

09-開-09

データベース構築促進及び技術開発に関する報告書

イメージファイリングの効率的活用を目指す 書誌情報データベース検索技術の構築

平成10年3月

財団法人 データベース振興センター

委託先 株式会社 会議録研究所

KEIRIN



この事業は、競輪の補助金を受けて実施したものである。

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および自由な情報流通の確保の観点からも必要である。しかしながら、現在わが国で流通しているデータベースの中でわが国独自のものは1/3にすぎないのが現状であり、わが国データベースサービスひいてはバランスある情報産業の健全な発展を図るためには、わが国独自のデータベースの構築およびデータベース関連技術の研究開発を強力に促進し、データベースの拡充を図る必要がある。

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および技術開発について民間企業、団体等に対して委託事業を実施している。委託事業の内容は、社会的、経済的、国際的に重要で、また地域および産業の発展の促進に寄与すると考えられているデータベースの構築とデータベース作成の効率化、流通の促進、利用の円滑化・容易化などに関係したソフトウェア技術・ハードウェア技術である。

本事業の推進に当って、当財団に学識経験者の方々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 東海大学教授 上條史彦氏）を設置している。

この「イメージファイリングの効率的活用を目指す書誌情報データベース検索技術の構築」は平成9年度のデータベースの構築促進および技術開発促進事業として、当財団が株式会社 会議録研究所に対して委託実施した課題の一つである。この成果が、データベースに興味をお持ちの方々や諸分野の皆様方のお役に立てば幸いである。

なお、平成9年度データベースの構築促進および技術開発促進事業で実施した課題は次表のとおりである。

平成10年3月

財団法人 データベース振興センター

平成9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 インターネット型先進材料DB活用プログラムの開発	(財)次世代金属・複合材料研究開発協会
	2 インターネットを利用したイベント関連情報に関するデータベースの構築	(社)日本イベント産業振興協会
	3 オーサリング型地図付地域ガイドデータベースのプロトタイプ構築	(財)地図情報センター
	4 高齢者在宅介護情報のデータベース構築	(株)フォワード
	5 筑波研究学園都市研究便覧インターネット対応化事業	(株)筑波出版会
	6 建築行政指導要綱のHTMLデータベース構築	日本建築法令(株)
	7 中小小売業のための商品仕入れ情報データベースプロトタイプ構築	(財)店舗システム協会
地域活性化	8 新聞記事・画像データベース構築	琉球新報社
技術	9 イメージファイリングの効率的活用を目指す書誌情報データベース検索技術の構築	(株)会議録研究所
	10 やわらかい分散オブジェクトシステムに関する調査研究	(株)シネ・ジャーナルプロダクション

目 次

1	概要	1
1. 1	目的	1
1. 1. 1	目的	1
1. 2	行政情報化推進基本計画に基づく文書管理のあり方	2
1. 2. 1	行政文書活用に求められる機能	2
2	イメージファイリングシステムの今後のあり方	4
2. 1	イメージファイリングシステムの現状	4
2. 1. 1	イメージファイリング専用システム（電子ファイリングシステム）	4
2. 1. 2	OCR 機能を有したイメージファイリングシステム	4
2. 2	既存システムと比較した本システムの利点	5
2. 2. 1	マルチプラットフォーム対応可能でオープンなファイル形式	5
2. 2. 2	文書イメージデータ容量のコンパクト化	6
2. 3	PDF ファイル	7
2. 3. 1	PDF ファイル形式の特徴とポイント	7
2. 3. 2	PDF による電子化文書作成の流れ	8
2. 4	イメージファイリングシステムの今後のあり方	9
3	文書ファイル形式としての SGML と PDF の対比	10
3. 1	SGML とは何か？	10
3. 1. 1	SQL データベースから SGML データベースへの変遷	10
3. 1. 2	SGML とは何か？	12
3. 2	PDF とは何か？	14
3. 2. 1	イメージデータのデファクトスタンダードとしての PDF	14
3. 2. 2	PDF ファイルに関わる関連ツール	15
3. 3	SGML と PDF の対比	15
3. 3. 1	行政文書分野における SGML 化すべき点	16
3. 3. 2	行政文書分野における PDF 化すべき点	17
4	本事業におけるイメージファイリングシステムの特徴と活用方法	19
4. 1	本イメージファイリングシステムの特徴	19
4. 1. 1	PDF ファイル形式の採用	19
4. 1. 2	PDF 対応文書管理システムの採用	19
4. 1. 3	本イメージファイリングシステムの活用・運用分野	23
5	イメージファイリングシステムの概要	24
5. 1	本システムの特徴	24
5. 2	本システムの機能	24
5. 3	システム全体の流れ	25
5. 3. 1	登録処理	25
5. 3. 2	検索処理	26
6	イメージファイリングシステムの機能	27
6. 1	イメージ登録	27
6. 1. 1	画面遷移図	27
6. 1. 2	メイン画面	28
6. 1. 3	スキャナ入力画面	29

6. 1. 4	データ入力画面	3 2
6. 1. 5	分類作成ダイアログ画面	3 4
6. 2	文書登録	3 5
6. 2. 1	画面遷移図	3 5
6. 2. 2	TOP 画面	3 7
6. 2. 3	索引メニュー画面	3 8
6. 2. 4	検索画面	3 9
6. 2. 5	類別一覧削除画面	3 9
6. 2. 6	シソーラス辞書展開画面	4 0
6. 2. 7	検索結果一覧削除画面	4 0
6. 2. 8	地所メンテナンスメニュー画面	4 1
6. 2. 9	システムカスタマイズ辞書不要語登録画面	4 2
6. 2. 10	ユーザ辞書新規作成画面	4 4
6. 2. 11	ユーザ辞書修正メニュー画面	4 5
6. 2. 12	ユーザ辞書修正グループ追加画面	4 6
6. 2. 13	ユーザ辞書修正グループ削除画面	4 7
6. 2. 14	ユーザ辞書修正グループ内用語修正画面	4 9
6. 2. 15	ユーザ辞書削除画面	5 1
6. 2. 16	ユーザ辞書一覧参照画面	5 1
6. 3	文書検索	5 3
6. 3. 1	画面遷移図	5 3
6. 3. 2	TOP 画面	5 4
6. 3. 3	検索画面	5 5
6. 3. 4	分類別一覧表示画面	5 7
6. 3. 5	シソーラス辞書展開画面	5 8
6. 3. 6	検索結果一覧画面	5 9
7	本事業開発によるイメージファイリングシステムの効率的な運用	6 0
7. 1	本イメージファイリングシステムの運用方法	6 0
7. 1. 1	シソーラス辞書	6 0
7. 1. 2	シソーラス辞書の構成	6 1
7. 1. 3	シソーラス辞書の内容	6 1
7. 1. 4	システムシソーラス辞書とユーザーシソーラス辞書の関係	6 2
7. 2	本イメージファイリングシステムの運用の姿	6 4
7. 3	本イメージファイリングシステムの運用分野	6 5
7. 3. 1	条例の制定／改定	6 5
7. 3. 2	意見書・請願書・陳情書	6 6
7. 3. 3	その他の行政文書分野における本システムの活用	6 6
8	行政分野以外の本イメージファイリングシステムの運用例	6 9
8. 1	民間分野における本イメージファイリングシステムの活用	6 9
8. 1. 1	製造業	6 9
8. 1. 2	金融機関	7 0
8. 1. 3	その他分野	7 0
9	本イメージファイリングシステムの発展性	7 1
9. 1	コンテンツ構築としての SGML、メディア出力としての PDF	7 1

1 概要

1. 1 目的

1. 1. 1 目的

行政機関における課題点とは、これまでの業務において大量に発生した行政文書の保管・再活用である。行政情報に関しては情報公開法に基づき、ある一定期間の文書保管が義務づけられ、どの自治体においてもその保管スペースに苦慮している。また行政文書とは、紙文書のままでは再活用・閲覧を行う際に多くの手間が発生する。紙文書の中から求める情報を検索・抽出してくることは容易ではない。そこでこれらの観点から行政文書の保管と今後の活用を目指し、行政情報化によるペーパーレス化運動の気運が各自治体で高まり始めている。

この背景として、平成6年12月に閣議決定され、平成9年12月に改定された「行政情報化推進基本計画」が示されており、この計画は行政改革の流れとあいまって、中央省庁と地方自治体連携による行政文書情報化、そして国民と行政機関との情報化連携による新たな国民参加型行政の実現が目指されている。本計画ではこれまでの紙文書による行政活動の非効率性を改め、しかもインターネット・イントラネットといったパソコン環境における既存の大型汎用機・オフコンといった大掛かりなシステム化よりもスリムでシンプルな情報化環境の実現を目指している。

つまり1人1台パソコン環境を実現し、すべての行政業務は紙文書から文書データへと置換し、業務の効率性と業務成果の効果的な活用を実現することをコンセプトとしているものである。

また本基本計画では、文書のデータ化実現のため、構造化文書による文書管理体系である「SGML (Standard Generalized Markup Language)」の行政文書分野における採用を決定している。ただこの手法は今後新たに作成される行政文書分野においてはツールの整備や共通的な文書構造規程の整備によって容易に進めることも可能であるが、問題となるのが過去の行政文書の取扱に関してである。

過去においてワープロ等を利用して作成された行政文書とは、文書の構造化を意識した文書作成がなされていたものは少なく、しかもその多くがワープロパソコン等の電子的データで既に存在しておらず、紙文書という形態でしかないものが多いはずである。(数十年前のものであれば和文タイプによるまったくアナログな文書でしか保存されていない。)

しかも行政情報化の過渡期である現在、関係省庁からの通達等すべての業務情報が電子的な形態で供給されているわけではない。まだまだ紙文書で供給され、アナログファイリングにより整理・保管されるものも多いはずである。そしてその他にも新聞や雑誌等の記事に関して資料文書として保存する場合にも、やはり紙で供給されるもののデータ化という課題は残ってしまう。

ではこれらアナログな紙による文書についても、果たしてSGML化することが得策であろうか。紙文書になってしまったものをSGMLのような構造化文書にするこ

とは大変難しい。単純な文書変換作業のみでこれは終わるものではなく、人間が介在し、文書構造に関する判断を加える作業がなくてはSGML文書変換は実現できない。そこで今後の行政業務に必要となる文書のみをSGML化し、その他のものに関してはイメージデータによる保存とするという見解が大方の流れである。

しかしイメージデータとは、データ容量が大きくなりすぎ、しかも文字も画像として認識されることから紙文書がその形態を変化させたのみで検索性等、電子化のメリットはそこには実現できない。しかも既存のイメージファイリングシステムの多くは専用システムとなっているものが多く、多額の投資をしても、そのイメージデータによる文書ファイルを他のアプリケーションなどにおいて活用することは難しい。

そんな中でAdobeより「PDF (Portable Document File)」というイメージファイル形式が提唱された。これは文字部分を文字データとして認識することができるものであり、これにより文字データの再活用とファイル容量のコンパクト性も実現することができる。しかもOCR (Optical Character Reader) 機能を付加することによって、過去の紙文書も文字をデータ認識させた形態で電子データ化することも可能となる。

そこで今回は、PDF 技術を利用した行政文書分野における新たなイメージファイリングシステムの機能開発を目的とすることとする。

1. 2 行政情報化推進基本計画に基づく文書管理のあり方

1. 2. 1 行政文書活用に求められる機能

行政機関における業務を現すものとは、行政文書にあるといっても過言ではない。担当部署において起案され、それを委員会等の組織の中で検討し、担当責任者が承認を行い（または国会に法案提出され）、通達・告示がなされる。これらの活動はすべて文書という形に現せられ、それが保管・再活用なされるものとなる。しかし紙文書によってこれが流通している間においては、各々の文書を担当者が作成し、その文書の関連性（リンク）という点については、文書を見る人間の判断に頼っている部分が多分にある。この点を肯定したままの行政文書情報化とはまったく無意味である。つまり文書内容に関して必要な情報テーマを容易に検索できる環境、そして関連する文書が時系列でリンクし、その文書内容を追いかけるのみで、その業務の流れ・過程が理解できることが電子行政文書に求められるポイントである。

そこで最適な文書形式とは構造化文書となっているSGML文書ではあるが、前述のような事由から考慮すると今回テーマとする過去の紙文書データ、紙文書で配信される通達文書、紙で得られる新聞・雑誌等の資料文書分野に関しては、すべてをSGML化することは難しい。

かといって単純にOCR入力を行い、TXTファイル化することだけでもあまり意味はない。行政文書には往々にして、システム図やイラスト等の図形データも含まれる。その際にOCRのみでは、結果としてTXTファイルと図形ファイルが別々のファイル化されてしまい、その双方を同時に活用しようとする場合には、そのファイル

のリンク付けと双方のファイルを異なるアプリケーションによって閲覧・加工しなければならない状況となってしまう。

つまり「ファイル検索の容易性」「検索用インデックスファイル作成の手間軽減」「イメージデータ容量のコンパクト化」「PDFファイルというオープンファイル形式による閉鎖性からの開放」等の点から、イメージデータ化された図形ファイルとTXTデータ化されたTXTファイルを1つのファイルの中で管理・閲覧できるPDFファイルの有効性は大きいものと考えられる。

紙文書から電子文書への移行期である現在、過去の文書蓄積を少しでも今後電子的に活用できることを望むのであれば、このコンセプトは間違いのない方向性であると考えられる。

2 イメージファイリングシステムの今後のあり方

2. 1 イメージファイリングシステムの現状

既存のイメージファイリングシステムとは次の2種の考え方による機能製品に集約される。

2. 1. 1 イメージファイリング専用システム（電子ファイリングシステム）

これはオフィス内の紙書類を簡単に保管・整理するために、文字も図形も単純にイメージデータ化する種類の製品であり、過去のイメージファイリングの主体であったマイクロフィルム（マイクロフィッシュ）によるファイリングシステムから発展し、電子化されたものを中心である。

しかし紙文書データをすべてイメージデータ化することから、1枚当たりのイメージデータ量が大量となり（最近の製品でも1枚当たり最低140KBが必要）システムが登場した時代のフロッピーディスクによるファイル管理を行っていた初期型パソコンでは取り扱うことが難しかった。そこで専用システムとして開発されたものが多く、数百万円単位でのシステム導入投資が必要となっていた。そのため役所内部においてもすべての部門・部署で導入されることはなく、統計部門や通達告示文書を管理する文書部門のみがシステム導入できる程度のものであった。

また専用システムであることから、独自のファイル形式によって文書保管がなされ、初期の製品においてはパソコン等の他のアプリケーション上においてデータ閲覧や加工ができないものであった。パソコンの処理速度が大幅に向上し、MOやCD-ROM等の大容量メディアが一般化してきている現在、このようなシステムコンセプトはオープンネットワーク化環境を目指す動きとは異なるものとなっている。

（最近のシステム製品事例）

「ソニー データイータ・PDF-V55」

- メディアにMDを採用し、コンパクトな専用システムとなっている。画像編集機能は有しているものの、文字に関してはイメージ認識機能しか有していない。
- しかしパソコン等との連携に関しては、PDFファイルによってデータ転送する機能を有し、これまでの専用システムの課題を解決したものとなっている。

2. 1. 2 OCR 機能を有したイメージファイリングシステム

前述の専用システムにおける次のような課題を解決するために OCR 機能を有し、文書内の文字データ認識が可能なワークステーション・パソコンによる第2世代のイメージファイリングシステムが登場した。

【OCR機能付与によるイメージファイリングシステムの利点】

1. 文書1枚当たりのデータ容量が大量となってしまう。
文字をOCR機能によってデータ化することができ、文字部分のデータ圧縮が可能。
2. 文書内の文字部分について文字検索ができない。
既存の専用システムでは、各ファイルに検索用のキーワードを付与してファイル検索を行っていた。しかしOCR機能によって文字データ化されたことにより、全文検索エンジン機能を付加し、検索・抽出が可能

また、UNIXやWINDOWS NT等のサーバー・クライアント方式によるネットワーク型システムも登場し、全庁的ネットワークや部門内におけるグループウェアと連携した文書管理システムも可能となった。

しかしこれらシステムも各システムの中でのみ閲覧・利用が可能なものや、特定の文書アプリケーションやブラウザが必要なものであり、オープン化しているとは言い難い。

(最近のシステム製品事例)

「リコー 読取物語キャビネット」

- Windows95 環境でのシステムであり、スタンドアロン環境の中で高速全文検索機能によってキーワードによる文書抽出が可能。
- パイリンガルOCR機能により英文・和文ともに対応可能。
- 本システムについては他のアプリケーションによる文書加工が可能なように「一太郎」「Word」「Excel」へのファイル出力機能を有す。

「松下電器 ドキュメントファイリングシステム」

- UNIXやWindows NTをベースとし、グループウェアと連携したサーバークライアント方式でのイメージファイリングシステム。
- オプションによりWWWブラウザによる文書閲覧も可能で、インターネット・イントラネットにも対応可能。

