

文書構造を利用する組版システムの設計と実現

1 J-7

遠瀬雅宏 伊藤聡 早川栄一 並木美太郎 高橋延匡

(東京農工大学)

1. はじめに

人間が文書を作成する場合、章や節、またはその見出しなどの文書構造を意識しながらその作業を行なう。そのような行程を組版システムに反映することにより、より高度な組版処理が可能となる。

そこで我々は、文書構造を表わすコマンドを含む組版指令を行なう言語を用意し、そのコマンドが埋め込まれた文書をもとに組版処理を行なうシステムを作成した。

本稿では、作成した組版システムの概要について述べる。

2. 本システムの特徴

本システムの特徴は、次のとおりである。

(1) 科学技術文書を対象とする

本システムは、対象文書を科学技術文書としている。そして、章、節1、節2、節3からなる階層構造とそれらに属する見出し、段落、図、表を本システムでサポートする文書構造として定義する。それぞれの構造は、文書のコラム幅で構成されるものと仮想する。また、脚注のような呼び出し位置と出現位置が異なるものは、処理の対象外とする。

(2) 書式情報や組版情報を別に与える

文字の大きさ、文書のフォーマットといった書式情報や禁則文字の設定などの組版情報を文書ファイルに持たせるのではなく、それぞれ別のファイルで設定し、それらを本システムの入力ファイルとする。これにより、文書ファイルの変更なしに、様々な書式の文書が作成可能となる。

(3) 簡単なレイアウト機能を持つ

本システムは、バッチ処理方式で組版処理を行なう。したがって、文書作成中にユーザがその処理結果を意識することは困難である。そこで、(1)で述べた文書構造を利用したレイアウト機能を持たせることにより、ユーザの手間を軽減する。

3. システムの全体構成

本システムの全体構成を図1に示す。

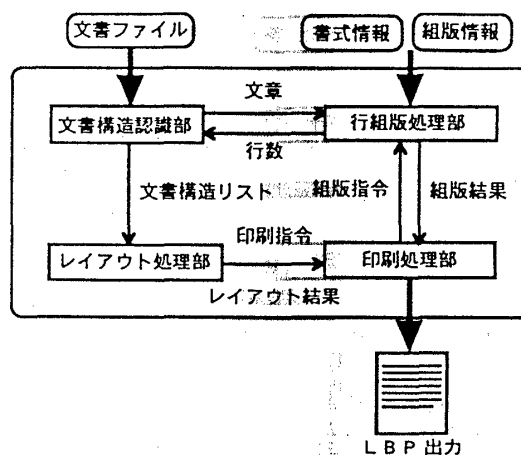


図1. システムの全体構成

本システムは、文書構造認識部、レイアウト処理部、行組版処理部、印刷処理部からなる。それぞれの概要は次のとおりである。

・文書構造認識部

前述した文書構造を文書ファイルから抽出し、それぞれの構造にレイアウトのための情報を付加する。

・レイアウト処理部

文書構造認識部で付加された文書構造情報をもとにレイアウト処理を行なう。

・印刷処理部

レイアウト処理部の結果に従い印刷を行なう。その際、文章に関しては行組版処理部へ文字列を与えてその処理結果を利用する。

・行組版処理部^[1]

JIS X 4051「日本語文書の行組版方法」^[2]で規定される行組版規則に従った組版処理を行なう。

4. 文書構造認識部の設計

文書構造認識部では、コマンド混在の文書ファイルから文書構造だけを抽出し、文書の流れを表わす文書構造リストを作成する。そして、文書構造リストの各要素に、次のような情報を付加する。

Design and Implementation of a Formatting System
Using Document Structure

TOOSE Masahiro, ITOU Satoshi, HAYAKAWA Eiichi,
NAMIKI Mitarou, TAKAHASHI Nobumasa
Tokyo University of Agriculture and Technology

(1) 属性

その構造がどのようなもの（見出し、文、図、表など）なのか、どの階層構造に属する構造なのかを示すものである。

(2) 高さ

その構造が持つ高さがどれくらいなのかを示すものである。

また、文書構造リストの各要素間では次のような情報を付加する。

(1) 不可分性

注目している構造が次の構造と分離されてよいかどうかを示すものである。この情報により、章や節の見出しがページ下端に配置されたり、図表の見出しが図表本体と違うページに配置されることを防ぐことができる。

(2) 交換可能性

処理している構造がそのコラムに配置不可能になった場合に、その次の構造との順序の入れ替えが可能かどうかを示すものである。この情報により、たとえば図表がページに入りきらない場合に、その図表を次のページへ送り、余ったスペースを文で埋めることができる。



図2. 文書構造リストの情報の例

5. レイアウト処理部の設計

レイアウト処理部では、文書構造リストの情報をもとにレイアウトを行なう。レイアウトする構造の単位は、文書構造リストの不可分性で不可分とされる構造の集合（以下、構造群とする）である。そのレイアウト規則を次に示す。

- (1) 対象構造群がそのコラムに入るならば、そのまま配置する。
- (2) 対象構造群がそのコラムに入らない場合、次の

構造群との交換可能性を見て、交換可能でその構造群がそのコラムに入るならば、構造群を交換しそのコラムに配置する。

- (3) (2) で交換不可能な場合や次の構造がそのコラムに入らない場合には、次のコラムに展開する。

(1)～(3)の処理を繰り返し行ない、ページを溢れるときに印刷処理部に印刷の指令を出す。

6. 行組版処理部の設計

行組版処理部は、印刷処理部からの組版指令を受けると入力ファイルで与えられる書式情報と組版情報をもとに JIS X 4051 に従った行組版処理を行なう。前述したとおり、本システムではすべての構造をコラム幅と仮想してレイアウト処理を行なうので、インデントやセンタリング、右寄せなど文章の配置に関する処理はここで行なわれる。

また、本システムでは段落ごとに最適な文字配置を計算しているため、前後の行でバランスのよい組版処理を行なうことができる。そして、段落に対して図3のようなインデント設定を行なうことにより、多様な段落の整形を行なうことができる。

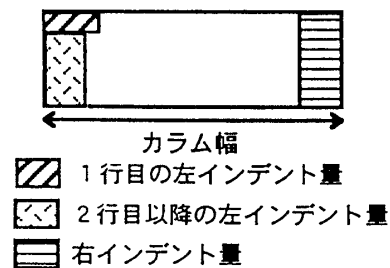


図3. 段落に対するインデント設定

7. おわりに

現在、本システムは我々の研究室で文書出力に実用している。今後は、さらに多くの文書を処理することにより、この方式の妥当性を検証する必要がある。また、文書構造の不可分性を用いたより複雑なレイアウト処理方法を検討していく予定である。

参考文献

- [1] 遠瀬, 他: “JIS に基づいた行組版機能の試作”, 情報処理学会第 50 回全国大会, 4N-4, 1995.
- [2] 日本規格協会: “JIS X 4051 「日本語文書の行組版方法」”, 日本工業規格, 1994.