

解説

2. 各界から見たDTP—ニーズの把握のために—



2.3 OA業界から見たDTP†

山川 正†† 川端 洋 一††
田村 秀 行††

1. はじめに

デスクトップパブリッシング(Desktop Publishing; DTP)という用語は、米国Aldus社のPaul Brenard氏の命名したものであるといわれている。同社のPageMakerに代表されるDTPソフトウェアは、従来の電算写植(Computerized Typesetting)に頼っていた高品位な文書作成を、パーソナルコンピュータ(PC)とレーザビームプリンタ(LBP)からなる手軽なシステムで実現するものとして注目を集めた。特に、PCビジネスソフト市場で大きな成功を収めるところとなった。

1988年ころから、我が国でも日本語DTPという言葉が良く用いられるようになってきた。オフィス内で手軽に高品位な文書出力を得るという意味は同じようであるが、製品の形態や市場規模は米国とはかなり様相を異にしている。その大きな原因としては、何よりも複雑で多数の文字を必要とする日本語フォントの取り扱いの困難さにあるといえる。ホストコンピュータの処理能力、補助記憶装置の容量、CRTやLBPの解像度のいずれをとってもPCクラスのシステムでは、日本語DTPの要求水準を満たすには、やや不十分であるとされている。また、PCを中心とした米国のビジネスソフト業界に対して、我が国のオフィス内需要では、すでに日本語ワードプロセッサ(WP)の専用機が大きなシェアを持っていたことも別の要因としてあげられよう。

本稿に与えられた表題は「OA業界から見たDTP」であるが、上述のように、技術要素や市場動向とは別に、DTPという言葉だけがある種の流行語として輪

入された傾向も否定できない。現時点では、日本語DTPは、従来の比較的小規模な電子出版システムから、出力の高品位化をねらったWPの上位機種までを含むものとして捉えられている。あるいは、本特集写のようにコンピュータ文書処理までもその技術基盤として考えることも少なくない。

一方、OA(Office Automation)業界というのも狭い意味での(社)日本事務機械工業会を構成する事務機業界から、オフィス業務の電子化に関連するコンピュータ業界全般までを含めて考えることができる。ここでは、OA、DTPともに広い意味で捉えるものとする。

以下では、現行のDTPシステムを中心に業界の状況を述べることから始め、次世代のDTPシステムへの発展、将来のコンピュータ文書処理の基幹となる要素技術とその動向、さらに筆者らの研究を中心に文書処理統合環境について述べる。

2. OA業界から見た日本語DTPの現状と市場動向

日本語DTPの市場が米国のそれとはかなり異なることは前にも述べたが、事務機業界では電子出版の市場を(1)印刷業向けシステム、(2)一般企業向けシステム、(3)ハイエンドワープロ(LBP付き)、の3つのジャンルに分けて考えている。そして、この中で(2)項と(3)項を一応日本式「DTPシステム」と定義している。この意味での日本語DTPの市場形成予測を図-1に示す。

(2)項と(3)項の違いは、日本語文書処理についての技術基盤の違いであり、メーカ各社の製品系列・開発体制にも由来している。実際、事務機械工業会では電子パブリッシング部会とワードプロセッサ部会が存在している。(2)項の企業内印刷システムの大半は、従来からの電算写植システムの流れをくむものが多く、高度な組版機能を実現している。日本語DTPブーム以

† Desktop Publishing Viewed from Office Automation Industries by Tadashi YAMAKAWA, Yoichi KAWA-BATA, and Hideyuki TAMURA (Information Systems Research Center, Canon Inc.).

†† キヤノン(株)情報システム研究所

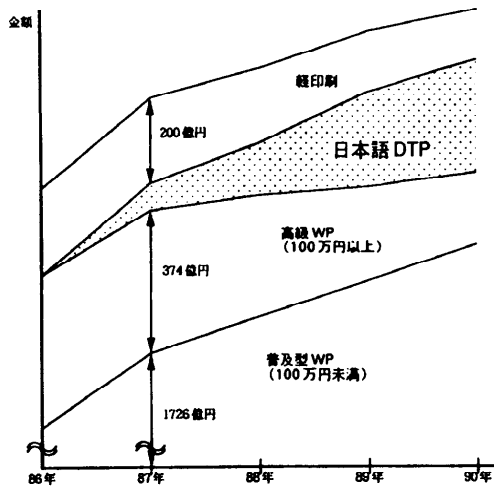


図-1 日本語DTPの市場形成予測¹⁾

前から、専用システムとして開発されていて、企業内だけでなく軽印刷業自体が簡易版下作成システムとして利用している例も多い。

一方の日本語ワープロの高級機は、印字品質の向上という点から、高精細LBP出力やアウトラインフォント対応等でDTP化が進められてきた。組版能力や操作環境の点では、(2)項と(3)項にかなり差があるものの、市場としては日本語DTPの名のもとに1988年頃から急速にさまざまな製品が現れて活況を呈したのも事実である(図-2参照)。実際、DTPブームは低価格競争で出荷額実績が伸び悩んでいた日本語ワープロの市場を、機能重視の高級機へ方向へも転じる役目を果たしたともいえるだろう。

