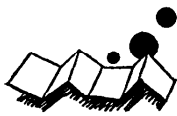


解説



OA システムと文書データベース†

田口和男** 坂下善彦***

1. はじめに

近年、半導体技術の進歩にともない、安価で性能の良いマイクロプロセッサが組み込まれたワードプロセッサやパーソナルコンピュータが多数開発され、その技術革新のスピードの速さには、目を見張るものがある。

また、これらの機器を使って文書作成や個人データの処理を行ういわゆるオフィスオートメーション(OA)指向も強く、一般企業での導入や活用も盛んである。

一方、これらのOA機器を単体で使用するだけでなく、企業ごとに導入されている汎用コンピュータやオフィスコンピュータに連結させてファイル転送を行い双方のデータ資源を一層有効に、かつ柔軟に利用するマイクロ・メインフレーム・コネクション(またはリンケージ)を中心としたOAシステムの開発も盛んに行われるようになった。このようなOAシステムでは、個別OA機器により作成された文書をホストコンピュータ上の大容量ディスク上に集中管理し、共同利用を促進する電子キャビネットや文書データベースを提供することを目的として、開発されている。

本稿では、上で述べた背景を踏まえて、文書データベースの基本概念、OAシステムでの位置づけ、文書データベースに要求される機能とそれを実現する技術について解説する。

2. オフィス業務と文書データベース

2では、一般的なオフィスで行われる文書作成や文書管理作業について、OAシステムの視点から概観し、その発展経緯や背景について述べる。

2.1 オフィス業務と文書

一般のオフィスで行われる文書業務は、各種の連絡文書、報告文書、説明資料、議事録などの作成、編集、コピー、ファイリング、及び社内外への文書発信などである。

オフィスオートメーション(以下OA)導入の第1ステップとしては、まずこれらの文書の電子ファイル化、すなわち、紙ではなくて磁気媒体上に作成し、その編集、管理、発信などをワープロ、パソコン、ファックスなどのOA機器を使って行う。また、文書業務のOAとともに、日常的に発生する数値データについて集計、統計、及びグラフ化などを行う数値データ処理を中心としたOAも行われている。

実用的なOAシステムの条件としては、一般のオフィスワークでも十分に使いこなせる簡便なユーザインタフェースをもったハードウェア、ソフトウェアが提供されることが必須条件である。

2.2 電子キャビネットと文書データベース

電子ファイル化されたオフィス文書は、ワープロやパソコンなどのOA機器の外部記憶(フロッピーディスクや固定ディスク)上に保存され、必要に応じて検索、取り出し、追加、修正などが行われる。

文書の検索、取り出しは、文書を保存する際に指定される識別用のシンボル(文書名、作成者名、作成日など)により行われる。

最近では、文書ファイルを単にOA機器上で保存、再利用するだけでなく、汎用コンピュータやオフィスコンピュータに転送して、大容量ディスク上に集中管理する電子キャビネット機能を提供するOAシステムが開発されるようになった。

さらに、OA機器上で作成された文書が単に一箇所に集中管理されるだけでなく、登録された文書の内容について、論理的な情報の単位での共同利用を促進する文書データベース機能の提供が要求されている。

文書データベースを実現する上で、解決しなければならない技術的課題は、ホストコンピュータ上で管理

† OA Systems and Document Database by Kazuo TAGUCHI and Yoshihiko SAKASHITA (Computer Works, Information Systems and Electronics Development Laboratory, Mitsubishi Electric Corporation).

