

## 解説



## 画像データベース

## 文書画像データベースシステム†

飯 沢 篤 志††

## 1. はじめに

オフィスでは、社内外への連絡用文書、報告書、管理用文書、会議議事録、伝票、説明資料、パンフレット・カタログ、手紙など多種多様な文書が活用され、文書が重要な位置を占めている。また、複写機やファクシミリ (FAX)、ワープロ (WP) という OA 機器が普及するにつれてますますオフィスで生成される文書の量が増えている。

文書を整理するために、キャビネットなどが利用されてきた。しかし、検索手段が、タイトル、分類、所有者程度のものしか提供されないために、既存の文書を検索するのに手間がかかり、情報の共有化や再利用が難しい。そのため重複文書や不要文書が多くなり、スペースを浪費してしまう。そこで、文書を電子的媒体に置き換え、データベース (DB) で管理し、共有化と省スペース化を図り、オフィスの作業効率を高める動きがある<sup>1)</sup>。

文書 DB を構築する上で画像との関係を考えるに次のような二つの立場がある。

- 文書を画像として蓄積する
- 文書の構成要素として画像が含まれる

前者の立場には、光ファイリングシステムの例がある。後者は、文書をテキスト、図、写真などの構成要素に分解し管理するマルチメディア DB (MMDB) として管理しようという立場である。

光ファイリングでは、文書をすべて画像として入力し、現実のキャビネットに類似した内部構造をもつ光ディスク装置に蓄積する。検索手段としては、キャビネット同様のキャビネット、ドロア (引出し)、フォルダという階層構造をたどる操作

とキーワードやタイトル、作成者、日付などの付加情報による方法が提供される。しかし、文書を画像データとして扱うだけではなく、文書作成システムとの融合や、WP のファイル管理との連動など、次第に MMDB としての文書 DB が要求されてきている。

最近ローカルエリアネットワーク (LAN) が普及しつつあり、パーソナルコンピュータ (PC) やワークステーション (WS) が LAN に接続され、LAN を通して他のマシンのファイルをアクセスすることができるようになった。この分散環境では、文書も分散して管理し、効率よく共有化や再利用することが望まれる。

本稿では、上記の背景を踏まえ、オフィスにおける文書の特徴を考察し、それを蓄積、管理するデータベースに要求される機能やその実現上必要とされる関連技術について解説する。最後に文書データベースのシステム例を紹介する。

## 2. オフィス文書の特徴とデータベース機能

オフィス文書の特徴と、紙を媒体とする問題点、DB を利用する長所、データベース管理システム (DBMS) にとっての課題などを説明する。

## 2.1 マルチメディア

オフィス文書は、テキスト、図形、画像、グラフなどのマルチメディアデータ (MMD) により構成され、また、手書き、WP、机上出版 (DTP)、FAX など多様な手段により生成、獲得される。DB を構築する上では、MMD に対する効率のよい蓄積方法や高速のアクセス方法とデータの入りや表示、印刷の手段が問題となる。

一般に入力、蓄積、出力それぞれでデータの表現方法が異なるので、DBMS としては、その表現形式に独立に管理できなければならない。異質なデータを同質化し統一するにはすべてをオブジェ

† Document Image Database Systems by Atsushi IIZAWA (RICOH Co. Ltd., Software Div., Software Research Center).  
†† (株)リコーソフトウェア事業部ソフトウェア研究所

クトとすると扱いやすいので、MMDBのベースとしてオブジェクト指向データベース(OODB)を利用することが多い<sup>2)</sup>。

OODBでは、文書をタイトル、要約、本文、図などの部分で構成された集約(aggregation)階層を表現する複合オブジェクトとして扱うことができる。ただ、個々のオブジェクトの構造ではなく、文書スキーマとしての構造を規定する方式を導入し、そのスキーマ自身の版管理や更新管理を行う必要がある<sup>3)</sup>。また、文書間や各構成要素間での参照関係をいかに表現するかも問題となる。

DBMSのアーキテクチャ上の課題としては、長大データの効率的取り扱い方があげられる。実用システムでは、長大データをDB内ではなく外部ファイルに蓄積し、DBにはそれに対するポイントだけ格納する方式がよく使われる。しかし、この方式ではDBと外部ファイルとの整合性をDBMSでは管理できないので、アプリケーションプログラムに負担をかける。DBMSが直接長大データを管理し、同時実行制御やバックアップ、アクセス権制御というDBMSの基本機能を長大データに対しても提供すべきである。