2. 2 既存システムと比較した本システムの利点

前述のとおりイメージファイリングシステムもパソコン等のハード・ソフトインフラの成長とあいまって、その機能を向上させてきた。

しかし特定のシステムやアプリケーションに依存しないオープンな環境のシステムへの展開という面では既存システムには多くの課題がある。そこでこれらの問題点を解決するために今回の事業において、新たなコンセプトでの「イメージファイリングシステム」の機能開発を行った。そのポイントとして今回開発の本システムの特徴となる利点を列挙する。

2. 2. 1 マルチプラットフォーム対応可能でオープンなファイル形式

これまでのイメージファイリングシステムとは、専用システムとして展開しているものが主体であり、パソコン環境のシステムに進化してもオープンシステムとしての課題は解決されなかった。その中で「リコー 読取物語キャビネット」は一太郎・

Word・Excelといったメジャーなアプリケーションへのファイル出力が可能ではあるが、文書閲覧者においてこれらのアプリケーションを必ずしも有しているとは限らず、しかも長期間の文書保管が求められる行政文書においてこれらのアプリケーションが将来にわたってもメジャーでありつづけるとは限らず、数年後にこれらの文書ファイルを利用することが不可能という状況も考えられる。また「松下電器 ドキュメントファイリングシステム」においては、オプション機能において WWW ブラウザによる閲覧を可能としている。たしかにこれはマルチプラットフォーム (Macintosh、Windows) への対応につながっているが、やはり Netscape Navigator や Microsoft Internet Explorer 等のブラウザがメジャーでありつづけるという保証もないと考えられる。

つまり特定アプリケーションファイルへの出力という点は現状を考えると有効性は高いが、将来性については疑問が多い。そこで特定のアプリケーションに偏らないイメージデータの採用が求められる。しかもこれは単純なイメージデータファイルではなく、文字部を文字データとして認識できるファイルでなければ今回の目的に合致しない。

以上のような点から考慮すると Adobe によって新たなイメージデータファイル形式「PDF (Portable Document Format)」が最も有効であると考えられ、本システムにおいてはこの PDF をファイル形式として採用する。(ADOBE より PDF ファイルの閲覧ソフト「Acrobat Reader」が無償提供されており、イメージデータ分野においては、世界的にデファクトスタンダード化する動きがある。)

2. 2. 2 文書イメージデータ容量のコンパクト化

イメージデータファイルとはその内容が文字であろうが、図形であろうが、写真であろうがすべて同様のイメージデータとしてしか認識してくれない。そこでワープロ等の文書アプリケーションにおけるテキストのみの文書ファイルとイメージデータ化された文字物文書ファイルのデータ容量を比較すると数倍～数十倍の容量を後者は前者に比較して有してしまう。

この点を考慮して第2世代のイメージファイリングシステムはOCR機能を付加し、文字部分に関してデータ化を図り大幅なデータ容量の縮小を実現できた。しかしこのイメージファイルを、特定のワープロ等文書作成アプリケーションのデータ化を図る場合、やはり文書管理情報等のファイルが付加されてしまい、トータルとしてはデータ容量を増やしてしまう。

その点 PDF ファイルでは、ネットワーク上でのドキュメントデータ配信を目指して開発されたファイル形式であることから、図表・イラスト等の画像データがイメージデータの中に入ってきた場合にも、それを作成したアプリケーション上でのファイル容量よりも大幅にコンパクト化することが可能となる。

イメージファイリングシステムの命とは、少ない容量でいかに大量のページ登録が可能かという点にある。今後これらの点をまったく考慮しなくても良い超高容量記憶メディア (テラバイトクラス) が登場し、それらメディアへのアクセス時間が超高速

化すればこの点も意識する必要はないであろうが、その過渡期である現在においては、やはりいかにファイルをコンパクト化するかを十分に考慮しなくては、最適なイメージファイリングシステムの開発には至らないものと考えられる。

2. 3 PDF ファイル

2. 3. 1 PDF ファイル形式の特徴とポイント

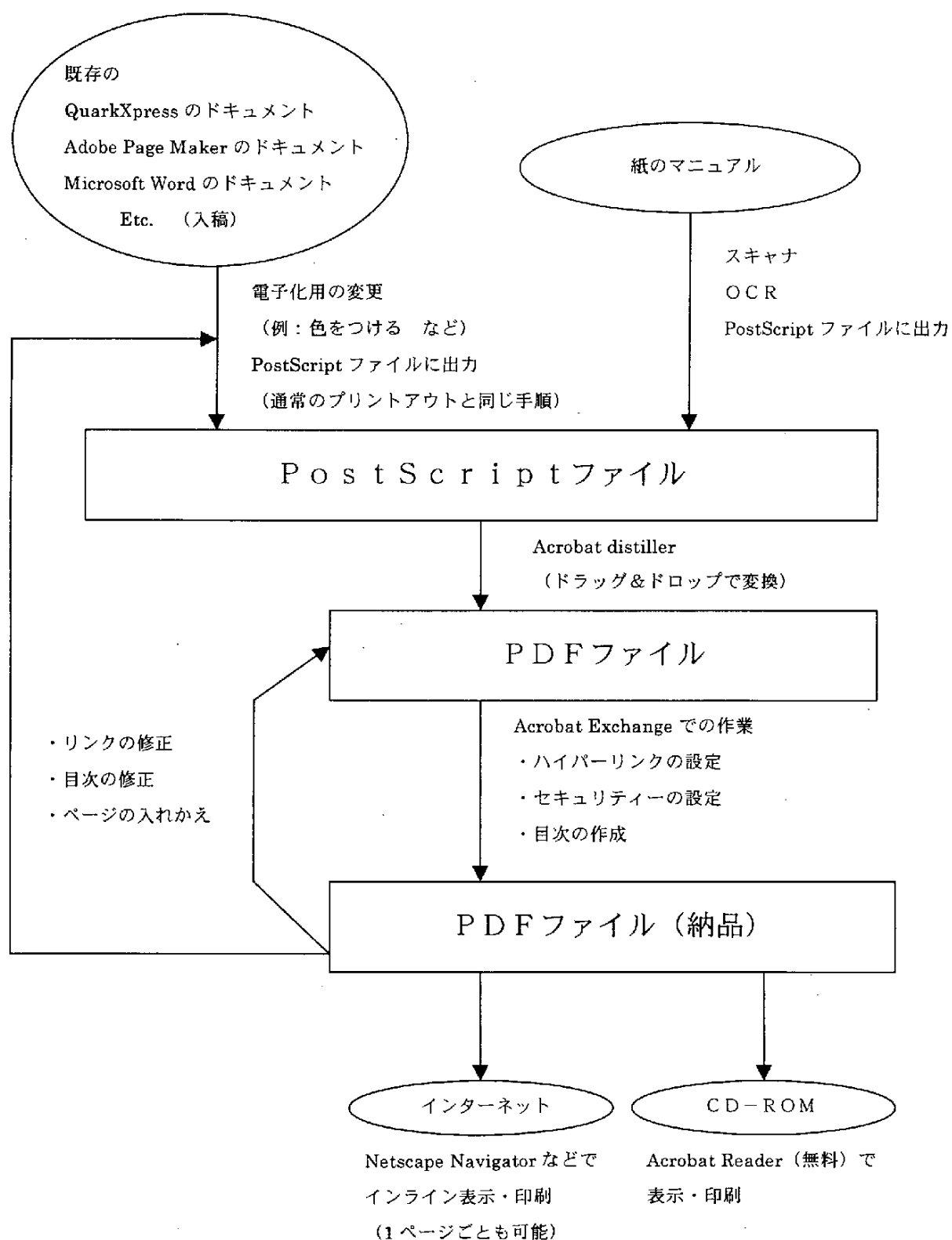
PDF (Portable Document Format) とは、Adobe Systems Incorporated が提唱し、一般に開示されている電子ファイル形式である。最大の特徴は「どこへ持っていても必ず同じように表示・印刷できる」点にある。

プラットフォームが異なっても、作成元アプリケーションがなくても、フォントがなくても、ディスプレイやプリンターが異なっても PDF ファイルを読み込むためのアプリケーション「Acrobat Reader」(Adobe のサイト等において無償配布) さえ有せば、オリジナルドキュメントと同じレイアウトで文書閲覧が可能となっている。

またこの PDF ファイルとは、インターネット環境が進展する中で特定のアプリケーションに頼らなくとも文書閲覧が可能という機能のみならず、いかにドキュメントファイルをコンパクト化し、短時間で容易に配信するかという点も考慮されている。

実際あるカラーパンフレットのイメージデータファイルを比較したところ、ある DTP アプリケーションで作成したデザインデータが数十メガバイトであったのに比較して、それを PDF ファイル化したところ 226 キロバイトまで圧縮することができたという事例もあげられている。

2. 3. 2 PDFによる電子化文書作成の流れ



(図) PDFによる電子化文書を作成する流れ

2. 4 イメージファイリングシステムの今後のあり方

以上のような点から考慮すると今後のイメージファイリングシステムに求められる機能とは次の点に集約される。

- ① 特定のアプリケーションに偏らないオープンなファイル形式の採用
- ② イメージデータ容量の小型化
- ③ 文字部分のデータとしての認識
- ④ 検索キーワードを付与しなくとも高速に検索可能な全文検索機能
- ⑤ 各文書の関連性を考慮した階層構造を持った文書データベース構造
- ⑥ 文字データのみデータ抽出が可能となり、他のアプリケーションによる2次加工可能
- ⑦ 文書データを構造文書化し、SGML 文書化することが可能

本システムとは①～⑤までの部分に関して実現するものである。この点までであっても現在市販されるイメージファイリングシステムにはない新たなコンセプトのシステムと評価できるものとする。

なお⑥及び⑦を実現するためには、⑥に関しては文字認識を持った PDF ファイルから文字部分を抽出して TXT ファイルへ出力させることで対応可能である。しかし⑥のみでは、レイアウトイメージが損なわれてしまうという点とプレーンなテキストファイルとなってしまう点から文書構造が不明確となってしまうという問題点が発生する。前者に関しては、元のアプリケーションによって作成された文書ファイルに立ち戻って手をつけることがベストであるが、とにかく紙でしか存在しないファイルに関してはレイアウト指示を誰かが与えなくてはならない。また後者に関しても、レイアウト文書化なされた時点で文書ファイル上、文書構造は不明確になってしまっている。つまりこれもシステムの問題ではなく、やはり文書構造指示を誰かが与えなくてはならない。(また一定のレイアウト文書であれば、そのレイアウト規程を分析し、ある特定のアプリケーションへの出力機能、SGML ファイルへの出力機能に関して自動変換プログラムを作成し対応することは可能である。しかしこの場合1枚でも規程外のレイアウト文書があれば、それは手作業にて対応しなければならない。)

つまり各行政機関において予算等の都合により、閉鎖的環境の過去のワープロ専用機を利用し、とりあえず紙による文書交換で進めなければならないのであれば、このような自動変換プログラムをもって一括変換できるよう、文書作成規程を再度徹底させ、同様の文書構造となる文書作成を末端の担当者に行わさせる努力を組織的に行うべきである。このような対応が無ければ、⑦とはシステム開発力のみで対応できる課題ではないと考えられる。

3 文書ファイル形式として SGML と PDF の対比

いま日本国内では、文書ファイル形式に対する意見が二分されている。一つは文書構造のみを規程し、レイアウト属性を持たない非常にプレーンな文書交換性や文書データベースやマルチメディア化への展開が容易な「SGML (Standard Generalized Markup Language)」であり、もう一つは今回テーマにする「PDF (Portable Document Format)」である。

双方の共通点は特定のアプリケーションに偏らないファイル形式であるという点であるが、SGML は文書構造のみを規程し、レイアウト情報を持たないドキュメントファイルであり、PDF はその逆にレイアウト情報のみを持ち、文書構造を規程しないドキュメントファイルである。

そこで本章では、この双方の特性を分析し、行政文書分野における双方の役割を明確化することとする。

3. 1 SGML とは何か？

3. 1. 1 SQL データベースから SGML データベースへの変遷

これまでの文書データ管理技術は、文書アプリケーション毎の閉鎖的環境の中で進展してきた。各々のアプリケーションが機能によって勝負し、自らのアプリケーションへの囲い込みを目的にファイル形式を異なる形態とした完全な相互交換が不可能な状況にあった。これは各文書アプリケーションの創成期には、自らのシェア獲得とマーケットの寡占化を目的に仕方のないことではあるが、これで最も苦労したのが我々ユーザーサイドではなかっただろうか。(各アプリケーション間でのデータ変換プログラムも各サードパーティから発売されているが、詳細な部分でレイアウト機能が異なることから、完全なイメージでのデータ互換は実現できていない。)

しかも1つのアプリケーションに頼ることは文書保管期間が長期間となる行政機関にとって大変危険なことである。例えば10年間の文書保管が必要な文書についてWORD や一太郎等の特定のアプリケーションで文書作成した場合、本当に10年後もその文書を電子的にファイル閲覧や加工が可能であるのかという問題である。つまりそのアプリケーションが10年後もメジャーであり続けるという保証はどこにもないのである。数年前、Windows の台頭と1人1台パソコン化の流れを受けて、多くの自治体においてワープロ専用機からパソコン上のワープロソフトへ文書作成アプリケーションが移行するケースが増加した。この際に発生した問題点も同様である。この際も付加作業として多くの予算を割いたものがワープロ専用機ファイルからパソコンワープロファイルへのファイル変換作業である。すべての業務は行政文書によって進められるという原則から考慮するとこの作業も致し方ないことであるが、今後もメジャーなワープロソフトが変化するとこのような作業を行っていたのでは、多くの無駄な予算が必要となってしまう。

そこでデータ交換という点のみを考えると、これまでは MS-DOS テキストファイ

ルによるデータ交換が主体となっていたが、これはあくまでもレイアウト情報を持たないプレーンテキストであり、データ交換は大変容易ではあるがレイアウトや文書構造定義がファイルの中身には含まれないことから、文書作成者の意図がその文書ファイルを得た第3者に完全に伝わらないという課題を持っていることとなる。

その後 RTF (Rich Text Format) ファイルという書式付き文書ファイルが定義され、異なるアプリケーション間での文書データ互換が可能とはなってきたが、これはパソコン用ワープロソフトの間での文書互換が主体となっており、ワープロ専用機は取り残された形となっており、これにおいてもやはり完全なものであるということは言い切れない。

そこでアプリケーション固有のレイアウト情報を省き、「章タイトル行」「作者名行」「日付行」などといった文書構造を示す識別情報を「タグ」という印で示し、文書構造情報が盛り込まれたテキストファイルとして生まれたものが「SGML」である。この SGML 文書ファイルを作成する場合、SGML エディター等を利用して文書作成を行うが、この際にほとんどのエディターの機能には必ずデータ入力が必要な項目（行）やある行を入力した際に次に出現する入力項目（行）が規定されていることから、誰もが同じ文書構造の文書作成を行うことができ（文書標準化の実現）、データ交換を行う際にもその SGML 文書作成の約束事となる「SGML 宣言」とその文書の文書構造を示す「DTD (Document Type Definition)」を SGML 文書ファイル (SGML 文書インスタンス) に付しておけば、互いに同じ環境でその文書データを取り扱うことが可能となる。

ただ例えば SGML 文書を紙に出力する場合には、レイアウト情報が付加されていないことから、組版指示を行う別アプリケーションが必要となる。(Adobe から 1998 年 2 月に日本語版が発表される DTP アプリケーション「Frame Maker + SGML」を利用すると DTP ファイルと SGML ファイルがリンクしており、各々を修正すると双方のファイルとも自動的に修正され、SGML とレイアウト情報の親和性が高まるものとなった。)

しかし SGML により、文書情報データベース用のファイルと各種メディア出力用のファイルが同一のファイルとしてストックできることから、その活用の幅は広がり、SGML とは非常に自由度の高いファイル形式であると考えられる。

そこでこれにより既存の SQL 形式 (RDB) では難しいと考えられていた文書データベース構築も、SGML によって解決できるものとなった。SQL データベースもデータ項目の非固定長化や並列 SQL 処理による高速検索性などにより、そのデータベースアプリケーション上では高速処理を行うことも可能となったが、アプリケーション間データ互換やエディター等を利用した簡易な文書編集を行うことができない点など課題点は残されている状況にある。

情報システムオープン化の中で、容易にデータ交換ができ、世界的な文書管理体系として ISO 規格化がなされている「SGML」は今後の文書管理の標準として、メジャーなものとなっていくべきと考えられる。

3. 1. 2 SGML とは何か？

SGML とは「SGML 宣言」「DTD」「文書インスタンス」の要素から構成されている。SGML 宣言とはどのような SGML 文書であるかを指示する宣言文であり、DTD とは文書構造を示す採用されるタグ（印）の規則であり、文書インスタンスとはタグが付与された文書ファイルである。

