録することにより、利用者間の情報交流を活発化させることを目的としている。

更に、将来的には海外に対しても、わが国の砂漠化防止に関する活動状況を知る有益な情報源となるべく整備拡充を目指す。

13.1.2 期待される効果

(1) 砂漠化防止関連プロジェクトの調査研究・計画立案業務の効率化

砂漠化防止は農学、工学を始め幅広い分野に関係しており、かつその情報源も各所に分散して把握しづらい。本データベースを使用することにより砂漠化防止関連の専門家が必要とする国内・海外の主要な情報源へのアクセス方法を簡単に知ることができ、必要な情報収集を網羅的、効率的に行うことが可能になる。

(2) 砂漠化防止に係わる研究者・技術者間の情報交流の活発化

本データベースは「人の活動情報」という形で主要な専門家の調査研究歴を収録し、プロジェクト情報に関する効率的な情報収集が可能である。砂漠化防止のように関係する専門家の幅が広い分野にあっては、本データベースは関係者間の情報交流を活発化させる有効なツールになりうる。

13.1.3 実施内容

本データベースは以下の三つのサブデータベースから構成される。

(1) 既存データベース所在情報

砂漠化防止関連の専門家にとって有用と思われる国内・海外のデータベースに関する所在情報を収録する。

・主に以下の種類のデータベースを対象とする。

- ① 開発援助関係のデータベース
- ② 海外の農学・工学系のデータベース
- ③ 気象関係データベース
- ④ 国内試験研究機関などのインハウスデータベース
- ⑤ 国内の商業データベース

・主に以下の項目を収録する。

- ① データベース名
- ② 所在地
- ③ 連絡先
- ④ 公開の程度
- ⑤ アクセス方法

(2) 既存文献所在情報

砂漠化防止関連の専門家にとって有用と思われる国内・海外の資料等文献に関する所在情報を収録する。

・主に以下の種類の文献を対象とする。

- ① 海外機関の技術レポート類
- ② 学会論文報告
- ③ 講演会・シンポジウム資料
- ④ 調査報告書（各省庁等）
- ⑤ 試験研究機関の研究報告

・主に以下の項目を収録する。

- ① 文献名
- ② 発行機関名
- ③ 連絡先
- ④ 公開の程度
- ⑤ 入手方法

(3) 人の活動情報

主要な砂漠化防止関連の専門家の活動情報を収録する。対象となる専門家は日本砂漠学会、日本砂丘学会、日本緑化工学会等関連学会の会員を中心に、専門委員会の意見を取り入れながら選定する。

・主に以下の項目を収録する。

- ① 所属機関
- ② 著書、論文
- ③ 調査研究活動歴

13.1.4 実施項目

(1) 入力情報調査・収集

・入力情報

- ① 既存データベース所在情報
- ② 既存文献所在情報
- ③ 人の活動情報

・作業内容

- ① データ収集フォーマット作成

- ② 資料調査, データベース検索
- ③ 特定の機関, 専門家に対するヒアリング調査
- ④ アンケート調査等

(2) システム開発

パソコン通信網から利用可能な, EWS上で稼働する検索システムの開発

- ① システム基本設計: 画面等入出力設計, データベース設計を行う。
- ② プログラム作成: システム詳細設計～プログラミング～テストを行う。
- ③ データ整理・投入: 収集された情報を整理分類し, システムにデータ投入する。
- ④ 運用テスト: 本番環境での運用テストを行う。

(3) 専門委員会の開催

調査対象, 調査項目等の妥当性を判断するため専門委員によるレビュー。

本データベースの構築を進めるに当たって, 学識経験者, 専門家からなる専門委員会を設置し, 調査対象, 項目の妥当性やシステムへの提案など専門的な検討を行った。

13.2 実施結果

実施項目に従って入力情報調査・収集, システム開発, 専門委員会の開催を行った。

13.2.1 入力情報調査・収集

(1) データ収集フォーマット作成

①既存データベース所在情報, ②既存文献所在情報, ③人の活動情報, ④機関情報について, 4種のデータフォーマットを作成した。

(2) 資料調査, ヒアリング調査

海外情報(主にアメリカ)及び国内情報に区別して, それぞれ調査を実施した。

① 海外情報の収集概要

・機関情報

「砂漠」「砂漠化」に関連する研究に積極的に取り組んでいて, 研究者との連絡が可能なものを抽出した。また主な研究分野別一覧表を整理した。

・文献情報

文献情報として文献の所在と入手方法を整理した。文献のコピーサービスもインターネット等で利用可能な範囲で整理した。

- ・データベース情報

データベースリストの作成を行った。

- ・個人情報

30名前後のリーダー的な研究者をリストアップし、著作物、活動を収録した。

② 国内情報の収集概要

- ・データベース情報

公的な機関の情報電子化の状況、外部への情報提供の可否、利用条件を整理した。

- ・文献情報

砂漠化防止に関連する事業の整理を行い、報告書等の公開性、利用条件を整理した。

(2) アンケート調査

人の活動情報について、国内の人物を対象にアンケート調査を実施した。

13.2.2 システム開発

(1) 基本要件

① パソコン通信により、利用者自らが検索できること。

現在パソコン通信は急速に普及し、オンライン検索手段として一般的なものになりつつある。提供側も情報提供が自動的に行え、人的負担が軽減される。

② 利用者電話料金負担の軽減として、商用VANのノードが利用できること。

商用VAN (FENICS) を利用することにより、全国約 180ヵ所のアクセスノードを経由し、本データベースにアクセス可能となる。

(2) システム構成

既存のハードウェア、RDBシステムを活用。

① ハードウェア：S-4/2GX メモリ32MB ディスク2GB

② ソフトウェア

アプリケーション：検索ソフトウェア（略称：ANDeT）

RDB : INFORMIX-ONLINE V4.1

OS : SunOS 4.1.2

*新規開発部分：アプリケーションソフトウェア

13.2.3 専門委員会の開催

本データベース構築に関し、専門委員会を開催し専門的な立場から砂漠化防止関連情報の範疇及びデータベースシステムや運用に関して検討を行った。

(1) 第一回専門委員会 平成6年9月29日

・議題及び審議内容

- ① 平成5年度実施のアンケート結果及び調査結果に関して
- ② 平成7年度以降の取り組みに関して
- ③ 砂漠化防止技術データベースの扱う情報に関して
- ④ 調査対象及び項目に関して
- ⑤ 基本システムに関して

(2) 第二回専門委員会 平成7年2月16日

・議題及び審議内容

- ① 事業の進捗状況に関して
- ② 調査結果概要に関して
- ③ システムの設計、プログラム作成に関して
- ④ 運用計画に関して

13.3 今後の課題

調査結果及び今後のデータ取り扱いに関して、以下のように課題を整理した。

(1) 調査結果に関して

平成5年度実施の情報ニーズ及び情報シーズの調査アンケートの回答者との情報が途切れないように連絡を行う必要と情報ニーズの再確認を行うことが課題である。また、平成6年度実施の調査結果に関しては昨年度情報の再検討と合わせ、いかに使い易いデータベースを構築していくか、利用者の側にたった情報提供形態を検討することが課題である。

(2) 今後の取り組み

情報の商品化を前提とした収集データの評価、段階分け、価格帯の設定を行い重要情報の蓄積を進めることが課題である。また、情報源としての「人」を重視し、情報の新鮮さを保つことが課題となる。

(3) システムに関して

文字列のみの情報提供形態に止まらずマルチメディア対応を検討し、より質の高い情報提供とインターネット接続の検討が課題である。

14 メディア変換型感性データベースの 構築方法の研究

財団法人 イメージ情報科学研究所

14. メディア変換型感性データベースの構築方法の研究

14.1 ネットワーク・マルチメディア・メディア変換

14.1.1 インターネットとマルチメディアデータベース

最近、マルチメディア検索アプリケーションMosaicの登場により、インターネットブームがまき起こり、わが国においても昨年あたりから、一般の家庭レベルでインターネットに接続できるサービスが普及の兆しを見せている。

このように一般の多くの人々が、ネットワークを通じて容易に大規模なデータベースサービスが受けられる状況が急速に整いつつある今日においては、よりヒューマンフレンドリーな観点に基づくデータベースシステムを構築する必要性が高まっている。

14.1.2 マルチメディアとメディア変換

本研究では、このようなマルチメディアネットワーク社会に対応し、また、人間のマルチモーダル知覚を積極的に活用すべく、感性をベースに置き、メディア変換技術の導入により画像と音響を統合的に取り扱う、データベースの構築方法について検討する。