DBMSで通常のレコードと長大データを同時に扱う場合は、バッファ領域を長大データで占めないようにバッファリングの機構を考えなければならない<sup>5)</sup>。また、長大データに対して、テキストの全文検索などの特殊な検索を必要とする場合には、それに適したファイル構造<sup>6)</sup>を構築し、その検索手段をDBMSに組み込めるような仕組が必要となる。

## 2.2 分散管理・分散処理

オフィスでは、通常つきのように文書が作成され、管理される。個人または共同作業により原案が作成される。作成された原案は、回覧や会議などでコメントをもらい、修正される。何度か修正されて完成した文書は、上司を含む複数人の検印を受けて、正式に承認され、運用・管理される。文書の重要度に応じて、個人、課、部、全社など組織階層内のどのレベルで管理されるかが決まる。頻繁にアクセスされる文書は、複製をなるべく手もとの近い場所に置き、アクセス時間を短くするという運用もされる。その際紙による管理では、実際に複製するので、重複文書を管理することになり、整合性をとる作業も膨大である。DB

化する最大の長所は、この分散管理をなるべく省力化することにある。また、作成段階からDBに格納することにより、作成、回覧、検印の効率化の可能性もでてくる。

最近では、課などの小さい単位で文書を管理し、ネットワークを介して接続するような分散システムを構築することが多い。それに利用するDBMSでは、分散したマシン上のデータを管理する機能が必須である。また多量の文書を管理する必要があるので、1カートリッジあたり600MB程度の光ディスクや光磁気ディスク、さらにそのメディアを複数管理するジュークボックス型の格納装置をネットワークを通して利用できる必要がある<sup>7)</sup>。

分散環境では、常に複数人がDBにアクセスする可能性がある。文書に含まれるデータのうち画像データなどは、容量が大きく、1回あたりのアクセス時間も長いので、同時実行制御に気を配らないと待ち時間が長くなる。

## 2.3 保存期間と廃棄

オフィス文書は、種類によって保存期間が定められているか、登録時に保存期間を定める。保存期間が過ぎた文書については、継続して保存するかどうかをその時点で見直すことになる。このような方法によって、不要文書の増加を抑えている。

DB化することで紙に比べて1文書あたりの空間占有率は小さくなるので、DBでは紙より保存期間を長くできる。しかし、1文書あたりの空間がゼロではないので、不要文書データを捨てる操作が必要となる。データ生成時にその寿命を定め、廃棄時期を予測することで、保管場所を決定するためのパラメータとしたり、全体の文書の増加量を見積もることができる。

## 2.4 機密ランク

オフィス文書には、カタログなど社外に公開して構わないものから、社外秘のもの、特定部署の人のみアクセス可能なもの、特定ランク以上の人のみアクセス可能なもの、などさまざまな機密方式が定められている。

文書にアクセスする人が、その文書に対してどのような権限をもっているかを確認し、アクセスを制限しなければならない。そのためには、ユーザ管理と識別、各文書の機密ランクを厳密に管理

する必要がある。ユーザ識別には、現状では社員証などの磁気カードを使うことが実用的だが、機密を重要視する場合には、もっと高度な識別技術を使う必要がある。また、期限付で特別なアクセスを許すという制御も必要であろう。

## 2.5 版管理

文書の版管理では、文書が時系列に従って改訂されていく版 (historical version) の管理と、一部だけが異なり同時に使われる版 (alternative version) の管理がある<sup>9)</sup>。改訂が続けられる文書に対しては、ある程度古い版の文書もアクセスできる必要がある。古い版は通常アクセスする必要がないが、なんらかの理由でアクセスされることがあるので、古い版をどの時点で廃棄するかは運用上かなり難しい問題である。

版管理の対象となる文書は、異なる版で共有できる部分が多いので、版の差異を管理すればよい。ところが、紙では、差異のみで管理しにくいので、版ごとに全体を置き換えることが多く、不要な文書を増やす原因となる。

文書 DB では、文書を構成要素に分解し木構造で表現すると、部分木を共有することで異なる版を表すことができる。共有する形態としては、その時点での複製と考えるか部分の参照と考えるかで共有部分を変更するときの操作が異なる。