(A) SGML 宣言

SGML 宣言とは、その SGML 文書の内容条件を示したものであり、次の 6 項目で構成される。

- ① Charset : 使用している文書文字集合の記述 (ex.ISO646IRV)
- ② Capacity : 許容できる文書の大きさなど文書容量の指定
- ③ Scope : Syntax の適用される範囲
- ④ Syntax : 具象構文の詳細に関する指定
- ⑤ Features : 使用している SGML 機能指定
- ⑥ Appinfo : アプリケーション固有情報

```
<!SGML"ISO 8879-1986"
  CHARSET          BASESET "ISO 646-1983//CHARSET International
                        Reference Version(IRV)//ESC 2/5 4/0"
                        DISCSET 0      0      UNUSED
                        9      2      9
                        11     2      UNUSED
                        13     1      13
                        14     18     UNUSED
                        32     95     32
                        127    1      UNUSED
CAPACITY          PUBLIC "ISO 8879-1986//CAPACITY Reference//EN"
SCOPE              DOCUMENT
SYNTAX            PUBLIC "ISO 8879-1986//SYNTAX Reference//EN"
FEATURE           MINIMIZE DATATAG NO  OMITTAG YES
                        RANK  NO  SHORTTAG YES
                        LINK  SIMPLE NO IMPLICIT NO EXPLICIT NO
                        OTHER  CONCUR NO SUBDOC NO FORMAL NO
APPINFO           NONE
>
```

【SGML 宣言サンプル (ISO08879)】

これら宣言により、使用されるコード種や要素名などの最大文字数、タグマーク (ex.<>)、SGML が提供する機能をどの程度活用して文書インスタンスが作られているのかなどが示されている。つまり文書交換を行う際に使用しているコード系などを認識することができることが SGML 宣言の目的である。

なお ISO では、標準 SGML 宣言として「ISO8879」の中で定義されたものも存在

し、多くの場合この標準的な SGML 宣言を利用している。

(B) DTD (Document Type Definition／文書型定義)

DTD によって定義されるものが文書構造である。文書インスタンスの中で構成される文書タグ (印) を定義し、その文書がどのような文書構造 (タグ) によって構成されるものであるのかが示されている。

また DTD とは、異なる文書種類によって各々の DTD が定義されている。(論文用・マニュアル用など、そして行政分野では「白書・報告書用」の DTD が定義されている。) SGML の目的が文書交換にあることから各々が勝手な DTD を採用することはあまり意味がない。そのため目的とする文書分野や業界において標準 DTD が定義されている場合にはそれを採用するべきであるし、DTD が存在しない文書分野においてはこれら標準 DTD を参考として自ら文書構造分析を行い、DTD 設計を行う必要がある。

(C) 文書インスタンス

一般的に SGML 文書と呼んだ場合、この文書インスタンスをそれは示している。つまり DTD によって定義された文書型にしたがって作成されたタグ付き文書が文書インスタンスである。

```
<report>
<head>
<hd>営業報告書</hd>
<date>平成9年9月1日</date>
<to>〇〇営業部長殿</to>
<author>営業企画部企画2課  △△</author>
</head>
<body>
<p>新規開拓した得意先につきまして、その打合内容をご報告致します。</p>
<sl>
<li>顧客名    □□株式会社</li>
<li>業種      電化製品メーカー</li>
<li>訪問日時   1997年8月30日  15:00~16:00</li>
<li>訪問先住所 東京都文京区●●町 1-2-3</li>
<li>顧客担当者名  販売企画部  ☆☆太郎氏</li>
<li>打合テーマ   製品カタログ制作に関して</li>
</sl>
<p>既存の印刷物による製品カタログ(約200P)とともに、同様の内容をベースとした CD-ROM
カタログを制作したいとの要望であった。そこでこれまでの弊社の「印刷+CD-ROM」によるカ
タログ制作事例を示し、その制作方法を提案した。</p>
</body>
<close>以上</close>
</report>
```

【SGML 文書インスタンス サンプル】

営業報告書

平成9年9月1日

〇〇営業部長殿

営業企画部企画2課 △△

新規開拓した得意先につきまして、その打合内容をご報告致します。

顧客名	□□株式会社
業種	電化製品メーカー
訪問日時	1997年8月30日 15:00~16:00
訪問先住所	東京都文京区●●町 1-2-3
顧客担当者名	販売企画部 ☆☆太郎氏
打合テーマ	製品カタログ制作に関して

既存の印刷物による製品カタログ（約 200P）とともに、同様の内容をベースとした CD-ROM カタログを制作したいとの要望であった。そこでこれまでの弊社の「印刷+CD-ROM」によるカタログ制作事例を示し、その制作方法を提案した。

以上

※ただし、SGML 文書構造 (DTD) の種類毎に組版指示をした結果の出力である。

【SGML 文書インスタンスの組版例】

3. 2 PDF とは何か？

3. 2. 1 イメージデータのデファクトスタンダードとしての PDF

PDF とはインターネットの発展の中で生まれてきたものであるといっても過言ではない。WWW ブラウザが登場し、HTML によってホームページデザインを行う際に、そのデザイン的な表現の限界を克服したいという点や印刷用にデザインされた DTP データをそのままインターネットを通じて提供したいという点などデザイナーサイドの要望によって生まれてきたものがこの PDF である。

ワープロ・DTP アプリケーションでデザインされたイメージをそのまま再現できる点とファイルサイズが小さくネットワーク配信に有利という点については前述したが、もう一つ特筆すべき利点がある。それはマルチメディア対応であり、インタラクティブであるという点である。

PDF とは容易にいろいろな種類のファイルとリンクを貼ることができる。例えば画像や音声ファイル、そしてメモファイルもリンクさせることができる。よって自治体におけるホームページを作成する場合にも、「祭り」を紹介する場合、静止画写真とともに、祭り風景の音声や動画などをリンクさせ、より臨場感あふれるものにすることができる。（ただし音声・動画の再生用プログラムは必要となる。）

また文書の中に「フォーム」を作することもできる。例えばアメリカの自治体では市

民からの各種申請様式をインターネットにて配信し、それを得た市民は必要項目を入力し電子申請を行うことができることが実用化されている。

例えば報告書を制作する場合、印刷物と共に、PDF ファイルによる電子ファイル化を行うことにより、目次から該当する章や内容にリンクすることはもちろん、該当する内容を紹介するインターネット URL を付しておけば、ホームページへリンクすることも可能であり、内容に関する問合せや質問もフォーム機能によってインターネットを通じて自動的に受け付けることも可能となる。つまり平面的な文書ファイルではなく、マルチメディアでインタラクティブな電子本を作成するという可能性を秘めているものである。

他にこれに類するイメージファイル形式が存在しない現在、フリーウェア閲覧用アプリケーションが供給されていることから考えても、PDF ファイルはイメージデータファイルのデファクトスタンダードといっても過言でないものといえるであろう。

3. 2. 2 PDF ファイルに関わる関連ツール

PDF ファイルに関わるツールは、このファイルの提唱者である Adobe の「Acrobat」というアプリケーションにより提供されている。(Windows 及び Macintosh) その構成は次のとおりである。

- | |
|---|
| <ul style="list-style-type: none">① Acrobat Reader : PDF ファイル閲覧用アプリケーション (フリーウェアとして流通)② Acrobat Exchange : PDF ファイルの編集やリンクの追加等、PDF 特有の機能を付加するアプリケーション③ Acrobat Distiller : PostScript ファイルの PDF ファイル変換アプリケーション④ Acrobat PDFWriter : プリンタドライバにより各アプリケーションから PDF ファイルを出力するアプリケーション |
|---|

PDF ファイルを作成する場合、この「Acrobat」というアプリケーションがなくては実現できないが、他の環境と大きく違う点とは閲覧用の Acrobat Reader がフリーソフトウェアとして無料で自由に入手できるという点である。Netscape Navigator などの WWW ブラウザが世界的に普及したこともこのフリーソフトウェアとして流通した点にあった。そこで Acrobat Reader も同様に普及するものと考えられ、この点からも PDF ファイルがデファクトスタンダード化するものと考えられる。

3. 3 SGML と PDF の対比

ここまで双方の機能や特徴を紹介する中で、この双方は対極的に対比すべきものではないという点についてご理解いただけるであろうか。SGML とは文書データベースを構築するための文書データそのものの管理体系であり、PDF とはその文書データをデザイン加工した後のメディア出力用ファイル形式の1つである。

つまり最良の文書管理体系とは、SGML によって文書データを作成し、文書データベースを構築し、それを目的に応じたデータのみを抽出し、受け手側に理解されやすいようにデザイン処理をあるデザイン・DTP アプリケーションによって行い、それを誰もが閲覧できるように PDF ファイルへ変換し、電子配信を行うという流れとなる。

しかし紙文書から電子文書への業務変革の移行期である現在、このような最良の流れに業務を移行することは難しいことから、それぞれの特性を理解した上で、次のようなすみ分けを行うべきであると考えられる。

3. 3. 1 行政文書分野における SGML 化すべき点

閣議決定された「行政情報化推進基本計画」において、今後の行政文書はすべての行政機関で交換が容易にできるよう SGML 化すべきであると定義されている。

【行政情報化推進基本計画の要旨】（平成6年12月閣議決定／平成9年12月改定）

① 行政情報化の推進

● 行政情報の提供

- ・報道発表資料等の行政情報に関してインターネット発信を推進し、国民からの意見等もインターネット受付を実施
- ・白書／年次報告書等の行政情報に関してインターネット／CD-ROM 等の電子媒体での提供を推進
- ・各種統計情報の電子媒体での提供
- ・行政情報の提供促進を目指す「クリアリングシステム（所在案内システム）」の整備
- 申請・届出等手続きの電子化（24時間対応・アクセスポイント拡大）
- 総合行政サービスシステムの実現（ワンストップサービス）
- ・インターネットを利用した各種行政手続の実現（行政手続案内、申請様式提供など）
- ・電子的アクセスポイントの設置（ex. 郵便局）
- ・特定分野の手続に関して、関係機関とのオンライン化
- 調達手続（入札等）、歳入歳出事務の電子化（国と民間との決裁手続の電子化）

② 事務・事業の簡素化・効率化、行政運営の高度化（りん議決裁システム・行政情報DB）

● 文書管理システムの整備（文書のライフサイクル・文書間のリンクを含む）

- ・省庁内の文書管理システム化
- ・省庁間の文書交換（本省庁／地方部局／特殊法人／地方公共団体間も含む）

● 情報共有化の推進

- ・法令／通達、白書／年次報告書、統計情報などのデータベース化
- ・政策支援データベース等既存データベースの LAN 活用
- ・クリアリング機能（文書所在案内）の整備

● LAN による業務効率化

- ・グループウェアの活用による共同作業化
- ・情報通信技術活用によるサテライトオフィス勤務、在宅勤務に関する調査研究

● 各種許認可等に係るシステム間の連携、高度化（内部管理業務の簡素化・効率化）

● 情報システムの運営管理に関する民間への外部委託（アウトソーシング）

③ 行政情報化推進のための基盤整備

- 庁内 LAN 整備、霞ヶ関 WAN の活用による総合的・広域的行政情報ネットワーク整備
- オープンシステム化の推進（汎用機→クライアントサーバーシステム）
- 個人情報保護などセキュリティの充実
- 通信プロトコル（TCP/IP）の採用、電子行政文書の SGML による標準化推進
- 申請・届出等の電子化手続を円滑とする共通的な記載事項の標準化
- 省庁内情報化推進体制の確立
- 電子文書の原本性、申請・手続等の個人・法人認証、手続オンライン実施のための手数料納付方法に関する課題解決（IC カード／印紙の代用方法・電子キャッシュ）

そこでこの基本計画にもあるとおり、法令・通達、公報・広報、白書・年次報告書、統計書といった行政情報をオープンにする種類の文書に関して、今後文書作成するものについては SGML 化すべきである。ただ単純にこれらの原稿データのみを SGML 化することではこの基本計画の目的である根本的な行政改革にはつながらない。つまりこれらをまとめる経過となる検討過程の文書（各自の意見や議事録等）に関しても SGML 化を図らなければならない。

既存の紙文書社会では、議論の過程を見直し、最終的に発表する行政情報を年度末の時期に苦勞して取りまとめるというアナログ的な作業を行ってきた。しかし電子ネットワーク時代の業務のあり方とは、その議論の過程がデジタルコンテンツとして蓄積され、最終的な報告書とは、その蓄積されたデジタルコンテンツが自動的にまとめあげられ、年度末時期に苦勞しなくとも出力されるという姿でなくては意味がないのである。

そこで単に行政文書を SGML 化するだけではなく、その文書を整理・管理するための文書管理システムのあり方も併せて考慮し、特定の担当者に頼らなくとも、すべての業務成果が SGML 文書管理システムの中に表現されるような管理体制を整備しなくてはならない。

また SGML 文書ファイルを制作する際、既存の SGML エディターでは、タグが見えるなどワープロソフトのようにレイアウトを確認しながらの文書作成作業ができないことから、文書作成種類ごとに専用の入力フォーマットをテンプレート化して文書作成者に配布するか、前述の「Frame Maker + SGML」のような SGML との親和性に富んだ DTP ソフトによる文書作成を行うことをお勧めする。（この際にも文書フォーマットテンプレートを付与すべきである。）

3. 3. 2 行政文書分野における PDF 化すべき点

PDF 化すべき文書分野とは、すでに SGML 的な意識を持たずに制作された過去の行政文書分野である。

人間がワープロ等によって文書作成する場合、次の 2 つの判断をおこなっている。その 1 つは「この行はタイトル行である」「この行は日付行である」という文書構造に関する判断であり、もう 1 つは「タイトル行は目立たせるようにセンタリングする」「日付行は右詰する」という文書レイアウトに関する判断である。この中でワープロ

文書というのは結果として後者の文書レイアウトに関する判断の結果しか見て取ることはできない。

つまりワープロ文書とは、それを基に誰かしらが文書構造に関する判断を加えなくては、SGML 文書に戻すことは不可能である。また文書作成者が明確なものは、その人に文書構造判断をさせれば SGML 化も可能であろうが、見た目では文書構造が判断できないもので、しかも 10 年前に作成された文書が 1000 頁もあるとすれば SGML 化は多額の費用をかけなくては不可能なものである。

過去の文書をこれから構築される文書管理システムのファイル形式に合わせて SGML 化する。しかし今後利用する機会があまりないと考えられる文書まで、このような多額な文書構造判断コストをかけることは無意味である。そこでこのような文書は PDF でレイアウト情報も付加されたまま管理しておくことが得策であろう。

また情報化移行期の現在、電子文書化が進んでいるとはいわれているものの、中央省庁や関係機関からの通達文書は紙文書のものがまだまだ多い、また新聞・雑誌のように紙でしか存在しない資料文書、そしてワープロファイルが既に存在しないワープロ文書出力などに関しても使用頻度や重要度が高いもの以外は PDF で十分であろうと考えられる。(これらのような文書分野を今回のイメージファイリングシステムのターゲットと考える。)

そして PDF の利用範囲はこのような文書ファイリング分野に限らず、行政文書の電子配信の分野でも、その利用分野は期待できる。SGML とは前述のとおりレイアウト情報を有していない。つまりデザイン処理を行い、市民にとって理解しやすいレイアウトを DTP やワープロ等で行った場合、すべての人がそのアプリケーションを有していないと考えられることから、それを PDF 変換して配信することは大きな意味を持っている。

4 本事業におけるイメージファイリングシステムの特徴と活用方法

4. 1 本イメージファイリングシステムの特徴

3章において述べたとおり、本イメージファイリングシステムのターゲットとは、紙でしか存在しない文書の電子文書化にある。しかもそれを特定の専用システム・アプリケーションの中でのみ利用できるだけではなく、インターネットやイントラネット等の環境においてマルチプラットフォームで移用できることを目標としている。

そこで次のような特徴を有したシステムとなっている。

4. 1. 1 PDF ファイル形式の採用

マルチプラットフォーム化と PDF ファイルのデファクトスタンダード化の流れを考慮すると、PDF ファイル以外のファイル形式は考えられない。ファイル圧縮率の面も含め、電子配信まで考慮しても最適なファイル形式であるといえるであろう。

4. 1. 2 PDF 対応文書管理システムの採用

これまでの文書管理システムとは、大別すると次の2種のものしか存在しなかった。

① 全文検索エンジンをベースとしたプレーンな TXT ファイルによる文書管理システム

文書構造や文書レイアウトを規定する属性を文書ファイルに含まず、キーワードを指示することによりそれに該当する語句を含む文書箇所を抽出・閲覧できる文書管理システム（議会の会議録検索システム等／弊社製品「Voices」）

② SGML 文書等に対応する文書構造まで管理可能な文書管理システム

SGML 文書等に対応した文書管理システムであり、例えば「文書作成者が〇〇氏の文書」「タイトル行に△△が含まれる文書」といった①よりも詳細な文書検索指示が可能であり、しかも文書閲覧者ごとに閲覧可能な文書構造の階層範囲が指示できるなど文書構造機能を十分に生かした発展的文書管理システム

一方、PDF ファイルによる文書管理システム・全文検索エンジンはこれまで存在しなかった。しかしインターネットやイントラネット上における文書管理が進展する中で、NTT アドバンステクノロジーより PDF 対応文書管理システム「PDF Manager (UNIX/Windows NT)」が発売され、これにより PDF ファイルをベースとした文書管理システムの可能性が多いに広がったことから、今回はこのシステムコンセプトをベースとして行政文書分野における PDF イメージファイリングシステムの開発に着手することとした。