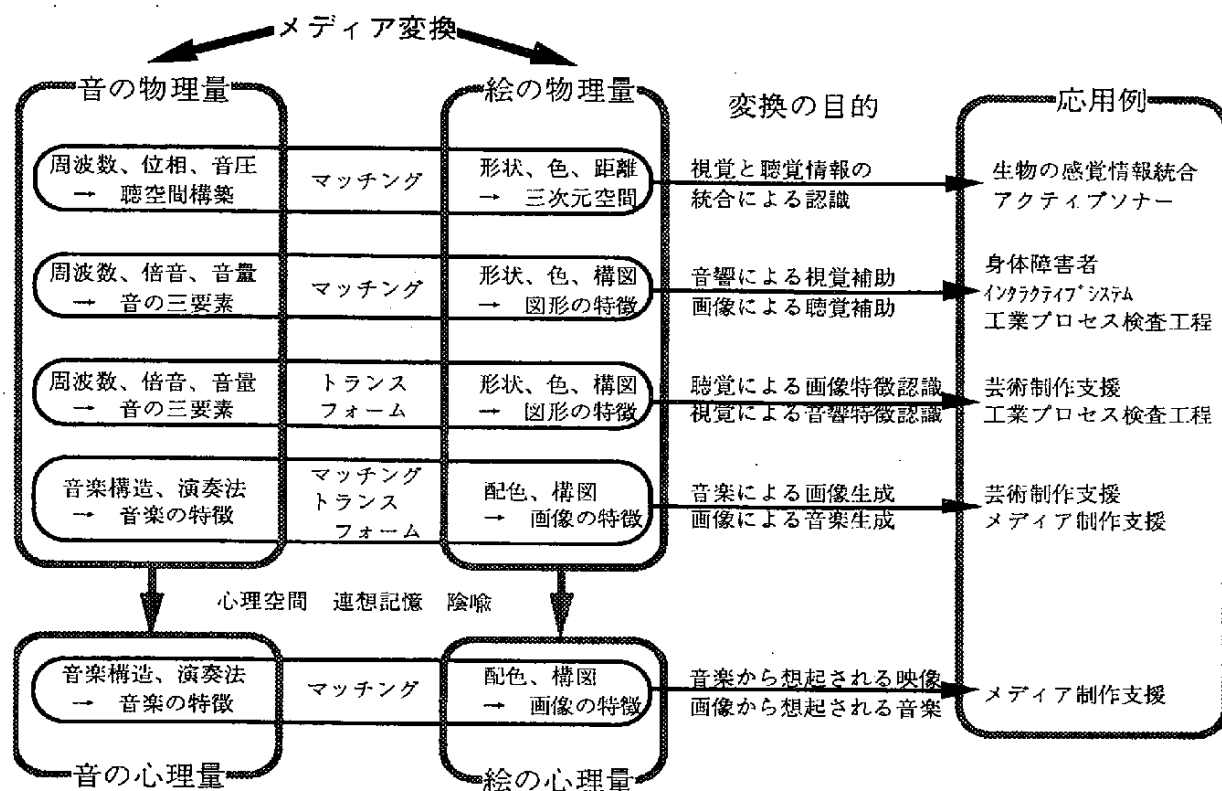


図14-1 メディア変換の概念

14.2 産業・研究分野でのニーズと従来の関連研究

14.2.1 印象語による相互変換

印象語を用いて、画像・形容詞間と音・形容詞間の変換をある一定の規則に基づいて行うことができれば、次のような応用が考えられる。例えば、無音で絵画を見ているときの印象と曲を聞きながら見ているときの印象の違いを、関数で表現することが可能であれば、その関数を逆に用いることによって、音に応じて絵画の方を変化させることができる。すると、音楽のリズムや音色によって絵画が変化し、難聴者に「音楽を見せる」ことができる。

また、デザインや映像・音楽作品の制作などの分野において、素材の検索に大いに活用でき、作業の高度化・効率化を支援できるものと思われる。

14.2.2 知覚情報による相互変換

画像からの色彩情報や四角形・三角形といったオブジェクト情報を、個人の感性若しくはニューラルネットで学習させることによって、音のリズムやメロディへ対応付けることができれば、抽象的な音の生成が可能となる。

また、人間は音によって見えないものを想像することができる。この過程をコンピュータで表現できれば、監視カメラの設置が非常に困難若しくは不可能であるような場所の状況を、あたかも直接監視しているかのように把握することが可能となるため、さまざまな業界でのニーズが期待できる。

14.2.3 信号による相互変換

映像における周期性を利用することにより、視覚的に理解しづらい微細情報を、音信号として表現することができる。目視検査のように常に監視が必要とされる場面においては、この技術を用いて音による情報を与えることにより、見落としや見間違いを防止したり、他の作業を平行して行えるようになる。

また、音の特性を図形化して目で確認することができれば、例えば超音波探傷試験の結果を三次元画像表示することで、解釈を容易にすることができる。

14.2.4 画像と音の統合

画像と音を同時に示すことにより、臨場感が増し理解が深まる。工業応用の例では、騒音の中でも話声を聞き分ける視聴覚センサフュージョンシステムがある。また、テレビや映画の世界では、臨場感を増すためだけでなく、例えば、ドラマのシチュエーションの切り替え

時に、映像よりもBGMの切り替えを先行させることにより、不自然さが出ないようにする、などの工夫に用いられている。

14.2.5 具体的なニーズ事例・研究事例の紹介

いろいろな大学や企業での、メディア変換若しくはそれに類するニーズや研究の内容を、幾つか取り上げて具体的に紹介する。（ここでは、紙面の都合上省略）

14.3 メディア変換の要素技術の現状

14.3.1 画像から音響への直接変換

二値図形は、 $P(\theta)$ 曲線表現法やフーリエ記述子表現法などの手法により、時系列信号化することができる。しかしこれだけでは、ただ単に図形を一次元信号即ち音にしたいだけなので、次の画像全体の複雑さに注目して音の基本的なパラメータを決定する。例えば、日常において、われわれが一枚の画像を見るときにまず注目するのは、その画像の中にある特徴的な要素である。普通は、大きなもの・特別な形をしたもの・色の鮮やかなものなどの刺激の著しい内容を先に認識し、それに伴い位置などの情報を認識する。そして、画像中の図形の種類や数が多いほど、全体を認識するのに要する時間は長くなる。よって、このような人の画像認識特性を、音の基本的なパラメータに反映させるものとする。

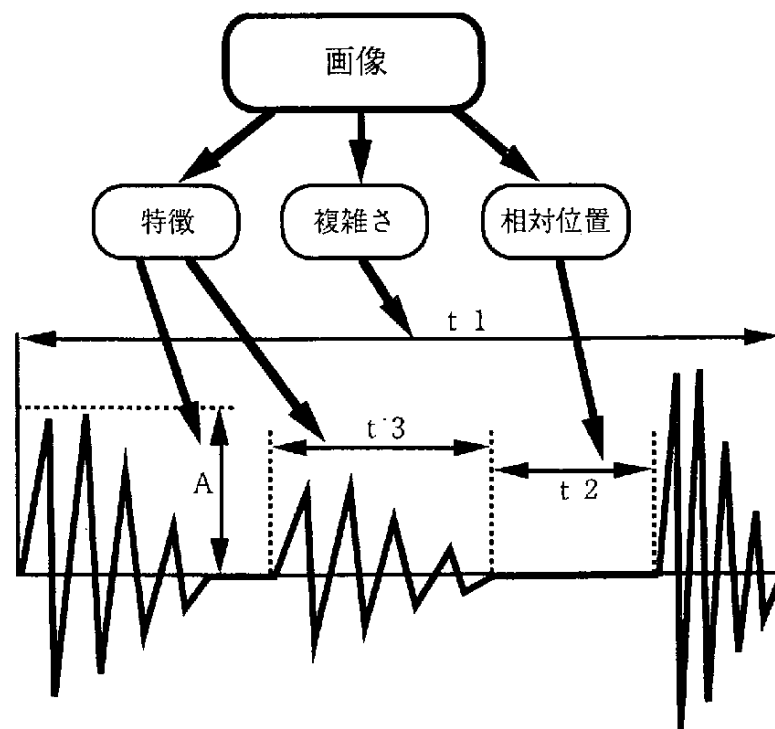


図14-2 画像から音響への直接変換

一方、濃淡画像については、リングオペレータを用いて直接音に変換することができる。その場合、生成された音は画像の変化により振幅が変化する。即ち、画像の内容が変化しないところでは音の振幅が0になり、変化が大きいところでは振幅も大きくなる。周波数についても、リング上のデータの変化に応じてパターンが変化する。その結果、生成された音には濃淡値の変化部分が反映され、音によって動的な印象が与えられる。

14.3.2 物理モデルからの映像と音の同時生成

このシステムは、中間表現（具体的には物理モデル）を操作することにより、画像と音を同時に生成・変化させることを目的としている。したがって、中間表現である物理モデルをコアとしてシステムが構築されており、すべての入力はず、まず最初に物理モデルに入り、そしてその中で計算された結果が画像生成部と音生成部に送られて、最終的に画像と音が出力される、という経路をたどる。ここで入力は、システム内に設定された仮想空想の状態を変化させるものであり、このとき各状態に割り当てられる値を環境パラメータと呼ぶことにする。入力の形態としては、主にマウスやキーボードからの入力と考えられるが、他にもマイクやMIDIあるいはライン入力も可能とする。

さて、処理全体の流れの中で重要なことは、同一のパラメータによって画像と音が同時に変化するという点である。したがって、物理モデルから画像と音がそれぞれ生成される過程においては、二つの処理が並列に行われなければならない。

