

文章・図形・画像処理の現状と動向

Present Status and Trends of Office Document Processing

 オフィスの生産性向上を考えると、文書の作成・管理の能率化は重要な部分を占めており、日本語ワードプロセッシングシステムが実用化されたことにより、大きな変貌を遂げつつある。本稿では、文書を構成する文章及び表、図形、画像のそれぞれの処理の現状を概観するとともに、今後の動向について論述する。

 これらの各処理技術は、文書作成・管理として融合化が図られるとともに、データ処理機能との融合、大規模データベースとの接続、ローカルネットワークとの接続が期待される。更に翻訳機能など、文章の意味処理技術をベースとした幾つかの機能の開発とその組み込みが試みられよう。

松岡 潤* Hiroshi Matsuoka

津田順司* Junji Tsuda

1 緒 言

 一般にオフィスでは、どのような種類の情報が扱われているかを一概に言うのは困難である。極めて模式的に考えるならば、オフィスは図1に示すように、トップマネージャーの意思決定のための情報の収集・整理と下位レベルの意思決定の機関とみることができるであろう。しかし、この過程は単純なものではない。オフィスの中には、通常複数の人間が存在し、それらが上下又は横の関係をもって、相互間に、またオフィス外部と複雑に情報の交換を行なっている。そして、そのような情報の飛び交う過程のほとんどすべてが意思決定

表1 オフィス内で生きている情報 意思決定につながる多種類の情報があるが、文書は記録される情報の主要部である。

情 報 の 種 類	例
(1) 紙面に記された情報	(a) 机上のメモ (b) ファイルに収納された文書 (c) 回覧中の文書
(2) 音 声 情 報	(a) 話されている音声 (b) 通話中の電話線内の音声電流 (c) 録音された音声
(3) 情 景	(a) 人間の動作・表情 (b) 物の動き、状態の変化の情景 (c) 景 色 (d) 上記(a)～(c)などの録画
(4) 人間の頭脳に記憶されている情報	(a) 思い出して利用できる情報 (b) 経験によって得られたノウハウ (c) 性格、好み