** 三菱電機(株)計算機製作所

*** 三菱電機(株)情報電子研究所

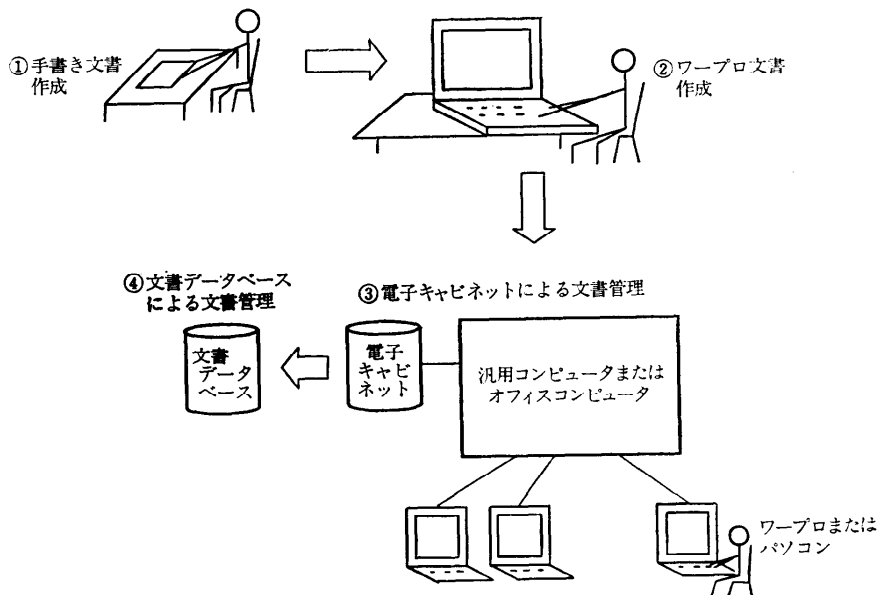


図-1 文書管理の発展

される文書データベースの構造設計, OA 機器との間で行われる文書交換のための標準的な手順や文書ファイル形式の規約化, 文書データベースをアクセスするための適切なユーザインタフェースの提供などである。

図-1 に, オフィスにおける文書作成や文書管理の発展を概念的に示す。

3. 要求機能

3 では, オフィス文書処理を中心とした OA システムに対する要求機能について述べる。

3.1 オフィス文書の作成と編集

取り扱える文書の種類としては, 一般的なオフィスで日常的に発生する文書, すなわち, 社内外への連絡文書, 報告書, 説明資料, 製品仕様書, 及び会議議事録などである。これらの文書の特長は, 単に文章だけにとどまらず, 説明用の図表や絵, 写真などが文書の中に含まれている点である。

オフィス文書の作成と編集機能については, まず OA 機器の上で, 日常的に使われている日本語のワープロ機能が必要である。たとえば, 各種の文字入力, 漢字変換入力, 文字サイズ/ピッチ, 行ピッチなどの指定, 禁則処理などの基本的な文書入力機能や, 文字単位, 行単位での追加, 削除, センタリング, アンダラインなどの文書編集機能である。

3.2 マルチメディア文書

すでに述べたように, 一般のオフィスで日常的に取り扱われている文書は, 単に日本語の文章だけでなく説明用の図表や, 絵や写真などが含まれている。また, 電子化された文書の中に文書の内容や図表を説明するための音声データを埋め込みたいという要求も出てきている。このように, 図形, イメージ, 音声などのマルチメディア情報を容易に入出力するためのハードウェアとそれらを有効に活用するためのソフトウェアが必要である¹⁾。

マルチメディア文書処理のためのハードウェアとしては, イメージリーダー/プリンタ, FAX, 音声入出力装置, 大量データを蓄積しておくための光ディスク装置などが必要である。

また, ソフトウェアとして, マウス/アイコン/タブレットなどを使って行われるデータ入力やコマンド入力を支援する機能や, 複数業務の動きを一つの CRT 画面上に表示するマルチウィンドウ機能などを含むユーザ・フレンドリなインタフェースを提供するソフトウェアが必要である。とくに, マルチメディア文書を効率良く柔軟に作成するための機能として, 入力された各種メディア情報を文書の 1 ページ上でどのようにレイアウトするかを CRT 画面上で簡単に指示できるソフトウェアが必要である。

3.3 文書管理

文書管理の第1ステップとしては、現実のオフィス環境と同じキャビネット／バインダ／文書という保管階層を通して文書を管理する電子キャビネット機能が要求される。キャビネットの種類には、OA 機器上のローカルキャビネット、ホストコンピュータ上の専用キャビネット、及び共用キャビネットがあり、共用キャビネットについては、アクセスの種類（参照／更新／削除）に基づく機密保護、安全保護などの機能が必要である。

電子キャビネット内の文書を検索する機能としては、登録時に指定する文書ごとの属性情報（作成日、タイトル、作成者名、文書番号など）による検索と、文書の中に含まれるキーワードによる検索（完全一致、前方一致、曖昧検索などを含む）が行えることが必要である。また、キャビネット内の文書を探す機能として、キャビネット／バインダ／文書の階層をたどるルート検索やキャビネットやバインダに格納されている文書の一覧表を表示して、その中から必要な文書を指定して取り出す一覧表検索も必要である。