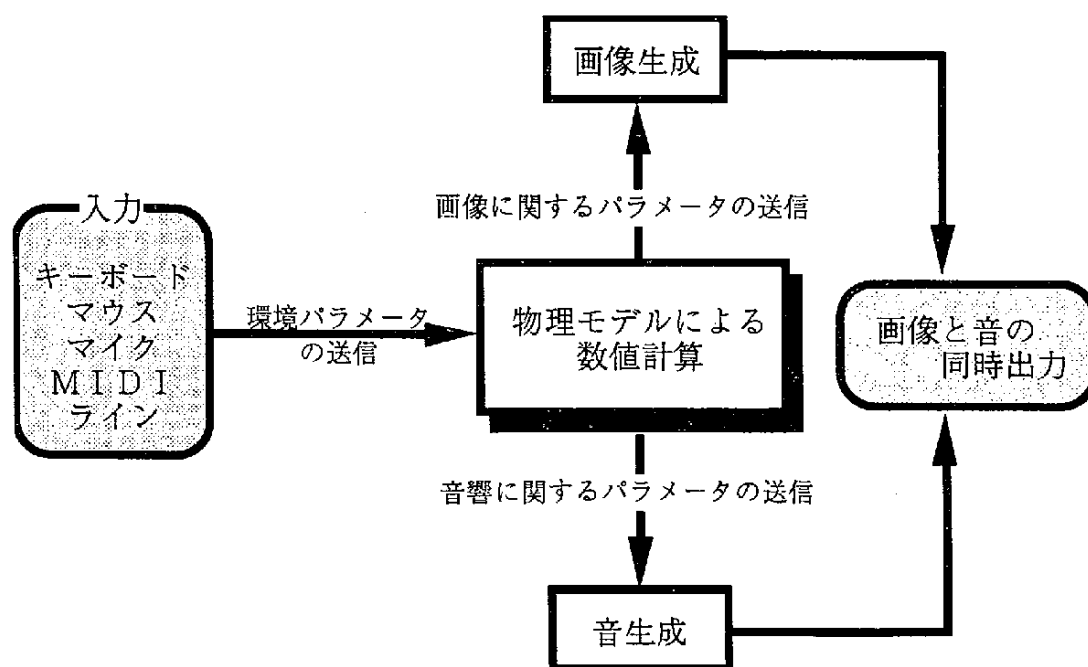


図14-3 画像と音の同時生成処理

画像の生成については、入力として与えられる環境パラメータが、すべて画像内の物体の運動に関与するものであるから、それに基づいて運動方程式を解けば、対象物体の運動の様子を容易に求めることができる。

一方、音の変化については、運動方程式のような確立した関係式が存在しないため、環境パラメータだけで音を生成するのは非常に困難である。そこで本システムでは、基本となるサンプリング音をデータベース上に用意し、それを環境パラメータのさまざまな変化に応じて変形させて出力する。

14.4 メディア変換型感性データベースのプロトタイプの概要

14.4.1 イメージに基づくイラスト画の検索システム

本システムでは、特定作者の作品に限定した作者心理をキーとしたイラスト画のイメージの解析及び検索を行う。データベース化されたイラスト画からは、絵画の重要な表現手段となる色とシンボルに関する特徴量が計算される。その特徴量から推測される作者心理と実際の作者心理との一致度から、作者心理を推定するのに有効な特徴量を見出し、効率的なイメージ検索を行う。

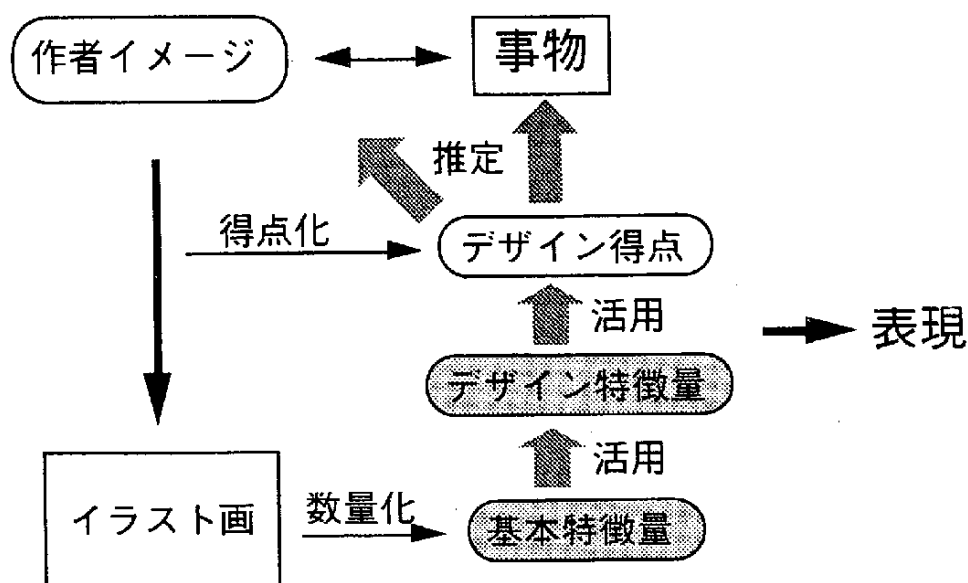


図14-4 イメージの解析

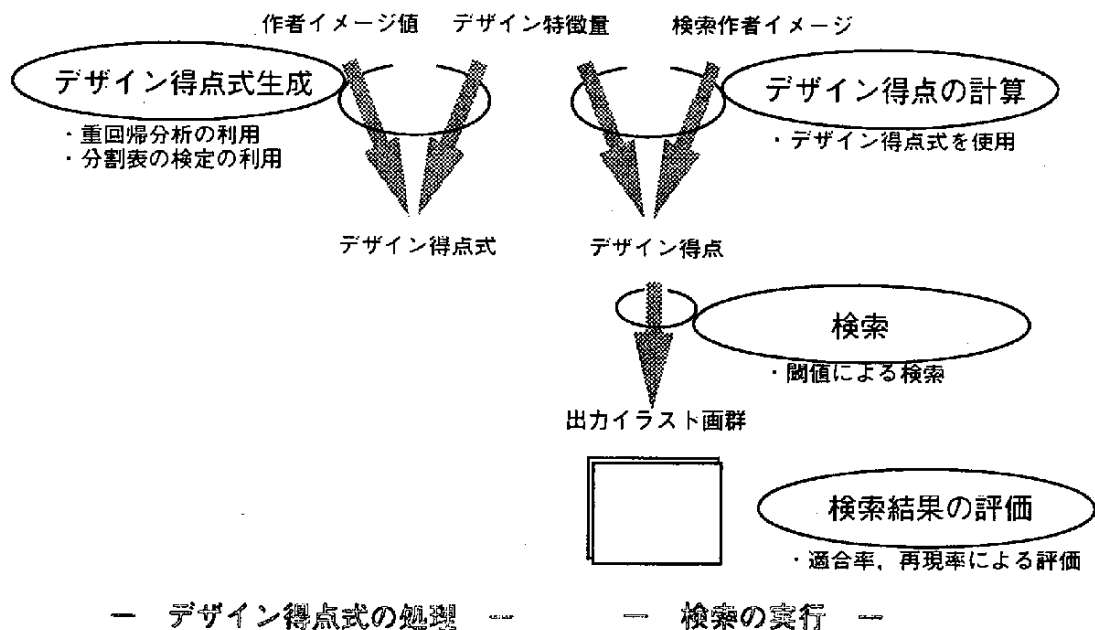


図14-5 検索処理の流れ

14.4.2 感性情報に基づくイラスト画の自動生成システム

本システムでは、感性情報に基づくイラスト画の自動合成を試みる。これは、ユーザから感性情報として形容詞が入力され、この入力された形容詞に対応するさまざまなイラスト画がコンピュータ内で自動的に生成される、というものである。具体的には、画像データベースに集積された特定作者のイラスト画群から、感性情報とキーカラー、キャラクタの関係を画像処理手法により抽出し、コンピュータでイラスト画を描くための技法について、描画の技法を幾つか実現した。また、複数の画像に対し縮小・拡大・移動を協調的に行うことで、構成要素間の相対関係を自動的にアレンジすることを考えた。