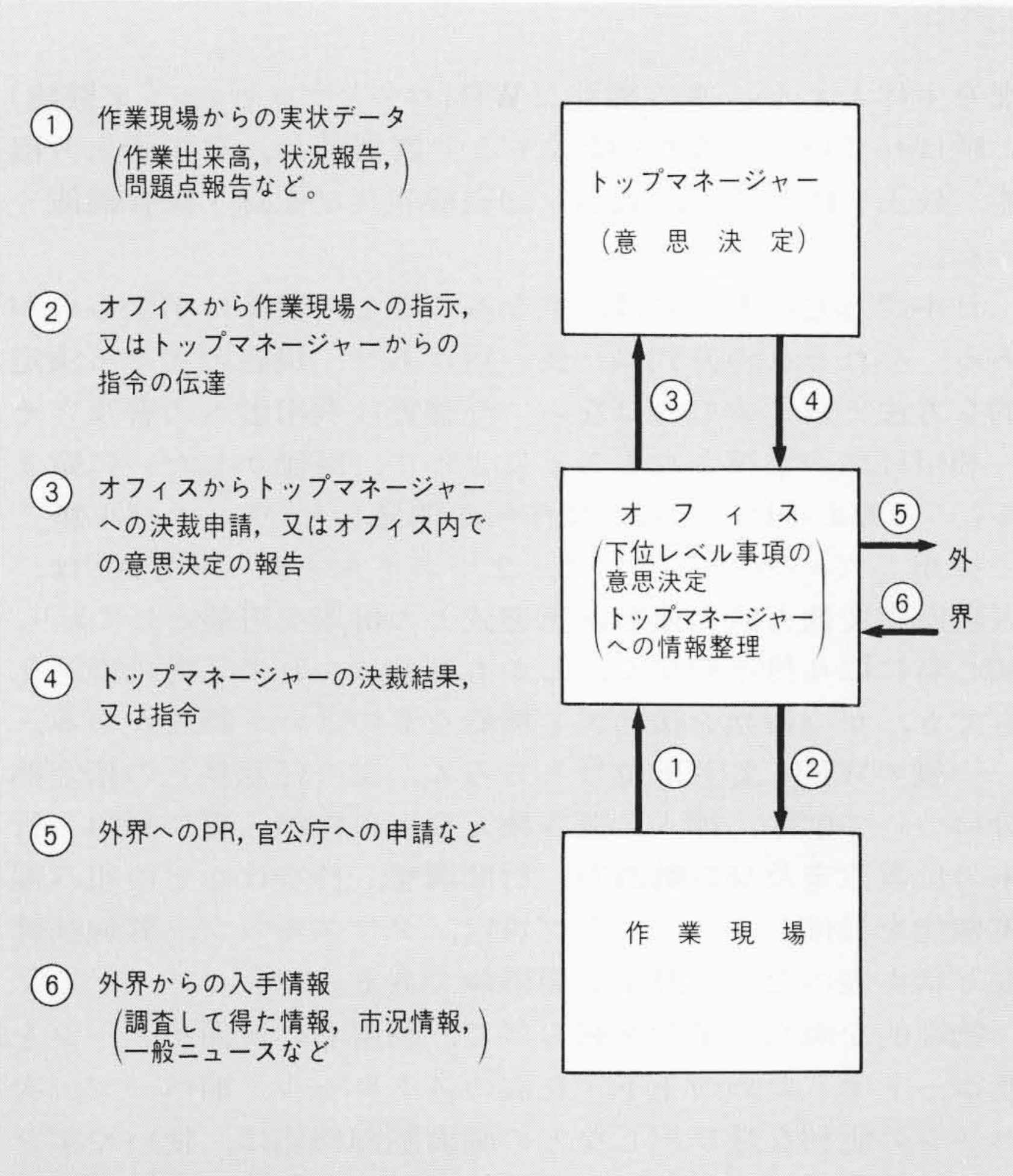


図1 オフィスの情報入出力模式図 オフィスはトップマネージャーの意思決定を補佐するために、トップマネージャー、作業現場、外界との間で、またオフィス内でダイナミックに情報の授受を行なっている。

 に向かっての下地作りと考えることができる。例えば、複数の人間が相互に折衝の上、妥結点を見いだし、意見を集約する、というたぐいのことが機関としての意思決定の重要な過程であると言える。

 オフィスの中での情報の形態は、表1に示すようにみることができる。同表の項目(4)は、オフィスの機能上重要な要素であるが、工学的扱いの対象とはなり得ず、本稿の対象外である。本稿では、記録された情報に主眼を置くこととし、その主要部を占める(1)紙面に記された情報(すなわち、文書)に対象を絞って論ずることとする。

2 文書処理技術の現状と動向

2.1 文書の主要要素

 オフィスで作られ、利用される文書を計算機処理の立場からみると、一般に文章、表、図形、画像から成るとみることができる。これらの要素の区別は、表2に示すとおりである。

* 日立製作所システム開発研究所

表2 文書の主要構成要素 文書の処理のためには、文章、表、図形、画像のそれぞれの処理と、それらを自由に組み合わせる技術が必要である。

文書を構成する 主 要 要 素	内 容	例	コンピュータ内 での 情 報 形 式
文 章	文字列(数字,記号, アルファベット, 仮名,漢字)	報告書などの本文	コード化情報
表	罫線と文字列	一覧表,帳票	同 上
図 形	線分の組合せ	棒グラフ,折れ線 グラフ,説明図	同 上
画 像**	2値画像*,濃淡画 像,色彩画像など	挿絵,印影,署名	画素ごとの白黒情 報の集合,画素ご との濃度情報の集 合,画素ごとの色 彩情報の集合