一方、汎用ワークステーションやパソコンとDTPソフトの組み合わせも徐々に伸びてきている。日本国内のオフィスでは専用機の需要が依然として強いが、パソコンの能力向上、LBPの低価格化とともに米国製有力ソフトの日本語化も整い、DTP市場もこれからの発展が期待されている。

日本事務機械工業会発行の報告書¹⁾では、普及型ワープロで原文を作成しておいて、DTPで作成者自身が編集を行っている利用者が多いことが指摘されている。また、一般ビジネス文書や通達などの作成にはワープロが主に用いられているのに対し、プレゼンテーション資料・マニュアル・仕様書・各種資料の作成にはDTPが用いられるようになったことが指摘されている(表-1)。これらは、DTP導入以前の外注処

表-1 オリジナル文書作成方法の統計¹⁾

文書	ワープロ		DTP		手書き		その他	
	件数	%	件数	%	件数	%	件数	%
社内向け一般ビジネス文書	16	30.0	3	15.0	3	15.0	0	0.0
社外向け一般ビジネス文書	16	30.0	6	30.0	0	0.0	0	0.0
通達文書	15	75.0	4	20.0	4	20.0	0	0.0
プレゼンテーション資料	14	70.0	10	50.0	0	0.0	2	10.0
マニュアル	10	50.0	14	70.0	0	0.0	0	0.0
仕様書	12	30.0	9	45.0	0	0.0	0	0.0
見積書・契約書	14	70.0	3	15.0	5	25.0	1	5.0
稟議書・起案書	13	35.0	2	10.0	7	35.0	0	0.0
技術資料	10	50.0	8	40.0	4	20.0	0	0.0
販売資料	9	45.0	8	40.0	3	15.0	0	0.0
その他	1	5.0	2	10.0	0	0.0	0	0.0

表-2 DTP導入以前の外注処理文書の統計¹⁾

回答項目	件数	%
1. 社内向け一般ビジネス文書	0	0.0
2. 社外向け一般ビジネス文書	2	10.0
3. 通達文書	0	0.0
4. プレゼンテーション資料	3	15.0
5. マニュアル	3	15.0
6. 仕様書	1	5.0
7. 見積書・契約書	0	0.0
8. 稟議書・起案書	0	0.0
9. 技術資料	3	15.0
10. 販売資料	2	10.0
11. その他	2	10.0
12. 不明	12	60.0

理文書の内訳と一致しており(表-2)、従来のワープロでは対応できなかった分野の市場がDTPにより生み出されつつあることが伺える。

なお、増設・買い替えの理由では、適用業務の拡大が70%と最も多く、増設・買い替え機器は80%がDTPレベルを望んでおり、今後のDTP市場の拡大も伺える。

3. 次世代オフィスシステムと文書処理

3.1 DTPシステムの高機能化

まず、ここでは、前章で定義した日本語DTPシステム製品の次世代機に望まれている機能、近い将来に達成するであろう機能を列挙しておこう。

- 記憶の大容量化 — ハードディスクの大容量化に加えて、書き換え可能な光磁気ディスクが普及することが期待されている。大量の文書データの持ち運び、保存用に適している。

- カラー化 — カラー複写機、カラープリンタの普及により、オフィス文書やプレゼンテーション資料中にカラーが登場することが増えている。DTPシステムとしても、カラーCRT上での操作・カラー文書出力

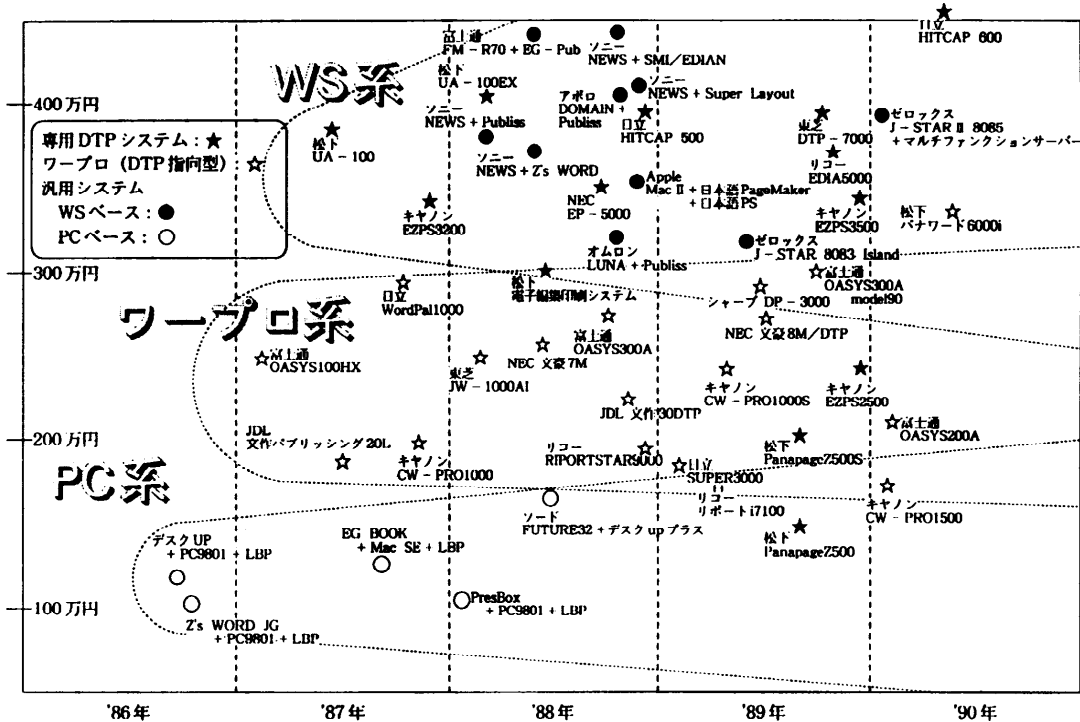


図-2 DTP製品化動向

の備わったものが望まれている。

- 通信機能の充実 — オフィス内のDTPシステムの増加とともに、他システムとの文書データ交換や文書データベース化の需要が増え、ネットワーク機能の充実の要求が強くなっている。

