文書処理とヒューマンインタフェース 20 － 3
（１９８８．９．８）

文書構造を用いた自動レイアウトシステム

福井 美佳* 土井 美和子* 竹林 洋一* 山口 浩司* 岩井 勇* 大黒 和夫**

* （株）東芝 総合研究所 ** （株）東芝 情報通信システム技術研究所

我々は、文書の構造を用いる自動レイアウトシステムを開発し、日本語ワードプロセッサJW-1000AI上に実装した。

従来の、ユーザが文書の構造を指定する文書整形システムでは、ユーザの負担が大きかった。本システムでは、文書の形態的情報やキーワードを用いて、標題，章，節などによる階層構造と、本文と図表間の参照構造などの、文書構造の自動抽出を行う。さらに、文書構造に基づいた書式とレイアウト知識を用いて、文書のフォーマッティングや図表の割り付けを行う、自動レイアウト処理を実現した。また、ユーザの指示と組み合わせることにより、ユーザの持つ文書構造の多様さやレイアウトの好みにも対応している。

A Document Layout System Using Automatic Document Architecture Extraction

Mika FUKUI, Miwako DOI*, Yoichi TAKEBAYASHI**
Kouji YAMAGUCHI, Isamu IWAI*, Kazuo OHGURO***

* Research & Development Center, Toshiba Corporation
** Information and Communication Systems Laboratory, Toshiba Corporation

* 1 Komukai-Toshiba-cho, Saiwai-ku, Kawasaki 210, Japan
**2-9 Suehiro-cho, Ohme 198, Japan

A document layout system based on automatic extraction of document architecture including logical and reference structures has been developed for reducing user's effort concerning document layout and text formatting, and has been implemented in a highly equipped Japanese word processor. The extracted logical structure consisting of logical elements such as chapters and sections is used for automatic document formatting. The extracted reference structure which indicates relations between referring words and referred figures/tables is used for automatic layout of the text, figures and tables. Automatic text element recognition is performed by morphological analysis using key words. Through in-line(one sentence) and inter-line(relations between sentences) analysis, logical and reference structures are obtained. The layout system automatically performs not only automatic layout but also text formatting using the extracted document structures. User definable structure and analysis error messages have been introduced for reliable user-interface.

1．はじめに

ワードプロセッサの高機能化に伴い、より高度な文書処理システムに注目が集まっている。

例えば、複雑なレイアウトと美しい印刷を可能にするデスクトップパブリッシング（ＤＴＰ）や、自然言語解析技術を用いた文書推敲支援システムなどが開発されている。また、文書を複数のデータ（ノード）とデータ間の関係（リンク）による構造で表現するhypertext の概念も、検索や文書作成支援などに応用されている。[1][2]

現在では、ＤＴＰ等の文書のレイアウトを行うシステムにも、標題，章，節，図表等の文書構造を取り扱えるものが増加している。しかし、現状のシステムでは、文書構造を指定する作業でのユーザの負担が大きく、ユーザが機能を十分に使いこなすのを妨げている。そこで、ユーザインタフェースをより容易にするために、文書構造を自動的に抽出するものを考えた。

本論文では、文書の標題，章，節，箇条書きなどの論理構造と、本文と図表の参照関係等の参照構造を自動的に抽出し、その情報を用いて自動的にレイアウトを行う、文書自動レイアウトシステム[3]～[10]について報告する。なお、本自動レイアウトシステムは、東芝の日本語ワードプロセッサJW-1000AI 上に、知的レイアウト機能として実装されている。

2．文書処理システムのユーザインタフェース

2．1　従来の文書処理システム

近年、文書処理の分野にも、マウス等を用いて、印刷に近いイメージを直接編集・レイアウトできるインタフェースが、採用されるようになってきた。従来のコマンド指定などに比較してもわかりやすく、一般のユーザに適したインタフェースとされている。[11]

このインタフェースにおいて、ユーザの編集の対象となるのは、文書全体あるいは文字単位のこともあるが、段落，見出し，章，図表などの論理的なまとまりであったり、頁やフレームなどのレイアウト上のまとまりであったり、そのときのユーザの目的に応じて違っている。従来のシステムでは、操作の対象が変わるたびに、その範囲（始点・終点）を、ユーザが指定しなければならないものが多かった。

また、これらのまとまりの間にはなんらかの関係があり、編集の影響が関係のある他のまとまりに及ぶことが多い。例えば、章節などの階層関係や図表などの参照関係があるとき、「*2章*を削除する」という操作の影響で、以下のような新たな操作の必要が生じる。

① *2章*で参照していた図表を、他に参照している章があれば移動し、なければ削除する。

② *3章*以下の変更すべき章，節，図表番号をつけかえる。

③ 変更になった章，節，図表番号を、参照している箇所の記述も変更する。

従来は、これらの操作をすべて、ユーザが指示していた。

2．2　文書構造を持つ文書処理システム

従来のシステムの問題点を解決するような、まとまりと関係（文書構造）を、ユーザが指定することによって取り扱える、文書処理システムは増加しており、文書推敲支援や文書整形などに用いられている。これらのうち、代表的な文書整形システムのユーザインタフェースについて述べる。