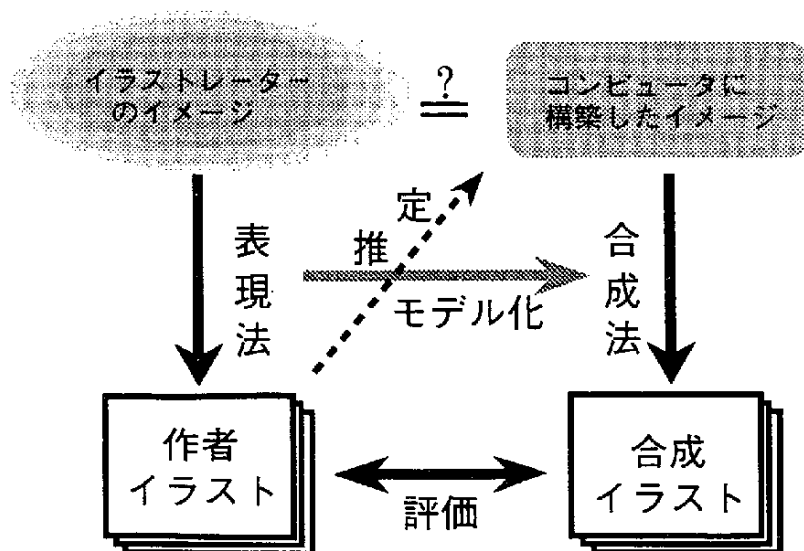


図14-6 システムの概念

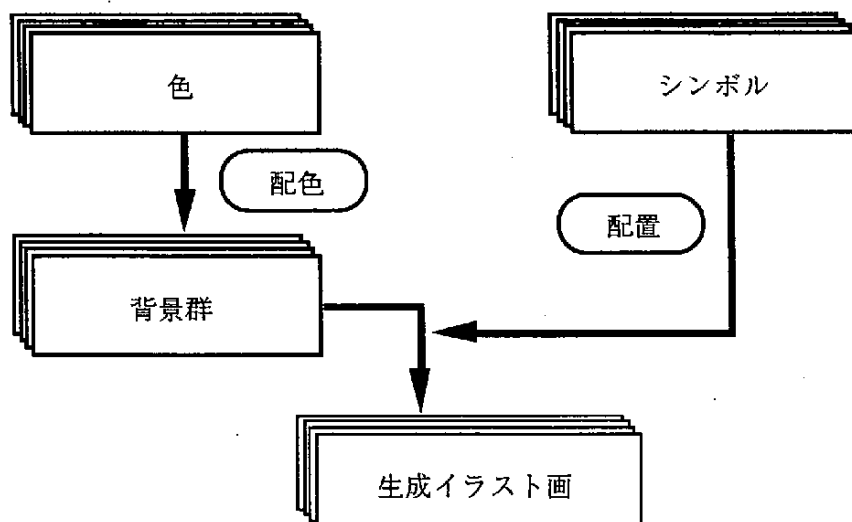


図14-7 処理の流れ

14.4.3 音楽からの感性情報の抽出

本システムでは、ある曲のノート情報を与えて、その拍子・調性・コード進行・メロディなど曲を構成する要素を認識させ、それを形容詞とリンクさせて感想文を出力させようというものである。

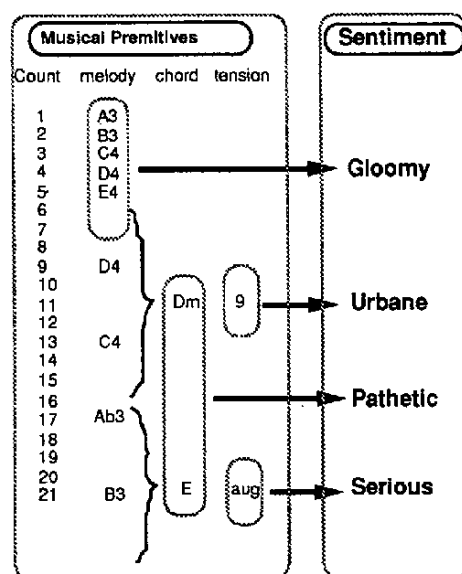


図14-8 感性情報抽出のためのリンクの例

図14-9には、村松健作曲「封じ込められた光と影」という曲に対し、システムが生成した感想文を示す。ここでは、曲全体に散らばったセンチメントのうち、多いものから順にピックアップすることにより、曲を支配するセンチメントを調べ、文を作るテンプレートに形容詞を埋め込む方法を用いて感想文を作成している。

Whole atmosphere of this music is
 slow and leisurely.
 sorrowful.
 smart and serious. Level is 18.
 urbane. Level is 7.
 rural. Level is 6.

There is pathetic mood on chord from 5 to 33.
 There is pathetic mood on chord from 37 to 65.
 There is hopeful mood on chord from 69 to 97.
 There is melancholy mood on chord from 105 to 116.
 There is pathetic mood on chord from 133 to 153.

On release at 75, there is hopeful, urbane, smart, serious mood.

図14-9 生成された感想文

14.4.4 感性パラメータを利用した作編曲支援システム

ここでは、感性情報を用いてより音楽性豊かな曲を作成できるシステムの構築を目指し、まずは次のような入出力を考えた。

Input : ・モチーフになるコード進行
 ・感性パラメータ（形容詞に代表される感性語）

Output : ・入力された感性パラメータに応じた曲

そして、これをもとに検討した結果、図14-10に示すようなブロックダイアグラムをもった、感性パラメータを利用した作編曲支援システムを実現した。

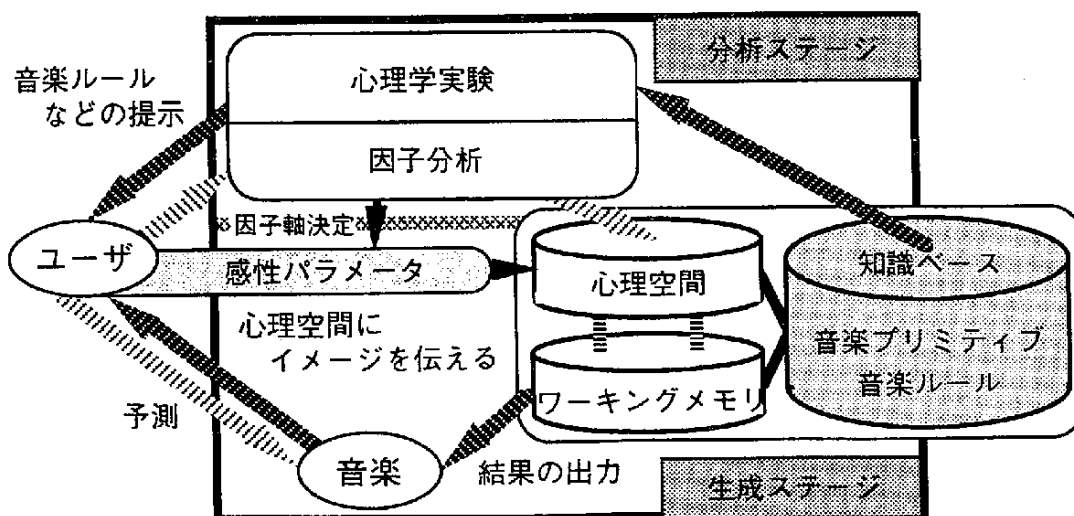


図14-10 感性パラメータを利用した作編曲支援システム

分析ステージは、イメージを具体的なコード進行や作編曲作業におけるルールの組み合わせ集合に対応付けるための、心理空間（Mental Space）を構築することを目的としている。具体的には、感性語と具体的事象の相関性について心理学実験を行い、この結果を因子分析することによって、感性パラメータによって形成される心理空間を得るというものである。ここで、感性パラメータとは、一般に、感性量と物理量を結ぶ要素のことである。本システムでは、音楽を構成するコード進行などの要素を心理的イメージで置き換えるための尺度を指す。この感性パラメータを軸として張られる多次元の空間が心理空間である。

生成ステージは、感性パラメータによって与えられたイメージからコード進行を生成することを目的としている。具体的には次の2通りの生成法がある。

- (1) ユーザが与えた基本的コード進行（4小節相当）のケーデンスを解析し、ルールによって8小節のコード進行を複数生成する。そして、これらのカデンツ（終止形）の形式を分析して心理空間に配置する。
- (2) ユーザが入力した感性パラメータから、システムが予め内部にもっている基本的なコード進行が選択される。