- マルチメディア化 — オフィス文書の電子化やコンピュータネットワークの利用が進むとともに、文書自体も動画や音声データを含むマルチメディア文書が話題になってきた。

- 複合商品化 — 文書の作成・出力を主体としたDTP機能に加えて、電子ファイルやFAX送受信機能と一体化となった複合商品も現れ始めている。

- 高付加価値化 — 人工知能分野の研究成果を利用した校正支援機能、自動翻訳もしくは翻支援機能を利用した検索等の高付加価値化がすでに一部のDTPシステムでは実現されている。

3.2 コンピュータデータ処理との統合化

ここでは、研究レベルも含め、もっと広い立場からコンピュータ文書処理の未来像について論じる。

いうまでもなく、文書作成や管理はオフィス業務の最

重要課題の1つである。我が国の場合、日本語入力が大きな障害であったところから、日本語ワープロの出現とはほぼ同時にOAなる言葉も広く用いられるようになった。日本語DTPが、日本語ワープロの高機能化あるいは電算写植分野との境界領域であるなら、DTPの発展は、文書処理業務のコンピュータ化が一段と進行する方向に向かうと考えられる。

文書を全面的に電子化すべきか印刷出力も併存すべきかについては議論が分かれるところだが、文書データの電子化の進展自体は次世代オフィスの必須条件と考えられる。この場合、文書に盛り込む情報をすべて電子化して管理するとともに、情報の選定から文書としての表現形態の決定までを一貫して取り扱うことが望ましい。

また、大規模な文書処理システムは、CADやスプレッドシートなどの既存のコンピュータデータ処理の結果を文書情報として同等に取り扱えるだけでなく、コンピュータネットワークを介して複数のユーザがデータを共有し文書化を進めていく形態になっていくと予想される。

EMP	NAME	LOC	MGR
__n		__l	LEE

NOTE
NAME: __n
LOC: __l
Subject: Vacation Plans This is to inform you that I'll be going on vacation from 5/5/79 to and including 5/15/79. David Jones will be acting manager in my absence, and all questions should be directed to him.
Earl Lee

COMMANDS
S. NOTE TO __n

図-3 OBEのプログラム例

この種のコンピュータ文書処理は、欧米ではオフィスシステム研究の基幹技術として古くから地道な研究がなされている。OAという言葉自体、ワークステーションやLANを生み出したように、コンピュータ利用の理想環境を目指す研究と密接に関連している。

代表的な研究例として、OBEは^{2),3)}、データ処理・文書作成・電子メールを統合して扱うためのエンドユーザ向けの2次元言語を提供するアプローチをとっている。言語形態は、Query-By-Example(QBE)データベース管理システムのスーパーセットにしている。端末上に表示される枠の欄にデータや問い合わせのための変数(「_」で始まる名前)などを記述することで、関係データベースに保管した情報を利用して文書作成・電子メールの発信などを指示することができる(図-3)。

トロント大のOfficeaid⁴⁾は、文書の作成・編集・配布・保管等を統合して管理するシステムである。関係データベース管理システムMISTRESSを文書保管・取り込みに用いて、このうに文書を操作するユーティリティ・ユーザインタフェースを組み上げた形態を取っている。

DTPシステムにこうしたデータベースからのアプローチを融合させることにより、オフィス情報を統一的に取り扱って、文書化を進めることのできる環境が提供できる。また、最近注目を集めているHypertext⁵⁾は、非線形な文書を操作・表示できる新しい機構であり、更にマルチメディアデータを扱うhypermediaなども、広い意味での電子出版(Electronic Publishing)

の形態とされている⁶⁾。

こうした次世代の文書処理システムを現実的なものにするには、現在の文書管理形態との併用、もしくは円滑な移行が図られなければならない。文書管理形態は、業務内容や組織の大きさなどにより適性が異なるうえ、管理者の好みにも左右される。だからといって、一般のオフィスのそれぞれに対してシステムを作り直すのは得策でない。ある程度、共通のシステム構成部品を用意しておき、適当なものを選択して組み立てるようになるのが現実的と考えられる。それには、文書処理ツールをパッケージ化して統一的に扱える統合環境が重要性を増してくる。

文書処理統合環境を実現するには、多様な文書処理への要求に耐えうる文書構造と文書処理モデルを確立する必要がある。次章では、その要求技術として文書編集・文書加工・文書管理・文書交換の最近の話題を述べ、第5章では文書処理統合環境構築の試みの例として筆者らの研究を紹介する。