## 3. 文書データベースの関連技術

文書データには、多種のデータが使用され、その間の相互変換を行う必要がある。また、同種のデータであっても表現形式を変換する必要がある。文書 DB を構築するためには、DBMS 以外にも重要な技術が必要となる。以下では、主な関連技術を概説する。

### 3.1 画像処理

オフィス文書の特性から、扱う画像はモノクロデータが多いが、今後カラーの入出力機器が普及し、格納装置の容量も大きくなるとオフィス文書でもカラー画像を利用するようになることは必ずである。

文書 DB に関連する画像データには次の操作が必要となる。

- 拡大, 縮小, 回転, 切り貼り, 左右反転
- 2 値画像, 多値画像, カラー画像間の変換
- 画質向上, 輪郭線抽出など

文書を表示、出力するためには、各機器の解像度に画像を変換しなければならない。たとえば、ビットマップディスプレイでは 75 から 100 dpi, レーザプリンタでは 240 dpi, 300 dpi, 400 dpi, FAX では G3 タイプで 200 dpi, G4 では 400 dpi などさまざまな解像度をもつ機器に画像を出力しなければならない。この操作は、画像の拡大・縮小と同等である。写真のような中間調やカラーで表現すべき画像を画素密度によって表している 2 値画像については、擬似的な濃度を補完するのが困難なので、画質を保ったまま拡大・縮小することがかなり難しい。

生の画像データは大きいので、格納・伝達する際は圧縮することが必須である。FAX の規格では、完全に元のイメージを復元できるデータ圧縮法を用いている。画像データの圧縮伸張方式については、CCITT (国際電信電話諮問委員会) が規定しており、MH (Modified Huffman) 方式、MR (Modified Read) 方式、MMR (Modified Modified Read) 方式という三つの符号化方式がある<sup>8)</sup>。これらの方式では、通常本文書原稿においては、走査線上には黒画素や白画素の連続することが多いことを仮定し、連続する画素数を可変長コードに変換している。

写真のような多階調画像を 2 値画像として表現するために中間調処理や誤差拡散処理を施した画像に対して、これらの圧縮方式を適用すると符号化効率が極端に悪くなり、多くの場合、元の画像より圧縮したほうがデータ量が増えてしまう。そのため CCITT と ISO の統合機関である JBIG (Joint Bi-level Image Group) で、2 値画像の新しい符号化方式を検討中である<sup>9)</sup>。

カラー画像を蓄積・伝達するには、圧縮技術が重要な課題である。カラーの自然画像については、JPEG (Joint Photographic Expert Group) で検討されている符号化方式が、早ければ 91 年末に国際標準として登録される<sup>10)</sup>。この方式では、データ量を 20 分の 1 以下に圧縮できるが、情報の欠落により元のイメージは完全には復元できない方式である。

### 3.2 画像認識

文書を DB に入力するとき、原本が紙で提供される場合は、スキャナで入力する方法がある。また、FAX で送られてくる画像データを直接に DB

に入力する場合もある。そのような場合、画像を認識しコードデータに変換することで、入力処理の自動化や省力化、データ容量の減少、検索機能の強化を図ることができる。ここでは、オフィス文書にとって重要な OCR (Optical Character Reader) と書式認識の技術を説明する。

OCR は、画像データからテキストデータへ変換する技術である<sup>11)</sup>。現在、活字の日本語と英数字が実用化されている。手書きの英数字も実用化段階に達しているが、使用条件が限定されていることが多い。手書き日本語は今後の研究課題である。

大量の文書を画像で入力し、OCR を使って全文 DB を作成する試みもある<sup>12)</sup>。その場合問題となるのは、OCR の認識率である。文字の形だけでは判別が難しい場合 (たとえば、「一」と「|」) もあるので、認識された文字列を自然言語処理を使って訂正する必要がある。また、1種類のアルゴリズムによる OCR では、認識率に限界があるので、複数の OCR を同時に実行し多数決をとることで、誤りを訂正し認識率を上げるという提案もある<sup>13)</sup>。しかし、認識率を 100% にすることは難しいので、実用的には OCR で得られた結果を効率よく修正するエディタを組み合わせる必要がある。

オフィスで使われる文書の中には、伝票や登録用紙などの定型書式がある。その書式に記入されたデータを DB に自動入力するためには、入力された画像データがどの書式かを認識する必要がある。書式が認識できれば、必要なデータの欄が分かり、OCR を使ってその欄のデータを自動的に抽出することができる。書式を認識する方式としては、以下の方法が考えられる。