レイアウト情報を有す PDF ファイルであることから本イメージファイリングシス

テムは上記の①と②の中間に位置する機能を有するものであると考えられる。ユーザーがWWWブラウザを利用してインターネットへアクセスするような感覚で気軽に利用できるという「PDF Manager」の主な機能とは次のとおりである。

(A) 文書ファイリング機能

業務で発生する各種文書データを PDF ファイル化し、任意の作成する階層構造とそ
の中のフォルダで管理することが可能。(フォルダに関しては運用体系や業務形態に
合わせてユーザーが任意に設定可能)

(B) アクセス権限管理機能

フォルダ単位、ファイル単位にユーザーに対するアクセス権限が設定でき、業務や運
用にあわせ、ユーザーごとやグループごとのアクセス権限が付与できる。

(C) 文書履歴管理機能 (バージョン管理)

文書データに対する変更、更新情報をバージョン及びリビジョンにて管理する機能を
有しており、最新の文書データを利用することはもちろん、古いバージョンの文書の
参照や文書履歴に応じた文書内容の変化に関して閲覧することも可能。

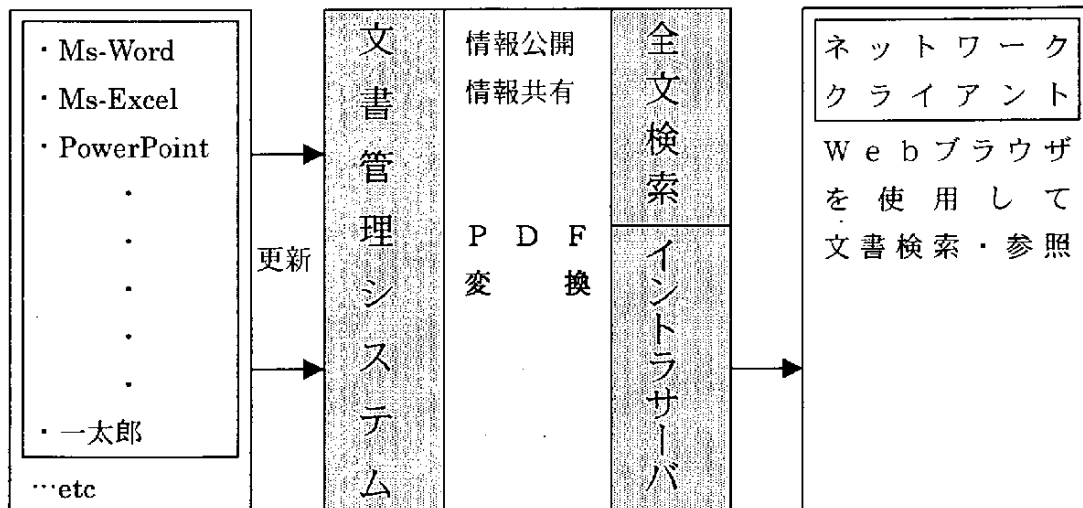
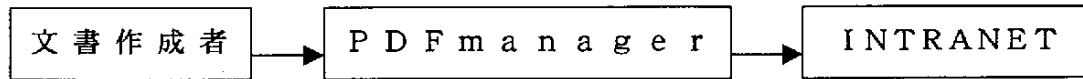
(D) ワークフロー及びイベント通信機能

文書の共同編集や承認作業を行う場合のワークフロー機能を有しており、また文書の
更新・削除等が発生した場合の任意ユーザーに対して自動通知発信を行うイベント通
信機能も有している。

(E) イン트라ネットシステムとしての機能

フォルダに登録された PDF 文書ファイルは特別な作業を必要とせずにインターネッ
ト・イントラネット環境で検索・参照・ダウンロードが可能。これにより WEB サイ
トでの管理が不要となり、システム運用コストが削減できる。また Acrobat による
PDF 自動生成機能を有することから各アプリケーションで作成・編集された文書を
PDF 変換することもでき、その他全文検索機能による必要文書の抽出や文書ファイ
ルに対して行われた参照・更新・削除等のアクション履歴に関しても管理することも
可能。

PDFmanagerの概要

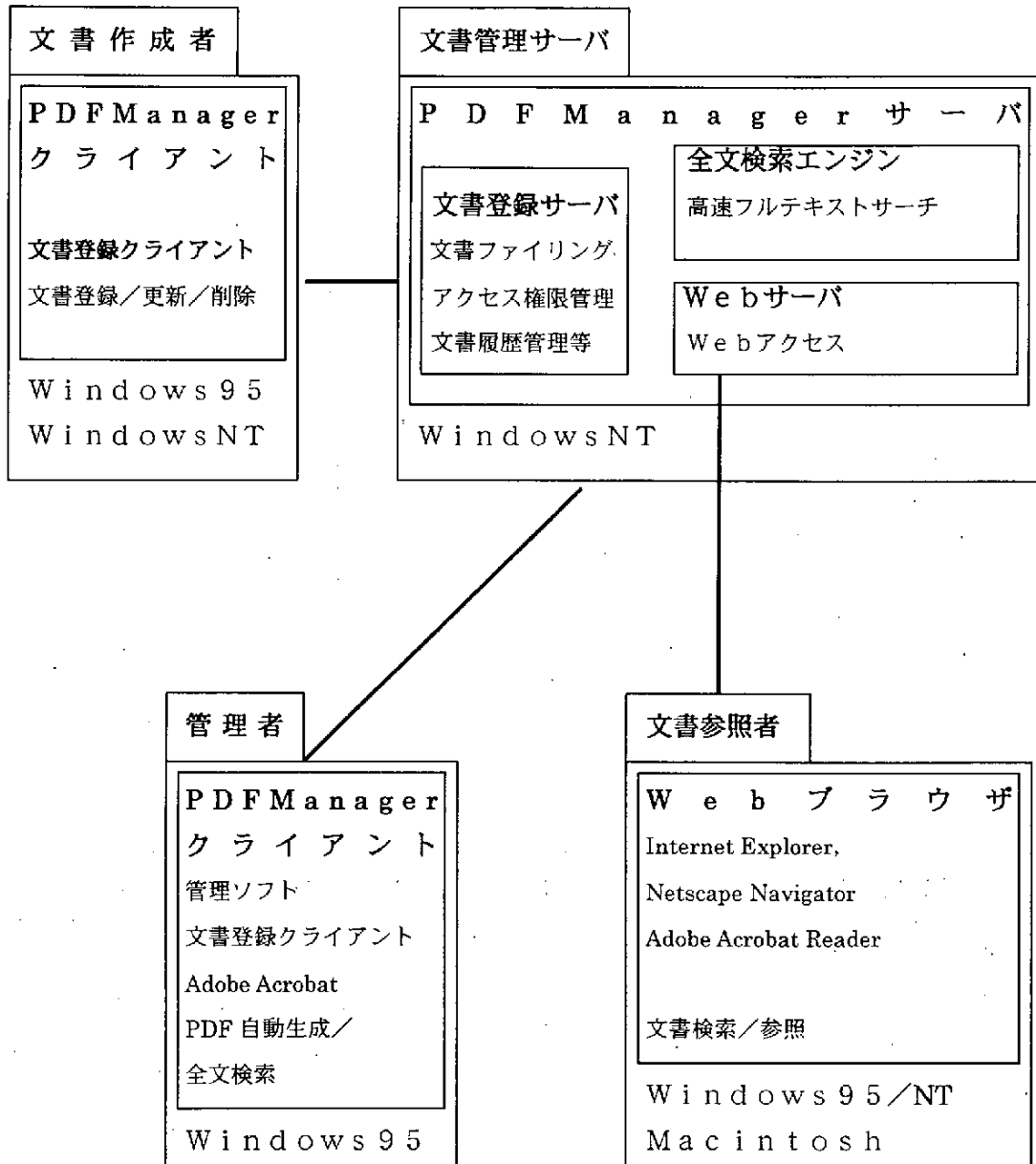


文書管理システムとして
文書ファイリング機能 アクセス権限管理機能 文書履歴管理機能（バージョン管理） ワークフローおよびイベント通知機能

イントラネットシステムとして
・Webアクセス機能 ・PDF（Portable Document Format）自動生成機能 ・高速テキストサーチ機能 ・アクセス履歴管理機能

（図）PDF Manager の概要

PDFmanagerシステム図



(図) PDF Manager システム図

4. 1. 3 本イメージファイリングシステムの活用・運用分野

本システムは PDF Manager のシステム概要からもわかるように、まずは SGML 化までの過渡期として発生するワープロ等の文書アプリケーションによって制作された文書ファイルの文書管理に対応可能である。そして今回は OCR 機能と合体させることから紙文書として既に存在する過去の文書を PDF ファイル化して管理するという分野まで対応分野を広げることができる。

例えばその行政文書分野とは次のような内容である。

(A) 自治体内部で過去に作成され、既にファイルが存在していない文書

ネットワーク型パソコン管理体系になる以前は、フロッピーディスクによる文書管理が主流となっていた。しかも行政文書は紙文書として保管するという規定しか示されていなかったことから、これら文書データは軽視され、異動時に引き継がれていないケースが多い。そこでこれら紙文書の再電子データ化がこの分野である。

(B) 中央省庁から自治体に提供される通達・告示文書

「行政情報化推進基本計画」も中央省庁からその整備を始めるとはされているが、ある一部分の文書分野のみしかそれらは進展しておらず、地方自治体に対する通達・告示文書に関しては、その多くが未だに紙文書によって伝えられている。しかしこれら文書を基として各自治体では自らの政策や法令を整備することから、自治体内部の文書が電子文書化される中で、この中央省庁からの文書が紙のままでは支障をきたす。そこでこの分野においても対象と考えられる。

(C) 新聞・雑誌等から抽出する資料文書

自らの自治体に関して紹介された新聞・雑誌記事や行政運営上参考となる新聞・雑誌記事などについても、その後の活用や整理・保管を考慮すると電子的なファイリングが必要不可欠である。(これまでの行政機関におけるイメージファイリングシステムの利用分野はこの分野が主体であった。) そこで単純なイメージデータとしての保管のみではなく、その文書内容をデータ化して文書データ検索が容易な形態とすることや記事文書データと行政内部文書のリンクを実現することは有効であると考えられる。

5. イメージファイリングシステムの概要

5. 1 本システムの特徴

本システムでは、スキャナから取り込んだイメージファイルを OCR でテキスト化し、全文検索を行う。全文検索の長所として、思いついた任意の用語で検索することができるが、短所として、意味は同じでも資料が異表記で書かれていた場合や、OCR で誤認識された場合に検索漏れが発生してしまうという短所もある。この短所を考慮し、本システムでは同義・類義語を集めたシソーラス辞書、用語の一致度が指定できるあいまい検索を取り入れる事で、漏れの少ない高速全文検索が実現される。

5. 2 本システムの機能

本システムは以下の3つ機能に分けられる。

(1) イメージ登録

- ・ スキャナから資料を読み込み、PDF ファイルを作成する。
- ・ OCR を利用し、テキスト化を行う。

(2) 文書登録

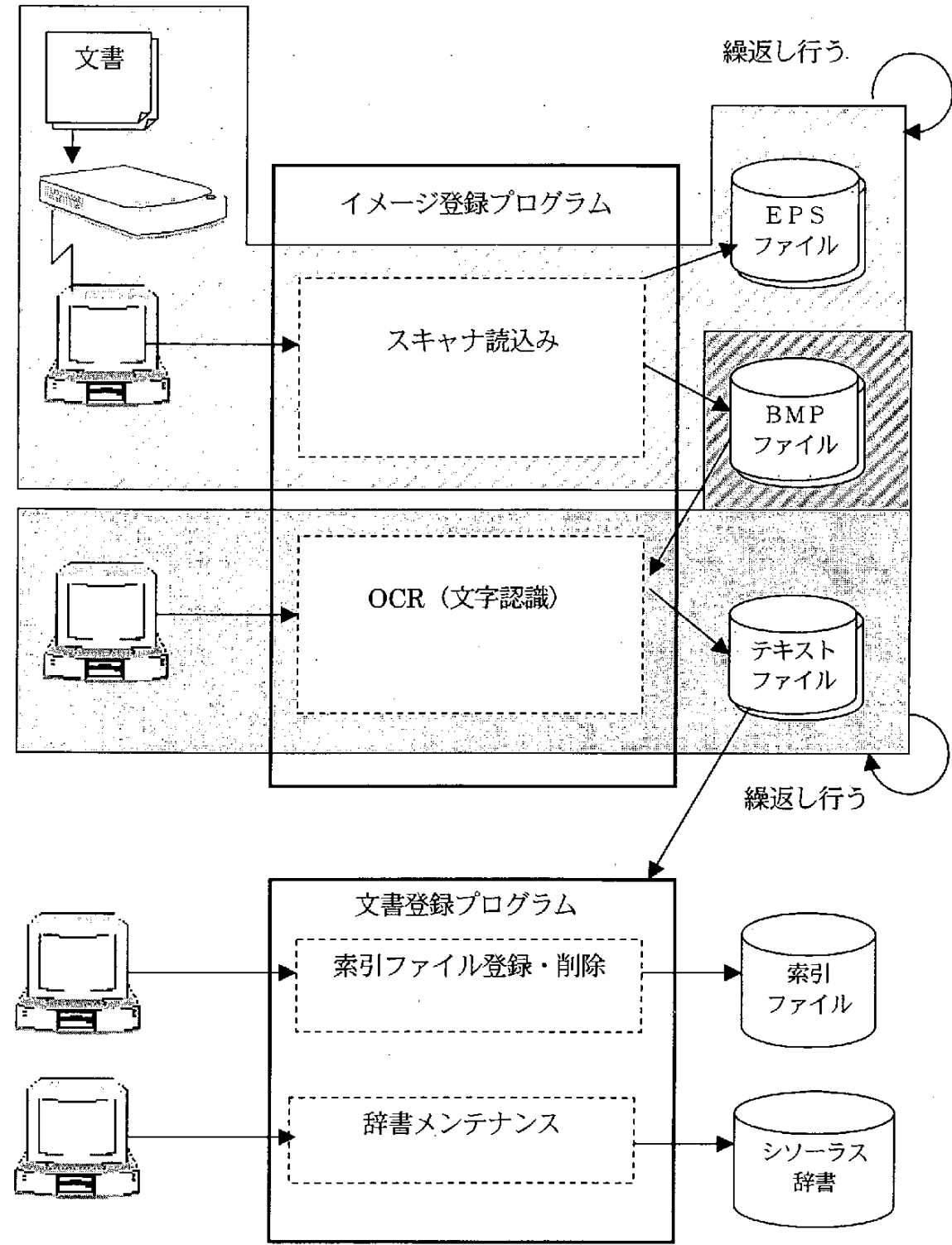
- ・ テキスト化した文書を、高速検索する為の索引ファイルへ登録を行う。
- ・ 不要データを索引ファイルから削除する。
- ・ シソーラス辞書の不要語設定、ユーザ独自の辞書の構築を行う。

(3) 検索

- ・ 任意のキーワードで全文検索を行い、結果を取得する。
- ・ 分類別に資料を閲覧する。
- ・ Acrobat Reader を利用し、PDF ファイルを表示する。