4. 文書処理の要素技術と技術動向

4.1 文書編集

現在のDTPシステムのほとんどは、直接操作WYSIWYG型の編集方式を採用している。直接操作WYSIWYG型とは、文書のレイアウト結果が常に表示されており、カーソルを用いて変更箇所を直接指定して適用する処理を指定すると即座に編集結果が表示に反映されるという文書編集方式である。

この編集方式は、文書のレイアウト方法(多段組、センタリング、インデント等)や文字の表現方法(フォントの種類・大きさ、下線、網掛け等の指定)などのスタイル情報の取り扱いによって、さらに2つに分けることができる。

(1) スタイル情報埋め込み型 — ユーザが、レイアウトや文字表現に関する指示を直接操作しながら文書編集する方式。

(2) スタイル情報分離型 — 文書の構成要素(見出し、脚注、強調したい部分等)の表現方法をスタイル情報として取り出し記述しておく方法。スタイルシート等が用意され、パラメータを与える形式が多い。

(1)の方式は、ほとんどすべての日本語WPが採用しているように、操作が単純で解りやすい。反面、一括処理が施しにくいので、大量の文書作成には適さない。これに対して、(2)の方式は、スタイル情報を独立させているので、パラメータを変更するだけで、文書

の一括修正が容易である。このためには、文書の構成要素を逐一マークアップする煩わしさがある。

DTPブームの引き金となったPageMaker(Aldus社)やこれに続いたReady,Set,Go!(Manhattan Graphics社)など、初期のDTPシステムは(1)の方式を採用していた。その後Ventura Publisher(Ventura社)など、(2)の方式を取るシステムが多くなってきている。日本語DTPの場合、当初から、電子出版の流れをくむシステムに(2)の方式が多かった。WPに比べてDTPシステムが使いにくいと言われるのは、マークアップ操作の煩雑さによるものである。しかしながら、日本語DTPの場合も、次第に(1)型から(2)型システムへと移行する傾向にある。

スタイル情報分離型の採用が増えているのは、高機能な文書処理が要求されているためと考えられる。実際、この方式では、一段組から多段組へのレイアウト変更や、見出しを集めての目次自動作成などが、容易に実現できる。これは、文書編集の容易さよりも文書構造がマークアップされたことによる文書加工の可能性が重要視されてきたことを意味している。

複雑な構造を持つ文書に対しては、文書構造を視覚的に表現することが望ましい。しかし、これを直接操作WYSIWYG型編集方式と両立させることはそう容易ではない。文書構造を表現する記号が混在することにより、WYSIWYG性を保ちにくくなるからである。WYSIWYG性を重視するシステムでは、文書原稿のデータ形式として線形リスト構造(段落の並びに限った構造)を採用することが多い⁷⁾。この場合、比較的単純な文書構造を記号表現を用いず、見出しなどの表示形式を変えるなど、レイアウトで文書構造を表現していることになる。

しかし、文書加工を発展させようとする、文書構造をもっと詳細に表現することが要求される。そこで、木構造・有向グラフ構造といった構造が文書構造のモデルとして注目されてきた。Furutaは、必須文書要素で文書内容が埋まっているもの/未定のもの、省略可能な文書要素で内容が埋まっているもの/未定のもの、といった違いで、文書の入れ子関係の表示のon/off・文書内容のon/offを切り替えるなどの方法を提示している⁸⁾。この場合、文書内容とともに文書構造を表現する記号が表示されるので、WYSIWYGを保てない。

文書構造に関する別の角度からのアプローチとしては、従来の編集方法で文書作成を行った後、人工知

能等の技法を用いて構造情報を抽出する試みもある^{9),10)}。

構造情報などの文書内容の蔭に隠れた情報の入力方法及び表示方法の改善は今後の課題として残されている。

4.2 文書加工

文書加工には、(i)テキスト(文字列)の挿入・削除・移動・置換・整列などの文書内容の加工と、(ii)レイアウトや文字属性など文書スタイルの加工、の2つの側面がある。

前者の典型例は、従来からUNIXシステムで利用されている各種テキスト処理言語による文書加工である。この場合、単純な文字列からなる(即ち、構造を持たない)文書データを対象としているので、sed、awk等のツールを用いて比較的簡単に文書加工ができる。

一方の文書スタイルの加工に関しては、まず、最も簡単なレベルとして、前説で述べたスタイルシートのパラメータでレイアウトを変更する方法がある。このレベルでは、予め用意されたレイアウトスタイル内で、サイズ・文字種等を選択できるにすぎない。

もう少しレイアウト形式を柔軟かつ強力に記述できる代表例としては、 \LaTeX ¹¹⁾があげられる^{*}。 \LaTeX では、論理構造を表現するコマンドのマクロ定義を書き換えることによって文書のスタイルを変更することができる。ただし、この場合、 \TeX ¹²⁾の基本文書整形コマンドの記述形式の制約を強く受けている。

もっと一般的かつ強力な文書スタイル加工をめざした研究が活発になってきている。一般的に言えば、文書の論理構造を導入した文書処理系では、文書出力に先立って、論理構造からレイアウト構造へのマッピング(以下、これを構造マッピングと呼ぶ)が行われる。文書スタイル加工とは、このマッピング関数を変更することに相当する。

