

複合文書処理における文書構造とプロトコル

柳 建一郎
((株) 東芝 青梅工場)

1. はじめに

統合ソフトウェアとして、文章のみならずグラフ、イメージを扱うものが最近増えつつある。

当社でもオフィス作業の機械化を推進するものとして、高性能ワークステーション (J5070) 上に複合文書処理サブシステム (OAEEDITOR) を開発した。ここでは複合文書処理サブシステムの特徴、文書構造を中心に述べる。

2. 複合文書処理システムの概要

2.1 ハードウェア構成

高性能ワークステーション (J5070) のシステムアーキテクチャは、図. 2に示すようにシステムバ

スを中心とした各種の専用途をもつ複数のマイクロプロセッサが効率良く処理できる方式をとっている。

(1) MCP (メインコントロールプロセッサ)

ワークステーション制御の中心となるプロセッサであり、アドレス空間は16MBの16ビットCPUを使用している。メモリ方式はマルチウィンドウ制御時に動的に生成・消滅されるウィンドウデータを効率的に処理するため、ページング方式がとられている。さらにソフトウェアでのマルチプロセッシングの効率化のため、複数のアドレス変換用マッピングテーブルを装備している。

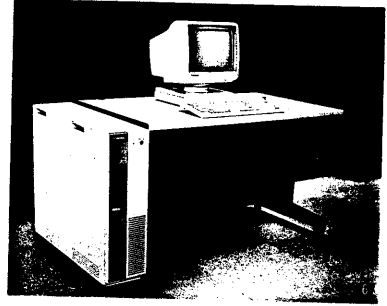


図. 1 高性能ワークステーションの外観

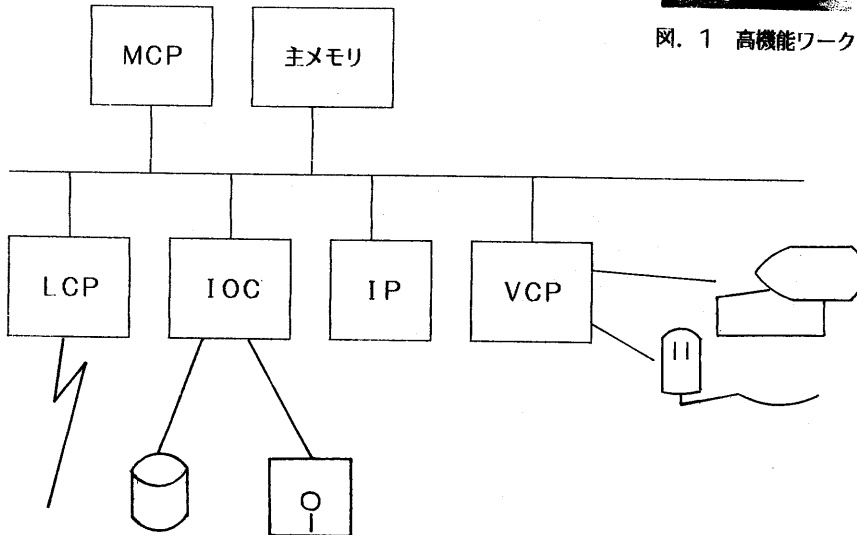


図. 2 高性能ワークステーションの構成

- ①デスクトップモデルの採用
- ②オブジェクト先行型の操作
- ③アイコンによる表示選択方式
- ④プルダウンメニューによる操作の指示
- ⑤プロパティシートによる属性の変更
- ⑥仮想キーボードによる特殊文字の入力
- ⑦関係の部分を参照できるヘルプメッセージがある(図. 4)。

3. 複合文書処理サブシステムの文書構造

3.1 文書データ構造の設計方針

前述のような高度な文書処理機能を実現するための文書構造の決定の方針として、以下の考慮を行った。
 (1) 文章、グラフ等の異なるメディアを融合するために必要な基本技術を確立する。