コマンド型文書整形システムには、Scribe[12]のように、標題，著者，章，図表，参考文献など各まとまりの論理属性や参照関係を指定する、コマンドなどを文書中に埋め込むことにより、文書構造に基づいたレイアウトや、目次，参考文献の作成が可能なものがある。これらのインタフェースは、コマンドの挿入というよりプログラム作成に近いので、ソフトウェアの知識をもつものには理解しやすいが、一般ユーザには使いにくい。すなわち、ユーザが文書として思い浮かべるイメージとは違う形で文書を作らねばならず、WYSIWYG なシステムに比較しても、初心者が文書構造を指定する際の負荷が大きい。

一般的なＤＴＰシステムには、ワークステーション上で稼動し、画面上で直接印刷イメージを確認できるWYSIWYG型のものが多い。例えば、東芝のワークステーション AS-3000 上のＤＴＰシステムAS-Documentsのように、マウスを用いて、パラグラフ（段落）単位の論理属性の指定やグループ化、ページやフレーム等のレイアウト構造の指定を行うことにより、文書構造による操作を行えるシステムがある。これらのインタフェースは、印刷イメージが目に見えて直接操作できるため、初心者にもとっつきやすいという利点がある。しかし、複雑なレイアウトの文書を作成するためには複雑かつ繁雑な操作が必要になり、レイアウト構造が理解できないユーザには、コマンド指定と同程度に負荷が大きくなってしまう。

2．3　自動レイアウトシステムの位置付け

以上のシステムでは、文書構造やレイアウトに関する知識を十分には持っていない一般ユーザにとって、文書全体の構造を矛盾なく指定することは、困難で時間のかかる作業になる。また、ユーザが自分でレイアウトを決定する場合、レイアウトに関する経験的知識や美的センスによって、完成度が左右されるきらいがある。

このような問題点を解決し、ユーザの負担を軽くするため、文書構造とレイアウトに関する知識を持ち、文書の整

図１　起動時画面例

形を自動的に行う、自動レイアウトシステムを開発した。このシステムは、各文書内のまとまりや関係の抽出を、ユーザの特別な指定作業なしで行う文書構造抽出と、図表の配置などを自動的に決定し、文章を書式に従って整形する自動レイアウトによって、構成されている。

また、本システムは、文書の構造やレイアウトに関する詳しい知識を持たない一般ユーザを、対象としている。

３．システムの機能説明

自動レイアウトシステムはAS-3000 上に構築されており、また、JW-1000AI に知的レイアウト機能として実装されている。

知的レイアウト機能は、ワープロ上で従来通りの方法によりユーザが作成した文書に対して、自動レイアウトを行うものである。以下のような機能をもつ。

- 原文１文書に対して、図表４文書をレイアウトできる。
- 書式データは、目的別に複数用意されており、ユーザが選択して使用する。（例えば、用紙サイズＡ４，Ｂ４，Ａ５，Ｂ５，Ａ３毎に１段組，２段組などの標準書式のほか、情報処理学会，電子情報通信学会など）
- 表紙や目次の自動生成

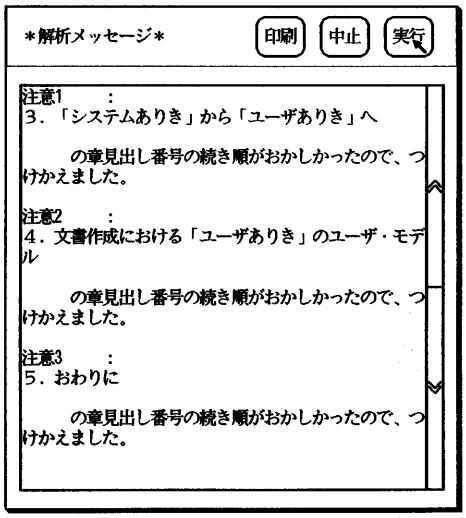


図２　解析メッセージ

- 見出し番号誤りの自動訂正
- 制御コードによる部分的なユーザ指定も可能
- 解析メッセージの表示

制御コードとは、レイアウト前に文章中に挿入し、ユーザが直接、文書構造や図表の割り付け位置を指定するものである。部分的に指示したい場合に、有効である。

解析メッセージとは、システムの持つ文書構造モデルに合わない文書が入力されたとき、ユーザにその原因を知らせるメッセージである。レイアウトが思うようにできない場合に、原因を知るためだけでなく、文書の校正支援としても利用できる。

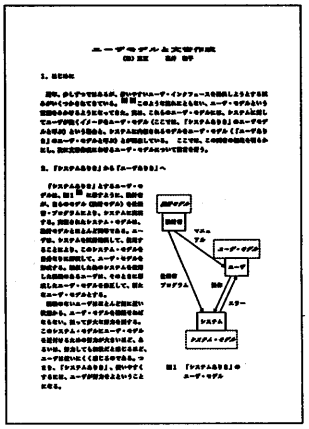
図１に、知的レイアウト機能起動時の画面例を示す。レイアウト前の原文を編集中に、プルダウンメニューを開いて、知的レイアウトを選択し、書式データ名と図表文書名（図表がある場合のみ）を指定して、実行する。