5. 3 システム全体の流れ

5. 3. 1 登録処理

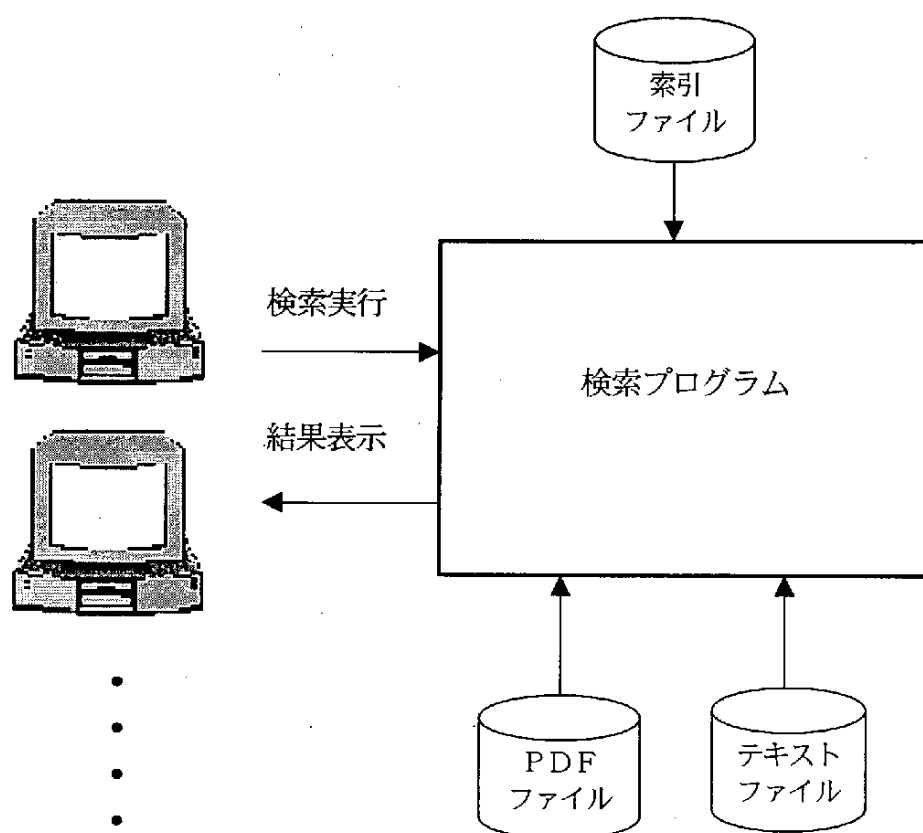


(図) システム全体の流れ



(図) PDF ファイル変換

5. 3. 2 検索処理



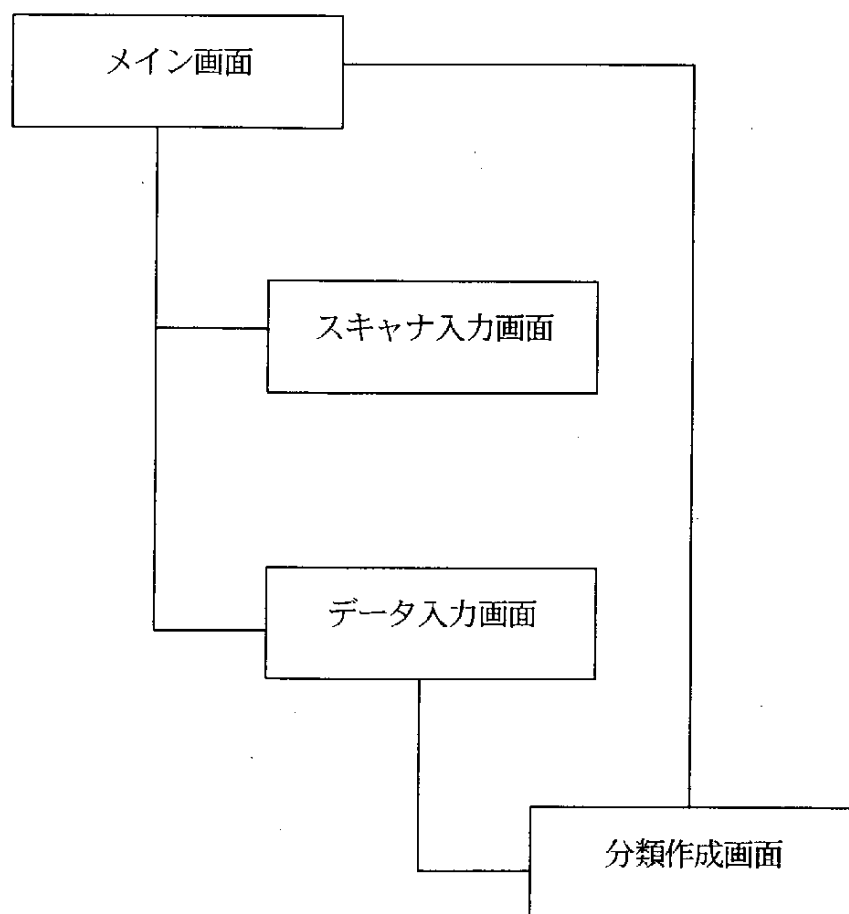
※ PDF 表示には Adobe Acrobat Reader を使用

(図) 検索処理フロー

6 イメージファイリングシステムの機能

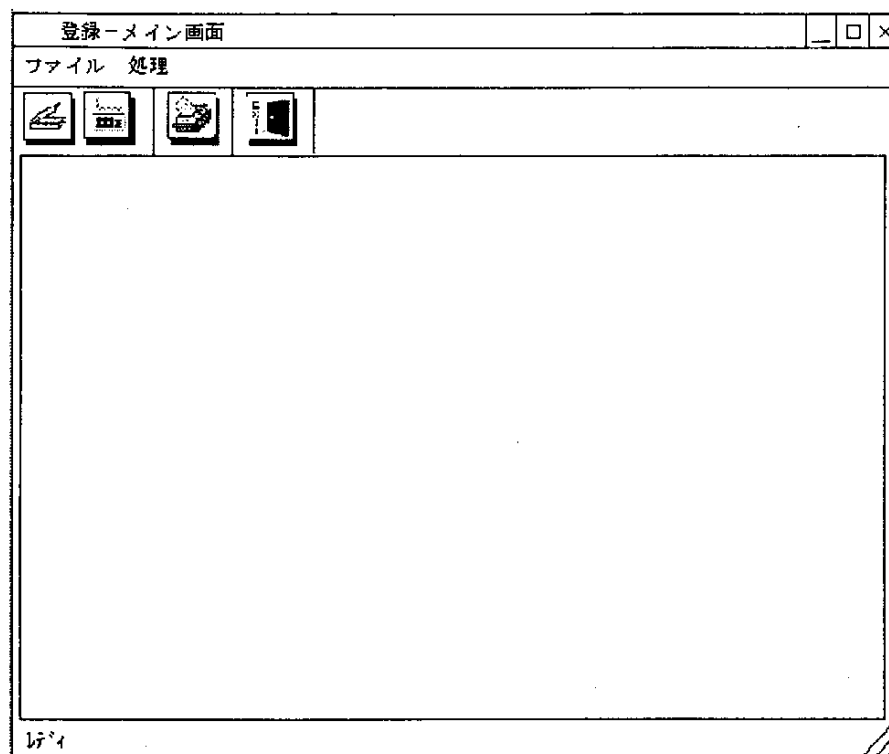
6.1 イメージ登録

6.1.1 画面遷移図



6. 1. 2 メイン画面

メイン画面では各処理画面の選択を行う。



(1) メニュー機能

ファイル	
終了	イメージ登録プログラムを終了

処理	
スキャナ入力	「スキャナ入力」画面を表示
データ入力	「データ入力」画面を表示
分類作成	「分類作成」ダイアログ画面を表示

(2) ボタン機能



「スキャナ入力」画面を表示



「データ入力」画面を表示



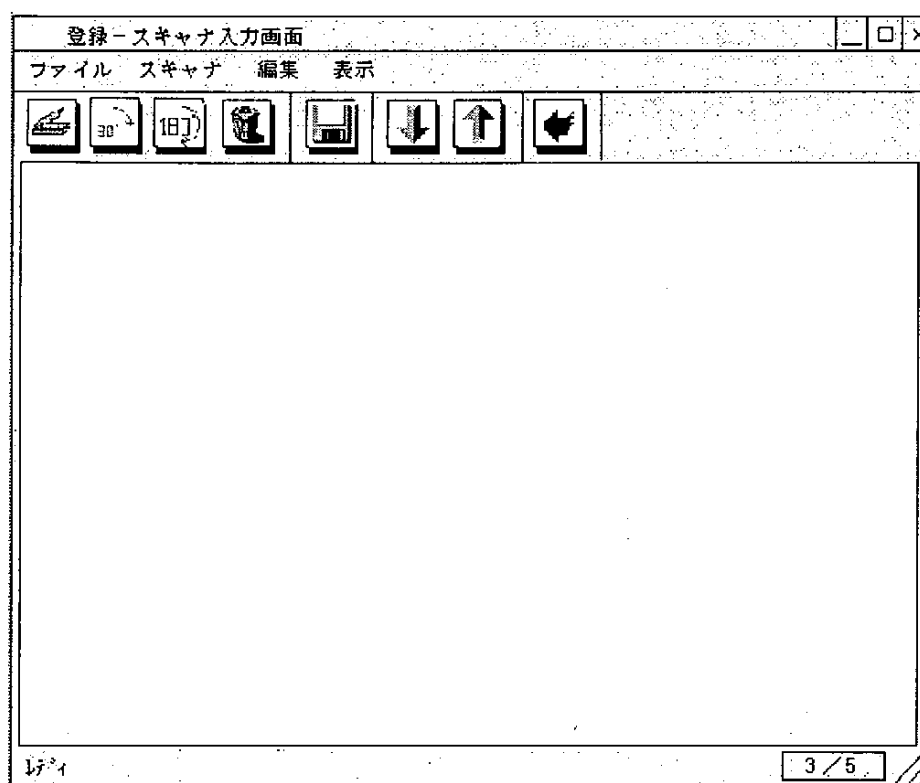
「分類作成」ダイアログ画面を表示



イメージ登録プログラムを終了

6. 1. 3 スキャナ入力画面

スキャナ入力画面では、1文書単位に複数のイメージを読み込み、ファイルへ保存する。



(1) メニュー機能

ファイル	
保存	1文書分のイメージファイルを保存
終了	イメージ登録プログラムを終了

スキャナ	
スキャナ入力	スキャナからイメージの読み込み

編集	
90° 回転	表示中の画像を90° 回転
180° 回転	表示中の画像を180° 回転
削除	表示中の画像を削除

表示	
次画像	次の画像を表示
前画像	前の画像を表示
メイン画面	「メイン」画面を表示
データ入力画面	「データ入力」画面を表示

(2) ボタン機能



スキャナからイメージの読み込み



表示中の画像を90°回転



表示中の画像を180°回転



表示中の画像を削除



1文書分のイメージファイルを保存



次の画像を表示



前の画像を表示



「メイン」画面を表示

6. 1. 4 データ入力画面

スキャナ画面で読込んだイメージファイルにタイトル等の書誌情報を付加する。この際、要旨欄は画像データから指定された範囲を文字認識し、入力することが可能。

(1) メニュー機能

ファイル		
保存		書誌情報を保存
終了		イメージ登録プログラムを終了

編集		
文字認識		選択した部分を文字認識し要旨欄に張付け
データ削除		表示中の文書を削除
分類作成		「分類作成」ダイアログ画面を表示

表示		
----	--	--

次画像	次の画像を表示
前画像	前の画像を表示
次文書	次の文書を表示
前文書	前の文書を表示
メイン画面	「メイン」画面を表示
スキャナ入力画面	「スキャナ入力」画面を表示
全データ表示	イメージ登録された全データを表示
未入力データ表示	データ入力されていないデータを表示
入力済データ表示	データ入力されたデータを表示