- (1) 特定の位置の文字列を OCR で認識する
- (2) 特定の位置のバーコードを認識する
- (3) 書式全体の特徴により認識する

使う書式を新しく作成してよいなら、(1)、(2)の方式が有力な候補である。既存の書式を認識する場合には、(3)のように書式の特徴を抽出しなければならない。そのひとつとして、縦横の罫線から特徴を抽出する研究がある<sup>14)</sup>。

### 3.3 文献検索

オフィス文書は、保存期間の過ぎた文書は削除されることや厳密に機密保護を管理する必要があ

るなど、文献検索の対象となっている文献とは異なる性質をもつ。しかし、付加情報を利用した検索を考えるとときには、文献検索の分野の成果は大いに参考になる。

文書に対して、キーワードを見つける作業は、主題の把握、概念の把握、キーワードの付与というステップを踏む<sup>16)</sup>。文書の本文からキーワードを自動的に付与するという研究も多い<sup>17)、18)</sup>。しかし、文書中にまったく現れない語をキーワードとするべき場合もあり<sup>19)</sup>、完全な自動化は今後の課題である。

キーワードを一定の語彙から選択するという方法もある<sup>15)</sup>。その語彙集をシソーラスと呼ぶ。シソーラスは、語の同義関係、階層関係、およびその他の関係を示したものであり、検索において標準的な語彙を与える。

ただし、一般オフィスでは、文書のキーワードを付ける際に、シソーラスで定める語を使うのは難しい。文献検索においては、文献の登録は専門家が行うが、オフィス文書を登録するのは専門家でないためである。そこで、オフィス文書 DB を構築するときには、キーワードは自由につけるようにし、それらのキーワードからシソーラスを構築していくという運用を考えるべきである。そうすることによりシソーラス自体も共通の DB として運用・管理し、日常業務での更新の対象となる。

## 4. システム例

文書を扱うデータベースのシステム例を紹介する。ここで紹介するシステムは、オフィス文書に限ったものではないが、一般文書データベースの実例を見ることで、今後、オフィス文書にどのように適用できるかを考える基礎にできると考える。

### 4.1 ホスト WS 連携のシステム

安田らは、ホスト計算機で文書管理をし、それに接続された多数の WS で文書作成、編集を行うというシステムを実現している<sup>20)</sup>。ホストには、マルチメディアデータを統一的に管理するための論理的な格納構造としてキャビネット、ドロア、フォルダという階層構造をもつ文書ファイルが装備されている。そのデータを組み合わせて、複数人で長大文書の作成や大量文書の管理・運用

を行うシステムを実現した。

文書の性質と文書管理の形態によって以下のよ  
うな三つの運用モデルにパターン化している。

**構造化文書モデル**：長大文書を複数人で共同作成  
するモデルである。全体を部分構造に分割して別  
別に管理し、最後にそれらを結合する。

**体系的文書群モデル**：管理者が社内規格にあった  
文書群の体系を決定し、その文書群を作成してい  
くモデルである。

**流動的文書群モデル**：多数の文書群が作成、蓄積  
された後、試行錯誤的文書体系を作成していくモ  
デルである。

この運用モデルをもとに長大文書、大量文書の  
文書管理に必要な共通機能として次のような機能  
を抽出している。なお、ここでいう「クラス」と  
は論理的分類のことである。

(1) 単一文書の操作

- 文書の登録、取出し、置換、削除、複写、  
移動

- 文書同士の関係付け

(2) 格納構造の操作

- 格納構造の作成、変更、削除、複写、移動

(3) クラス単位の操作

- クラス単位の登録、取出し、置換、更新、  
削除

- クラス単位の複写、移動、分離、結合

- クラス単位の文書管理情報の集計

(4) 検索

- 階層的検索
- 管理情報による検索
- テキスト検索

(5) 版数管理

- 版数の更新
- 版数指定の文書取出し、置換

(6) 文書資源の保全

- 文書資源の退避、復元
- 文書資源の機密保護
- 文書資源の異常時のリカバリ

文書データベースは、論理構造、格納構造、物  
理構造の3階層で構成される。論理構造は、リ  
レーショナルDBの部品展開表の考え方によって  
実現され、文書の階層構造の管理、文書間の関係  
付け管理、版管理を行っている。格納構造は、文  
書の構成要素を論理的な入れ物に対応付ける構造