処理が終わると、図２のような解析メッセージが現れる。この例では、図１に示した原文の、２章の章番号が誤って３になっていたので、整形文書で以降の章番号のつけかえを行ったという、メッセージが表示されている。他にも、参照箇所に対して対応する図表がないときなど、同様なメッセージが現れる。

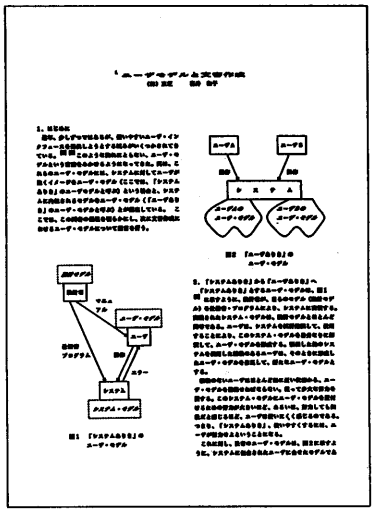
最後に、整形された文書の印刷イメージが画面に表示されるので、整形された文書の印刷や保存などの指示を行う。図３に、２種類の書式を用いて印刷した、レイアウト例を示す。図３（ａ）は、Ａ４縦１段組の書式によるレイアウトの第１ページで、図３（ｂ）は、Ｂ４縦２段組の例である。

本システムの構成図を、図４に示す。

本システムは、文書の階層構造や参照構造などの文書構



（a）A4　第1ページ



（b）B4　第1ページ

図3　レイアウト例

例えば、学会の予稿集のような技術文書を調査したところ、標題，著者名，所属，概要，章，節，項，箇条書き，例，式，定義，図，表，参考文献などの、論理的なまとまりがあることが判明した。そこで、技術文書の文書構造モデルを、図5のように定めた。

実線で表わされている階層的な構造（階層構造）と、それにからむように点線で表わされている網目上の構造（網構造）を合わせて、文書の論理的構造とする。図5（a）にみる階層構造は、章同士，箇条書き同士，段落同士の兄弟関係や、章節間の親子関係などによる、階層（レベル）を持つ。図5（b）の網目構造は、本文中から図表などを参照する参照関係による、参照構造である。

造を抽出する部分とレイアウト処理をする部分に分かれている。以下、4章で文書構造の抽出、5章で自動レイアウト処理について説明する。

4．文書構造の抽出

4．1　技術文書の文書構造

本システムで自動抽出する文書構造は、文書が用いられる分野によって異なっている。また、ユーザの負担を軽くするために、より一般的な形であることが望ましい。

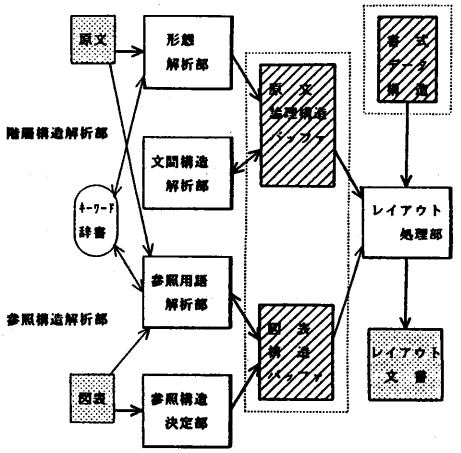
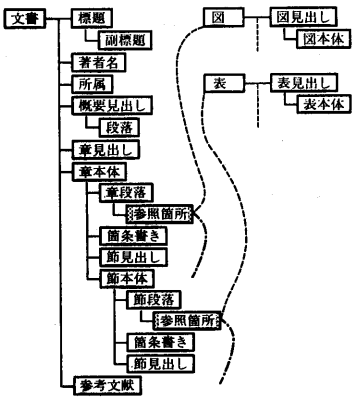


図4　システム構成

4．2　形態による文書構造抽出

一般的には、文書の構造を自動抽出するためには、構文解析による自然言語処理や、文脈理解による意味構造の解析等が、不可欠であると考えられている。

しかし、図5に示したような構造であれば、文章の内容まで読んで理解しなくても、ざっと目を通した段階で、ある程度は抽出できる。このとき、人間が抽出に用いる情報を調査した結果、行あけや段下げ，センタリング，フォントの変更，各種文字修飾などのレイアウト情報や、見出しに含まれている英数字や記号などの形態的情報、よく使われるキーワード、著者名や所属に含まれる固有名詞などが、用いられることが判った。



(a) 階層構造　　(b) 参照構造

図5　文書構造（技術文書）

このうちレイアウト情報については、技術論文などの原稿を調査すると、文書全体で統一されたレイアウトが行われていないことも多く、むしろあいまいな情報であることがわかった。また、本システムではレイアウトされていない文書でも、取り扱えるようにするため、基本的にはレイアウト情報は利用しないこととした。