注： * 2値画像は、各画素が白か黒かの2値だけを取り得る画像である。
* * 文章、表、図形のいずれも画像としての情報形式で表わすことも可
能である。

表3 官報資料版における図・表などの構成比の例 表、図、画像を合わせると、資料版全体の半分近くを占めている。

構 成 要 素	面 積 比(%)
文章(一般記事の領域)	55
表	25
図	14
標 題(画像)	6

注：本データは参考文献1)～3)の平均値である。

表2の(注**)に示すように、コンピュータ内での情報形式として、文章、表、図形とも画像と同形式で表わすことも原理的に可能であるが、そうすることは一般には情報量が大きくなることと、その後の処理で計算機の情報理解が困難となることの理由で、可能な限りコード化情報で扱うのが普通である¹⁾。例えば、A4サイズ1ページの文章は、コード化情報とすれば(40字×40行の場合)3.2kバイトで表わされるが、2値画像とすれば(画素を0.1mm×0.1mmとする場合)0.7Mバイトを要することとなり、実に約220倍となるのである。

これらの構成要素の割合は、文書の種類によっても異なるものと考えられる。一つの代表例として、官報資料版の幾つかのもの^{2)~4)}を例に挙げると、表3に示すような割合となっている。「標題」は号の標題領域、各記事の標題領域などの合計であり、この部分は全部ではないが図案化された文字や、挿絵を含んでおり、画像としての取扱いを要する部分である。この表で、画像はそれほどではないが、図形や表が大きな割合を占めていることが分かる。今後オフィス効率化の推移、テレビジョン普及に伴うパターン理解形人間の増大により、ますます表、図形、画像が文書中に占める重要性が増してくる傾向をもつと考えられる。

2.2 処理技術の現状と動向

(1) 文章及び表

日本語ワードプロセッサが実用化されて以来、文章編集処

表4 代表的な日本語入力方式 適用分野、目的に応じて適した入力方法を選択利用する必要がある。

入 力 方 式		入 力 方 法	初心者の 利用可否	習熟後の入力 速さ(字/分)
直接 入力 方式	多 段 シ フ ト 方 式 (漢字テレタイプ方式)	キー上に4~15個の文字が 割り当てられている文字キ ーと、シフトキーとで該当 文字を選択する。	△	70~90
	タ ブ レ ッ ト 方 式	タブレット上にマトリッ ク状に配列された文字を、 ペンで触れて入力する。	△	30~70
仮名 キー ボ ー ド 方式	仮 名 漢 字 変 換 方 式	読みで仮名データを入力す る。ソフトウェアで漢字仮 名まじりに変換される。	○	50~100
	連 想 記 憶 方 式 (2ストローク方式)	漢字の意味や字形などから 連想できる仮名2文字から 成るコードを、各文字にあ らかじめ与えてあり、これ を暗記して入力する。	×	120~200
	漢 字 コ ー ド 方 式	漢字コード2バイト分をそ のまま16進数4桁で入力す る。コード表を見ながらの 入力となる(外字入力などに 利用)。	○	(非常に遅く、 習熟効果ほと んどなし。)

注：○(利用可能)、△(利用困難)、×(利用ほとんど不可能)

理を主体とする一連の機能はWP(ワードプロセッシング機能)と呼ばれている。WPを構成する主要機能は、日本語入力機能、校正・編集機能、表示・印刷機能及び記録・保管機能である。

日本語入力の方式には、主なものとして表4に示すものがある。それぞれの方式に一長一短があり、現在のところ決定的な方法と言えるものはない。今後更に利用者への普及とその利用経験が集積されることによって、評価がしだいに定まるものと考えられる。日立製作所が開発した「ワードパル20」⁵⁾で採用している一つの方式、2ウェイキーボード方式⁶⁾は、仮名漢字変換方式と連想記憶方式との併用を可能としており、初心者に取り付きやすく、しかも習熟につれて入力の高速化もでき、かつ習熟を援助する機能をもつという特長がある。

一般のWPは文字単位はもちろん、文の任意長さの指定部分について削除、挿入、置き換えなどの機能、更に行頭・行末の位置設定及び自動改行、行間調整、枠空けなどの組み編集機能を具備している。タブ設定、タブスキップ、罫線制御などは作表のための最小必要機能である。

物理的な画面の限界を和らげて、利用者の紙面イメージをサポートするための上下・左右のスクロール、前ページ・次ページの便利な呼び出しなどの画面制御機能は、使いやすさの面で重要である。更に、出来上がり文書上では離れた位置にある情報の部分を、一時的に画面上で並べてみるための機能、すなわち横スプリット、縦スプリットや、マルチウィンドウ機能も期待される機能である。