こうした考え方に基づく研究例としては、論理構造の強力な表現能力をもつSGML(Standard Generalized Markup Language)¹³⁾で記述した文書原稿を扱うPrice^{14),15)}、Van¹⁶⁾らの方法が代表例である。Priceの開発したMARKUPシステムでは、上述の構造マッピングを図-4のように記述している。いま、SGMLで記述された原稿(図-5)が入力されると、

^{*}本稿の版下は、日本語 \LaTeX 及び弊社EZPSを複合的に用いて作成したものである。ここで、文書スタイルを情報処理学会の様式に合わせるため、日本語 \LaTeX のマクロ定義を変更している。

```

<ELEMENT GLOSSARY>
  <START-STRING>\input glossmac
</START-STRING>
  <END-STRING>\bye
</END-STRING>

<ELEMENT TITLE>
  <START-STRING>\title</START-STRING>
  <END-STRING>}
</END-STRING>

<ELEMENT TERM>
  <START-STRING>\defterm</START-STRING>
  <END-STRING>}
</END-STRING>

<ELEMENT DEFINITION>
  <END-STRING>
</END-STRING>

<ELEMENT XREF>
  <START-STRING>{\it </START-STRING>
  <END-STRING>}</END-STRING>

```

図-4 論理構造からレイアウト構造へのマッピングの記述

Glossary of Animals

Aardvark
The first animal listed in a dictionary.

Cat
A carnivorous mammal long domesticated and kept by man as a pet or for catching mice (Webster's New Collegiate Dictionary, 1973).

Dog
A domesticated canine, probably descended from the <xref>wolf</xref>.

図-5 MARKUPへの入力原稿

```

\input glossmac
\title{Glossary of Animals}
\defterm{Aardvark}
The first animal listed in a dictionary.

\defterm{Cat}
A carnivorous mammal long domesticated and kept by
man as a pet or for catching mice (Webster's New
Collegiate Dictionary, 1973).

\defterm{Dog}
A domesticated canine, probably descended from
the {\it wolf}.

\bye

```

図-6 MARKUPの出力(TeXへの入力)

SGMLパーザが論理構造解析を行い、マッピング記述に基づいて、文書整形系TeXへの入力データ(図-6)を発生する。

MARKUPでは、文書要素の直前や直後に挿入する文字列(文書整形コマンド)を指定する以外に直前・直後に実行するプログラムコードを指定できることが大きな特長である。

これに対し、Vanの提唱している構造マッピン

グの記述方法では、文書要素のそれぞれにおける変換を、IF文・WHILE文・READ/WRITE文・SUSPEND/RESUME文などを用いてプログラミングする。

以上のいずれでも、文書整形系の如何にかかわらず、構造マッピングを記述することができる。すなわち、構造マッピングを変更することにより、同じ文書原稿を異なった文書整形系を用いて文書出力することが可能になる。文書整形系に依存しない文書スタイルの定義方法は、今後の研究課題である。

文書内容及び文書スタイルの加工方法を定義できると、これらの加工手順を指示する文書プログラミングによって、さまざまな文書作成を自動化できる。京嶋ら¹⁷⁾は、さらに論理構造間のマッピングを扱える文書プログラミングを提唱している。オフィス業務の自動化を目指すうえで、文書処理方法の記述言語及びその処理系は今後の大きな課題のひとつである。

4.3 文書管理

電子化した文書を統一的に管理するためには、文書データベースを構築し、その上で文書処理を行う必要がある。元来、文書は、文字列のほかに図形・画像データを含んでいる。最近、これらに音声・動画等を加えたマルチメディア文書といった考えかたも現われてきた。しかし、これらの文書要素間の関係は簡単な構造では表現しきれない、このような文書を取り扱うために、マルチメディアデータベースの必要性が説かれている^{18), 19), 20), 21)}。なお、文書データベースの研究動向は佐藤¹⁹⁾、田口²⁰⁾に詳しい。

Woelk²¹⁾は、マルチメディアデータベースシステムにおけるデータモデルの要求事項として、1) 集積化(Aggregation)の階層の取り扱い、2) 一般化(Generalization)の階層の取り扱い、3) 構成の自由な定義と変更、4) 手続きにより作成するデータの利用、5) 動的な結合関係をあげ、オブジェクト指向データモデルが、これに適合することを指摘した。実際、オブジェクト指向パラダイムを用いたマルチメディアデータベースシステムがいくつか提案されてきている^{22), 23)}。これは今後の文書データベースの基盤になっていくものと考えられる。

オブジェクト指向を採用することで、言語処理系を文書情報の問い合わせ処理に限定せず、文書データの変換などの文書処理一般に拡張できる。たとえば、オブジェクト指向データベース管理システムGemStoneの標準言語であるOPALはSmalltak-80を拡張する

ことにより、かなり強力なプログラミング環境を実現している²⁴⁾。これは文書プログラミングの一形式にもなりうることを示している。

文書内容による検索を重視した文書管理形態にフルテキストデータベース²⁵⁾がある。管理対象は文書のテキスト部分に限るが、これを全部蓄積しておき、文字列のパターンマッチングを全文に対して行うことで文書の検索を行うものである。この方式の文書管理に対してSGMLを活用できることが紹介されている²⁶⁾。すなわち、文書構造情報を生かした検索の問い合わせが可能になる。