そこで、レイアウト情報以外の形態的情報やキーワード等から、文書構造を抽出する方式を開発した。４．３節に階層構造、４．４節に参照構造を、具体的に抽出する方法について述べる。

４．３　階層構造抽出処理

４．３．１　処理の流れ

本システムでは、改行コードまでの１パラグラフを一文と定義して、取り扱っている。図４に示すように、階層構造抽出処理は、一文形態解析部と文間構造解析部の２つに分かれている。

一文形態解析部では、キーワード辞書に登録されている、よく使われる語句や英数記号の抽出を行って、一文の構成要素を明らかにする処理を行っている。これによって、見出し（標題，章見出しなど）と内容部（章段落，節段落など）の判別を行う。

文間構造解析部では、一文形態解析情報を用いて、見出しの形態のマッチングやキーワード情報などにより、文と文の間の階層構造を決定する。

すなわち、文書の形態素にあたる一文の解析を行ってから、文書構造構文木に当てはめる、ボトムアップ方式に近い階層構造決定処理を行っている。

４．３．２　一文形態解析部

著者，所属，参考文献部などを除く、通常の技術文書の一文は、図６のような構文図で表わすことができる。この構文図を用いて、以下の３種類の一文構造が解析できる。

① 「見出し」のみの文（例　『１．はじめに』）

② 「内容部」のみの文（例　『　ワードプロセッサの高機能化に伴い、…。』）

③ 「見出し」＋「内容部」の文（例　『１．はじめに　ワードプロセッサの高機能化に伴い、…。』）

「見出し」はさらに、図７のように形態的に分類できる。「前見出し記号」は、英数記号等で構成される見出しの番号やレベルを表わすものである。また、「主見出し本体」は見出しの文字部分を表わし、一般的には、名詞句や見出しによく用いられるキーワード（見出し予約語）で終わっている。

一文の構成要素の最少単位であるキーワードは、「英数字部」「後置部」「見出し予約語」等のカテゴリに分類されて、あらかじめキーワード辞書に登録されている。

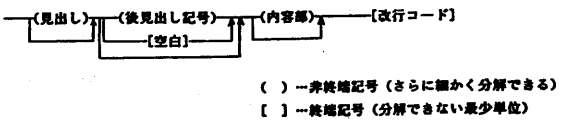


図６　一文構成要素構文図

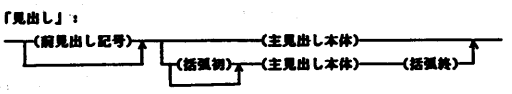


図７　見出し構成要素構文図

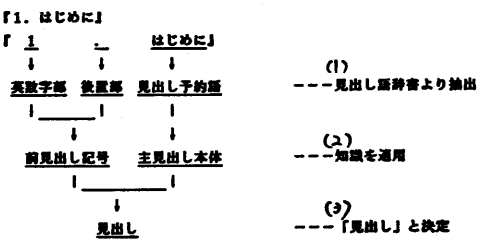


図８　一文形態解析処理

```
(ルール　章見出し１
        (見出し予約語を含む)
        (前方に章見出しがない)                    条件部
→
        (章見出しである)
        (構成要素を章見出しパターンとする)        動作部
)
(ルール　章見出し２
        (見出し予約語を含まない)
        (前方に章見出しがある)
        (章見出しパターンと一致する)              条件部
→
        (章見出しである)
        (章見出しパターンのオーダを１増す)        動作部
)
```

図９　階層構造決定ルール

図８に、①の例文『*1．はじめに*』の、具体的な一文形態解析部の見出し決定処理を示す。処理は、以下の３段階からなる。

（１）　文頭など必要な部分でのみ、キーワード辞書と抽出規則を用いて、キーワードの抽出を行う。

（２）　前見出し記号や主見出し本体などの構成知識とのマッチングを行う。

（３）　見出しの構成知識（図７）を用いて、「見出し」の決定を行う。

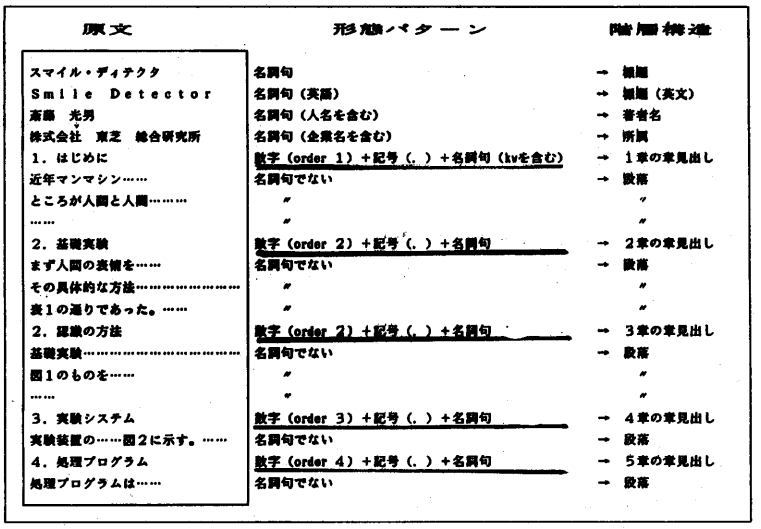


図１０　階層構造解析処理

４．３．３　文間構造解析部

図５（ａ）の階層構造を抽出するため、階層構造知識を実装している。この知識は、具体的には、約３００の文間構造決定ルールで記述されている。図９に決定ルールの１部を示す。