14.5 将来に向けての課題と展望

14.5.1 システム提案

前節では、メディア変換型感性データベースを構成するための以下の四つの要素を取り上げた。

- ・画像から感性情報を抽出する。
- ・感性情報から画像を生成する。
- ・音楽から感性情報を抽出する。
- ・感性情報から音楽を生成する。

これらを同一の形容詞空間上で結合したものが、本研究で提唱するところのメディア変換型感性データベースであり、その概念を図14-11に示した。

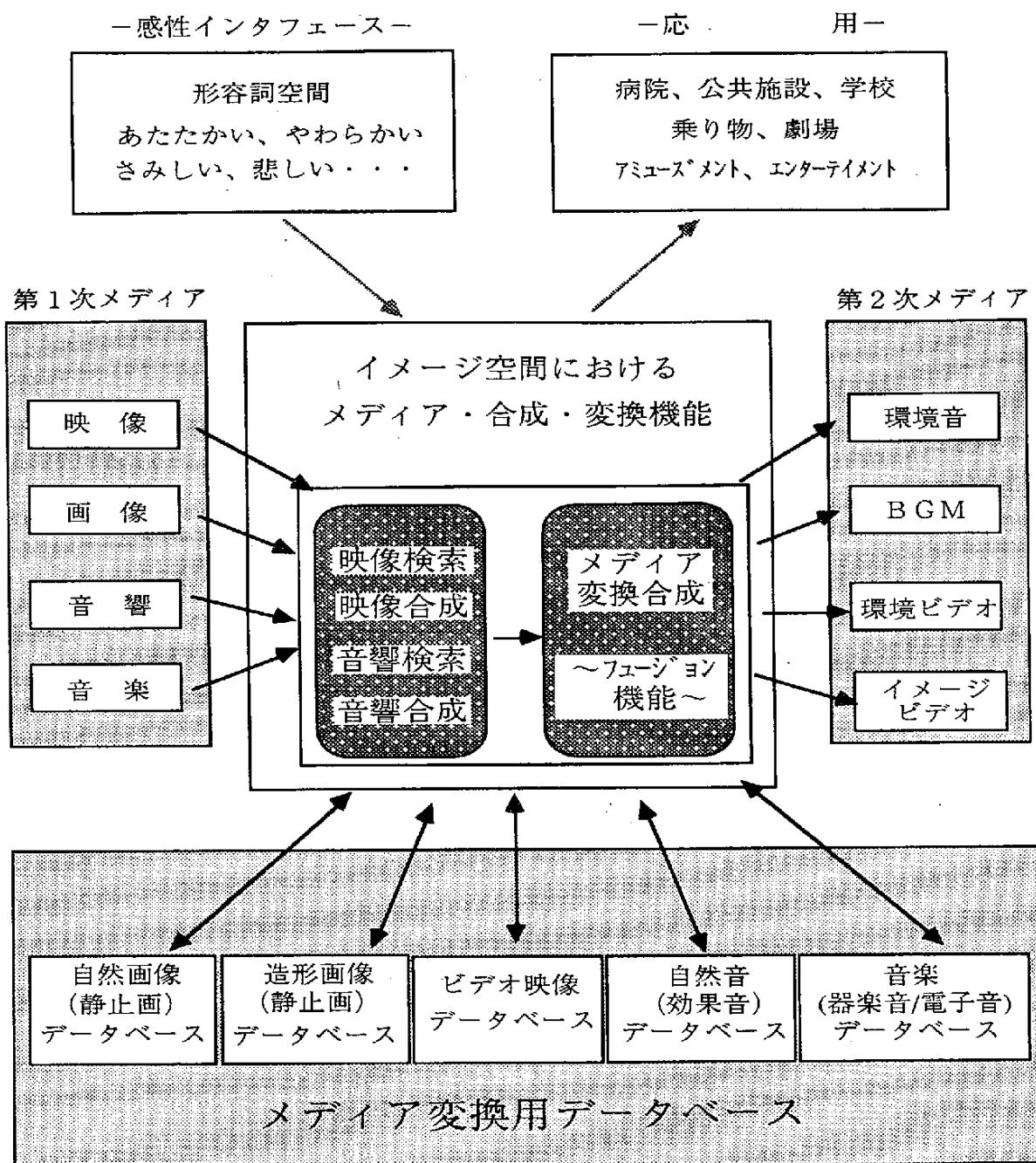


図14-11 メディア変換型感性データベースの概念図

14.5.2 これからの課題

今後の課題としては、以下のような4点が考えられる。

- (1) 扱う対象の制限
- (2) 処理過程構築の効率化
- (3) グローバルな形容詞空間の構築
- (4) 公共的感性和個性的感性の共存

15 フォールトトレランスな多データベース
 サーバシステムに関する調査研究

株式会社 シネジャーナルプロダクション

15. フォールトトレランスな多データベース サーバシステムに関する調査研究

15.1 概 要

15.1.1 目 的

近年のワークステーションと通信ネットワークの発展により、データベースシステムを複数のワークステーションから構成することが、コストパフォーマンスの点から、可能となってきた。また、通信ネットワークは、これまでのEthernetと公衆通信ネットワークを中心とした構成から、光ファイバとATM交換機を用いたGbps程度の通信速度の高速ネットワークへと変化してきている。今後の情報システムは、こうした高速ネットワークを用いて、複数のワークステーションに分散したデータベースシステムを中心として構成されてきている。現在のデータベースシステムは、クライアントとサーバを基本要素として、これらが、異なったワークステーションに機能分散された形態をとっている。光ファイバ等を用いた高速ネットワークは、高信頼となってきた。これに対して、現在のワークステーションは、従来の大型汎用コンピュータと比較しても、その信頼性と可用性は高くない。複数のデータベースサーバを含んだシステムでの一つの課題は、各ワークステーションが低信頼である下で、どのようにシステム全体の信頼性を高めていくかである。

本調査研究では、データベースシステムが、高速ネットワークにより結合された複数のデータベースサーバから構成されるシステム形態の下で、システム全体をフォールトトレランスとする方式について調査研究する。本調査研究により、わが国におけるデータベースシステムの発展に寄与することを目的としている。

15.1.2 実施内容

本調査研究における主な実施内容は、以下の項目に示すとおりである。

(1) システムのフォールト（障害）の種類の検討

分散システムがこうむるフォールトとして、どのようなものがあるかを検討する。ワークステーションのフォールトとして、従来から検討されている停止障害に加えて、各サーバとクライアントが誤動作するビザンティン障害も検討する。

(2) 多重化方式

複数のサーバから構成されるデータベースシステムで、信頼性と可用性を向上するためには、データベースシステム内のデータを複数のサーバに冗長に配置することである。デ

データベースサーバの種々のフォールトに対して、信頼性と可用性を向上するためのデータの冗長配置方法について検討する。特に、多重化することにより生じる全処理量と通信量を低減する方式を検討する。

(3) 同期方法

分散したデータベースシステムでは、フォールトが発生する下で、データベースサーバ間、クライアントとサーバ間、クライアント間で、種々の同期をとる必要がある。データが、冗長に配置されている場合の同時実行方式を検討する。各トランザクションが、従来のように一時に一つのオブジェクトを要求する場合だけでなく、 n 個を同時に要求し、これのすべてを待つ場合、この中の一つの獲得を待つ場合等でのデッドロックの解決方式も研究する。

(4) フォールトトレランスな分散データベースサーバシステムのアーキテクチャ

サーバとクライアントの種々のフォールトに対して、データベースへのアクセスを提供するシステムのアーキテクチャを検討する。

15.2 実施成果の要旨

15.2.1 データベースシステムの動向

まず、フォールトトレランスな分散データベースシステムを考えると必要となるデータベースシステムの技術動向についてまとめた。大型の汎用コンピュータに集中的に構成されてきたデータベースシステムが、ローカルエリアネットワーク（LAN）等の通信ネットワークにより相互接続された複数のワークステーションにより実現されてきている。特に、データベースシステムの機能をデータベースに対する基本的な操作を行うサーバ、ユーザに対するインタフェースを提供するクライアントとの二つに、機能分割し、各々を複数のワークステーションに配置するクライアントサーバモデルが中心となっている。ここでは、こうしたクライアントサーバモデルのデータベースシステムについてまとめている。またデータベースシステムとして、テーブルをデータ構造としたリレーショナルモデルについて考えている。

データベースシステムのサーバのフォールトに対処するためには、データを複数のサーバに配置することが必要となる。データベース内のデータ（テーブル）を複数のサーバに分散する方法として、水平断片化と垂直断片化について述べ、各々の特徴をまとめている。また、これらの断片単位に、複数のサーバに冗長に配置する方法を示している。

15.2.2 フォールト

分散データベースシステムを含めた分散システムがこうむるフォールト（障害）として、どのようなものがあるかを検討した。従来の大型汎用コンピュータに対して、ワークステーションの信頼性は高くない。複数のワークステーションからシステムを構成するためには、各ワークステーションがこうむるフォールトについて検討する必要がある。

ワークステーションのフォールトとして、大別して、停止障害とビザンティン障害がある。停止障害とは、ワークステーションが一定時間以上動作を停止するフォールトである。いわゆるシステムのダウンである。この停止障害に加えて、各サーバとクライアントが誤動作するビザンティン障害がある。一般に、データベース管理システムのバグにより、ワークステーションが正常な動作を行わなくなるフォールトである。

こうしたワークステーションのフォールトに対して、分散型データベースシステムのサーバをユーザに正常に提供するために、本調査研究では、データベース内のデータを複数のサーバに冗長に配置する多重化方式を検討することとした。冗長に配置されたデータを、レプリカ(replica)とする。

15.2.3 トランザクション

次に、分散データベースシステムでの処理の基本となるトランザクション(transaction)について検討した。トランザクションとは、複数のサーバ内のデータのレプリカを操作する複数のreadと write演算の実行状態で、原子的な実行単位である。ここでは、複数のデータベースサーバから構成されるシステムでのトランザクションの概念を明確にし、各トランザクションが複数のサーバ上で並列に実行される場合を検討した。特に、複数のサーバに冗長に配置されたレプリカを更新する方法を検討した。複数のサーバが同一のデータをもつことから、各冗長レプリカ間での一致性を保つ必要がある。このための方式として、(1)すべてのサーバ内のレプリカをロックしてから更新する基本方式、(2)一つのサーバを主サーバとしてロックは主サーバに対してのみ行う主サーバ方式、(3)すべてのサーバではなく、ある一定数のサーバを更新するクォーラム方式を示した。

更に、データのロックを行うときに生じるデッドロックについて検討を行った。冗長なレプリカが存在する場合には、ロック方式で示したように、複数のレプリカのすべてのロックを行えなくてもよい場合がある。ここでは、各プロセス間でのデータの待ち関係を示す待ちグラフを作り、これが有向辺をもつときデッドロックとする従来の概念では、デッドロックを定義できない。このために、互いに待ち関係にあるプロセスの集まりを結び(knot)として、この中に含まれているプロセスをデッドロック状態とした。このようなデッドロックを

明確にし、その検出と解除方法を示した。

トランザクションが複数のサーバを操作することから、すべてのサーバの更新が完了したときに、トランザクションが完了したとすることを補償する必要がある。ここで、一つでもサーバの更新を行えなかったとき、すべてのサーバの更新を無効とする必要がある。このための制御として、コミットメント制御がある。ここでは、コミットメント制御の方式として、二相コミットメントプロトコル、三相コミットメントプロトコルについて示し、その性能を示した。

15.2.4 多重化方式

フォールト（障害）に対して頑強な分散データベースシステムのモデルについて検討を行った。分散データベースシステムのあるプロセスにフォールトが起こると、利用者はシステムを使用できなくなる場合がある。特に、データベースサーバのプロセスにフォールトが起きた場合、そのサーバの提供するデータベースが使用不能となる。このために、サーバを多重化し、各々異なったコンピュータに配置する。データベースサーバを多重化したものをレプリカとする。フォールトトレラント分散データベースシステムでは、あるレプリカがフォールトしても残りのレプリカによりサービスを継続して提供できる。