(2) ボタン機能



選択した部分を文字認識し要旨欄に張付け



表示中の文書を削除



1 文書分のイメージファイルを保存



次の画像を表示



前の画像を表示



次の文書を表示



前の文書を表示



「メイン」画面を表示

6. 1. 5 分類作成ダイアログ画面

データ入力で選択する分類の追加、修正、削除を行う。

分類作成画面

×××××××室

新規追加

修正

削除

×××××室
×××××課
×××××課

新規追加

分類を新規に追加

修正

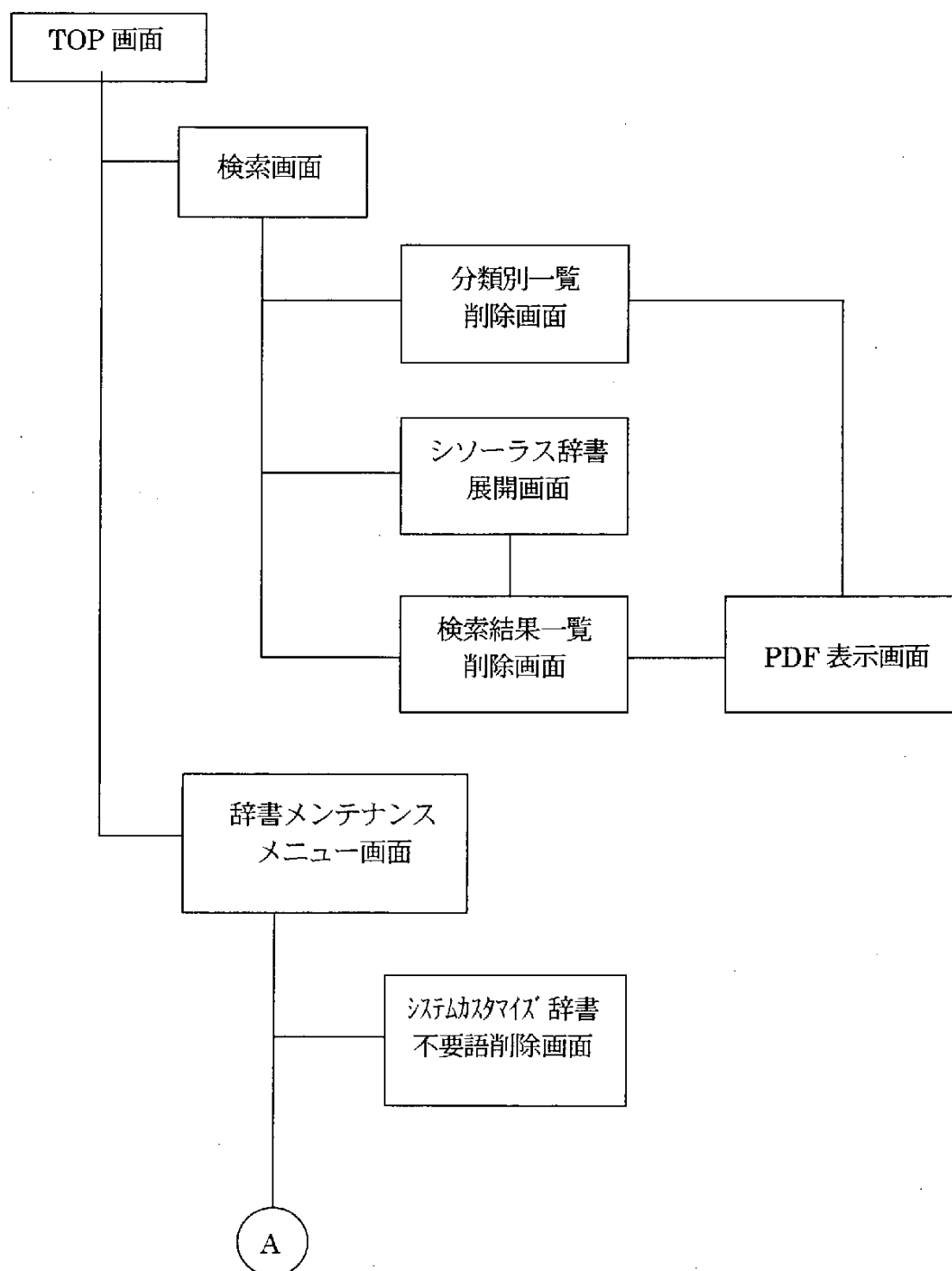
選択中の分類名を変更

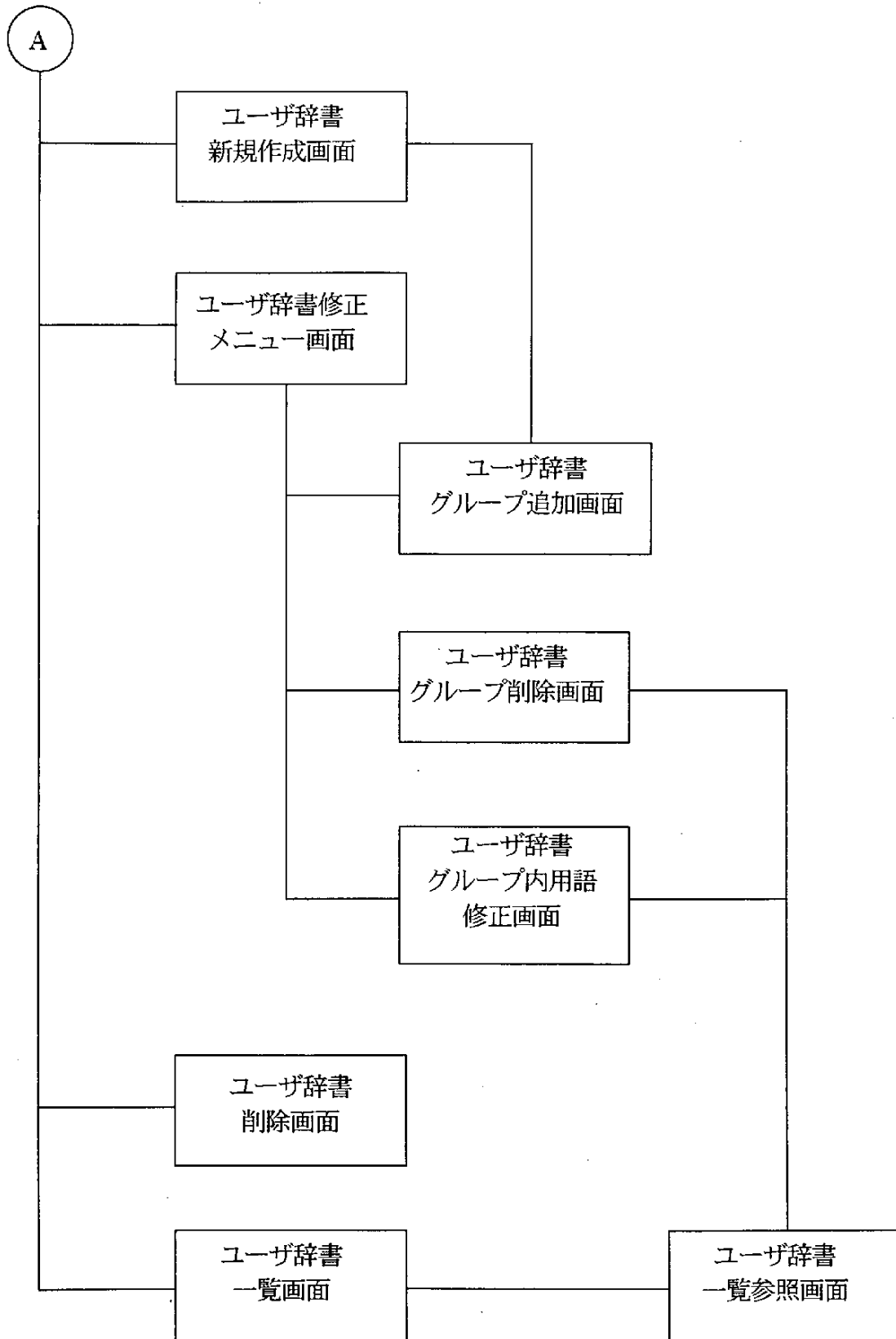
削除

選択中の分類を削除（分類内にデータが存在する場合は削除不可）

6. 2 文書登録

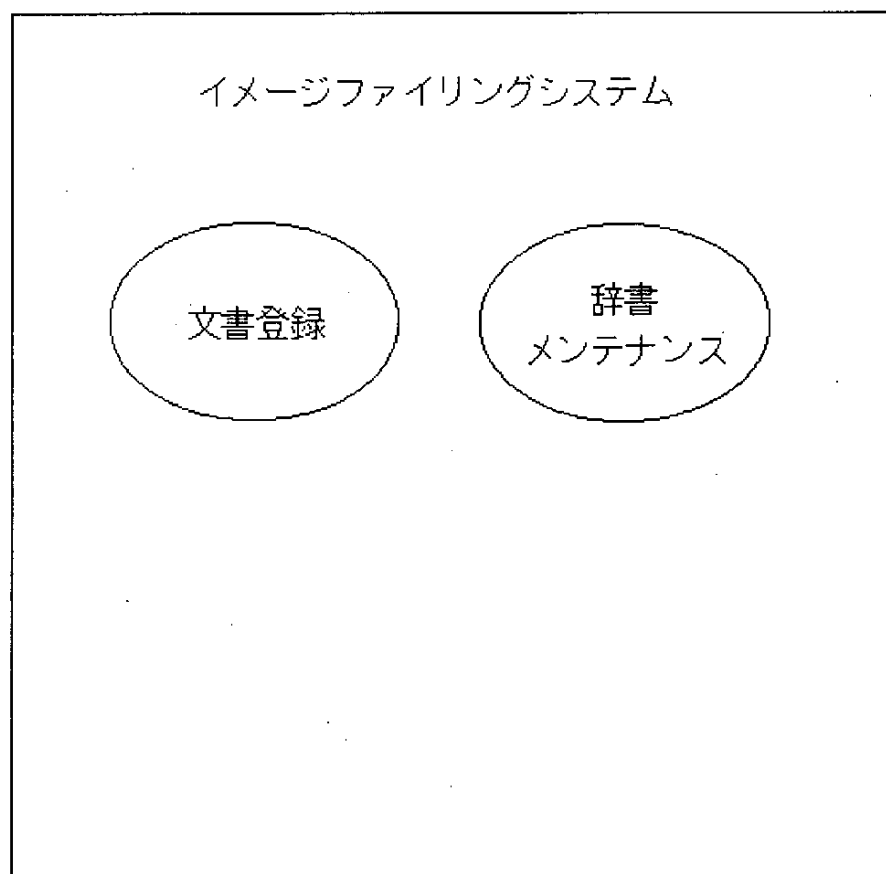
6. 2. 1 画面遷移図





6. 2. 2 TOP 画面

TOP 画面では、各処理の選択を行う。



(1) ボタン・選択機能

文書登録 : 「索引作成メニュー」画面を表示

辞書メンテナンス : 「辞書メンテナンスメニュー」画面を表示

6. 2. 3 索引メニュー画面

索引ファイルへの追加登録・削除を行う。追加登録は、ボタンを押下するとデータ入力まで終了しているデータを自動的に登録し、PDF ファイルを作成する。

The screenshot shows a rectangular window with a black border. Inside, there are three rectangular buttons stacked vertically. The top button is labeled '登録' (Register). The middle button is labeled '文書の削除' (Delete Document). Below the buttons, the text 'TOP 画面へ戻る' (Return to TOP screen) is displayed as a link.

(1) ボタン・選択機能

登録 : データ入力されたデータを索引ファイルへ追加登録

文書の削除 : 「検索」画面を表示。
索引ファイルへ登録された特定文書を削除

TOP 画面へ戻る : 「TOP」画面を表示

削除する文書を分類別、又は、キーワード検索で絞込みを行う。検索画面の機能については、6. 3. 3 検索画面を参照とする。

分類別にタイトルが左側フレームに表示され、タイトルの前にチェックされたデータについて削除を行う。左側フレームのタイトルを選択すると右側フレームに書誌情報が表示される。

(1) ボタン・選択機能

 PDF ファイルを Acrobat Reader にて表示

シソーラス辞書から同義・類義語を展開し、表示する。検索からはずす用語にチェックし検索実行することで、検索のノイズを減らす効果がある。シソーラス辞書展開画面の機能については、6. 3. 5 シソーラス辞書展開画面を参照とする。

検索条件に一致した文書のタイトルが左側フレームに表示され、タイトルの前にチェックされたデータについて削除を行う。左側フレームのタイトルを選択すると右側フレームに書誌情報が表示される。

40

(2) ボタン・選択機能

削除実行 : チェックされたデータを削除



PDF ファイルを Acrobat Reader にて表示

6. 2. 8 辞書メンテナンスメニュー画面

シソーラス辞書には2種類あり、本システムではこれを「システム辞書」「ユーザ辞書」と記述する。システム辞書とは、本システムに組込まれた提供辞書を言い、ユーザ辞書とは、ユーザが独自に作成する辞書を言う。

システム辞書メンテナンス

不要語の削除

ユーザ辞書メンテナンス

新規作成

修正

辞書の削除

一覧表示

[TOP画面へ戻る](#)

(1) ボタン・選択機能

不要語の削除 : 「システムカスタマイズ辞書不要語削除」画面を表示

新規作成 : 「ユーザ辞書新規作成」画面を表示

修正 : 「ユーザ辞書修正メニュー」画面を表示

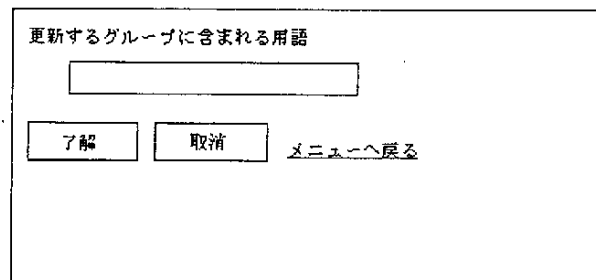
辞書の削除 : 「ユーザ辞書削除」画面を表示

一覧表示 : 「ユーザ辞書一覧」画面を表示

TOP 画面へ戻る : 「TOP」画面を表示

6. 2. 9 システムカスタマイズ辞書不要語登録画面

システム辞書から、検索に不要な用語を登録する。この登録する辞書を「システムカスタマイズ辞書」と言う。不要語が含まれるグループを参照する為の用語入力し、展開された用語一覧から不要語をチェックし登録する。



更新するグループに含まれる用語

[メニューへ戻る](#)

画面1 用語入力

(1) ボタン・選択機能

了解 : 画面2 不要語登録画面を表示

取消 : 画面を初期状態に戻す

メニューへ戻る : 「辞書メンテナンスメニュー」画面を表示

グループ番号	用語	削除
		☐
		☐
		☐
		☐
		☐
		☐
		☐
		☐
		☐
		☐

[前ページへ戻る](#)
[メニューへ戻る](#)

画面 2 不要語登録画面

(1) ボタン・選択機能

: チェックされた用語をシステムカスタマイズ辞書へ登録

: 画面を初期状態に戻す

[前ページへ戻る](#) : 画面 1 用語入力画面を表示

[メニューへ戻る](#) : 「辞書メンテナンスメニュー」画面を表示

6. 2. 10 ユーザ辞書新規作成画面

ユーザ辞書ファイルを新規に作成。ユーザ辞書は最大15個までを作成可能とする

辞書ファイル名

コメント

辞書ファイル名一覧

↑

↓

[メニューへ戻る](#)

(1) ボタン・選択機能

作成実行 : ユーザ辞書ファイルを作成し、「ユーザ辞書修正グループ追加」画面を表示

取消 : 画面を初期状態に戻す

メニューへ戻る : 「辞書メンテナンスメニュー」画面を表示

6. 2. 11 ユーザ辞書修正メニュー画面

作成したユーザ辞書へ登録、削除、修正を行う。

修正するユーザ辞書ファイル

↓

グループの追加

グループの削除

グループ内用語の修正

[メニューへ戻る](#)

グループの追加 : 「ユーザ辞書修正グループ追加」画面を表示

グループの削除 : 「ユーザ辞書修正グループ削除」画面を表示

グループ内用語の修正 : 「ユーザ辞書修正グループ内用語修正」画面を表示

メニューへ戻る : 「辞書メンテナンスメニュー」画面を表示

6. 2. 1 2 ユーザ辞書修正グループ追加画面

ユーザ辞書へグループの追加を行う。追加後、画面が初期状態に戻り、繰返しグループを追加することが可能。

辞書ファイル名 ××××××××××××

グループ

追加実行

取消

[ユーザ辞書修正メニュー画面へ戻る](#) [メニューへ戻る](#)

(1) ボタン・選択機能

追加実行 : ユーザ辞書へグループの追加登録

取消 : 画面を初期状態に戻す

[ユーザ辞書修正メニュー画面に戻る](#) : 「ユーザ辞書修正メニュー」画面を表示

[メニューへ戻る](#) : 「辞書メンテナンスメニュー」画面を表示

6. 2. 13 ユーザ辞書修正グループ削除画面

ユーザ辞書からグループを削除する。削除するグループに含まれる用語を入力し、展開された結果から削除するグループ番号を選択し削除する。

画面1 用語入力

辞書ファイル名 ××××××××××××		
削除するグループに含まれる用語		
了解	取消	一覧表示
ユーザ辞書修正メニュー画面へ戻る メニューへ戻る		

(1) ボタン・選択機能

: 画面2 削除画面を表示

: 画面を初期状態に戻す

: 「ユーザ辞書一覧参照」画面を表示します。

[ユーザ辞書修正メニュー画面に戻る](#) : 「ユーザ辞書修正メニュー」画面を表示

[メニューへ戻る](#) : 「辞書メンテナンスメニュー」画面を表示

6. 2. 14 ユーザ辞書修正グループ内用語修正画面

ユーザ辞書のグループへ用語の追加、削除を行う。修正するグループに含まれる用語を入力し、展開された結果から、グループ内の修正を行う。

辞書ファイル名 ××××××××××××

修正するグループに含まれる用語

了解

取消

一覧表示

[ユーザ辞書修正メニュー画面へ戻る](#) [メニューへ戻る](#)

画面1 用語入力

(1) ボタン・選択機能

了解

 : 画面2 修正画面を表示

取消

 : 画面を初期状態に戻す

一覧表示

 : 「ユーザ辞書一覧参照」画面を表示します。

[ユーザ辞書修正メニュー画面に戻る](#) : 「ユーザ辞書修正メニュー」画面を表示

[メニューへ戻る](#) : 「辞書メンテナンスメニュー」画面を表示

辞書ファイル名 ××××××××××

グループ番号	用語	削除
		☐
		☐
		☐
		☐
		☐
		☐
		☐
		☐

用語の追加

↓

[前ページへ戻る](#) [メニューへ戻る](#)

画面 2 修正画面

(1) ボタン・選択機能

: グループから用語を削除

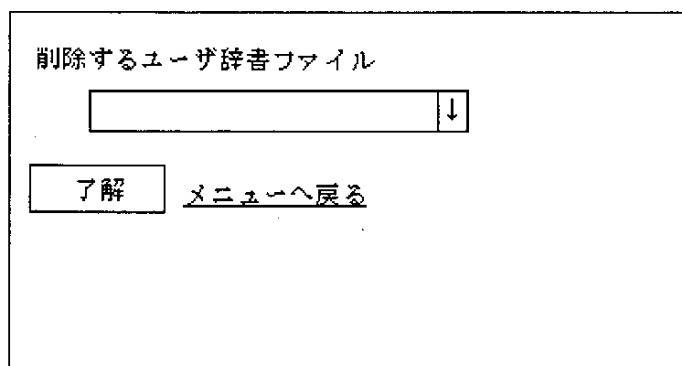
: グループへ用語を追加

[前ページへ戻る](#) : 画面 1 用語入力画面を表示

[メニューへ戻る](#) : 「辞書メンテナンスメニュー」画面を表示

6. 2. 15 ユーザ辞書削除画面

選択されたユーザ辞書ファイルを削除する。



削除するユーザ辞書ファイル

↓

[メニューへ戻る](#)

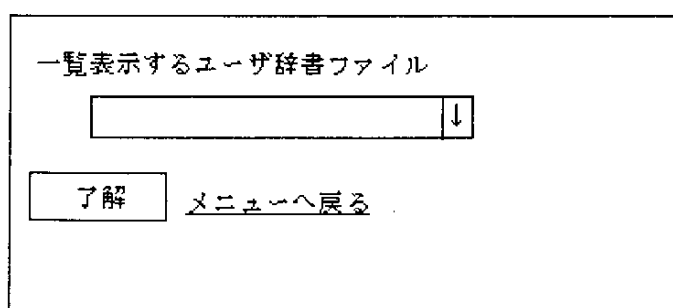
(1) ボタン・選択機能

: グループから用語を削除

[メニューへ戻る](#) : 「辞書メンテナンスメニュー」画面を表示

6. 2. 16 ユーザ辞書一覧参照画面

選択されたユーザ辞書ファイルへ登録されている用語を一覧表示する。



一覧表示するユーザ辞書ファイル

↓

[メニューへ戻る](#)