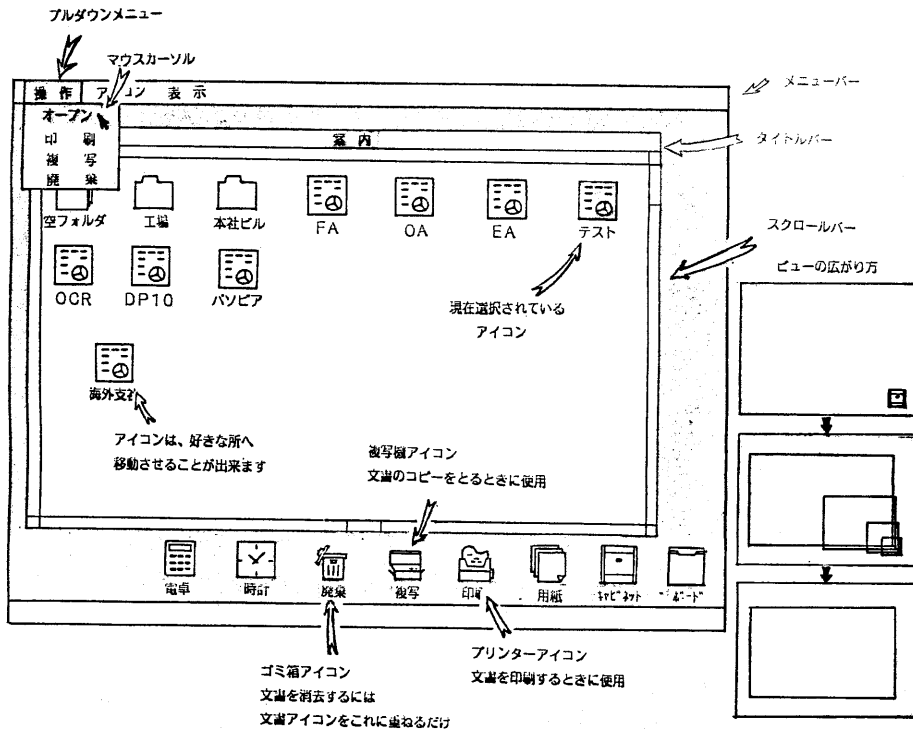


図. 4 アイコン操作におけるマンマシンインタフェース

(4) 表示と印刷の完全一致

高機能ワークステーションにより画面表示上も行ビッチ、文字サイズが可変にでき、グラフ等との融合においても印刷結果と同じものが得られる。これにより校正編集の結果をただちに確認することができ、文書作成の能率が向上する。

(5) ワークステーション主導の構成

会話形システムであるため、操作に即時に応答することが要求される。このため編集機能をワークステーションに負担させ、必要なデータを全てワークステーション内に置き、性能を維持している。本体はファイルサーバとして位置付け、文書管理、ヘルプメッセージの維持等も行っている。

単体機能としての文章、グラフ等のデータ構造の場合と融合されたデータ構造は大きく異なると思われる。またイメージデータ等の新しいメディアの扱いに関しても基本技術を確立する必要があった。

(2) 各製品間のデータの互換性、転送方式の確立を図る。

(1) で得られたノウハウを活用し各製品に移植することにより互換性の高いものとする事ができる。各製品により個々の機能を付加する場合があるが、文書データ構造としては、これら要求のフルセットとして実現をすることにより、各製品間との上位互換を考えるものとする。

3.2 文書の全体構造

文書データの全体構造は、大別すると文書ヘッダ部と文書本体部より構成される。文書ヘッダ部は、更に文書管理部、共通レイアウト情報部、注釈記述部より

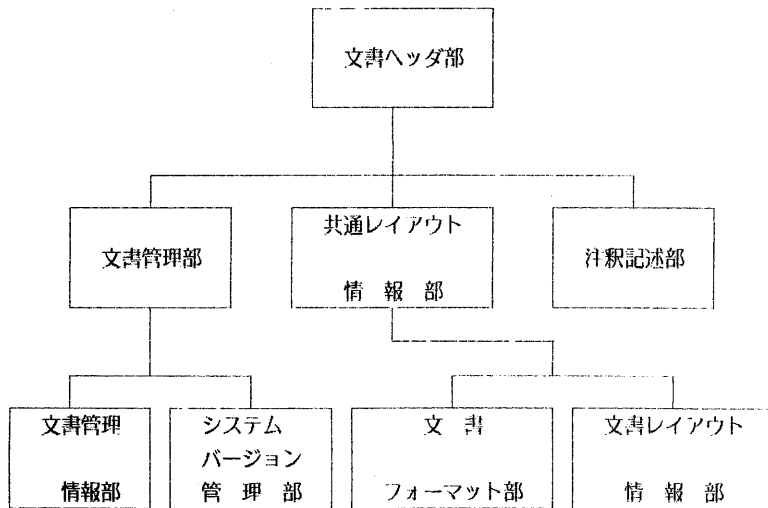


図. 5 文書ヘッダの構造

構成される。文書本体部は、複数の頁より構成され、頁をグループ化することも可能である。

文書ヘッダ部は、文書管理部、共通レイアウト情報部、注釈記述部より構成される。文書管理部は更に文書管理情報部とシステムバージョン管理部より構成される。共通レイアウト情報部は文書フォーマット部と文書レイアウト情報部より構成される(図. 5)。

