

4C-1

ODA自動割付けシステムの開発

長崎晶子 石井弘 大黒和夫

(株)東芝 情報処理・機器技術研究所

1. はじめに

ODA<sup>1)</sup>は、文書の交換を目的としてISO/IEC JTC1によって定められた国際規格であり、文書体系および内容体系について規定している。また、文書の編集、割付け、可視化といった一連の文書処理のモデルを提供している。この文書処理モデルに基づいて、文書の割付けを自動的に行うことが可能である。我々は、この自動割付け処理の実現を目指して、ODAの文書応用仕様の一つであるFOD11<sup>2)</sup>レベルの文書に対応するシステムを開発したので報告する。

2. 文書応用仕様FOD11の概要

FOD11は、ODAに関する国際標準プロファイルの中で最も単純なものである。FOD11文書の概要を次に示す。

- (1) 論理構造はルート、複合論理対象体 (Passage)、基本論理対象体 (BodyText) の3階層から成り、例えば、1つ以上の章があり、その中に1つ以上の本文のかたまりがあるといった文書を表現できる。
- (2) 割付け構造はルート、ページセット、ページ、枠および区画の5階層から成る。ページセットは、含まれるページのパターンの違いによって5種類用意されている。各ページは、任意選択のヘッダ枠およびフッタ枠並びに必須の本文枠から成る1段組である。
- (3) 文書内容としては文字内容のみが許される。
- (4) 指定可能な割付け指示属性の種類が少なく、適用できる論理対象体もかなり限られている。
- (5) 属性“論理源”によるヘッダフッタ内容の生成およびページ番号の自動生成機能がある。

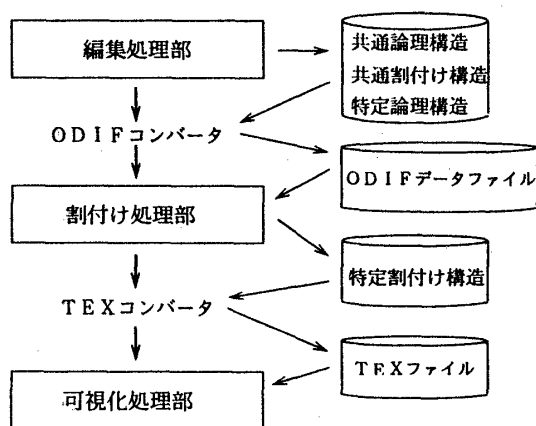


図1 本システムの処理の流れ

3. システム構成

本システムは、文書の編集処理部、割付け処理部および可視化処理部から成る。開発は、当社のUNIXワークステーションSPARC LT上で、C言語を用いて行った。画面表示などのユーザインタフェースにはXウィンドウシステムを利用している。図1に本システムの処理の流れを示す。

3.1 編集処理部

ここでは、特定論理構造およびそれに結びつく内容部の作成を行う。ODAの文書構造は木構造で表すと理解がしやすいため、文書構造編集エディタを用意した(図2)。これを用いて、あらかじめ共通論理構造および共通割付け構造の生成子完全集合を作成しておき、さらに共通論理構造をもとにして特定論理構造を作成する。このとき、共通論理構造の各対象体クラスに対して指定されている各種の属性は、特定論理構造の各対象体に継承される。また、特定論理構造の最下位の各対象体に対しては、それらと結びつく内容の作成を行う。特定論理構造を初めとして、すべての文書構造は、ODIFコンバータによってODAの文書交換形式ODIFに基づいたデータストリームに変換され、保存される。

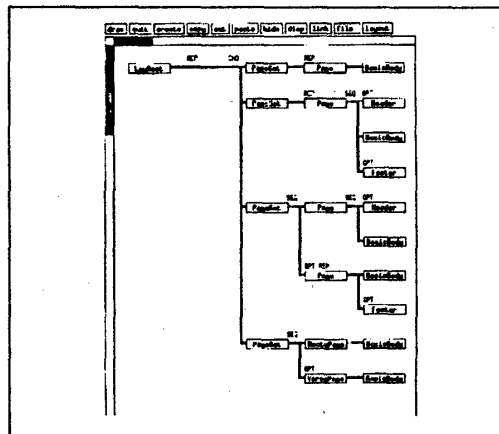


図2 文書構造編集エディタ画面

3.2 割付け処理部

ここでは、特定論理構造に対して指定された割付け指示属性をてがかりとして、共通割付け構造のインスタンスを生成することによって、自動的に特定割付け構造を作成する。

3.2.1 バックトラック処理

FOD11レベルでは、共通割付け構造が単純であり、割付け指示属性も制限されている。このため、本システムにおいては、基本的には、論理順番に従って一つ一つ論理対象体と結びつく内容を割り付けていくというやり方で、割付け処理を実現しており、割付け指示属性“同一割付け対象体”が

Development of Automatic Document Layout System Based on ODA  
Akiko NAGASAKI, Hiroshi ISHII, Kazuo OHGURO  
TOSHIBA Information Systems Engineering Laboratory