4.4 文書交換

従来、異なった文書処理システム間での文書交換は、(i)テキスト部の文字コードのみを引き渡す、(ii)PostScript²⁷⁾等のページ記述言語で表現されたレイアウト済の文書を送る、(iii)同一の文書整形系(T_EX, troff²⁸⁾等)を用いることを前提として整形コマンド付きの文書データを送る、等の方法が採用されている。これに対し処理系・整形系を特定しない一般的な文書交換のための標準案として、ODA(Office Document Architecture)²⁹⁾とSGML(Standard Generalized Markup Language)¹³⁾が提案されている。

SGMLは、文書の論理構造の記述を定義する言語である。その定義に従った方法で構造情報を記述した文書原稿とを合わせて送ることにより文書交換を行う。一方、ODAは、文書の論理構造とレイアウト構造の各々に対して、独立した表現形式を定め、その両者間の構造マッピングを定義する方法を与えている。

SGMLやODAに基づいて交換された文書は、いずれも整形・出力する前に再編集可能である。ただし、SGMLが論理構造表現に重きをおいているのに対し、ODAではレイアウト済の文書を交換する機構を重視している。また、SGMLが文書情報の記述言語であるのに対し、ODAは文書情報を抽象的に規定する文書モデルであり、両者は性格をかなり異にしている。

既存の異なる文書処理システム間でのODAを用いた文書交換の例としては、カーネギーメロン大学のAndrew Toolkit Document Systemとミシガン大学のDiamond multi-media systemの間で文書交換を行うEXPRESプロジェクト³⁰⁾がある。この研究プロジェクトの実行にあたって、ODAの文書データ形式であるODIFを取り扱うプログラムが開発されODA Tool Kitとして公開されている。この副産物は、これからのODA関係のシステム構築に役立つものと期待

されている。

欧米においては、SGMLの実用化がかなり進められている³¹⁾。欧州共同体の出版部門で開発したFORMEX(Formalised Exchange of Electronic Documents)システムでは、欧州共同体の出版物を電子的に取り扱うのにSGMLを採用している。また、オックスフォード大学出版局は、オックスフォード英語辞書(OED)全16巻分のデータベース化にSGMLを用いた。とりわけ、注目されるのは、合衆国の国防総省のCAL(SComputer-aided Acquisition and Logistic Support)プログラムである。これは、兵器の設計から保守に至る膨大な量のペーパーワークを電子的に統合する計画であり、この交換にSGMLが義務づけられている。

5. オフィスシステムにおける文書処理統合環境

以下、我々の研究例をあげ、文書処理統合環境の考え方とその応用例を紹介する。

5.1 文書処理統合環境DI_ET

文書処理統合環境とは、形態の異なる文書要素の統合した編集作業、文書データの複合利用・管理等を行う文書処理システムを構築するためのものである。こうした統合環境には以下の機能が備わっていることが望ましい。

- 基本的な文書処理機能を実現するツール類が用意され、それらを簡便な方法で組み合わせて利用できること。
- そのためには、個々のツール間で情報交換を统一的に扱える文書記述形式が定まっていること。
- この記述形式で表現された文書データの管理機能が充実していること。
- 他の文書処理システムとの文書交換が行いやすいように考慮されていること。

筆者らが研究を進めている文書処理統合環境DI_ET(Document Integration Environment & Tools)^{22), 23)}では、文書型・論理構造木・内容集合の3つ組を文書の構造化モデルとして採用している。このモデルに基づく文書記述形式(DI_ET形式)で記述した文書データの例を図-7に示す。

DI_ET形式で記述された文書データをDI_ETaryと呼ぶ。DI_ETでは、DI_ETaryを処理対象とする基本文書処理ツール、DI_ETaryの文書要素の抽出・置換等を行う

<通知文書><発行元><所属>情報システム研究所</>
 <氏名>山川 正</></>
 <標題>慰安会に関して</>
 <本文>
 <段落>秋の慰安会を10月7日(日)に行いますので、
 皆様この日を空けておいてください。</>
 <段落>場所や何を行うかの詳細は未定です。</>
 <段落>皆様のご希望を私までお知らせください。
 よろしくお願ひします。</></></>

図-7 DiF形式の記述例

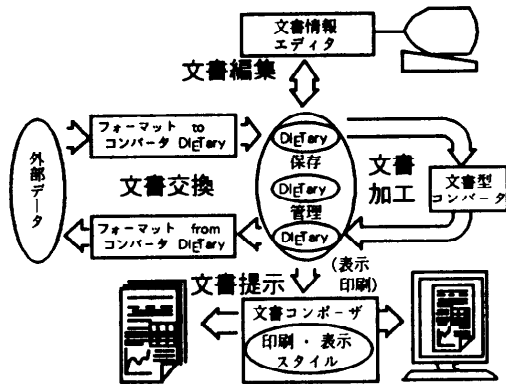


図-8 DiFにおける文書処理モデル

プログラムライブラリ、文書ツール自体を生成するツールジェネレータを提供している。

DiFにおける文書処理モデルを図-8に示す。ここでは、文書処理ツールを、(i)文書コンポーザ、(ii)文書型コンバータ、(iii)フォーマットコンバータ、(iv)文書情報エディタの4つに分類している。これらの文書処理ツールを組み合わせることにより、いろいろな文書処理システムを構築することができる。