現在ほとんどのWPは、文書中の表の作成・編集をなんらかの形で支援する機能をもっていると言えよう。しかし、一

般にこの面での機能は十分とは言えず、今後の充実が期待される。例えば、上記のスプリット機能やマルチウインドウ機能のほか、欄の中に数値の入るべき表で、横の集計、縦の集計などが簡単にできる機能や、任意に指定した文字、又は数値アイテムを第1キー、第2キー…などとして、ソートする機能が重要である。更に、従来から建設されてきている各種のデータベース内から必要なデータを検索し、これを表の指定の欄に配置する機能が、基本的な重要機能である。このような機能は、オフィスコンピュータの機能として実現されてきているものであり、WP機能とオフィスコンピュータ機能との融合が一つの大きな発展方向と言えるであろう。

文章作成・編集機能は、現在、スタンドアロン形、インテリジェント端末形及び処理をすべてホストコンピュータに負う端末形のいずれの形のものも開発されており、それぞれ用途に応じて利用されている。スタンドアロン形、インテリジェント端末形ともフロッピーディスクが情報蓄積媒体の主役として利用されている。オフィスの個々の文書を単位として扱う上で、この蓄積媒体は容量的にも、扱いの便利さからも適当と考えられる。また、オフィスの個々の人間の自発的な創意や業務遂行の支援の意味からも、安価なスタンドアロン機器や個人ファイル機能は大切である。しかし、OA(オフィスオートメーション)の普及、浸透が深まるにつれて、情報の共有性や連絡の容易性の重要性が再認識されることとなるであろう。その意味で、フロッピーディスクファイルのハードウェア的及びソフトウェア的互換性、これを支えるための標準化は、今後の最重要課題であると考えられる。また通信回線を通じて、自社のセンタファイル、ローカルネットワークとの接続、更には公益的データベース(例えば、技術情報サービス用や株価情報サービス用など)への問い合わせと情報取得機能が重要になるものと考えられる。

(2) 図形処理

前節2.1で述べたように、文書中での図形は、文意を効率良く伝える上で欠かせない情報である。現在図形情報の扱いは、利用分野によって大まかに二つの種類に分けてみることができる。それは、表5に示すようにビジネスグラフと設計図面である。このうち設計図面の処理は、設計の効率化〔CAD(Computer Aided Design)〕の1ステップとして扱うのが妥当であり、一般のオフィスの文書と区別して考えるのが現在のところ普通である。したがって、本稿でも設計図面の処理については除外して考える。

ビジネスグラフは、コンピュータ内で、幾つかのパラメータの一組みとして保持することにより、再現可能である。入力パラメータの入力であり、特別の操作を要しない。このようなパラメータのうち、あるものはデータベースから検索して取り出すことが必要であり、一連の操作に便利さが要求

される。

与えられたパラメータから具体的図形を作成し、表示及び印刷出力を行なうためのソフトウェアは、各種のものが実用化されている。現在までのところ、処理ソフトウェアはWPシステムとは独立なシステムであるものや、グラフ出力専用端末を要するものがほとんどであり、WPと融合されているものは少ない。

ページの一部分に図形を挿入すること、図の中に説明用の文字を配置すること、図形全体の大きさを指定の大きさに拡大・縮小することや、その際に図中の文字情報は本来の図形情報部分と異なる拡大率・縮小率にできること、などの機能開発が期待される。

図形の見やすさのためには、色の使い分けや、ある範囲の塗りつぶしも重要な機能であり、専用システムでは実現されているが、WP機能と一体となった利用可能化が望まれる。

(3) 画像処理

画像処理を広くとらえると、人工衛星が撮影した画像情報から資源探索に有効な情報を取り出すための処理や、人体のX線写真や胃カメラなどのとらえた画像から診断を下すための処理、すなわち生体画像処理、またテレビジョンやVTR(ビデオテープレコーダ)に代表される動画像を扱うものなどがある。一方、文書中に掲げる挿絵、写真、印影やサインなど、文書と直接かかわる画像がある。本稿では後者に限って述べる。