多重化の方式として、本調査研究では、(1)能動的多重化、(2)受動的多重化、(3)準能動的多重化の3方式を示した。能動的多重化では、全レプリカが同じ処理を行うものである。この方式では、全レプリカで処理が行われるために、全処理時間と通信量が問題となる。これに対して、受動的多重化では、ある一つのレプリカが主レプリカとして、サーバとしての処理を行う。主レプリカは、他のレプリカに定期的にデータベースの状態（又は状態の変化）を示すログを送信する。他のレプリカは、送信されたログにより、自分のデータベースの更新を行う。準能動的多重化では、全レプリカは同じ処理を行うが、クライアントへの応答は、主レプリカのみが行う。

以上の方式は、レプリカの停止障害に対しては、対処できる。レプリカのビザンティン障害に対しては、複数のレプリカからの応答を比較することによりフォールトを検出できる。このためには、フォールトしたレプリカ数を f とすると、 $2f + 1$ 以上のレプリカが必要となる。能動的多重化は、ビザンティン障害に頑強であるが、処理時間と通信量が大きくなる問題がある。一方、受動的と準能動的では、処理時間と通信量は減少できるが、応答を返す主レプリカが一つであるために、ビザンティン障害に対処できない。このために、本調査研究では、主レプリカを複数として、ビザンティン障害に対処できるとともに、処理時間と通信量を最小化する複合多重化方式を示した。

16 創出キーワードの自動付与に関する調査研究

株式会社 エレクトロニック・ライブラリー

16. 創出キーワードの自動付与に関する調査研究

16.1 概 要

16.1.1 背景と目的

データベース構築技術の最も基本となる索引では、近年、テキストデータからの索引語の自動抽出が主流になっている。文中に含まれる自然語を一定のルールに従って抽出して検索キーとするこの方式は、大量の情報を扱う上で効率的であり、コスト軽減効果が大きいばかりでなく、検索機能としても安定性、網羅性で優れている。反面、①索引語が膨大になってシステムに負荷がかかる、②文意を離れて抽出された索引語が作用して検索結果に「ノイズ」を生じる、③テキスト文中に現れない表記や、抽象的な概念が抜け落ち、検索時に「漏れ」を生じる、などの弱点を抱えており、共通の悩みとなっている。

(株)エレクトロニック・ライブラリー（略称「EL＝イーエル」）は、国内で発行されている新聞・雑誌の記事を対象に、1988年から記事単位の切り抜きイメージを蓄積・提供するデータベースサービスを行ってきた。印刷物そのものを素材としている関係から、コンピュータで処理できる文字データがなく、索引、入力など作業の大半を人手によっている。このため構築コストの負担が大きく、数年前から作業機械化の本格検討に取り組んでいる。

ELでは平成5年度に、(財)データベース振興センターの委託事業として「OCR（光学的文字読み取り装置）を利用したキーワードの自動抽出に関する調査研究」を行い、「既に商品化されているOCRと日本語処理ソフトをベースに、一定の補正機能と自動抽出キーワードの絞り込みルールを付加することによって、人手作業と遜色のないデータベースの品質を維持できる」との結論を得た。この成果の上に立って、今回は品質の向上の中身に一步踏み込み、自動抽出の対象となるテキストの文中に現れない重要キーワード（「創出キーワード」）を、自動的に付与する実践的な方法論を探ることとした。

16.1.2 実施内容と結果の概略

本調査研究では、テキスト（記事本文）中にそのままの形では現れない使用頻度の高い重要キーワードを、自動抽出キーワードに対置する「創出キーワード」と定義し、ELデータベースに蓄積された350万件にのぼる大量の記事と関連データを活用し、創出キーワードをコンピュータ処理で自動付与する実践的な「創出キーワード辞書」と付与論理式のプロトタイプを構築した。

その結果おおむね以下のような諸点が確認され、比較的簡便で柔軟な手法によって、テキ

スト文中に現れない概念的なキーワードを自動的に付与する実戦的な方法の確立への展望が開かれた。

(1) 「創出キーワード辞書」モデルの有効性確認

今回の調査研究の核になった「創出キーワード辞書」モデルは、主題概念を表す中心的な語(Core Term)を中心に、関係するキーワードや文字列の特徴を関連付けた上、それぞれの語、文字列の出現順序や位置、キーワードに付与されている情報などを組み合わせたもので、各部分の手法だけをとってみれば特に目新しいものはない。

しかし、これらを「記事把握機能」と「ノイズ排除機能」に整理して組み合わせることによって、自動付与ツールとして十分機能することが実証でき、今後の改善によって実用にたえうとの見通しを得た。

(2) 検索精度の向上と効率化に寄与

実験では自動付与の「平均再現率87.2%」の成果が得られた。これは同一データに対する人手付与の再現率(約40%)の倍以上になっており、不適キーワードによる「ノイズ」を考慮しても、全体として精度向上に寄与するとの見通しが得られた。一定水準に達したもののからキーワード自動抽出の補助ツールとして組み込むことによって相応の作業効率向上が期待できる。

(3) 柔軟な活用とメンテナンスが可能

「創出キーワード辞書」はシソーラスのように全体の体系にこだわらず、個別キーワードごとに対応できる形になっている。このため、利用頻度の高い重要キーワードで、付与結果に一定の信頼性が得られたものから順に組み込むといった柔軟な対応ができる。

また、構成要素も「記事把握機能」と「ノイズ排除機能」を縦軸、語彙や構文情報などの組み合わせを横軸にしたモジュール構造に整理されており、キーワードごとの個別条件の修正だけでなく項目ごとの追加・修正・削除も容易な仕組みになっている。新たな事象の出現、概念の変化に対応して、柔軟なメンテナンスができることも実戦的な特長といえる。

16.2 実施結果

16.2.1 実施方法の概略

本調査研究は、E Lデータベースに蓄積された3,000の新聞記事をOCRを使って文字データ化し、これを素材として3段階に分けて行った。

(1) 「創出キーワード辞書」モデルの作成

(2) 自動付与論理式の構築

(3) 自動付与の実験と結果の評価・分析

対象となる創出キーワードは、素材 3,000記事のうち日本語処理ソフトにかけて必要なデータ処理をした 1,000記事から抽出した「候補群」を母体とし、E Lデータベース辞書に蓄積されているキーワード単位の記事数データ、ユーザ検索式のキーワード登録頻度情報などを参照して、利用度の高い50キーワードを選定した。(表16-1)

表16-1 最終選定された50キーワード

項 番	KW	項 番	KW	項 番	KW
1	システム開発	18	建設工事	35	新会社
2	システム導入	19	研究開発	36	新製品
3	価格改定	20	現地生産	37	人 事
4	価格動向	21	合併会社	38	生産体制
5	会社決算	22	材料開発	39	製品開発
6	海外進出	23	市場開拓	40	相場動向
7	海外調達	24	市場参入	41	装置開発
8	環境対策	25	市場動向	42	体制強化
9	環境問題	26	施設建設	43	大気汚染
10	企業業績	27	事業強化	44	調査結果
11	企業合併	28	事業戦略	45	店舗展開
12	企業戦略	29	事業展開	46	日米関係
13	企業提携	30	需給動向	47	廃棄物処理
14	技術開発	31	需要動向	48	販売強化
15	金利動向	32	周年記念	49	販売計画
16	経営計画	33	出店計画	50	販売動向
17	建設計画	34	情報提供		

次に、それぞれのキーワードについて収集したサンプル記事を材料として、創出キーワードの概念を構成する語彙の相関関係や構文上の特性などを解析し、適合する記事候補を「把握する機能」と候補の中からノイズを排除する「絞り込み機能」に整理。これを縦軸とし、関係語彙や構文情報の組み合わせを横軸とするモジュール構造の「創出キーワード辞書」の

モデルを構築した。

自動付与論理式は、創出キーワード辞書に整理された情報を実行するため、①文字列、②キーワード情報や新聞紙名コードなどのフィールド情報、③出現順序や出現位置などの構文情報を論理演算できるプログラムのプロトタイプを開発した。

素材 3,000記事のうち 2,000記事を使って、パソコン上で自動付与の実験を行い、キーワードごとの付与結果の適否をE Lデータベース作成に係わっているメンバー 8 チームで評価した。その結果を調査研究チームのメンバーが総合的に分析し、問題点と課題を整理した。

16.2.2 「辞書」と論理式

(1) 「辞書」モデルの構造

「辞書」モデルは創出キーワードの意味を体現し、適合する記事を集める側に働く語彙群を縦軸とし、対象概念ごとの語彙群の組み合わせを横軸とする統一したモジュール構成に組み上げた(図16-1)。