図１０に、簡単な例を用いた、文間構造解析処理の例を示す。処理は、以下の２段階からなる。

（１）　形態解析部で得られた各見出しの形態の、パターンマッチングによって、章，節などの兄弟関係を決定する。（この例では、前見出し記号のパターンが、「数字」＋「記号」で一致しているので、章見出しの兄弟関係が決定される。）

（２）　（１）の結果と形態情報やキーワードにより、全体の階層構造を決定する。

４．４　参照構造抽出処理

４．４．１　図表の参照構造

図表を含んだ技術文書の場合、文章中にその図表を参照する箇所（参照箇所）が含まれている。参照箇所は、『*図１の…*』『*図1．3に示す。*』『*第3－2表参照*』というような、番号をともなった図表を表す語句であることが多い。この場合参照される図表にも、同様な語句を含んだ図表見出しがある。ここでは、これらの図表を表わす語句を参照用語と定義する。参照箇所と図表は、参照用語によって関係付けられている。

実際の技術文書では、１つの参照箇所から複数の図表を参照したり、１つの図が複数の章から参照されたり、参照構造が網目状に張りめぐらされている。よって、図表の参照構造は、参照用語によって関係付けられた、文章と図表の間の網状関係で表現される。

４．４．２　処理の流れ

図４に示すように、参照構造解析部は、キーワード辞書を用いた参照用語解析部と、参照構造決定部に分かれている。

キーワード辞書には、前節で示した参照用語、例えば「*図*」「*Fig.*」など図を表わすキーワードや「*(). －～*」等の区切り記号が、それぞれの役割に応じたキーワードのカテゴリに分類されている。

参照用語解析部では、階層構造抽出処理と同様に、文章と図表の文字部分から、キーワード辞書に登録されている参照用語と一致する部分を抜き出す。例えば、文章中に『*図１の…*』という表現があった場合、図１１のように、参照用語「*図*」「*1*」が切り出され、*1*という番号を持つ図への、参照箇所候補となる。

また、図表の文字部分から切り出された参照用語などから、図表の見出しにあたる文を決定する。例えば、図１１のように、『*図1　認識の流れ*』という文が図表の中に含まれている場合、図見出しになる可能性があることがわかる。同一図表中のすべての文字部分をこのように解析し、図表見出しになる可能性がもっとも高い文を、その図表の見出しと決定する。

参照構造決定部では、文章中の参照箇所候補と図表見出しを比較して、照合するものがあればその図表への参照箇所と決定し、リンクづけを行なう。図１１では、図１の参照箇所候補と図見出しの形態が一致するので、リンク付け

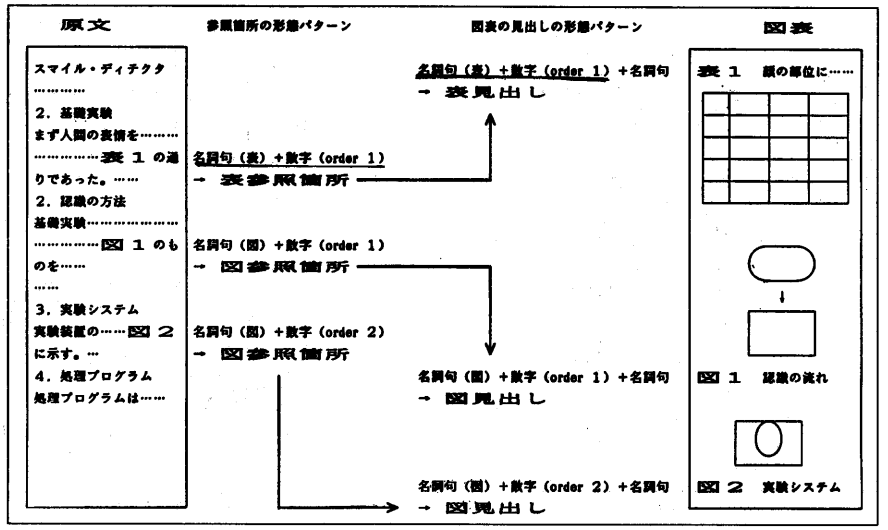


図１１　参照構造解析処理

を行っている。また、同一の図表への参照箇所が複数ある場合は、それらを候補リンクで連結する。

5．自動レイアウト処理

5．1　文書の自動レイアウト

文書のレイアウトは、ページや段組などの制限を考慮しながら、わかりやすく美しい配置を決定していく作業である。人間は、論理的なまとまり単位でレイアウトし、参照箇所の近くに図表を配置することによって、わかりやすいレイアウトを実現している。例えば、図表を頁のどこに配置するか決定する図表の割り付けや、見出しがフレームの下のほうに位置しないようにする見出しの追い出しなどのレイアウト知識は、ある程度一般的に記述できる。このような文書構造に基づいたレイアウト知識を、システムが持つことにより、画一的な流し込みとは異なる、構造に即したレイアウトが可能になる。