挿絵や印影などは、これを線分に分解してコード化することは實際上ほとんど不可能なため、画像として扱うのが普通である。一方、我が国では、従来から情報を記録するのに、まず手書きでメモや下書きを作る習慣がある。したがって、このような手書きの文書をそのままコンピュータに入力し、その後の編集や伝送処理ができることには大きな利点がある。更に、いったん印刷された文書も、それがコード化されていないものについては、やはり画像として入力できることは、文書の整理、蓄積、検索、伝送などの上で大きな利点となる。この理由と、光ディスクや大容量磁気ディスクが実用化されていることともあいまって、文書を画像として扱う技術が発展しつつある⁷⁾。フィルム利用のシステムも並存してゆくものと考えられる⁸⁾。

この技術とコード化された情報の処理とを組み合わせることにより、オフィスの文書処理及び管理に大きな便利をもたらすものと期待される。

3 文書処理の将来展望

3.1 データ処理機能と文書処理機能との融合

前章でも触れたが、文書処理の今後の一つの大きな方向は、文書作成・編集機能と、従来からのコンピュータ利用で発展してきているデータ処理機能との融合であると考えられる。ここに言う「データ処理機能」は若干広くとらえ、数値計算機能のほか、情報蓄積・検索機能や、対話形での各種プログラムの呼び出し・起動、又はプログラム作成・実行処理などをも含むものである。実際、文書の作成に当たっては、ほとんどの場合に過去に作成された文書や、データ又はそのデータを元になんらかの演算を施して得られるデータを必要とするものである。そして人間の側から、作成中の文書に修正や挿入を行なうのと全く同等に、コンピュータ内のプログラムからその作成中の文書に加工できる機能が不可欠である。このような機能構造には、最近のパーソナルコンピュータやオフィスコンピュータの発展形としても指向されてきている方

表5 利用分野による図形の類別 OAの主対象はビジネスグラフである。

	概 要	範 囲・特 徴
ビジネスグラフ	一般事務文書中の共通的な図形	棒グラフ、折れ線グラフ、円グラフ、ブロックチャートなど。
設 計 図 面	機械図面、電気回路図面、建築図面など。	(1) 一般にビジネスグラフに比べて複雑である。 (2) 一般に寸法や座標に正確さが要求される。

向であり、マンマシンインタフェースの基盤的なものと言えるであろう。

3.2 文章・図形・画像処理の統合

今後のもう一つの発展の方向として、文章処理、図形処理及び画像処理の融合が図られるものと考えられる。その一つの動きとして、端末での機能の複合化が行なわれ、ワークステーションの名にふさわしいものになると思われる。この統合が体系的に実現されるためには、システム全体、すなわち入出力、内部処理、ファイル体系及び伝送体系にわたって、これら異質な情報を矛盾なく効率的に扱うハードウェア・ソフトウェアの体系を構築する必要がある。これに向かって、幾つかの試みが実施されることとなるであろう。

これと並行して、処理の高速化や高機能化が図られる。例えば、カラー化の普及や高精細化がそれである。

3.3 意味処理技術の開発と適用

もう一つの本質的な進展の方向としては、高インテリジェント化である。例えば、従来の文章処理では情報を文字列としての扱いが主体であり、文の意味には無関係であった。今後、文の構造や意味にかかわる処理の技術^{9),10)}が開発され、徐々にではあるが実用に移されてゆくものと考えられる。例えば、誤字・脱字に対する警告、更には文としての意味不明確なものに対する警告の機能、また外国語への翻訳機能¹¹⁾などが考えられる。このような意味理解の技術は、最近研究が活発化されてきた知識工学の一分野であり、この技術が進歩することにより、マンマシンインタフェースも格段に改善されるものと思われる。各種のOA機器に対する指示方法は、煩雑さを減じ、人間に対する仕事の指示の仕方に漸次近づいてゆくものと期待される。

4 結 言

オフィスでの文書作成・管理システムを中心とした情報処理技術の状況と将来の動向について述べた。OAは急テンポで進展しつつあるが、まだそのスタートを切ったばかりであり、研究・開発すべき課題は山積している。日立製作所は、これらを精力的、かつ着実に解決してゆく考えである。