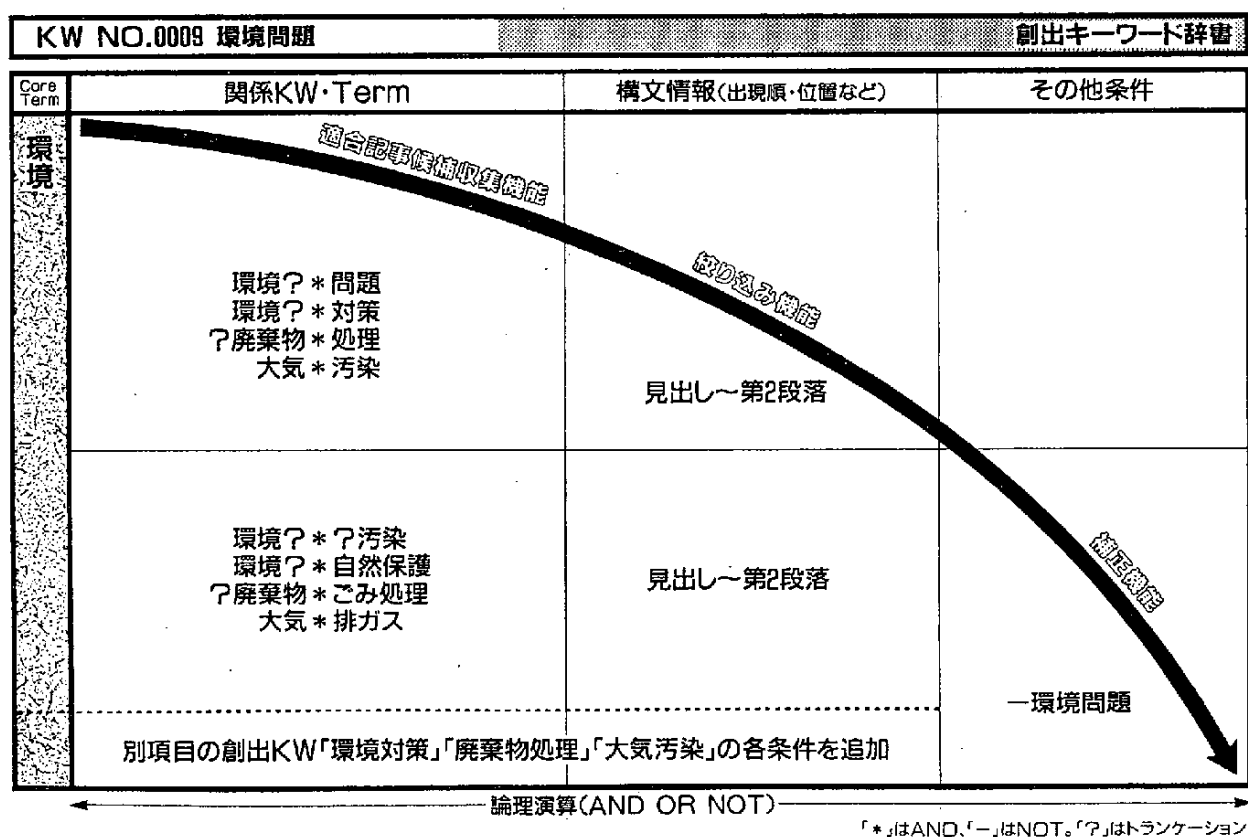


図16-1 「辞書」モデルの構造

(2) 論理式の機能

「辞書」の要請によって論理式は、以下の四つの機能を備えた仕様とした。

- ① 論理演算（論理和，論理積，論理否定）ができる。
- ② 文字列検索，キーワードの前方一致，後方一致，中間一致（トランケーション）指定ができる。
- ③ 新聞紙名コードなどE Lデータベースの検索フィールド記述子，日本語処理ソフト「INDEXER」のキーワード種別が使える。
- ④ キーワード及び文字列の出現範囲，出現順が指定できる。

16.2.3 自動付与の実験と結果

実験用素材として文字データ化しておいた 2,000記事を対象に，プリテストで検証した「辞書」の内容を論理式に記述した50の創出キーワードをかけ，パソコン上で自動付与の実験を行い，その結果を評価した。

(1) 結果の概要

実験の結果，2,000記事中1,850記事に創出キーワードが自動付与された。最も多かった「新製品」の183記事を筆頭に100記事以上が3キーワード，10記事未満が9キーワード。50キーワードの平均は37キーワードだった。

同じ記事に対しE Lデータベースのインデクサーが付与した結果と比較すると，総数では人手 578記事に対し自動は 1,850記事。内訳をみると，双方に共通のものは 384記事，人手だけのものは 194記事，自動でだけ付いたものは 1,466記事だった。

評価を集計した結果，適合記事は全体で 1,402記事（自動独自＝840記事，人手独自＝180記事，共通＝382記事）となった。これに対して不適合とみなされたものは 642記事（自動独自＝626記事，人手独自＝14記事，共通＝2記事）で，不適合記事の97%強が自動付与によるものだった。人手と自動の適合記事の合計を創出キーワードに対する適合記事の総数と仮定し，これに対する適合記事数の比率を「再現率」，自動，人手それぞれの総数に対する適合記事数を「適合率」として，精度，「もれ」などをみた。自動付与の「再現率」は50キーワード平均で87.2%（人手は40.1%），「適合率」は66.1%（人手は97.2%）となった（図16-2）。

各キーワードの実用度を計ると尺度として「再現率」80%以上，「適合率」70%以上のところで線を引いてみると，ほとんどはどちらかの線に達していることが分かった（表16-2の網掛け部分）。しかし，双方の条件を同時に満たしているキーワードは20にとどまった。

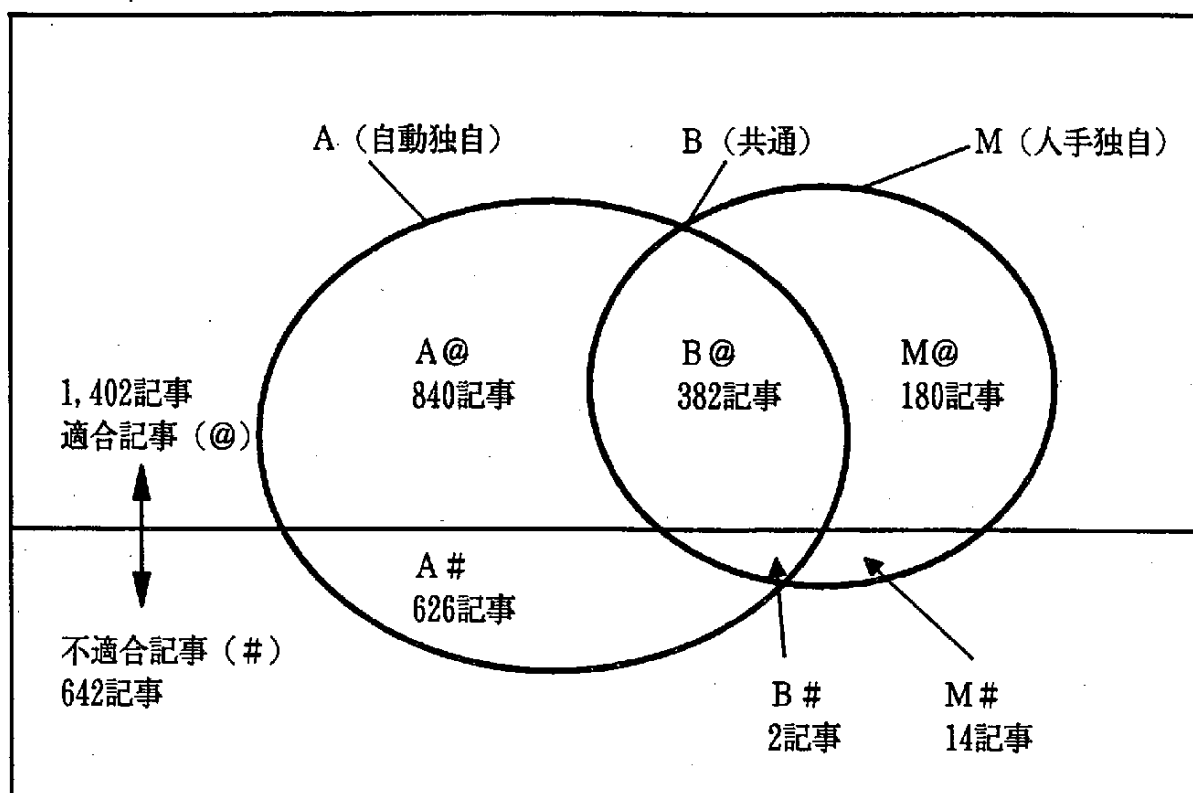


図16-2 創出キーワード付与記事数分布 (概念図)

(2) 「不適合」の原因

創出キーワード付与が不適合だった記事について、すべて原因を解明することは時間的余裕もなかったため、再現率、適合率がそれぞれ80%、70%の線に達しなかったキーワードの中から12キーワードを選んで分析した。

① 再現率の低かった7キーワード

価格改定, システム開発, 販売強化, 販売計画, 事業強化, 市場開拓, 情報提供

漏れた記事が8件あった「価格改定」「システム開発」をみると、「システム開発」では論理式中の「開発」に後方一致の「？」がついていなかったため、システムの共同開発が脱落した。また「実行予算書システム」「故障予知・診断システム」「工程表作成システムPATRA」などEL辞書登録語でないものは「？システム」で拾えなかった。「価格改定」では日本語解析の際、「値上げ」「値下げ」「引き上げ」「引き下げ」などが動詞と判定されたことが漏れの原因となった。

② 適合率がいま一步の5キーワード

出店計画, 海外調達, 会社決算, システム導入, 日米関係

「出店計画」では自動付与した18件の記事のうち、8件がノイズと評価されたので、論理式の絞り込みを厳しくした。企業名、出店、店舗など周辺語の3語の出現条件が「見出しから第2段落まで」とあったのを「第1段落まで」としたところ、6件の排除ができた。「日米関係」では、20件のノイズのうち12件が「日本時間」によるいたずらだった。辞書の内容を論理式に記述する際のミスマッチが原因だった。