一方、１ページの紙面上に文章をどのように割り付けるかを定義する枠（文章フレーム）と、各論理属性をどう展開するか等は、文書の目的などに応じて異なっている。よって、この様な情報は書式データに記述しておけば、書式データを多数持つことにより、必要な書式を選択するだけで、一つの文書を何通りにも利用できる。

レイアウト処理部では、上記のようなレイアウト知識を持ち、書式データに基づいて文章のレイアウトを行う。また、図表のレイアウトは、その図表が文章中から参照されたときに行っている。以下、５．２節に文章、５．３節に図表の自動レイアウト処理の説明を行う。

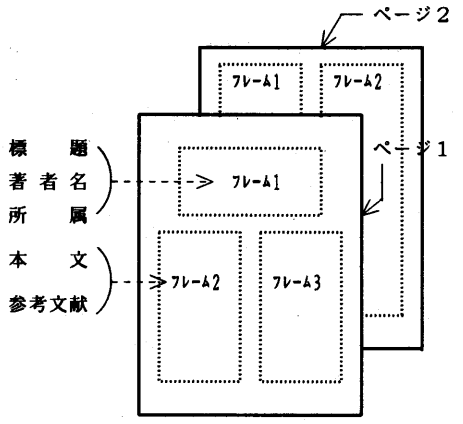


図１２　書式のデータ構造

5．2　文章の自動レイアウト処理

書式データは、文章を紙面上に物理的に割り付けるための規則であり、文書情報，ページ情報，文章フレーム情報が定義されている。

- 文書情報　　：　用紙サイズ，用紙方向など
- ページ情報　　：　実際の印字サイズなど
- 文章フレーム情報：　位置，サイズ，フレーム番号，展開論理属性など

文章フレームは１ページに複数定義でき、フレームの大きさ、位置により、いろいろなレイアウトの文書を作成することができる。図１２には書式データによって定義され

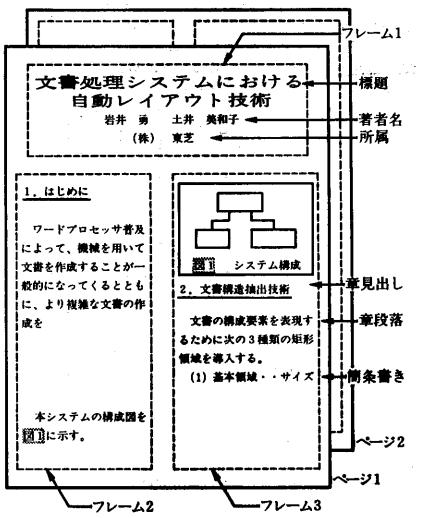


図１３　書式によるレイアウト例

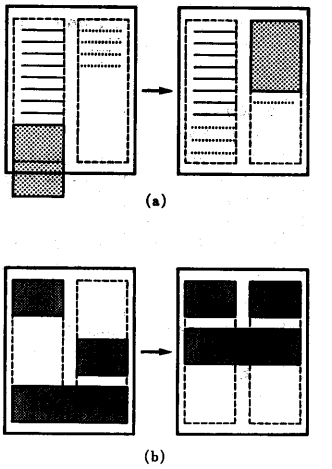


図１４　図表の割り付け例

た文章フレームの一例を示した。１ページの第１フレームはページの上部に１段組の文章フレームを定義し、第２フレームと第３フレームで２段組を構成している。２ページ以降は２段組の構成が定義されている。この文章フレームには、各フレームに展開すべき文を排他制御するための、論理属性情報を記述する。

これに対して整形処理部では、書式に基づいてレイアウトを行う、レイアウト知識を持っている。例えば、図１２に示すように、第１ページの第１フレームに標題，著者名，所属の論理属性を定義する。図１３に示す通り、原文に付けられた標題属性，著者名属性，所属属性を持つ文のみ第１フレームに展開し、それ以外の属性を持つ文は第１フレームから排除する制御を行う。同様に第２フレーム、第３フレームに章、節、箇条書き、参考文献の属性を定義することにより、第２フレーム、第３フレームに本文、参考文献が順次割り付けられる。２ページ以降の第１フレーム、第２フレームも同じ論理属性が定義されているので、文章は順に流し込まれる。