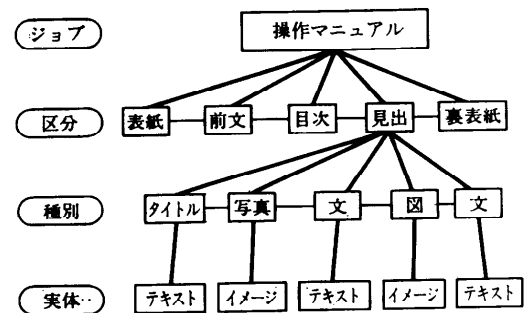
である。物理構造は、物理的資源と格納情報に対  
応付けている。格納構造と物理構造は従来の文書  
ファイルで提供される機能である。

文書処理を行うアプリケーションはWS上で稼働  
し、対象となる文書構成要素をホストで検索し、  
取り出して処理を行う。文書の修正はWSに取り  
出して行うので、修正中の文書構成要素は「更新  
状態」となるように管理される。従来のDBMS  
が提供するトランザクションだけでは、文書修正  
のような長時間にわたるトランザクションを実現  
することはできない。このシステムでは、文書の  
階層構造のノードの状態を変更することで実現し  
ている。

## 4.2 CAMMS

マニュアル作成・管理においては、改訂作業が  
多く、新しい版では70~80%が再利用できる  
ので、部分文書の共有化が重要な課題となる。  
CAMMS<sup>21)</sup>は、マニュアル原稿の構成部品の関  
連を管理し、ひとつのマニュアルに必要な原稿を  
組版システムに渡すという機能をもっている。  
CAMMSへの原稿入力には、SGML<sup>22)</sup>の考え方  
を取り入れた独自のマークアップ言語を使う。ま  
た、DB内の原稿をマージして新しい原稿を作成  
する機能もある。

CAMMSで扱う論理構造は、ジョブ、区分、種  
別、実体という4階層の構造をもつ(図-1)。ジョ  
ブはひとつのマニュアル全体に対応し、区分の列



□ インスタンスオブジェクト

○ クラスオブジェクト

— 親子関係

— 順序関係

図-1 CAMMSにおけるインスタンスオブジェクト間の関係<sup>21)</sup>

がマニュアルの論理的構成を表す。実体には、テキスト、イメージ、CAD 図面などのデータが対応する。区分と種別内のオブジェクトには、線型順序関係をもたせている。また、区分と種別においては、共有ノードを実現するために、他のノードを間接的に指すような仮想オブジェクトを導入している。

階層構造などの構造を表現するには、それを表現するのに適した DBMS を利用することが望ましい。CAMMS は、リンク指向 DBMS G-BASE<sup>23)</sup> 上に実現されている。G-BASE は、リレーショナルモデルを拡張したグラフデータモデル<sup>24)</sup>に基づいて開発された DBMS である。G-BASE では、実体を表すレコード同士をリンクで関連を付けることができ、Chen の実体-関連モデル<sup>25)</sup>と同等の表現力をもつ。CAMMS では、このリンクを用いて、階層の親子関係と同階層内での順序を表現している。

4.3 MASCOT

MASCOT<sup>26)</sup> は、自動車の整備マニュアルを作成管理するシステムである。整備マニュアルは、分量が多く、一定の構造をもち、複数のマニュアルで共有される部分が多い。また、他の言語に翻

表-1 文書アクセスメソッド (文献 26) より)

	分類	メソッド名	機能
基本機能	論理構造の操作	select-child	子ノード選択
		select-descendent	子孫ノード選択
		select-parent	親ノード選択
		select-ancestor	先祖ノード選択
		delete-node	ノード消去と関連リンク更新
		delete-tree	指定ツリー消去
		update-link	is-part-of リンク更新
		copy-tree	ツリー複製
		move-tree	ツリー移動
		create-child	子ノード生成
応用	ページレイアウト	layout	レイアウトルールを使い文書要素レイアウト OASYS-EP 形式で出力
機能	機械翻訳	translate	ATLAS と日米テキスト間のリンクを使い翻訳

訳されることも多い。このような文書を扱うためには、DBMS に次のような機能が必要である。

1. マルチメディアの文書構成要素の管理  
マルチメディアデータを格納し、操作すること。また、キーワードを使って文書構成要素を検索できること。
2. 論理構造  
階層構造を表現し、その階層をたどる操作が高速に実現できること