そのほかの文書管理機能としては、文書単位でほかのバインダやキャビネットに移動する機能、複写する機能及び任意の文書をほかのユーザへ送信する機能などが必要である。

電子キャビネット機能に対する要求機能を図-2 に示す。

文書管理の第2ステップとして、文書データベースが要求されている。電子キャビネット機能が文書単位での蓄積、検索、抽出を行うのに対して、文書データ

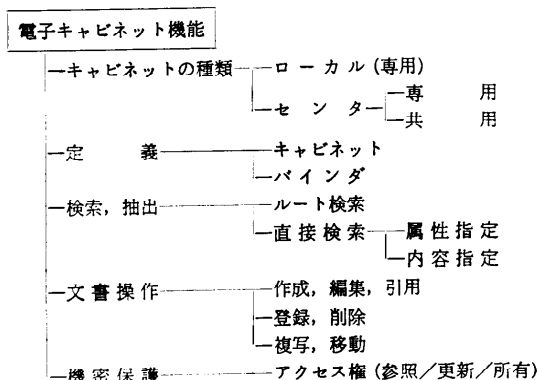


図-2 電子キャビネット機能

ベース機能は、文書内容を論理的なメディア情報単位で管理するので、文書の中に含まれている情報単位での流用、共用が柔軟に行える。また、実現方式の違いとして、マルチメディア情報を階層構造／関係構造（ポインタ）で管理するため、データの重複を少なくすることができる。しかし文書データベース構造を事前に定義しなければならないこと、文書単位での検索、抽出する際に、文書構成単位を集める必要があるため、電子キャビネット機能に比べて時間がかかるなどの課題も多い。ユーザ要求としては、文書単位の保管、検索とともに文書内容に対するマルチメディア文書構成単位での操作、管理が行える文書データベース機能の提供が望まれる。

文書データベースについては、汎用的なエンドユーザインタフェースとともにユーザプログラムからデー

表-1 文書データベースと電子キャビネットの違い

比較項目	文書データベース	電子キャビネット
利用目的及び操作単位	文書内容レベルでの共用と一元管理ができる。 文書内の論理（情報）単位（例、文章、図、表、など）で操作する。	文書レベルの保管、検索が中心。 文書単位で操作する。
検索方法	多種多様な検索方法としてキーワードによる直接検索、ブラウジング検索、パターンマッチング検索、意味検索、が行える。	文書保管階層をたどるルート検索（例、キャビネット→バインダ→文書）が主体。
内部構造	論理的階層構造。	物理的階層構造。
長所	- 文書の内容について共用レベルが高い。 - 重複が少ない。	- 現実のオフィスでの保管イメージに合っている。 - 文書内容を問わずに保管できる。
欠点	- 論理構造の定義が必要。 - 文書単位での操作効率が悪い。	- 文書内容レベルで共用不可。 - 文書内容レベルでの重複保持が発生するため外部記憶効率が悪い。

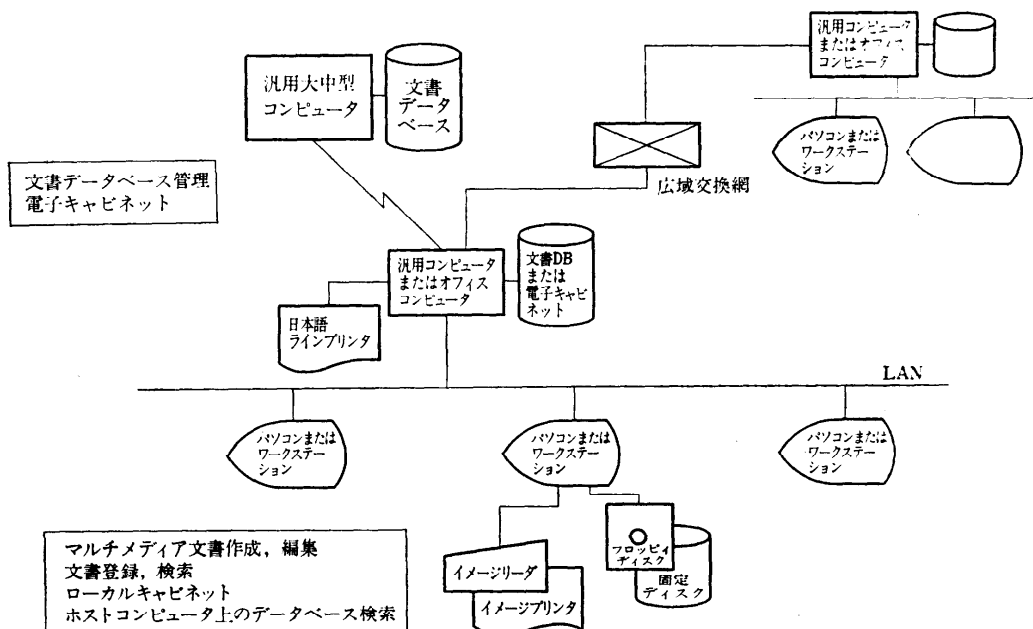


図-3 文書データベースの応用システム (概念図)