画面 1 辞書名選択

(1) ボタン・選択機能

了解 : 画面2 参照画面を表示

メニューへ戻る : 「辞書メンテナンスメニュー」画面を表示

辞書ファイル名 ××××××××××××	
グループ番号	用語

前ページへ戻る メニューへ戻る

画面2 参照画面

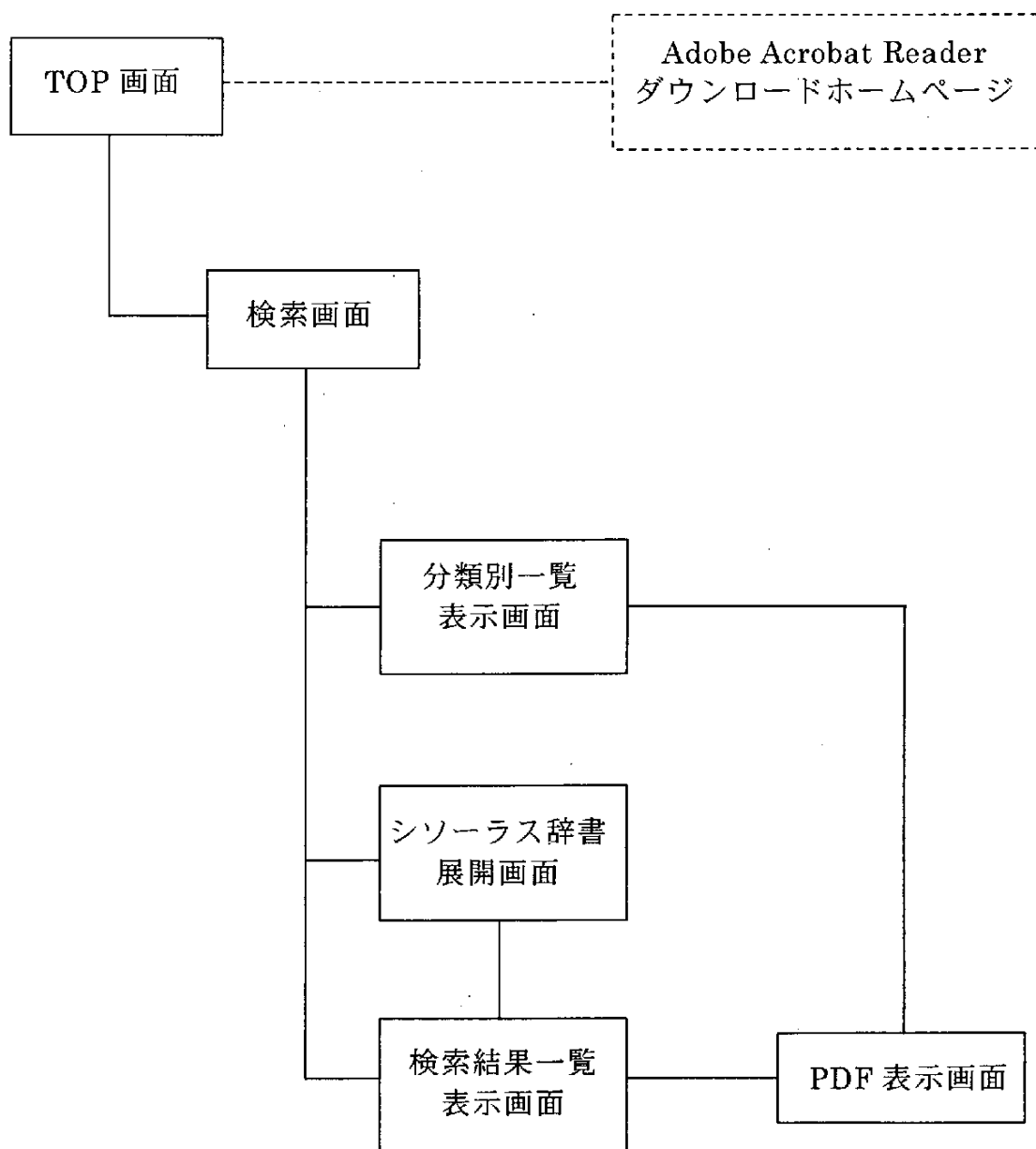
(1) ボタン・選択機能

前ページへ戻る : 画面1 辞書名入力画面を表示

メニューへ戻る : 「辞書メンテナンスメニュー」画面を表示

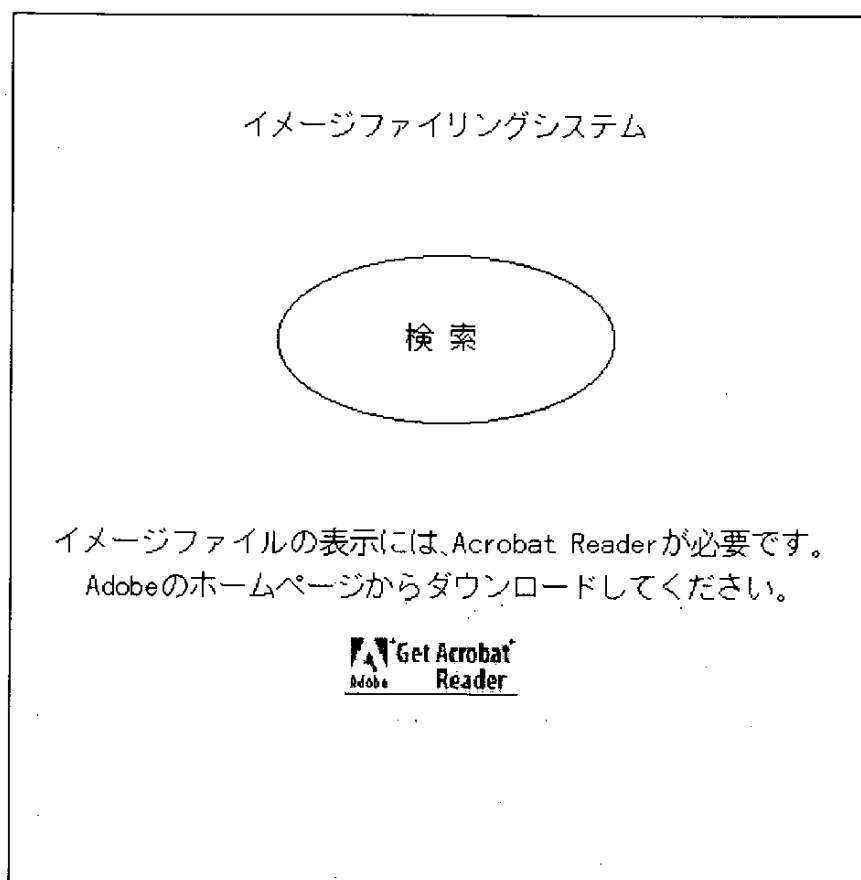
6. 3 文書検索

6. 3. 1 画面遷移図



6. 3. 2 TOP 画面

TOP 画面では、各処理の選択を行う。



(1) ボタン・選択機能

検索 : 「検索」画面を表示



: Adobe Acrobat Reader ダウンロードのホームページを表示ホームページの内容に従い、Acrobat Reader のインストールを行う。

6. 3. 3 検索画面

閲覧する資料を分類別、又は、キーワード検索で絞込みを行う。キーワード検索は更に資料が作成された日付で絞込むことが可能。

分類 ☒ 全分類 ☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G 分類別に表示 TOP画面へ戻る	検索条件 日付 年 月 日 ~ 年 月 日 検索対象 : ☐ タイトル ☐ 要旨 <table border="1"><thead><tr><th>キーワード</th><th>辞書展開</th></tr></thead><tbody><tr><td> </td><td>☐</td></tr><tr><td> </td><td>☐</td></tr><tr><td> </td><td>☐</td></tr><tr><td> </td><td>☐</td></tr><tr><td> </td><td>☐</td></tr></tbody></table> 検索式 AND OR オプション(要旨の検索のみ有効) あいまい検索 ↓ HITした頻度の高い順に並べる ☐ 検索実行 辞書展開 取消	キーワード	辞書展開		☐		☐		☐		☐		☐
	キーワード	辞書展開											
	☐												
	☐												
	☐												
	☐												
	☐												

(1) ボタン・選択機能

・右側フレーム

分類別に表示

 : 「分類別一覧表示」画面を表示

TOP画面へ戻る

 : TOP画面を表示

・左側フレーム

AND : 検索式を AND 条件で自動挿入

OR : 検索式を OR 条件で自動挿入

検索実行 : 入力された条件で検索を行い、「検索結果一覧表示」画面を表示

辞書展開 : 辞書展開にチェックされたキーワードの展開用語をシソーラス辞書から取得し、「シソーラス辞書展開」画面を表示

取消 : 画面を初期状態へ戻す

(2) 入力項目

日付 : 資料の作成された日付

タイトル／要旨 : 検索対象とする項目をチェックする。両方対象とした場合は OR 条件とする。

キーワード : タイトル 又は、要旨に含まれる用語を入力

辞書展開 : チェックされたキーワードの同義・類義語をシソーラス辞書から取得し、展開された用語を含めて検索する。

検索式 : 複数キーワードを入力した場合の条件を入力

あいまい検索 : 要旨の検索のみ有効。キーワードの一致度を選択する。

HIT した頻度の高い順に並べる : チェックされた場合は、HIT 率の高い文書の順に結果表示する。

分類別にタイトルが左側フレームに表示される。左側フレームのタイトルを選択すると右側フレームに書誌情報が表示される。

分類 A 	日付：××××年××月××日 タイトル：×××××××××××× 
B 	×××××××××××××××××××××××××××××××××× ××××××××××××××××××××××××××××××××××
C 	×××××××××××××××××××××××××××××××××× ××××××××××××
D 	×××××××××××××××××××××××××××××××××× ×××××××××××××××××××××××××××××××××× ×××××××××××××××××××××××××××××××××× ×××××
検索画面へ戻る	要旨

検索画面に戻る : 「検索」画面を表示

6. 3. 5 シソーラス辞書展開画面

検索画面で入力されたキーワードの同義・類義語をシソーラス辞書から取得し表示する。検索からはずす用語にチェックし検索実行することで、検索ノイズを減らす効果がある。

日付：××××～××××
検索対象：タイトル 要旨
検索式：a * b
オプション：あいまい検索 60% H I Tした頻度の高い順に並べる

検索キーワード	展開用語	検索無効
	-----	☐
	-----	☐
	-----	☐
	-----	☐
	-----	☐
	-----	☐
	-----	☐
	-----	☐
	-----	☐
	-----	☐

(1) ボタン・選択機能

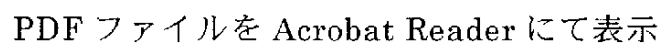
検索実行： 検索画面、シソーラス辞書展開画面にて入力された条件で検索を行い、「検索結果一覧表示」画面を表示

取消： 画面を初期状態に戻す

検索画面へ戻る： 「検索」画面を表示

検索条件に一致した文書のタイトルが左側フレームに表示される。左側フレームのタイトルを選択すると右側フレームに書誌情報が表示される。

(1) ボタン・選択機能



検索画面に戻る : 「検索」画面を表示

7 本事業開発によるイメージファイリングシステムの効果的な運用

7. 1 本イメージファイリングシステムの運用方法

イメージファイリングシステムの特長は、シソーラス辞書の採用による高い検索性にある。そこで運用方法の想定も含め、このシソーラス辞書の機能概要紹介を行う。

7. 1. 1 シソーラス辞書

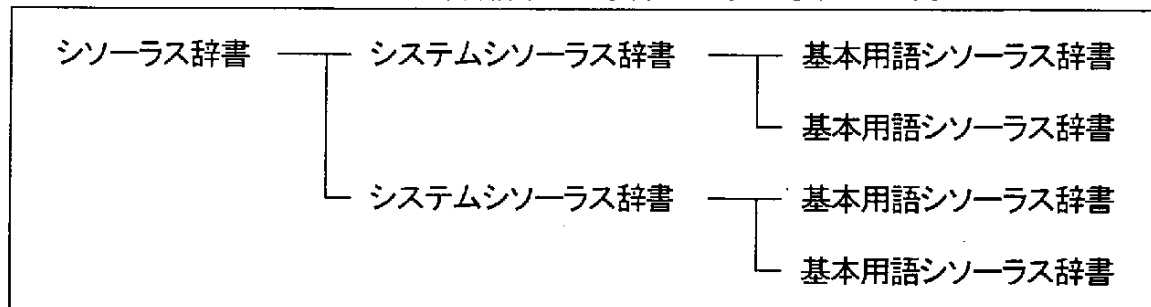
キーワード検索のポイントは検索漏れをなくす点にある。例えば「自動車」に関するキーワード検索を行う場合、単純検索では検索対象から「自動車」という言葉が含まれる文書のみを抽出してくる。しかし「自動車」を表現する言葉は、これ以外にも「車・クルマ・カー」等多数存在する。(表「シソーラス辞書基本語サンプル」参照)

文書検索者はこれら他のキーワードも含めて「自動車」全般に関わる文書の検索・抽出を希望するはずである。そこでこれら複数語の意味的な関連性を示し、それら語句をリンクさせる「シソーラス辞書」が文書検索システムの機能としては必要不可欠であることから、今回開発のイメージファイリングシステムでは5. 8万語のシソーラス辞書を搭載している。

シソーラス辞書基本語サンプル	
自動車 オートモービル オートモビル カー クルマ モーター・カー モーターカー 車	L. A. LA ロサンジェルス ロサンゼルス ロス ロスアンジェルス ロスアンゼルス 羅府
通信販売 通販 カタログ販売 メール・オーダー メールオーダー カタログ・ショッピング カタログショッピング	体育館 ジムナジウム ジム ギムナジウム 屋内運動場 屋内体操場 室内運動場
アメリカ アメリカ合衆国 合衆国 米国 米 亜米利加 U. S. US U. S. A. USA	LBO レバレッジド・バイ・アウト レバレッジドバイアウト レベレッジド・バイ・アウト レベレッジドバイアウト 会社転がし 買収先の資産を担保にした融資による買収 買収先企業の資産を担保にした買収

7. 1. 2 シソーラス辞書の構成

今回のシステムのシソーラス辞書構成は、以下のようになっている。



7. 1. 3 シソーラス辞書の内容

●システムシソーラス辞書（この辞書はユーザー修正できない。）

システムシソーラス辞書は次の二つに大別される。検索システムではこの辞書を自由に組み合わせて利用することが可能。

・基本用語シソーラス辞書

基本普通名詞（一般文書、報告書、新聞、雑誌などでよく利用されるシソーラス）と一般語（技術一般、ビジネス一般、地名のシソーラス）を集めた辞書。

基本用語シソーラス辞書は、一つのファイルとして提供される。

・専門用語シソーラス辞書

専門分野別にシソーラスを集めた辞書。ジャンル毎（※）に辞書ファイルは別々となっている。

※専門用語シソーラス辞書（14ジャンル）

経済・法令・公共用語／企業名／機関・団体・学校名／化学・石油工業用語／電力エネルギー用語／食品用語／資材用語／非鉄・金属用語／コンピュータ・情報・通信用語／自動車用語／機械用語／土木建築用語／医学・薬学用語／時事・スポーツ・芸能・放送用語

●ユーザーシソーラス辞書：ユーザーが辞書の内容を定義する辞書。

・システムカスタマイズ辞書

システムシソーラス辞書から削除したい用語を定義する辞書。

この辞書を利用することで論理的にシステムシソーラス辞書から用語を削除することが可能。システムカスタマイズ辞書はシステム内で1つのファイルとして構築可能。

・ユーザーコンテンツ辞書

ユーザー独自の用語を定義する辞書。

ユーザーコンテンツ辞書は用途に合わせて最大15個まで同時に利用することが可能。

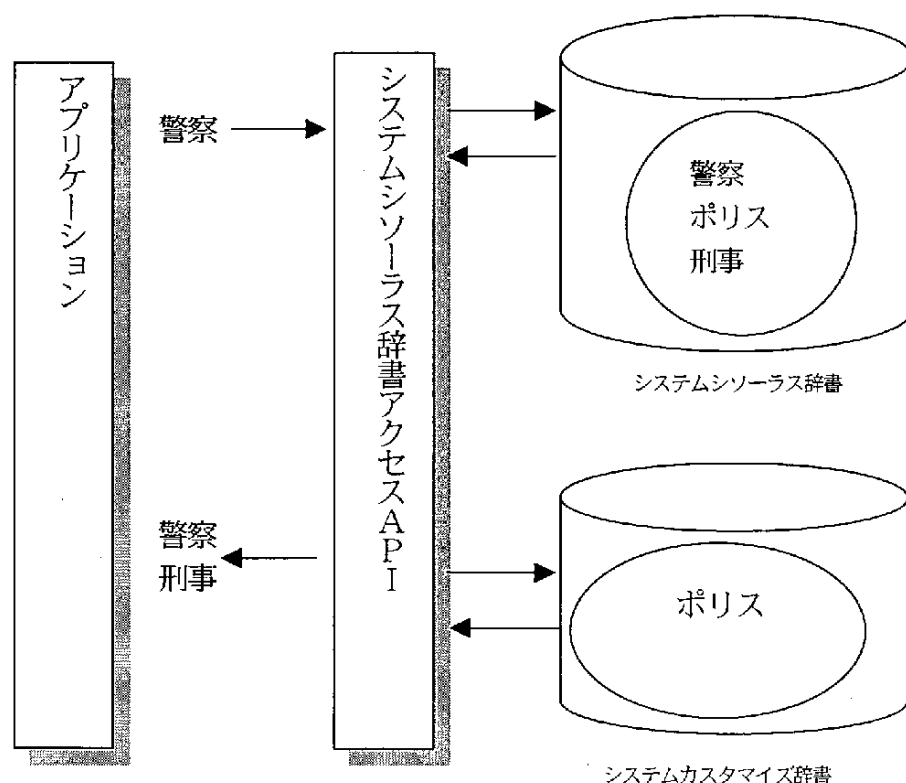
7. 1. 4 システムシソーラス辞書とユーザーシソーラス辞書の関係

● システムシソーラス辞書とシステムカスタマイズ辞書

システムカスタマイズ辞書には、システムシソーラス辞書の特定のグループから削除する用語が記述されている。システムシソーラス辞書アクセスAPIは、システムシソーラス辞書の実語の展開を行う際にシステムカスタマイズ辞書に削除用語が記述されていれば、その用語は展開されない。

(例)

この例では、システムシソーラス辞書に「警察」、「ポリス」、「刑事」が1つのグループとして登録されている。また、システムカスタマイズ辞書には、「ポリス」を削除せよという情報が記述されている。このとき、システムシソーラス辞書アクセスAPIを利用し、キーワード「警察」で用語の展開を行うと、結果として「警察」「刑事」を得ることができる。



● システムカスタマイズ辞書を利用するときの留意点

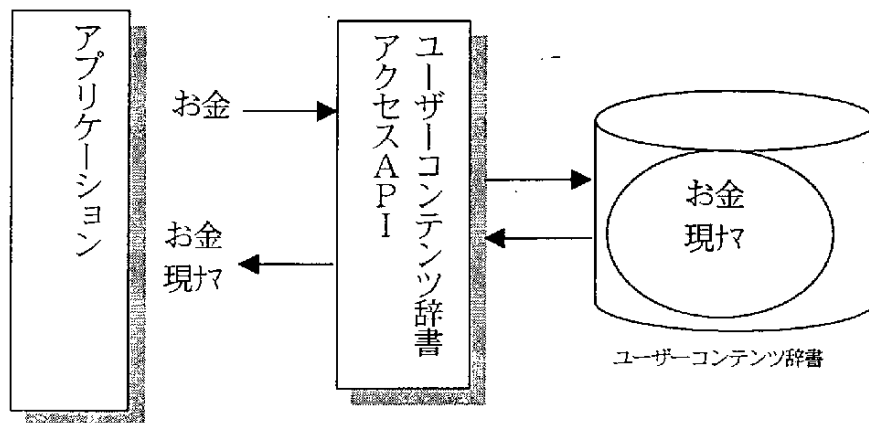
- ・ システムカスタマイズ辞書に用語が記述されているとき、システムシソーラス辞書の扱いは当該用語がない場合と全く同じです。たとえば上の例では、システムシソーラス辞書に「ボリス」が記述されていないこととまったく同じです。
- ・ システムシソーラス辞書の用語展開時にシステムカスタマイズ辞書を参照するかどうかは、システムシソーラス辞書アクセス API のパラメータで指定することができます。