また、このほかに書式データには文章をどのようなフォーマッティング形式でフレーム内に展開するかの情報を、論理構造属性ごとに記述する部分がある。例えば、フレーム１の標題の論理属性に対して、「センタリング」、「倍角」を定義する。フレーム２の章見出しの論理属性に対して「ゴシック」、「下線」のフォーマッティング形式などを定義する。この書式データを用いて整形処理部で各文ごとに展開すると、図１３のようなレイアウト処理が行われる。

このようにレイアウト処理は、書式データに基づいて行なわれるので、書式データを変更することにより、同一原文を目的に合ったレイアウトや形式の文書として出力することができる。また、レイアウト知識を持つことによって、ただ書式データに従って文章をそのまま流し込むのではなく、割り付けられるフレーム内の位置関係などを考慮したレイアウトができる。例えば、節見出しの前を一行空ける書式であっても、フレームの先頭行には空行を入れない、などの処理を行っている。

５．３　図表の自動レイアウト処理

自動レイアウトシステムでは、従来の文書の作り方や図表の再利用を考慮して、図表は原文とは別文書として作成することを前提としている。図表の自動割付けのために、４．４節で述べた文章中の図表の参照構造抽出処理の他に、文章中に実際に割り付ける図表領域の大きさの識別も行っている。

人間が図表のレイアウトを決定する際、文章のレイアウトを行いながら、図表を参照する記述のできるだけ近くに、図表を配置しようとする。その図表配置（割り付け）のプロセスは、以下のようにシュミレートされる。

① 参照位置と同じフレーム内に割り付け可能か？
② そのページの次フレーム内に割り付けられないか？
③ そのページのどこかに割り付けられないか？
④ 次ページの先頭フレーム内に割り付けられないか？
⑤ 次ページのどこかに割り付けられないか？

本システムでは、この割り付けプロセスに基づき、参照構造情報と図表領域の大きさから、文章中の参照箇所の近傍に図表の割付けを行なう割り付け規則を持つ。これによって、図表が文章フレームやページをまたがって割付けを行なわない（図１４(a)）、あるいは１ページに複数の図

表が割り付けられる場合に、図表の大きさや図表番号の順位性などを考慮して見易く割付けを行なう（図１４(b)）、などのレイアウトを可能にしている。

６．考察

６．１　本システムの持つ問題点

本自動レイアウトシステムのユーザインタフェースは、使用する書式と図表文書の指定以外は、基本的には全自動で処理が行われる。

しかし、対象とする文書にはとくに制約を付けず、ユーザが通常作成している文書をそのまま扱うため、文書の複雑さ，多様性あるいは自然言語の曖昧さから、１００%の構造解析は現在の技術では不可能である。そのため、ユーザの意図に反した解析やレイアウトが発生することがある。本システムでは、文書構造や図表の割り付け位置を、ユーザが直接指定できる手段（制御コード）を設けて、これに対応している。

そこで、一般のユーザの作成した学会予稿集や社内報告などの原稿の、解析およびレイアウト結果とユーザの反応を調査した。これによってあきらかになった本システムの持つ問題である、

① 文書の複雑さ，多様性への対応

② ユーザ指示が必要となる要因

の２点について述べる。

６．２　文書の複雑さ，多様性への対応

以下に、本システムが、文書の持つ複雑さや多様性に対応して、正しく解析できた例と、誤った例を示す。

（１）　正しく解析できた例

- 文の冒頭が『*1 はじめに*』のように、前見出し記号と主見出しの間に空白がなくても、区切ることができる。
- 逆に、一文の冒頭が、『*②は、*』『*3個の積み木*』のように、前見出し記号があり、見出しとまぎらわしい表現でも、後ろの接続をみて、前見出しの切り出しは行わない。
- 冒頭近くに、『*10：9の比率を*』『*bit fieldを*』のように、区切り記号や空白があっても、前後の接続をみて見出しとして切り出さない。
- 標題，著者名，所属，章，参考文献などは、ユーザの考える階層構造モデルと図５のモデルは一致しており、解析できた。
- 図表の参照構造は、ユーザの考えるモデルと一致しており、解析できた。

（２）　正しく解析できなかった例

- 一文に箇条書き複数含まれている場合、ばらばらに解析することができない。（例『*①高速性　②信頼性　③安全性*』）
- 一文に２つ以上の見出しが含まれている場合、１つの見出ししか切り出せない。（例『*1. はじめに　1. 1背景　近年、…*』）
- 節，箇条書きのような細かいレベルで、階層構造モデルの不一致がみられた。

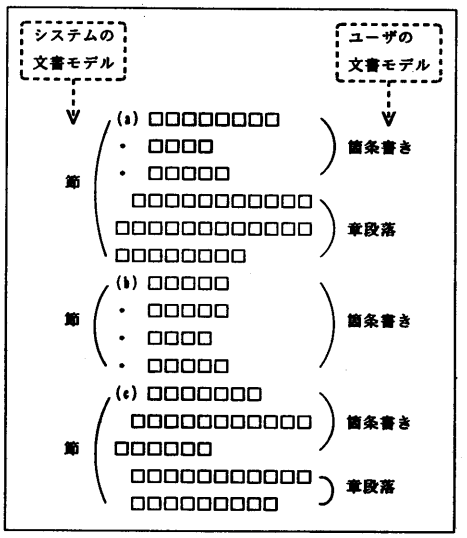


図１５　階層構造モデルの食い違い

一文の解析については、（１）で対応しているように、英数記号，数量詞等の接続関係を用いて、解析を行っている。ただし、構文木に一致しない文に関しては、解析を正しく行うことができない。