表-2 文書データベースの応用システムの機能

応用システム事例 (適用分野)	応 用 シ ス テ ム の 機 能
営業用プレゼンテーション文書作成管理 (流通業)	営業対象製品の写真 (イメージ)、説明図 (図形)、及び説明用の文章を含む客先説明用のプレゼンテーション文書の作成、編集、管理を行う。
技術文書管理 (製造業)	数値データ/図面 (図形)/写真 (イメージ)/説明用の文章、などを含むマルチメディア文書 (技術文書) の作成、編集、管理を行う。
社内一般文書業務の OA (一般企業)	通知、連絡用の文書作成、発信、会議室予約、内線電話帳作成、編集、配布などを行う。

データベース内容を検索できるプログラムライブラリの提供も要求されている。

表-1 に、電子キャビネット機能と文書データベース機能との違いを示す。

3.4 応用イメージ

上で述べた要求機能が提供された場合に実現可能と思われる応用システムの概念図を図-3 に、機能概要を表-2 に示す。

4. 実現技術

4 では、マルチメディア文書の形式、文書データベースの構造、文書データベースの検索と管理について、実現技術の事例と技術的課題についてのべる。

4.1 マルチメディア文書の形式

マルチメディア文書の表現形式としては、ISO TC 97/SC 18 で ODA/ODIF として標準化がすすめられている。ODA/ODIF の詳細については、本特集号の 3.1 で述べられているので参照されたい。ODA で表現される文書 (以下 ODA 文書) では、文章のほかに図形、イメージ、表データグラフ、音声などのマルチメディア情報を規定された表現形式で取り扱うことができる。

ODA 文書は、レイアウト構造と論理構造をもっている。レイアウト構造は、ページセット、ページ、フレーム、ブロックなどのレイアウト単位から構成される。ページセットは、連続したページの集まりで、論理的に関連するページをまとめたものである。ページ

は、通常の書籍などというページに相当し、フレームまたは、ブロックから構成されるく(矩)形領域である。フレームはページ内のブロックの集まりでフレームまたは、ブロックから構成される矩形の領域である。ブロックは、レイアウト構造の最小単位で、ただ一つだけのメディアからなる矩形領域である。図-4に ODA 文書のレイアウト構造の例を示す。

一方、論理構造は、序文、本文、あとがき、章、節、項、段落などの論理的単位から構成される。これらの論理的単位は、木構造によって表現される。図-5に ODA 文書の論理構造の例を示す。ODA 文書の論理構造は、レイアウトを意識することなしに、文書の作成、追加、削除、訂正などができるという利点をもたらす。文書データベース・システムでは、内部的には、要素単位で保管されているデータベース内容を、ODA 形式で構造化された文書の形式でユーザに見せることになる。

4.2 文書ファイルのデータベース化

4.2.1 データベース化の発展

(1) 初期段階

ワードプロセッサ (WP)、パーソナルコンピュータ (PC) 及びオフィスコンピュータ (OAP) などで作成された文書がホストコンピュータ上に格納され管理され、文書ファイルが集合したデータベースとして実現される。しかし文書ファイルの同一部分の情報は共有されずに文書ごとに複製され保管される。この段階が今日の多くの文書ファイルの実態である (図-6)。