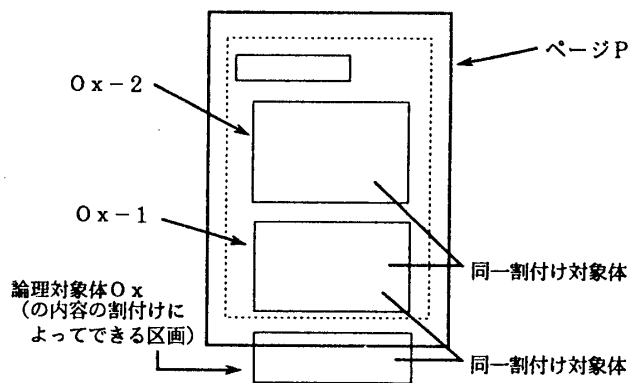


図3 バックトラックが発生する場合の例

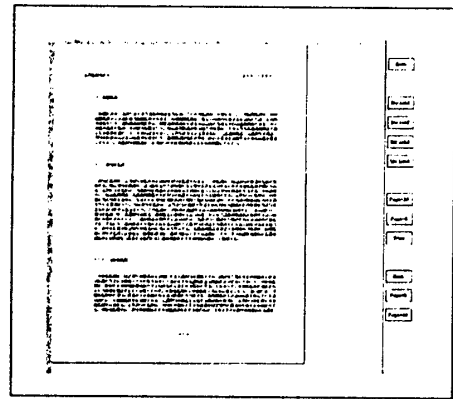


図4 特定割付け構造の可視化例

満たされない場合についてのみ、バックトラック処理を行っている。図3に示す例では、論理対象体 $O_x$ に対して指定されている“同一割付け対象体”に従って、 $O_x$ の内容をその直前の論理対象体 $O_{x-1}$ の内容と同じページPに割り付けようとしているが、このままでは不可能である。このような場合は、次の処理を行っている。

(1) 割り付け直す先頭の論理対象体を決める。

図3の例のように、 $O_{x-1}$ もまた、“同一割付け対象体”によって、その前の論理対象体 $O_{x-2}$ と同じページPに割り付けられている場合などは、 $O_{x-2}$ も同時に割り付け直す必要がある。このため、 $O_{x-1}$ と同じページに、 $O_{x-1}$ よりも先に割り付けられた論理対象体がある場合は、その割り付け指示属性を調べ、割り付け直す必要のある先頭の論理対象体 $O_{x-n}$ を求める。

(2) 新たにページを作れば割り付け可能かどうかを調べる。

共通割付け構造の従属生成子に従って新たに生成できるページの大きさを調べ、その中に $O_{x-n}$ から $O_x$ と結びつく文字内容が割り付け可能かどうかを調べる。このとき、 $O_x$ に対して“非分割性”が指定されていない場合は、 $O_x$ と結びつく内容が少しでも割り付けられればよいが、指定されている場合は、 $O_x$ と結びつく内容がすべて割り付けられなければならない。

(3) 割り付け可能な場合はページを生成する。

実際にページを生成し、 $O_{x-n}$ から $O_x$ と結びつく内容を割り付ける。このとき、(1)において $O_{x-n}$ がページPの先頭に割り付けられていた場合には、ページP全体を破棄する。先頭の論理対象体でなかった場合には、 $O_{x-n}$ が割り付けられていた区画から $O_{x-1}$ が割り付けられていた区画を破棄する。

(4) 割り付け不可能な場合は割り付け指示属性を無効とする。

$O_x$ に“非分割性”の指定があればまずこれを無効とし、それでも割り付け不可能な場合は、“同一割付け対象体”の指定も無視する。

### 3.3 可視化処理部

ここでは、割り付け結果を画面に表示する。割り付け処理部で生成された特定割付け構造をT<sub>E</sub>Xファイルに変換するT<sub>E</sub>Xコンバータを用意した。変換したファイルをXウィンドウシステム用のT<sub>E</sub>Xのプレビューアを用いて表示する(図

4)。また、PostScriptファイルに変換して、プリンタに出力することも可能である。

### 4. おわりに

以上に述べたように、FOD11の文書に対応した自動割付けシステムを開発した。このレベルであれば、割り付け処理が複雑にならず、実現できることがわかった。しかし、表現できる文書がごく単純なものに限られてしまうため、引き続きFOD26およびFOD36に対応する割付けアルゴリズムの検討を行っている。また、ODAに関する知識がない人にも容易に使うことができるシステムにするため、ヒューマンインタフェースの改良も行なっている。

### 5. 参考文献

- [1]ISO 8613 : Information processing - Text and office systems - Office Document Architecture(ODA) and interchange format(1989)
- [2]Proposed Draft International Standardized profile(ISP) FOD11 Office Document Format Profile for the Interchange of Basic Function Character Content Documents in Processable and Formatted Forms Part 1: Document Application Profile(1989)

- UNIXオペレーティングシステムは、AT&Tのベル研究所が開発したソフトウェアであり、AT&Tがライセンスしています。
- Xウィンドウシステムは、米国マサチューセッツ工科大学の登録商標です。
- T<sub>E</sub>Xは、American Mathematical Societyの登録商標です。
- PostScriptは、米国Adobe System社の登録商標です。