参考文献

- 1) 矢島，外：分散形図形処理システムHITAC G 730における図形表現法，昭54情処学会全国大会
- 2) 官報資料版No.1216：大蔵省印刷局発行（昭56-11-11）
- 3) 官報資料版No.1218：大蔵省印刷局発行（昭56-11-25）
- 4) 官報資料版No.1219：大蔵省印刷局発行（昭56-12-2）
- 5) 内田，外：普及形日本語ワードプロセッサ，日立評論，63，8，523～528（昭56-8）
- 6) 藤方，外：動的仮想テンキー方式漢字選択法とその和文ワードプロセッサへの応用，電学会論文C，101，1～8（昭56-1）
- 7) 田畑，外：文書管理システムにおけるイメージ処理技術，日立評論，64，4，249～252（昭57-4）
- 8) Y. Nitta, et al.: Information Retrieval for Design Review Support; 1980 Proceedings of Annual Reliability and Maintainability Symposium, pp.152～157（1980-1）
- 9) H. Kinukawa, et al.: Japanese Sentence Analysis for Automatic Indexing; Proceedings of COLING 80, pp.514～519（1980-10）
- 10) 松岡，外：語彙出現率に関する推定法；日本語情報処理シンポジウム報告集，325～331（昭53-7）
- 11) 新田，外：英和機械翻訳のための構文解析法；情処学会自然言語処理研究会資料，28-4（1981-11）

論文抄録



出力技術

日立製作所 岡田康行・斉藤 進

情報処理学会誌 22-10, 953～960（昭56-10）

将来のオフィスでの出力装置は、文字、図形、画像、音声などを複合したものとなる。低価格、簡便性、耐環境性及び高品位（画質など）の要件が満たされる必要がある。文書の高品位化は精巧な印刷技術を取り込むことで推し進められるであろう。出力技術にとって重要な事項として、データフォーマットの標準化がある。漢字についてはコード、フォントが標準化されようとしている。文章に関してはワードプロセッサ間の通信について、図形に関してはソフトウェアモデルの、画像に関してはファクシミリなどの標準化が進んでいるが、文章、図形、画像、音声の複合データフォーマットの標準化はこれからの課題である。

オフィス用のハードコピー装置には複写機、プリンタ、COM（Computer Output Microfilm）及びプロッタなどが考えられる。ジアゾ複写機は古くから広く用いられているが、主流はゼログラフィーに始まった電

子写真式複写機である。ファイバレンズアレイの導入などによる小形化、拡大、縮小及び両面コピー、紙処理の自動化などの高機能化、1成分トナーや圧力定着法の導入による保守の簡単化などの改良が行なわれている。プリンタ方式では、電子写真式レーザービームプリンタ、感熱プリンタ、インクジェット及び静電プリンタの4種が大部分を占めるであろう。電子写真式レーザービームプリンタは、レーザーダイオードを利用した小形化、2色印刷、両面印刷などの高機能化が行なわれている。レーザに代わり密着形の発光ダイオードアレイを用いるLEDプリンタの開発も行なわれ、薄形プリンタの出現に希望がもてる状況である。感熱プリンタ、静電プリンタ、ワイヤドットプリンタも小形化、高性能化、カラー化などが進められている。

ディスプレイ装置は、CRT（Cathode Ray Tube）が主流である。テレビジョン

受像機と同じ方式であるラスタ方式が主に用いられ、このほかランダムスキャン方式、ストレージ方式などが用途に応じて用いられている。高精細化は、ディスプレイ装置の側からだけでなく、高精細テレビジョン放送の技術開発の面からも行なわれており、いよいよ拍車がかかるであろう。CRTは、画面サイズに比べ占有面積、重量が大きく、平面ディスプレイの研究が盛んである。液晶、プラズマ、エレクトロルミネセンス、エレクトロクロミックなどの技術が有望である。電子会議に利用される投写形ディスプレイは、CRT投写方式、液晶のレーザ書込方式が有力である。音声出力技術については、PCM方式、PARCOR方式が実用となっているが、文章を分析した規則によって音声出力する研究に大きな期待が寄せられている。