(2) 文書データが共有されるデータベース化

前項に述べた文書構造もデータとしてデータベースの中に定義され、さらにおおのの情報本体はデータ

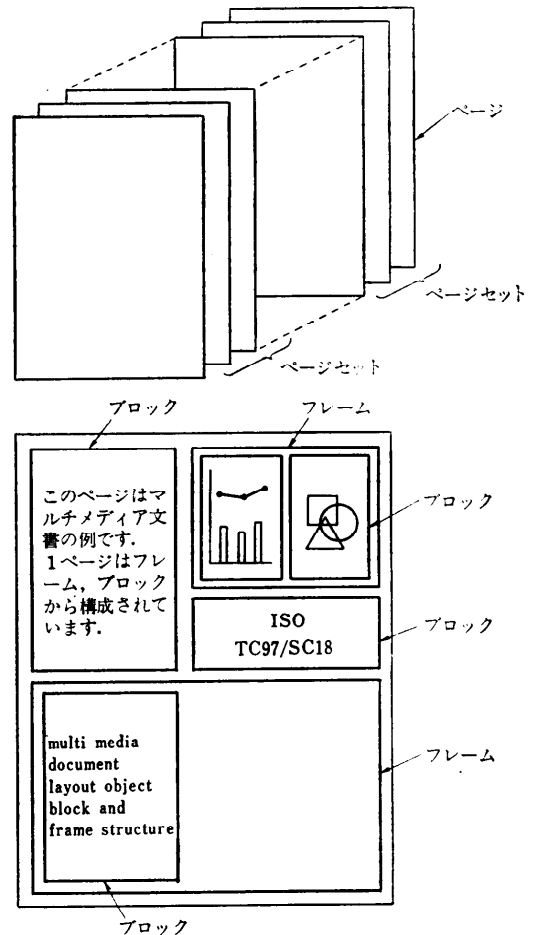


図-4 ODA 文書のレイアウト構造

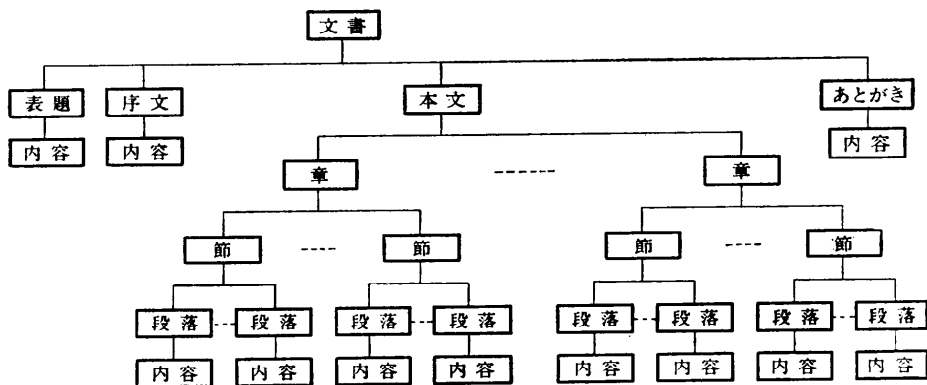


図-5 ODA 文書の論理構造

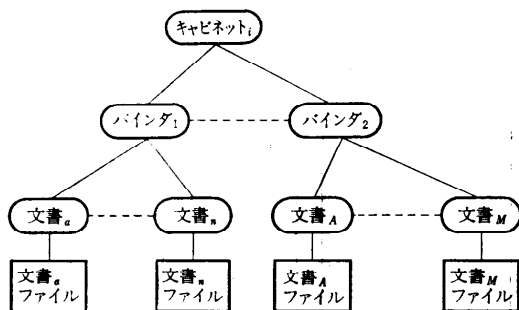


図-6 文書ファイル

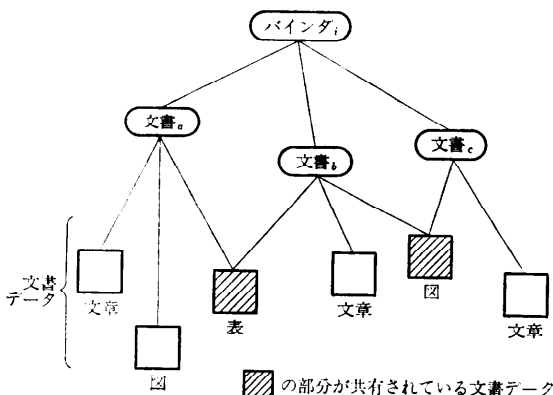


図-7 文書データが共有される文書ファイル

ベースの中に格納・管理される（図-7）。この段階では複数文書間で独立したデータを備える必要はなく、同一のデータを共有する。

(3) 構造を備えた文書のデータベース化

次の段階では、文書は前章の文書管理に述べた文書構成、すなわち論理またはレイアウトの構造を備え、この構造情報とマルチメディアを含む各種情報が結合された文書として格納・管理される文書処理システムが²⁾ネットワーク結合したワークステーションシステムの上に実現されている例がある。そのシステム構成を図-8に示す。この構造に適する図、イメージなどのマルチメディアを考慮したファイル構造が MIDOP³⁾にて提案されている。そこではおのおのの構成要素に対する各種情報はファイルまたはレコードの単位で管理されている（図-9）。しかしこの段階においても情報本体はおのおのの独立に作成、格納そして管理される。表は情報を効率良く表現する手段として多く利用され、そこにはあらゆる種類の形態をした情報が詰め込まれる。この表の構造とメディアを意識したファイル構造の提案は難しい。なぜならば、表自体の syntax を表現する手段と、表がもつ semantics を表現する