● システムシソーラス辞書とユーザーコンテンツ辞書

システムシソーラス辞書とユーザーコンテンツ辞書は互いに独立した関係にある。そこでアクセスAPIもそれぞれ別のものが用意されている。ユーザーコンテンツ辞書の利用法はアプリケーションプログラムで自由にアレンジすることができる。例えば、同じキーワードでシステムシソーラス辞書とユーザーコンテンツ辞書から用語を展開し、双方の用語を足し合わせたり、ユーザーコンテンツ辞書の用語をシステムシソーラス辞書より優先して取出すなどの利用の仕方が考えられる。

(例)

この例では、「お金」をキーワードにしてユーザーコンテンツ辞書からは「お金」、「現ナマ」を求めています。この後、この用語をどう利用するかはアプリケーションで自由にアレンジすることができます。



7. 2 本イメージファイリングシステムの運用の姿

行政機関における文書とは各部門ごとに発生し、文書による業務運営という側面から考えるとその対象分野は多岐にわたっている。例えばその中で議会事務局を想定するだけでも本イメージファイリングシステムで管理すべき文書分野は次のようなものとなる。

まずは本会議開催の際、最初に発生する文書が議案書である。この議案書は、本会議で議論される議題として提出されるものであり、各議題の詳細内容が記載されており、議員をはじめとする関係者に配布される。これらは開催ごとに作成されるものであるが、その量は膨大なものとなる。

また議案書とは、その内容が議会において議論をなされ、議決され、最終的には条例の制定や事業の計画・実施とその後の行動を起こすための第一歩の文書であるといえる。しかしこれまでの紙文書による議会事務とは、これら審議案件の過程に関する文書がリンクした形で保管・管理されておらず、情報公開法によって議案の審議過程とその結果や行政活動を公表して欲しいという市民の請求があったとしても、その公表までに時間がかかることもいたしかたないことだろう。しかし情報公開のニーズが高まり、しかも議会事務の効率化を目指すのであれば、これら文書を電子的にリンクさせた形式で管理しなければ行政改革は実現できないものである。

～イメージファイリング導入前～

議 案 上 程 ・ 関 係 資 料 調 査 依 頼



紙 の 関 連 資 料 か ら 調 査 を 行 う

ベテラン職員の記憶に頼りながら、調査を行う
ベテラン職員が存在しない場合、調査期間に数日を要する場合もある

～イメージファイリング導入後～

議 案 上 程・関係資料調査依頼



イメージファイリングシステムにより関連資料の調査を行う

ベテラン職員の有無に関わらず、提出日時、内容等から検索を実行することにより、調査資料を検索することが可能となる

しかも議題には「条例の制定／改定」「意見書・請願書・陳情書の提出」などの関連文書が付されることが多い。これに関しても現状においては各々別々に管理されており、紙文書によるアナログ的管理体制ではそのリンクの状況を見取することはできない。そこでこれら議案過程と共に、関連する文書に関してもリンクをさせ、総合的な文書管理及び文書検索を可能とすることが本システムの目的であるといえるであろう。

そこで本項では、議会事務局における本イメージファイリングシステムの運用を「条例の制定／改定」「意見書・請願書・陳情書の提出」の2点を対象として、その利用方法を検討するとともに、行政機関における本イメージファイリングシステムに最適と思われる文書データの事例を挙げてみることにする。

7. 3 本イメージファイリングシステムの運用分野

7. 3. 1 条例の制定／改定

地方自治体における条例の制定／改定にあたっては、首長名にて議会に議案書が上程される。この議案は議会審議が行われ議決後に施行される。議案提出から施行までの間、各種文書の電子的なリンクや対比が不可能な現状においては、関係者は旧条例の内容確認や新条例の修正等にかかなりの時間をさくことになる。また実際には在職期間の長い職員の記憶に頼っていることが多分に多く、異動・退職等によりその職員が存在しなくなる可能性から考慮しても、文書管理システムの整備を関連情報のリンクも含めて行い、特定の担当者に頼らなくともこれら作業がスムーズに進行できるシステム構築を行わなければならない。そこで本イメージファイリングシステムを有効に活用することにより、過去の記憶を持たない職員であっても必要な情報を引き出し、その過程や背景を理解することが容易に可能となり、事務の効率化は格段に向上する

もの考えられる。

7. 3. 2 意見書・請願書・陳情書

議会では、地方自治体が都道府県へ提出する意見書・決議書、議員の紹介により提出される請願書、住民から議長に対して提出される陳情書等の書類が議案に付され、議会審議が行われる。

過去に提出された意見書・決議書、請願書・陳情書を調査する際も、文書管理がなされていない現状の多くは、やはり条例のケースと同様に、在職期間の長い職員の記憶に頼っていることが多分に多く、異動・退職等によりその職員が存在しなくなる可能性を考慮すると本イメージファイリングシステムの導入が有効になると考えられる。

7. 3. 3 その他の行政文書分野における本システムの活用

(A) 地図情報

行政機関における管理情報として大量に存在するに「地図情報」も考えられる。公共工事の施工場所に関する管理データや公共施設等の記された地図も本イメージファイリングシステムの対象分野と考えられる。これらも現状は紙による図面として保管がなされており、その保管スペースや情報検索性から考慮するとイメージデータが必要不可欠であると考えられる。

なお図面データに関してはCADシステム等によって作画されているものが多いが、このCADデータであってもPDFファイル化は容易なものであり、少ないデータ容量によってデータ保管が可能である。

またこれまでの行政機関における地図情報管理の問題点とは、各々の部門ごと別々にファイル管理がなされていたという点である。例えば災害発生時の緊急対策を地図情報をベースとして行おうとする場合においても、河川に関しては建設局河川課、国道に関しては建設局道路管理課、鉄道に関しては運輸局、水道管に関しては水道局、消防署所在地に関しては消防局など、別々の地図管理がなされていた。本イメージファイリングシステムでは、1つの地図をベースとして、これら防災関連管理情報を階層的にリンクさせ、複合的な判断が可能な文書管理システムが構築可能である。

また地域再開発工事、拡張工事などの現在の施工場所の検索や工事終了期日を検索することやこれまで膨大なデータ量を要する地図情報を利用して作成していた公共施設の案内文書の作成や管理情報も本イメージファイリングシステムにおいて複合的に管理することが可能となる。

(B) 新聞・雑誌記事

毎日発表され、追加・更新が必要となる新聞・雑誌記事に関しても本イメージファイリングシステムは有効なものである。従来のイメージファイリングシステムでは新聞名・雑誌名、見出しをインデックスとして付与し、それをキーワードとして検索

するまでの機能にとどまっていたが、本イメージファイリングシステムでは記事内容までもOCRにより文書データ化することも可能であり、これらの記事の蓄積により新聞社、雑誌社のデータベースに匹敵するものとして記事データベースを構築することが可能となる。

新聞記事の切抜をファイリングする作業については毎日2～3時間要している自治体もあり、本イメージファイリングシステムの導入の際には、大幅に作業時間の短縮も可能であり、しかもその時間内に記事内容の文書データ化まで可能となり、大きな成果が期待できるものと考えられる。(しかもイントラネットを利用して、それら記事内容は全庁的に電子配信することも可能となる。)

(C) 指名参加願等の申請・届出書式

前述のアメリカにおける「フォーム」機能による申請届出様式の電子配信と電子的な申請・届出受付とその申請内容に関するデータ管理という面についても、本イメージファイリングシステムの対応分野である。(紙による申請・届出を受け付けたものに関しても、ワープロ等の電子的文書作成をなされ、それを紙に出力したものであれば、OCR機能によってデータ化が可能である。)

行政事務の多くは、これら申請・届出案件に関して、自らの審査承認用に再度データ入力を行うというような非効率的な作業に多くの時間が割かれている。しかし本システムを活用すれば、申請・届出というデータ発生源のデータがそのまま電子文書データとすることができ、その後の審査・承認作業も大幅に効率化でき、しかも申請・届出書類データとその後の審査承認関連文書をリンクさせることによって、その案件の審査過程も明確となり、この分野においてもその分野に関してノウハウを有した特定の職員に頼ることなく、業務の標準化が示されるものと考えられる。

(D) 予算書・決算書

予算書・決算書に関しても、本システムによってデータベース化することにより、毎年の書類をリンクさせておけば、各年度の収支情勢の推移を容易に知ることができる。しかもOCR機能によって過去の予算書・決算書のデータ入力を行えば、今年度からでも過去の予算・決算の推移を即把握することができ、再データ入力という手間のかかる作業を考慮しなくとも容易に分析が実行できる。

また予算書・決算書とは、ベースとなる書類の他に、各分野ごとの予算書・決算書やその予算書・決算書が完成するまでの関連文書が背景には存在する。これまでのアナログ的な文書管理体制では、これらをリンクさせて閲覧することが難しかったが、この分野においてもこれら関連文書をリンクさせることにより、より正確で詳細な予算書・決算書の中身に関する文書検索が可能となる。

(E) 広報・各種行政案内文書

広報等の各種行政案内文書分野に関しても本システムにおいて管理すべき文書分野であると考えられる。これまでの広報はアナログ的に庁内から行政案内案件を収集し、それを印刷物やインターネットホームページ上で公開し、そこまでで終わってしまう

ていた。しかし全庁的な文書管理を本システムをベースとして実行すれば、単純に広報の掲載情報をストックするのみではなく、その広報に掲載された案件の背景にある行政機関内部の内部関連文書とリンクさせ、その情報の背景となる部分を行政機関職員が検索できるものとするにより、この分野においても特定の情報を有した職員にその案件に関して頼ることは必要なくなり、職員誰もがその案件に関する背景に関して容易に理解でき、引継ぎ等も容易に行われるものと考えられる。

また住民から問合せがあったとしても、この背景情報がリンクしていることにより、正確な情報発信や対応が可能なものとなり、住民サービスの向上につながるものと考えられる。

8 行政分野以外の本イメージファイリングシステムの運用例

本イメージファイリングシステムの利用範囲とは、行政機関における行政分野のみではない。民間分野における利用の幅も十分に考えられる。そこでまずは本システムの発展性の1つとして他分野におけるシステム活用の姿について考えてみたい。

8. 1 民間分野における本イメージファイリングシステムの活用

民間分野においても大量の文書を取り扱っている業種・部門は数多く考えられる。そこでこれら行政機関と業務スタイルが類する業種を選択し、その業種企業におけるシステムの活用方法を考えてみる。

8. 1. 1 製造業

製造業の各業種においては、通産省がすすめる「N-CALS」の流れの中で、各種技術コンテンツの SGML 化が明確に提唱されている。また情報化先進国の欧米諸国からは、これまでの国内製造業のスタイルである「紙文書を中心とした業務スタイル」に対する批判が高まり、国際連携による製品製造が当たり前となる中で、早急なコンテンツの電子化が外圧として求められている。

しかしこの点に関しても、やはり行政機関の文書管理状況と同様に、今後生まれる技術コンテンツに関しては、SGML 文書管理体制の整備とあいまって、SGML 化への移行を目指すことができるが、過去に大量に発生した技術コンテンツに関しては、投資対効果に見合った取扱方法の答がどの業界からも求められている状況である。

そこで、製造業において過去に作成された技術コンテンツである「取扱説明書」「保守・整備マニュアル」「製造技術標準文書」「特許申請文書」等に関して、本システムにおいて PDF 化することが、まずは得策ではないかと考えられる。

SGML 文書とは、その文書データが1つの目的のみならず、複数の目的にその利用範囲が広がる場合には、その有効性は非常に高い。しかし紙文書の形態で存在する各種技術コンテンツに関して、その中からリンクする内容を検索・抽出し、共通したリンク情報を統合化して、その技術コンテンツのベースとなる SGML 技術文書データベースを過去の文書に関してこれから着手するのでは、多額のコストと多くの時間が必要となってしまう。そこで「取扱説明書」「保守・整備マニュアル」等、ある決まった文書利用目的が明確であり、しかもそれが既に紙の形で存在し、新たに SGML 的な統合的文書データベースを構築する必要のないものに関しては、文字を文字データとして認識し、それを全文検索エンジンで検索可能な PDF ファイルをベースとした本システムによって、個別分野の文書ファイルを行うことが現状では最良と考えられる。

8. 1. 2 金融機関

銀行・証券・生命保険・損害保険等の金融機関というのも、商品種類が多く、それを営業マンに理解させるための商品マニュアルというコンテンツを大量に抱えている。この商品マニュアルに関しても、製造業と同様の理由から考慮すると、過去の商品に関するコンテンツについては PDF ファイル化し、本システムによって管理すべきである。

そして金融機関の商品とは、やはりその複雑性から単純に文字のみで表現がなされた商品マニュアルはあまり多くない。つまり関連図を示し、他の商品との相違点を示すことによってビジネスにつなげているのである。そこで単純に文字データのみをデータベース化するだけではあまり意味がなく、システム図・関連図等の図形データとリンクした形式で管理を行うべきである。この点からも PDF ファイルによって管理することの有効性は大きいであろう。

また金融機関における最大の問題点は、各金融機関サイドでは外圧や文書データベース管理の有効性理解の高さから SGML 化が急速に進展しているにも関わらず、それを監督・指導する「大蔵省」サイドの電子文書化・SGML 化が進展しないという点にある。つまり現状大蔵省が各金融機関に対して行う、各種通達・指導に関する文書とは、紙文書が今でも利用されている。そこでそれを受け取った SGML 化を進めている金融機関においては、それをわざわざ再度 SGML 文書として入力しなおし、自社の金融商品関連文書とのリンクを図るという無駄な作業を取り組まざるを得ない状況にある。

大蔵省内部ではパソコン・ワープロ等電子的ツールによって、文書作成がなされているはずではあるが、それを紙文書でしか提供されない現状においては、やはり紙文書を OCR 入力し、本イメージファイリングシステムにおいて文書管理を行うことが有効であろう。

8. 1. 3 その他分野

本システムの有効活用が考えられる業種・分野とは、過去の大量紙文書が存在し、それをデジタル化して、今後保存・活用しなければならないというところである。この点から考えると次のような分野がその他にあげられる。

- ① 研究論文等の学術分野
- ② 保守管理・運行管理において大量のコンテンツが生まれる航空会社
- ③ マニュアル・仕様書等において大量のコンテンツが生まれるソフトウェア業

9 本イメージファイリングシステムの発展性

本イメージファイリングシステムは、ワープロ・パソコン等により過去に作成された紙文書を管理対象とすることをきっかけに考えられてきたが、やはり今後主流となっていく電子文書ファイル形式「SGML」との連携を考慮することにより、一層高機能な文書管理システムを実現することができる。そこで SGML との連記による今後の本システムの発展性に関して考えてみたい。

9. 1 コンテンツ構築としての SGML、メディア出力としての PDF

前述のとおり、SGML と PDF は対極的に比較するべきものではない。本システムにおいて PDF ファイルをベースとする意味も、過去の SGML 化なされていない紙文書をどのように電子的に取り扱うのかという観点から生まれたものである。そこで今後もベースとなる文書コンテンツの最良の形式とは「SGML」なのである。

ただ SGML とはレイアウト情報を有していない。そこで SGML 文書データをビューイングするためには、有償供給による SGML ビューアー (Panorama Pro 等) をユーザーが必ず持たなくてはならない。LAN 等のクローズドな環境の中で、文書利用者が明確に見えているのであればこの方法も得策であろうが、行政機関における情報公開を想定した場合、不特定多数の市民が WWW ブラウザを利用してその行政情報を検索・閲覧することを考えなくてはならず、その場合有償供給の SGML ビューアーに頼っていたのでは、すべて市民に等しい情報サービスの提供を行うことにはつながらない。

しかし PDF の環境では、PDF ビューアーである「Acrobat Reader」が無償供給であることから誰もが自由に入手でき、誰も同じようなわかりやすいレイアウト済みの文書ファイルを検索・閲覧可能である。

そこでこの SGML 及び PDF の長所・短所を考慮し、しかも電子文書化過渡期である現在の文書管理状況を紙文書による過去と SGML による今後の双方の姿も考慮した形で、行政文書管理及びその行政文書の電子メディア発信の姿をシステム図化すると次のようになると考えられる。(あくまでも過渡期としてのシステム連関図であるが、ここ数年はこのコンセプトで進まざるを得ない状況にあると考えられる。)

—— 禁無断転載 ——

平成10年3月発行

発 行 財団法人 データベース振興センター
東京都港区新橋2丁目13番8号
新橋東和ビル5階
TEL 03-3508-2430 (代表)

委託先 株式会社 会議録研究所
東京都新宿区市ヶ谷八幡町16-209
TEL 03-3267-6051 (代表)

印刷所 フュージョン株式会社
札幌市中央区15条西9丁目
TEL 011-551-8055 (代表)