5.2 DiFの応用例

文書処理統合環境の応用例として、研究報告統合処理システムを紹介する。このシステムは、研究報告のために作成した原稿を、各種印刷フォーマットの異なる報告・論文等に何度も再利用することを目的としている。図-9はその全体構成である。原著である研究報告の文書データは、構造エディタを用いて前述のDiF形式のデータ(DiFary)としてデータベース化される。換言すれば、多種多様な形態に変換できるよう考慮した文書DB用の記述形式がDiFaryである。ここでは、論理構造を持った文書データの作成に、前章で述べた「スタイル情報分離型」の編集方式を採用し

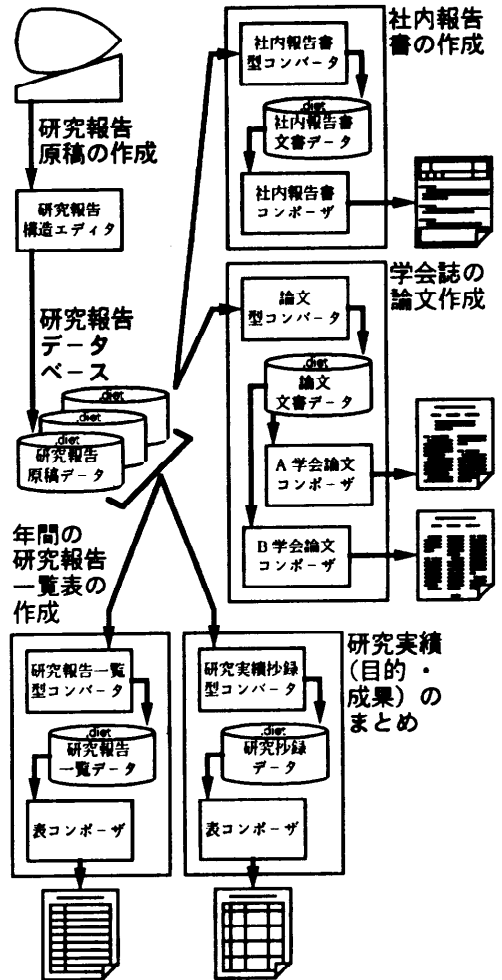


図-9 研究報告統合処理システム

ている。

社内報告書や学会投稿論文等の各種整形出力を作成するのが、各々文書コンポーザである。文書コンポーザは、印刷スタイルが異なる度に用意する必要がある。例えば、A学会とB学会の論文誌では、組版スタイル、参考文献の記法、英文梗概の有無等が違っている。これは、前章で述べた文書加工の「構造マッピング」が異なっていることになる。

しかしながら、印刷スタイルは異なるものの、大抵の論文は類似した論理構造情報を有している。したがって、原データベースから最終論文出力を行うコンポーザを作るより、一旦、論文型の文書データ(これもDiFaryである)を作成し、これを入力とする各種論

文コンポーザを作成するほうが容易である。

社内報告書の場合は、研究テーマ名・配布先等の学会論文にはない情報を付け加える必要がある。このため、論文型とは別の社内報告書型の文書データを作成し、これを入力とするコンポーザを用意するのが望ましい。即ち、構造マッピングに徹した文書処理ツールがコンポーザであり、原データベースから必要な情報のみをフィルタリングし型変換を行うのが文書型コンバータである。

文書型コンバータのフィルタリング機能を用いると、年間の研究報告の一覧表や研究実績抄録等を作ることができる。この型コンバータは、複数の研究報告原稿データから必要な情報のみを取り出し、一般の表コンポーザ用入力データへ変換する働きをする。

以上のように、文書処理統合環境では、文書型コンバータや文書コンポーザなどの文書処理ツールを組み合わせるにより、さまざまな文書処理機能を実現可能になる。

6. おわりに

OA業界におけるDTPの現状と、今後の展開について述べてきた。これまでのDTPは、印刷出力表現力の向上と文書レイアウト操作の簡易化に重点が置かれてきたが、今後は文書の加工や文書の管理など、オフィス情報を統合的に取り扱う文書処理システムへと発展していくことが予想される。

本稿では、かなり広い観点からコンピュータ文書処理の未来像について述べた。ここでは触れなかったが、秘密保護、バージョン管理、トランザクション処理などもデータベース技術と文書処理技術の統合したオフィスシステムでの重要な技術となっていくであろう。DTPからこうした文書処理への発展は、OA業界自体をますますコンピュータとネットワークを中心とした情報処理産業へと転換させていくものと思われる。