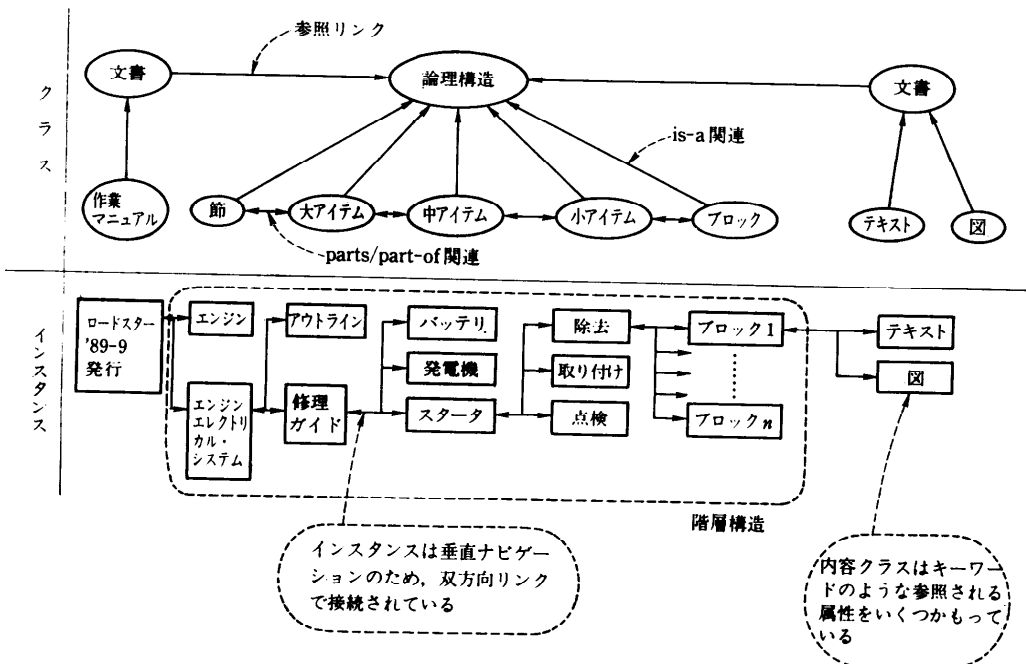


図-2 MASCOT における作業マニュアルのクラス構造<sup>26)</sup>

### 3. 参照構造

データの関連付けをし、関連するデータをたどること。参照するときに指定された変換をしたり、他種のデータベースをアクセスすることができること。

MASCOT では、これらの要求を満たす DBMS として Jasmine というオブジェクト指向 DBMS を利用している。

文書は、文書クラス、論理構造クラス、内容クラスから構成される (図-2)。論理構造のインスタンスは、階層構造を構成し、親子関係は双方向リンクで接続される。文書アクセスのためのメソッドとしては、表-1 が用意されている。テキストにデータ操作の手続を埋め込み、参照されたときに他種のデータベースにアクセスし、データを得る機能もある。特殊な変換として、日本語-英語間の機械翻訳をサポートしている。このように参照や変換操作をメソッドとして埋め込むことにより、複雑な構造をもつ文書の作成・管理を容易にすることができる。

### 5. おわりに

従来、紙で管理されてきたオフィス文書をデータベースに蓄積することは次世代 DBMS の大きな応用のひとつである。本稿では、その実現に必要な技術課題について解説してきた。課題には、DBMS に関するものとともに関連技術の課題も多い。最近になりマルチメディアを扱う技術が進歩してきたおかげで、オフィス文書のデータベースもやっと現実のものとなりつつある。今後は、実際のシステムを構築し、現実で利用してみた上で文書モデルや DBMS のアーキテクチャに反映させるような段階である。