文書本体部のレイアウト構造は、頁、領域の階層構造より構成される(図. 6)。

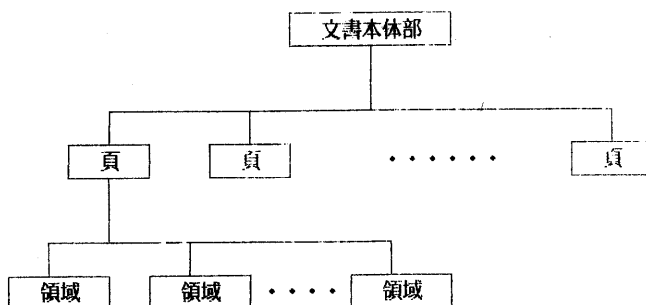


図. 6 文書本体の構造

3.3 頁内構造

文書の頁をながめてみると色々な異なるメディアが存在するが、矩形でそれらを区分し、別々に扱うことができることがわかる(図. 7)。

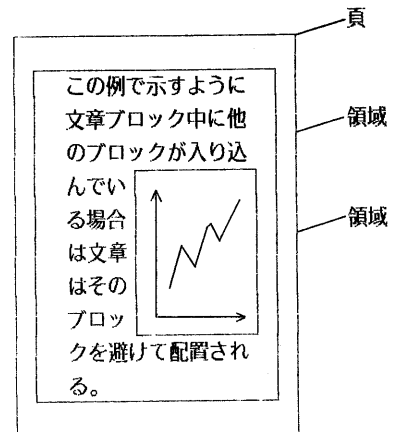


図. 7 頁内の構造

表. 1 領域の定義

属性	文章、表、ビジネスグラフ、図形、イメージの5種類とする
形状	矩形とする
位置	頁内であれば任意の位置、領域相互の位置関係も自由
大きさ	頁の大きさ以下
重ね	透明、不透明、挿入(文章領域に対してのみ有効で文章データを押しつける)
座標軸	4象限分
連結	複数領域にまたがる場合の文章の順序づけ

また複雑な階層構造もこの矩形をいくつも重ねてゆくことにより、最後は1つのみのメディアを含む矩形とすることができる。そしてその矩形を領域と呼ぶことにする。領域の定義は表. 1に示すようなものと考えられる。

3. 4 領域のデータ構造

(1) 文章領域

文章領域の内容は図. 8に示すように文字列、文字属性情報、書式定義情報からなる。

文字属性情報は文字列と独立に存在するのではなく、文字列中に特殊なコードに埋めこまれる形式となっている。文字属性情報としては文字サイズ、文字強調（アンダーライン、網かけ、囲い込み等）、均等割付、ルビ、添字等からなる。

書式定義情報は、行ピッチ、文字ピッチ、左マージン、タブ位置指定等からなる。

縦組、横組は文章の属性ではなく領域の属性としてとらえた。また文章は領域の重ね等により矩形にならないケースでも、残された部分のみで字詰めを行う。また挿入等により文字があふれた場合も領域の連結があれば自動的に文字送りがされる。

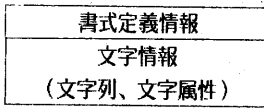


図. 8 文章データ構造

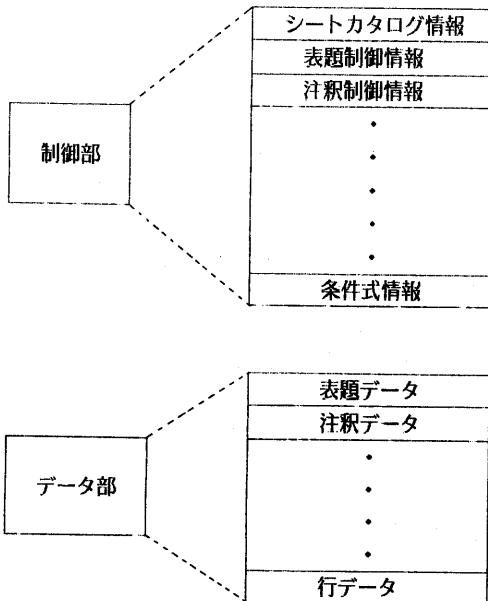


図. 9 表データ構造

(2) 表領域

表領域の内容は図. 9に示すように大きくは制御部とデータ部に分かれる。制御部はさらに表題、注釈、列見出し、……、条件式情報からなる。データ部はさらに表題、注釈、……、行データからなる。