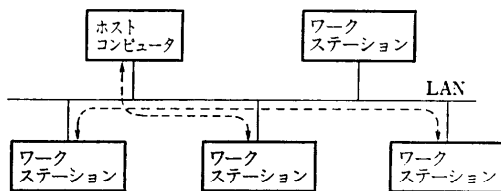


図-8 ネットワーク結合したワークステーション

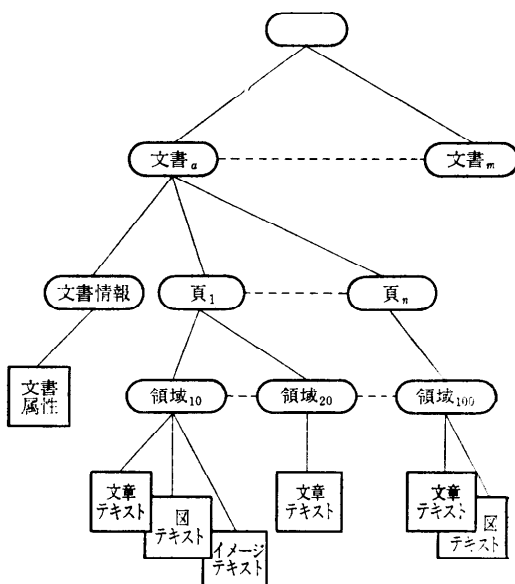


図-9 構造を備えた文書のデータベース

手段とをもち合わせなければならない。すなわち表の形式を定義することと、定義された表の中の各フィールド間の関係と、この各フィールドに表される各種メディアの情報の関係を統合的に表すことが要求される。

(4) 汎用的文書データベース

本来の文書データベースはデータベース内の文書データは文書という枠を越えて、一般のデータベース管理システムからの操作の対象ともなる（図-10）。すなわち文書処理の環境において一般のデータベース管理システムの環境で作成されたデータを利用して、文書という枠の中の情報として扱うことができ⁴⁾、またこの逆も可能である。

4.2.2 データベースの構築

データベースシステムを構築する構成概念を図-11に表す。システム[3]では基本文書処理サービスにおいては格納情報に関する属性情報を備え、この情報に基づいてデータをアクセスする（図-12）。また表の

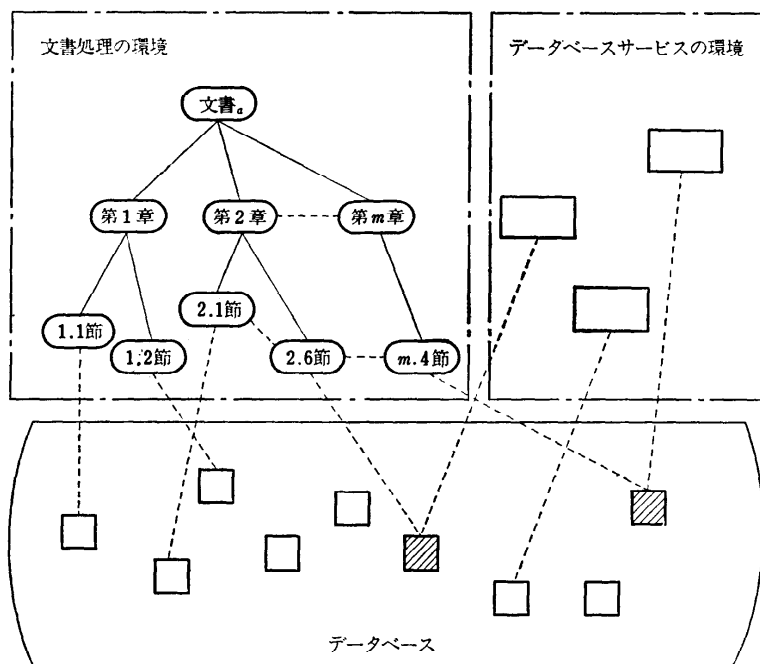


図-10 汎用的文書データベース

例のように情報自身が備えている構造をどのように表現するかはまだ十分に研究されていない。文書の構造を表す情報は辞書スキーマで管理されるか、または構造情報をデータとしてデータベースそのものの中に格納し、文書情報と関連付ける方式とがある。この際文書内容に対応する情報を文書とは独立に操作できるように実現する。