謝辞 本稿で用いた資料をご提供いただいた当社DTP事業企画部の柿田浩孝部長、宇田川栄一氏に深謝します。

参考文献

- 1) 電子パブリッシングシステムの使用状況に関する実態調査報告書、(社)日本事務機械工業会(1989.3).
- 2) Zloof, M. M. : *Office-by-Example: A business language that unifies data and word processing and electronic mail*, IBM Systems Journal, Vol. 21, No. 3, pp. 272-304 (1982).
- 3) Huang, K.-T. : *An Integration Framework for Office System Design*, Proc. of IEEE Computer Society Office Automation Symposium, pp. 190-199 (1987).
- 4) Lee, A. Woo, C. C. and Lochovsky, F. H. : *Officeaid: An Integrated Document Management System*, Proc. of ACM SIGOA Conf., pp. 170-180 (1984).
- 5) Conklin, J. : *Hypertext: An Introduction and Survey*, IEEE COMPUTER, Vol. 20, No. 9 (1987).
- 6) *Special Issue on Electronic Publishing Technologies*, IEEE COMPUTER, Vol. 21, No. 1 (1988).
- 7) Furuta, R. : *Structured Document Models and Representations*, Lecture Note of a Short Course with a PROTEXT IV: Proc. of the Fourth International Conf. on Text Processing Systems, pp. 7-22, BOOL PRESS (1987).
- 8) Furuta, R. : *Complexity in Structured Documents: User Interface Issues*, PROTEXT IV: Proc. of the Fourth International Conf. on Text Processing Systems, pp. 7-22, BOOL PRESS (1987).
- 9) 山口, 福井, 岩井 : 文書作成システムにおける構造化とレイアウト支援, 電子情報通信学会 ワークショップ — 電子出版の現状と課題, pp. 61-66 (1989.4).
- 10) 野村, 柳, 福岡 : オフィスワークステーションにおける報告書作成支援システム, 情報処理学会第38回全国大会, 1J-5, pp. 615-616 (1989).
- 11) Lamport, L. : *TeX: A Document Preparation System*, Addison-Wesley (1986).
- 12) Knuth, D. E. : *The TeXbook*, Addison-Wesley (1984).
- 13) ISO 8879 : *Information Processing - Text and Office Systems - Standard Generalized Markup Language (SGML)* (1986).
- 14) Price, L. A. : *A Parser Generator for SGML*, Lecture Note of a Short Course with a PROTEXT IV: Proc. of the Fourth International Conf. on Text Processing Systems, pp. 118-123, BOOL PRESS (1987).
- 15) Price, L. A. : *Evolution of an SGML Application Generator*, proc. of ACM conf. on Document Processing Systems, pp. 51-60 (1988).
- 16) Van, H. L. and Terreni, E. : *A Language to Describe Formatting Directives for SGML*

- Documents, T_EX for Scientific Documentation Second European Conference*, pp. 100-119, Springer-Verlag (1986).
- 17) 京嶋, 上林, N.V.ゴバル: 文書プログラミングについての一考察, 信学技法, DPHI 23-4, p. 1-9 (1989).
- 18) 増永良文: マルチメディアデータベース総論, 情報処理, Vol.28. No.6, pp. 671-684 (1987).
- 19) 佐藤, 絹川, 大町: オフィス文書の標準化と文書データベースの研究動向, 情報処理, Vol.28. No.6, pp. 710-720 (1987).
- 20) 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).
- 21) Woelk, D. Luther, W. and Kim, W.: *Multimedia Applications and Database Requirements, Proc. of IEEE Computer Society Office Automation Symposium*, pp. 180-189 (1987).
- 22) Schlageter, G. Unland, R. Wilkes, W. Zieschang, R. Maul, G. Nagl, M. and Meyer, R.: *OOPS - An Object Oriented Programming System with Integrated Data Management Facility, Proc. of Fourth International Conference on Data engineering*, pp. 118-125 (1988).
- 23) Fishman, D. H. Beech, D. Cate, H. P. Chow, E. C. Connors, T. Davis, J. W. Derrett, N. Hoch, C. G. Kent, W. Lyngbaek, P. Mahbod, B. Neimat, M. A. Ryan, T. A. and Shan, M. C.: *Iris: An Object-Oriented Database Management System, ACM Trans. on Office Information Systems*, Vol. 5, No. 1, pp.48-69 (1987).
- 24) Copeland, G. and Maier, D.: *Making Smalltalk a Database System, Proc. of the 1984 ACM SIGMOD conference*, pp. 316-325 (1984).
- 25) 有川, 篠原, 宮原, 武谷, 宮野, 竹田, 大島, 白石, 酒井, 山本: テキストデータベース管理システムSIGMAとその利用, 情報研報 89-FI-14-7, Vol. 89. No. 66, p. 1-8 (1989).
- 26) 芝野耕司: SGMLと全文データベース, 情報研報 89-FI-14-2, Vol. 89. No. 66, p. 1-8 (1989).
- 27) Adobe Systems Inc. (石田 監訳): *PostScript リファレンス・マニュアル*, アスキー (1988).
- 28) 長谷部紀元: UNIXの文書処理機能, 情報処理, Vol. 27, No. 12, pp. 1374-1382 (1986).
- 29) ISO/DIS 8613: *Information Processing - Text and Office Systems - Office Document Architecture (ODA) and Interchange Format, Part 1 - 6* (1986).
- 30) Rosenberg, J. Sherman, M. S. Marks, A. and Giuffrida, F.: *Translating Among Processable Multi-media Document Formats Using ODA, Proc. of ACM conf. on Document Processing Systems*, pp. 61-69 (1988).
- 31) Barron, D.: *Why use SGML?, Electronic Publishing*, Vol. 2, No.1, pp. 3-24 (1989).
- 32) 山川, 川端, 田村: キヤノンFOD_s計画とその統合文書処理環境, 情報処理学会 卓上出版シンポジウム報告集, pp. 223-232 (1988.7).
- 33) 川端, 山川, 出井, 田村: 文書処理ワークステーションと文書アーキテクチャ, 電子情報通信学会 ワークショップ - 電子出版の現状と課題, pp. 53-59 (1989.4).

(平成2年7月31日受付)