この構造により行、列単位の処理が高速に行え、一般のデータファイルとのデータのやりとりが簡単に行えるようになっている。

(3) ビジネスグラフ

ビジネスグラフ領域の内容は図. 10に示すようになっている。

ビジネスグラフ領域の特徴としては、データ情報とグラフ属性を分離することにより、グラフ種の変更が簡単に行えること、また表処理で入力したデータからグラフが描画できることである。

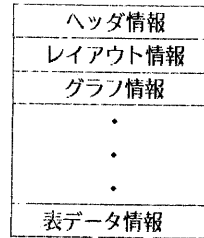


図. 10 ビジネスグラフデータ構造

(4) 作図領域

作図領域の内容は図. 11に示す構造となっている。作図領域の特徴としては、線、円等の個々の要素を単位とし、個々の図形をグループ化して一つの図形としての扱いが可能になっている点にある。

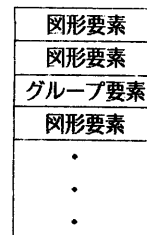


図. 11 作図データ構造

(5) イメージ領域

イメージ領域の内容は図. 12に示す構造となっている。イメージはイメージ入力出力の機器の特性を考慮してヘッダ部に圧縮方式、密度、階調等の情報を持っている。この情報を用いることにより、入出力機器に対してIPを使い変換を行っている。

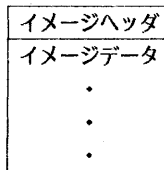


図. 12 イメージデータ構造

3.5 データ構造の表現形式

データ構造を具体的に表現する方式としては図. 13に示すような可変長データ方式を考えた。この方式の特徴は

- (1) イメージのような比較的長く不定長なデータや図形データのような比較的短く固定長要素の強いデータまで、無駄なくかつ無理なく表現できる。
 - (2) 先頭に識別子、長さという固定形式があるため取り扱い上扱わないデータが簡単にスキップできる。
 - (3) 識別子を増やすことにより機能強化、互換性の問題も対応しやすい。
- などである。

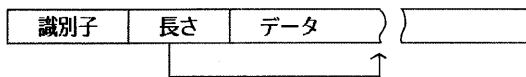


図. 13 可変長レコードの形式

4. 文書通信プロトコル

4.1 ホスト-ホスト間プロトコル

文書通信プロトコルとしては以下のものが考えられる。

- (1) 同機種間の文書転送
- (2) 異機種間の文書転送
- (3) 本体とワークステーション間の文書転送

(1)については分散処理コンピュータではDPNETという水平分散型のコンピュータネットワークがすでに実現されている。この機能としては通信回線を経由しているリモートコンピュータのファイルシステムが、応用層からはあたかも自ノードにあるファイルに見える機能NAF (Network Access Facility) が実現されている。またNSF (Network Station Facility) では、他ノードのステーションとしてネットワークを意識させない結合が可能である。

これらの機能により複合文書サブシステムではネットワークを意識することなく文書の転送ができることになる(図. 14)。

(2)のケースが一般に言う文書転送プロトコルに当り、文書転送サービスにおいては親展、同報、時刻指定等がある。

(3)については本システムではワークステーション側に負担を分散する方式で考えたため、本体-WS間で文書を転送するプロトコルが必要となった。この方式の場合、考慮すべき大きな問題は転送の単位を何とすべきかである。これには以下の2つの方式が考えられる。

①1文書単位で転送を行う

②1文書内の必要部分(例えば頁単位)で転送を行う。

①は最近の垂直分散を行うシステムで多く採用されている方式であるが、我々は②の方式を採用した。この理由は以下の通りである。

②の文書単位転送方式ではワークステーション側に、最大の文書を扱うだけのファイル容量が必要であること。①では一般的には文書全体を転送するまでは文書の展開表示が不可のため、高速の通信回線以外では文書の指定から表示までが遅くなる。