文書の目次を先に述べた論理構造に対応させて、目次を作成・編集の対象とし、文書内容は従来のデータベースに別途作成・編集をあらかじめ行っておき、この文書内容に対して結び付けるアウトラインプロセッサがすでに製品として存在する。

4.3 文書データベースの検索

(1) 文書の検索

(a) 文書属性の情報に対して

多くのワードプロセッサにある検索の方法はいわゆる完全一致によるもので、文書名、著者名、作成期日などを検索キーとして行う。これらの情報は文書属性の中にあるので、検索の際には当該検索キーと合致する情報をその中から探す。システム[3]では、キャビネット、フォルダそして文書の階層的レベルのファイル体系となっている。検索のユーザインタフェースの例を図-13に示す。ネットワークに結合されたワー

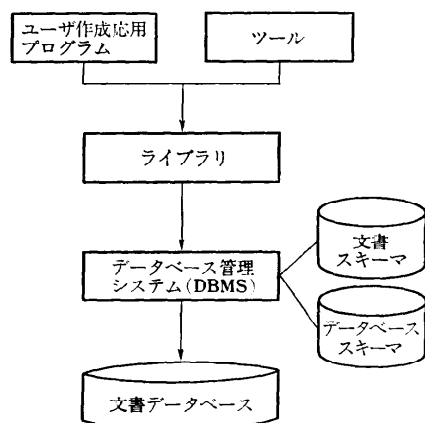


図-11 文書データベースのアクセス環境

クステーションシステムのため任意のワークステーションにある文書データへのアクセスが可能となっている。

(b) 文書内容に対して

文書に記載されている内容に関する検索が研究されている⁵⁾。今日ではいまだ形態的な処理のレベルであるが、文書表題、キーワード、文字列、図・絵の一部のパターンを検索キーとして文書内容に対して検索を

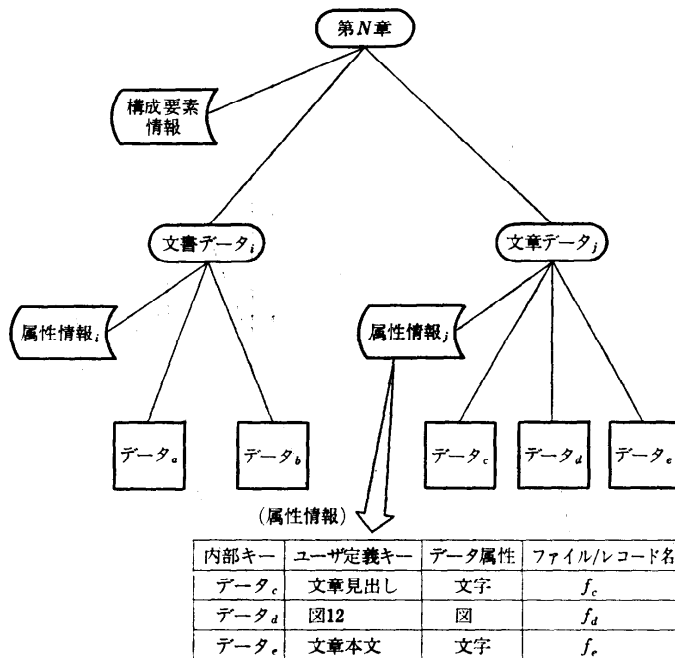


図-12 格納データの属性情報によるデータアクセス

実施する⁶⁾。これは検索結果を文書の編集及び推稿の作業における支援に大きな効果がある。将来的には文書内容そのものが保持している意味を把握し理解して、検索の質問も断片的な知識・情報から知識の体系化をととして検索できるようになると考えられる⁹⁾。