**謝辞** 本稿を丹念に読んでさまざまなご指摘をいただいた同研究所に勤務する白田由香利氏に感謝いたします。

### 参考文献

- 1) 田口, 坂下: OA システムと文書データベース, 情報処理, Vol. 28, No. 6, pp. 721-729 (1987).
- 2) 増永: 次世代データベースシステムとしてのオブジェクト指向データベースシステム, 情報処理, Vol. 32, No. 5, pp. 490-499 (1991).
- 3) Woelk, D. et al.: An Object-Oriented Approach to Multimedia Databases, Readings in Object-Oriented Database Systems, Zdonik, S. B. and Maier, D. (ed.), pp. 592-606, Morgan Kaufmann Publishers (1990).
- 4) Masunaga, Y.: Design Issues of OMEGA: An Object-Oriented Multimedia Database Management Systems, Journal of Information Processing, Vol. 14, No. 1, pp. 60-74 (1991).
- 5) 牧之内: オブジェクト指向データベース, 情報処理, Vol. 32, No. 9, pp. 1032-1040 (1991).
- 6) Faloutsos, C.: Access Methods for Text, ACM Computing Survey, Vol. 17, No. 1, pp. 49-74 (1985).
- 7) Shirota, Y. et al.: Image Database Construction Tools for RICOHBASE, Proc. of IEEE Pacific RIM Conference on Communications, Computers and Signal Processing, pp. 277-283 (1991).
- 8) CCITT: Standardization of G. 3 Facsimile Aparatus for Document Transmission, Recommendation T. 4 (1980, 1984).
- 9) 山崎他: 2値画像階層形符号化方式 (JBIG アルゴリズム), 画像電子学会誌, Vol. 20, No. 1, pp. 41-49 (1991).
- 10) 小野: カラー静止画符号化国際標準方式 (JPEG) の概説 (その2: 算術符号), 画像電子学会誌, Vol. 20, No. 2, pp. 113-120 (1991).
- 11) 森, 板倉: 画像認識の基礎 [I], [II], オーム社 (1986, 1990).
- 12) 堀池, 星野: 日本語テキスト・データベースの作成・利用ツールについて, 1990年情報学シンポジウム論文集, pp. 153-160 (1990).
- 13) Bradford, R. and Dickey, L.: How to Create Large Full-Text Document Databases, 1991 AIIM Conference Presentation (1991).
- 14) Pizano, A. et al.: A Business Form Recognition System, Proc. of COMPSAC 91, pp. 626-632, IEEE Computer Society Press (1991).
- 15) 上田, 三輪: 書誌データベース構築法, 丸善 (1991).
- 16) ISO 5963-1985: Methods for Examining Documents, Determining their Subjects, and Selecting Indexing Terms.
- 17) 内山, 中村: 重要キーワード抽出方式とその活用方法, 情報処理学会研究報告 91-DBS-84, pp. 151-161 (1991).
- 18) 木本: 日本語新聞記事からのキーワード自動抽出と重要度評価, 電子情報通信学会論文誌 D-I Vol. J74-D-I, No. 8, pp. 556-566 (1991).
- 19) 増永: オブジェクト指向マルチメディアデータベース—特にテキストデータベースにおける内容検索を考える, 情報処理学会研究報告 91-DBS-84, pp. 239-248 (1991).
- 20) 安田, 大山, 工藤: クラス階層モデルによる文書データベースの実現方法, Proc. of Advanced Database System Symposium '89, pp. 223-230 (1989).
- 21) 佐藤: データベース機能を持つマニュアル作成支援システム, 情報処理学会研究報告 90-DBS-78, pp. 93-100 (1990).
- 22) ISO 8879 Information Processing: Text and Office Systems—Standard Generalized Markup

Language (SGML) (1989).

- 23) 平岡：リンク指向 DBMS G-BASE におけるリンク機能の拡張とその応用, 情報処理学会研究報告 91-DBS-82, 82-5 (1991).
- 24) Kunii, H. L.: *Graph Data Model and Its Data Language*, Springer-Verlag Tokyo (1990).
- 25) Chen, P. P. S.: The Entity-Relationship Model —toward a unified view of data, *ACM Trans. Database Syst.*, Vol. 1, No. 1, pp. 9-36 (1976).
- 26) Kuwano, N. et al.: Applications of Object-Oriented Databases to Publishing Systems, *Proc. of DASFAA*, pp. 421-429 (1991).

(平成3年10月8日受付)



飯沢 篤志 (正会員)

1957年生. 1980年東京大学理学部情報科学科卒業. 1982年同大学院理学系研究科情報科学専門課程修士課程修了. 現在, (株)リコーソフトウェア事業部ソフトウェア研究所勤務. データベース管理システム, イメージデータベース構築ツールの研究・開発に従事. 日本ソフトウェア科学会会員.