(3) 改善テスト

原因追究のため選んだ12キーワードについて、修正した論理式でねらい通りの結果がえられるかどうか再度実験をした。その結果は、条件が異なり、適不適の判定も自前で行ったので、本番テストの数字と単純比較はできないが、改善効果があったといえる。

特に「再現率」は「情報提供」を除いて11キーワードが80%以上になった。「適合率」は「システム導入」「日米関係」「出店計画」「販売計画」で改善はみられたものの、70%台に乗せたのは2キーワードだった。「販売強化」「会社決算」は「再現率」で改善がみられた反面、「適合率」は落ち、両者がトレードオフの関係にあることを裏付けた。

改善テストの結果、7キーワードが同時に「再現率80%」、「適合率70%」の線を越えたことは、今後の論理式の精度アップに明るい見通しを与えた。

16.3 実験結果の評価と展望

16.3.1 評価

調査研究対象の50キーワードの平均「再現率」が87.2%に達したことは、短時間の実験結果としては高く評価できよう。

今回の実験では、同じデータでの人手付与に較べ倍以上の高い付与率が確認できた。論理式による付与は均質・網羅性ではすぐれており、ノイズの排除（「適合率」の向上）さえ伴えば、実戦に耐えうるとの展望を抱かせるに十分な数字であった。

12キーワードの追加改善テストで「再現率」が約18%アップすると同時に、「適合率」も約8%上がっていることは、改善への展望を裏付けているといえる。

創出キーワードは概念的要素が高く、それ自体検索効果の高いキーワードといえる。今回の調査研究の手法を踏襲して、利用頻度の高いキーワードから順に取り上げ、ブラッシュアップを繰り返す、一定の信頼性が得られ次第、辞書に組み込むことでもかなりの検索効率の向

上が期待できる。積み重ね方式の可能な辞書の構造が、このような実戦的で柔軟な活用を可能にする。

創出キーワードの付与は、インデクシングの中でも最も負荷の高い作業となっている。一定水準以上の信頼性をもった辞書と自動付与論理式が増えていけば、作業効率化に大きく貢献するはずである。

以上のような諸点から、今回の調査研究は、比較的簡便な辞書と論理式によって有効な創出キーワード自動付与システムへの展望を切り開くことができたと評価したい。

16.3.2 展 望

今回の調査研究に当たって、われわれのチームの合い言葉は「実戦的」の一言だった。「コンピュータに意味を解析させ、概念的なものを処理させる」ことは、古くて新しいテーマである。研究レベルではさまざまな成果が発表されており、データベースサービスの世界でも地道な努力が続けられているが、汎用的なツールまでにはいたっていない。

われわれは平成6年度の委託研究で、商品化されているOCRと日本語解析ソフトを組み合わせ、独自の絞り込み方式を加味することによって、比較的少ないキーワードで効率的な自動抽出システム構築への確信を得た。今回はこれを一歩進め、より検索効率の高いキーワードの自動付与の実験に取り組んだ。

今回の調査研究結果で、われわれはテーマとした創出キーワードの自動処理の実用化への手がかりを得たと確信する。

われわれは、電子メディアの急進展の中にあっても、なお、印刷物を素材とするデータベース構築技術の有効性を疑わない。簡便でより効率的な技術を開発、実用化することが真の情報化時代の一翼を担うものと考えている。

2年にわたる調査研究の成果をまず現実のデータベース構築に活かすと同時に、より汎用性の高い共通の財産へと高めていく決意を新たにしている。

表16-2 創出KW自動付与テスト結果集計 (全KW)

項番	KW	A@	A#	A計	B@	B#	B計	M@	M#	M計	自動計	人手計	再現率	適合率	適合記事
1	大気汚染	1	0	1	1	0	1	0	0	0	2	1	100.0	100.0	2
2	需給動向	7	1	8	0	0	0	0	0	0	8	0	100.0	87.5	7
3	企業合併	5	1	6	1	0	1	0	0	0	7	1	100.0	85.7	6
4	現地生産	10	2	12	1	0	1	0	0	0	13	1	100.0	84.6	11
5	相場動向	24	6	30	1	0	1	0	0	0	31	1	100.0	80.6	25
6	合併会社	10	3	13	0	0	0	0	0	0	13	0	100.0	76.9	10
7	企業提携	10	5	15	2	0	2	0	0	0	17	2	100.0	70.6	12
8	研究開発	14	6	20	0	0	0	0	0	0	20	0	100.0	70.0	14
9	海外調達	2	1	3	0	0	0	0	0	0	3	0	100.0	66.7	2
10	環境問題	27	40	67	0	0	0	0	0	0	67	0	100.0	40.3	27
11	金利動向	8	18	26	0	0	0	0	0	0	26	0	100.0	30.8	8
12	市場動向	56	46	102	4	0	4	1	0	1	106	5	98.4	56.6	61
13	販売動向	36	36	72	5	0	5	1	0	1	77	6	97.6	53.2	42
14	企業業績	29	97	126	11	0	11	1	0	1	137	12	97.6	29.2	41
15	会社決算	16	14	30	11	0	11	1	0	1	41	12	96.4	65.9	28
16	環境対策	26	29	55	1	0	1	1	0	1	56	2	96.4	48.2	28
17	体制強化	43	7	50	1	0	1	2	1	3	51	4	95.7	86.3	46
18	施設建設	12	4	16	6	0	6	1	0	1	22	7	94.7	81.8	19
19	価格動向	24	3	27	4	0	4	2	0	2	31	6	93.3	90.3	30
20	需要動向	25	34	59	0	1	1	2	0	2	60	3	92.6	41.7	27
21	企業戦略	70	6	76	8	0	8	7	0	7	84	15	91.8	92.9	85
22	周年記念	6	0	6	5	0	5	1	0	1	11	6	91.7	100.0	12
23	経営計画	41	43	84	2	0	2	4	0	4	86	6	91.5	50.0	47
24	建設工事	28	3	31	2	0	2	3	1	4	33	6	90.9	90.9	33
25	事業戦略	28	3	31	2	0	2	3	1	4	33	6	90.9	90.9	33
26	日米関係	7	20	27	2	0	2	1	0	1	29	3	90.0	31.0	10
27	店舗展開	21	18	39	5	0	5	3	0	3	44	8	89.7	59.1	29
28	出店計画	10	8	18	6	0	6	2	0	2	24	8	88.9	66.7	18
29	事業展開	24	18	42	0	0	0	3	0	3	42	3	88.9	57.1	27
30	システム導入	8	10	18	6	0	6	2	0	2	24	8	87.5	58.3	16
31	新製品	15	17	32	150	1	151	25	6	31	183	182	86.8	90.2	190
32	技術開発	11	3	14	2	0	2	2	1	3	16	5	86.7	81.3	15
33	海外進出	18	9	27	7	0	7	4	0	4	34	11	86.2	73.5	29
34	製品開発	52	11	63	19	0	19	12	0	12	82	31	85.5	86.6	83
35	材料開発	3	4	7	2	0	2	1	1	2	9	4	83.3	55.6	6
36	市場参入	6	3	9	2	0	2	2	0	2	11	4	80.0	72.7	10
37	人事	18	35	53	8	0	8	7	0	7	61	15	78.8	42.6	33
38	生産体制	6	3	9	1	0	1	2	0	2	10	3	77.8	70.0	9
39	事業強化	11	5	16	2	0	2	4	0	4	18	6	76.5	72.2	17
40	調査結果	17	16	33	50	0	50	21	1	22	83	72	76.1	80.7	88
41	廃棄物処理	2	0	2	1	0	1	1	0	1	3	2	75.0	100.0	4
42	装置開発	4	4	8	9	0	9	5	0	5	17	14	72.2	76.5	18
43	販売計画	7	3	10	2	0	2	4	0	4	12	6	69.2	75.0	13
44	建設計画	21	25	46	14	0	14	16	0	16	60	30	68.6	58.3	51
45	市場開拓	2	0	2	2	0	2	2	1	3	4	5	66.7	100.0	6
46	新会社	3	2	5	8	0	8	6	0	6	13	14	64.7	84.6	17
47	販売強化	7	0	7	0	0	0	4	0	4	7	4	63.6	100.0	11
48	システム開発	2	1	3	10	0	10	8	0	8	13	18	60.0	92.3	20
49	情報提供	4	0	4	2	0	2	5	0	5	6	7	54.5	100.0	11
50	価格改定	3	3	6	4	0	4	8	1	9	10	13	46.7	70.0	15
合計		840	626	1466	382	2	384	180	14	194	1850	578			1402
平均		16.8	12.5	29.3	7.6	0.1	7.7	3.6	0.3	3.9	37.0	11.6	87.2	66.1	28.0

注) ①網掛けは再現率80%、適合率70%以上のKW。 再現率の高い順でソート。

②自動計=A+B 人手計=M+B

③再現率= $\frac{\text{自動適合記事 (A@+B@)}}{\text{適合記事 (A@+B@+M@)}}$

適合率= $\frac{\text{自動適合記事 (A@+B@)}}{\text{自動計 (A+B)}}$

禁 無 断 転 載

平 成 7 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

TEL 03-3832-9571