(2) 文書データに対する情報の検索

文書とは独立に文書データとなる情報を通常のデータベース管理システムを利用して検索する。すなわち利用者はデータベースサービスを用いて特定の情報を利用することも、文書の環境からその情報を利用することもできる。ネットワークに結合したワークステーションを対象としたデータベース管理システムの研究も盛んである^{7), 8)}。

4.4 文書データベースの管理

(1) 文書の版の生成と管理

文書における版の管理はきわめて重要である。主な制御は次のものである。

- あらかじめ定める方法で版の名の決定を行う。
- 生成される各版の管理を行う。
- 版の指定及び参照に対応する。

文書データを文書間またはほかとの間で共有する文書データベースシステムにおける問題はデータの更新

の可否またその判断基準、さらには更新の履歴の保持方法、そして更新事象の通知先、などなど解決すべき課題も多い。

(2) 同時制御

同一の文書または文書内容に対する複数ユーザによる同時アクセスに対する制御を行う。この制御は一般のファイルシステムにおける排他制御と同様である。ただ制御の単位が文書に依存する、たとえば複数文書、単一文書、レイアウト構造または論理構造の構成要素文書内容を更に分割した要素などの単位で排他制御を行う。

(3) 情報の共有

データベースシステムを利用する最大の利点は情報の共有が可能となることである。同一の情報を利用する場合に、新たに複写し記憶域を専有することなく利用でき、また共有する情報が更新された場合に新たな処理をせずに更新された情報を得ることができる。

逆に欠点もある。更新の必要もないのに更新されるまたはその有無を意図的に知る必要を生じる。この場合は先に述べた版の指定を陽に指定または参照する必要がある。

文書処理

文書名/様式名 : _____
 フォルダ名 : B フォルダ
 キャビネット名 : C キャビネット
 ステーション名 : _____

文書作成	フォルダ作成	表示	複写
様式作成	キャビネット作成	印刷	移動
削除	アクセス権変更	終了	

↓ <表示>をクリック

文書一覧

ソートキー: 名 称 作成年月日 作成者 次ページ 前ページ 閉じる

種別	電子ファイル名	作成年月日	作成者
A1 文書	XXXXXXXX	YYY	
A2 文書			
A3 文書			

↓ <A1 文書>をクリック

文書処理

文書名/様式名 : A1 文書
 フォルダ名 : B フォルダ
 キャビネット名 : C キャビネット
 ステーション名 : _____

文書作成	フォルダ作成	表示	複写
様式作成	キャビネット作成	印刷	移動
削除	アクセス権変更	終了	

クリックした文書名がエコーバックして表示される。

図-13 文書データの検索ユーザインタフェース例

5. む す び

オフィスで作成される文書を中心にして、その文書データベース化についてユーザ要求を上げ実現技術とそこに存在する課題について論じた。データベース上に文書環境を構築する動きはまだ始まったばかりである。ホスト計算機において集中的に文書データが集積

格納し、データの管理は従来のデータベース管理システムが行っているのが大勢である。

計算機システムにおける情報表現の手段としての文書はアナログ的にも有効であり、より一層その価値は広がる。この意味においても文書データベースの存在価値は大きなものがあるが、その実現に向けた研究開発はこれから本格化するものと期待される。

参 考 文 献

- 1) 中村, 溝口, 田口: マルチメディア DBMS モデルの考察, 情報処理学会第 32 回全国大会論文集, 2B-5 (1986).
- 2) 田中, 土田, 坂下: マルチメディア文書処理 (論理構造), 情報処理学会第 32 回全国大会論文集, 2K-2 (1986).
- 3) Sakashita, Y. et al.: Document Processing System (MIDOP), ICTP '83 J1 (1983).
- 4) Woeik, D., Kim, W. and Luther. W.: An Object-Oriented Approach to Multimedia Databases, ACM SIGMOD, pp. 311-325 (1986).
- 5) 鈴木他: 日本語文書校正支援システム CRITAC, 情報処理学会研究会資料, 86-JDP-8 (1986).
- 6) 宇津宮他: 日本語文書の構造的作成支援環境構築の試み, 情報処理学会「日本語文書の入力と編集」シンポジウム (1985).
- 7) 藤澤他: 高度ファイリングの理念と要素技術—文書理解と知的ファイリング—, 情報処理学会研究会資料, 86-JDP-7 (1986).
- 8) 金森他: ワークステーションデータベース管理システム—WS・DBMS—, 情報処理学会データベースシステム研究会資料, 57-6 (1987).

(昭和 62 年 2 月 13 日受付)