逆に②の方式では小さな単位となるため制御がやや複雑になる。

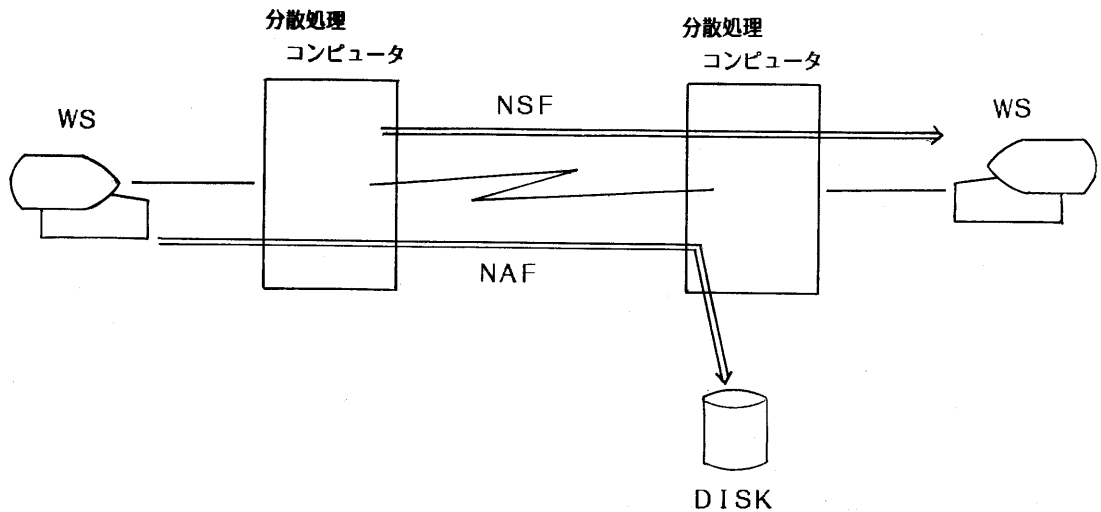


図. 14 DPNETの機能

4. 2 本体-WS間の通信プロトコル

複合文書処理システムでは表. 2のプロトコルを採用した。

表. 2 本体-WS間のプロトコル

- 文書フォルダのオープン
- 文書のオープン
- 文書フォルダ、文書のクローズ
- 文書リストの取り出し
- 文書フォルダの作成
- 文書の作成
- 文書フォルダの削除
- 文書の削除
- 文書の複写

このプロトコルによりワークステーション側から本体にある文書をファイルの階層構造をたどることにより検索できる。また文書の複写等もプロトコルさえ出せば本体側のファイルサーバ機能で実行できる。

5. 標準化動向

文書構造ならびに文書通信については、いろいろな機関で標準化が検討されている。

残念ながらすべてを網羅する決定版標準仕様はなく、いろいろな分野での標準化をボトムアップして行っているのが現状である。

文書データ構造についてはISO/TC97/SC18のODA (Office Document Architecture) に従った考え方とするのが最善と考えられる。しかし、詳細部分については未定部分もある。

データ構造の表現形式については日本語文章ではJIS C-6225、グラフィックスではGKS (graphics kernel system) 等があるが、ビジネスグラフについては標準がない。以上の点も考慮して可能な限り標準化に従うものとしたが、未定の部分等は独自の構造となった。

6. おわりに

複合文書処理システムの文書構造とプロトコルについて述べてきたわけであるが、この分野はまだ始まったばかりであり今後の発展が期待される分野と思われる。発展の方向として標準化が推進され異機種、異文書構造間で文書データが交換されるのが究極の結

論であるが、ここでは別の面からの発展を書いてみたい。

現在、WP機能、作表機能あるいはそれらの統合された機能があるがそれらはすべて独立した機能として存在していた。利用者がこれらの機能とデータのやりとりを行いたい場合、変換ユーティリティなどにより静的な変換しか行えなかった。今後OA機能が発展してゆくためには独立した機能であるよりもむしろ、いかに応用ソフトにその機能が提供できるかが大切であろう。我々はこの複合文書処理システムを単なる、独立したユーティリティとしての存在するではなく、以下に述べる2つのインタフェースで結合を図ろうとしている。

1つは文章、作図、イメージ処理については基本機能のコマンドとしてのインタフェースである。例えば応用ソフトから文章であれば文字列の挿入、削除が自由に行える。この基本コマンドを組み合わせることで、より利用者に最適の文章処理機能を独自に作成できる。

他の1つはユーザインタフェースを含めた高レベルインタフェースである。これはデータ構造、操作はこの統合OAサブシステムのものを使うことになるが、簡単にサブシステムを使用する場合有効である。

参考文献

- [1] “文書変換に関する国際標準化動向”、情報処理学会誌、VOL. 26、No. 1、pp33-41、1985
- [2] “日本語処理技術に関する調査研究”、日本電子工業振興協会、1985
- [3] “日本語処理技術に関する調査研究”、日本電子工業振興協会、1984
- [4] “Office Document Architecture Second Version of DP”、ISO/TC97/SC18/WG3 N391、October、1984
- [5] 棚橋他 “高機能ワークステーションの設計思想” 他、情報処理学会第30回全国大会、pp1809-1828、1985