同様に、システムの持つ階層構造モデルと、大きく異なる階層構造は、正しく解析することができなかった。例えば図１５のような形態の文書の場合、ユーザの意図する文書モデルとシステムの解析が食い違ってしまう。

このような問題を解決する方法として、階層構造の抽出に関しては、レイアウト情報の取り込みが、ある程度有効であると思われる。文書を作成中のユーザは、箇条書きや節など細かい階層構造を特に意識していないが、段下げや字下げなどのレイアウト情報を挿入することにより、作成者自身はある程度判断していると思われる。レイアウト構造の解明，抽出によって、節，箇条書き，表，例，式などの、細かい分類が可能になる。

しかし、一文構造の複雑化，階層構造モデルの拡張等は、慎重にせねばならない。細かい知識の増加により、前の知識との矛盾が生じて、拡張が不可能になることもある。現在は、多様な文書構造に対しては、部分的なユーザ指定情報を解析にいかすことにより、より精度のよい解析を行う

方向で進めている。

6．3　ユーザ指示が必要となる要因

自動レイアウトがユーザの意図と食い違うのは、前節で示したような、多様な文書モデルの解析が正しく行われなかった場合と、図表の割り付け位置や大きさなどが、ユーザの好みに合わなかった場合におこる。ユーザは、解析メッセージやレイアウト結果によってそれを知り、原文を直したり制御コードを原文中に挿入して対処する。

図表の割り付け位置やサイズは、システムによって一意に決められてしまうが、ユーザの好みやバランス感覚と一致せず、クレームがつくことがある。現在は、レイアウト前の文章中に、図表の位置，大きさを指定する、制御コードを挿入してもらうことによって対応している。しかし、バランス等を考慮することにより、よりよい図表の配置を自動的に実現するとともに、各ユーザの好みに応じたレイアウト処理が必要である。

レイアウトのチェックを行うユーザを観察すると、本システムで最初にターゲットとしたような、完全にシステムにお任せという態度のユーザ以外にも、レイアウトに関する注文が多く、自分の好みのレイアウトを追及するユーザもいた。

今後は、このようなユーザに対応して、自動を基本としながら、判別しにくい点，ユーザの好みをいかせる点では、ユーザの指示を簡単に受けられるように、より対話的なインタフェースを構築していく。また、ユーザごとに、解析やレイアウトの方法を変えられるようなシステムを目指す。

7．おわりに

文書の自動レイアウトシステムは、レイアウト前の文書から、階層構造と図表の参照構造を抽出し、自動的にレイアウトを行うものである。

本論文では、文書自動レイアウトシステムの背景，現状，問題などを報告した。

今後は、自動化を基本として、ユーザの好みなどに関しては指示を受けられるような、ユーザインタフェースを持つ、知的文書処理システムに発展させていきたい。

参考文献

[1] Conklin, J., "Hypertext: An Introduction and Survey",IEEE Computer, pp.17-41, Sept.(1987)

[2] Yankelovich, N. et.al., "Intermedia: The Concept and the Construction of a Seamless INformation Environment", IEEE Computer, pp.81-96, Jan.(1988)

[3] 土井、岩井、「知的文書処理における文書処理モデル」、電子通信学会オフィス・システム研究会９月(1986)

[4] 土井、岩井、「文書処理モデルに基づく知的文書処理システム」、第２回ヒューマン・インタフェース・シンポジウム10月(1986)

[5] Doi,M.,et.al.,"Research on Model Based Document Processing System",Proceedings of INTERACT87,p.101-1106(1987)

[6] 岩井他、「知的文書処理システムにおける自動フォーマッティング機能」、情報処理学会第34回全国大会３月(1987)

[7] 福井他、「知的文書処理システムにおける図表の参照構造抽出方式」、情報処理学会第34回全国大会３月(1987)

[8] 山口他、「知的文書処理システムにおける図表の割り付け方式」、情報処理学会第34回全国大会３月(1987)

[9] 岩井、土井、「文書の自動レイアウトシステム」、東芝レビュー、４３巻５号(1988)

[10]土井他、「文書構造の自動抽出」、人工知能学会第２回全国大会７月(1988)

[11]Shneiderman, B., "Designing the User, Interface",Addison-Wesley Publishing Company, (1987)

[12]Furuta,R.,et.al.,"Document Formatting System: Survey,Concepts,and Issues",Computing Surveys, Vol.14,No.3(1983)

[13]Peels,A.J.H.M.,et.al.,"Document Architecture and Formatting",ACM Trans.OIS,vol.3,no.4, pp347-369(1985)