

3C-9

マルチメディア文書交換形式(ODA)を既存DTPシステムに  
適用するための論理構造変換方式

柴田 顕男                      久保 徹                      秦 務  
松下電器産業(株) 情報通信東京研究所

1. はじめに

我々は、(財)情報処理相互運用技術協会(INTAP)が開発したマルチメディア文書交換形式(ODA<sup>[1]</sup>)実装規約<sup>[2]</sup> AE.1136のPDAの処理系を、既存DTPシステムとのコンバート方式により試作し、複数社のシステムと文書交換実験を行ない、異機種異システム間での相互運用性を確認した<sup>[3][4]</sup>。

本稿では、試作した処理系の論理構造変換方式について述べる。

2. ODAと対象DTPシステムの特徴

2.1 ODAの論理構造

本処理系では実装規約で規定するAE.1136に準拠したODA文書を扱う。AE.1136は、高度な論理的階層構造を持つODA文書を交換するためのプロフィールである。内容としては、文字内容のほかラスタ図形や幾何学図形を含むことができる。AE.1136の論理構造は次のような特徴を持つ。

- ・章、節など、高度に階層化された論理構造
- ・テーブル、フォームのサポート

2.2 対象DTPシステムの特徴

本処理系で対象とした既存DTPシステムには、次のような特徴があり、これらをODAへのマッピングの際に利用することとした。

- ・段落に対してパラグラフタグと呼ばれる名前を付与することができる
  - 論理構造を表現するために使用する(3.1節参照)
- ・段落単位で各種の属性を設定できる
  - 表示属性・割り付け指示属性にマッピングする
- ・位置が文字内容と連動するフレーム(アンカーフレームと呼ばれる)が使用できる
  - ラスタ図形、幾何学図形、テーブル、フォームなどを配置する

2.3 機能の使用に関する制限

ODA(AE.1136)の持つ機能と対象DTPの持つ機能との間では、互いに対応する機能をサポートしていない場合があるため、これらの機能の使用に関しては一定の制限を設けている。

また、本処理系のコンバート方式との関連で、制限を設けている場合もある。すなわち、DTP文書に関しては、後述するようにパラグラフタグの使用方法を制限しており、ODA文書に関しては、階層の深さや使用する論理構成体の種別などに制限を設けている。本処理系ではこの制限の範囲内のAE.1136サブセットのODA文書を処理の対象としている。本処理系で扱うAE.1136サブセ

ットのODA文書の論理構造の範囲を図1に示す。

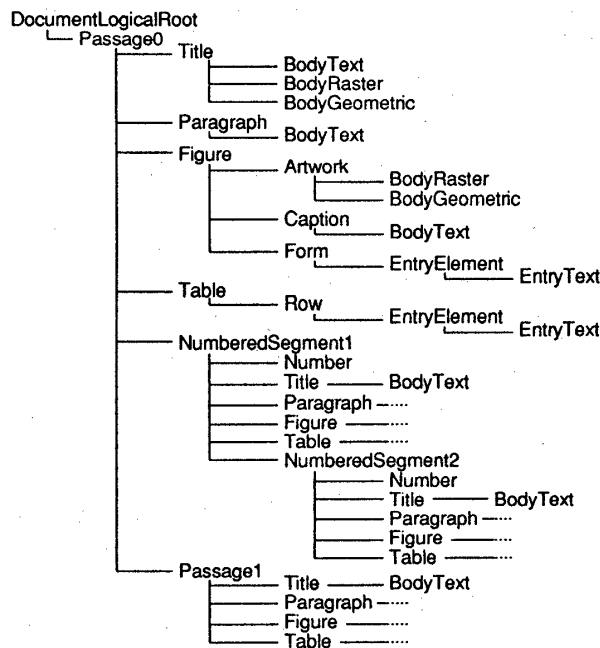


図1: 本処理系で扱う論理構造

3. ODA文書とDTP文書とのマッピング

3.1 論理構造のマッピング

ODA文書とDTP文書との最も重要な差異は、階層的な論理構造を持つか否かである。そこで、ODAの階層化された論理構造をDTP文書で表現するため、パラグラフタグに特別な意味を持たせ(表1)、本処理系で扱うAE.1136サブセットのODA文書の論理構造(論理構成体の種別や階層の深さのレベル)との対応づけを明示することとした。

3.2 属性のマッピング

DTP文書の各段落に対して設定された各種の属性のうち、ODAの表示属性・割り付け指示属性属性に対応するものについては、当該の属性にマッピングする変換を行なう。逆方向の変換の場合も同様である。

これらの主な属性の対応関係を、表2に示す。

4. 論理構造変換方式

4.1 DTP文書からODA文書への変換処理

DTP文書からODA文書への変換処理では、ODAの特定論理構造を生成する。以下、この変換処理について順を追って説明する。

(1) 変換処理の概要

変換処理では、DTP文書の段落を先頭から順次取り出し、そのパラグラフタグ名に基づいて処理を進める。このときアンカーフレームも段落と同様に扱い、パラグラ

タグ名 'Anchored' が付与されているものとして処理する。

DTP 文書の各段落は ODA の基本論理対象体に対応しており、そのパラグラフタグ名は表 1 に示すように ODA の論理構造に対応する意味付けがなされている。変換処理では、段落に対応する基本論理対象体を生成するとともに、必要に応じてその上位の複合論理対象体も生成する。例えば、パラグラフタグ 'PText' に対しては BodyText のみを生成するが、'Para' に対しては BodyText に加えてその直上の Paragraph も生成する。なお、アンカーフレームに対しては、その内容に応じてテーブル、フォーム、ラスタ図形、もしくは幾何学図形を生成する。生成した各論理対象体には必要な属性を設定する。

表 1: パラグラフタグの意味付け

パラグラフタグ	意味
Title	Passage0 直下先頭の Title 直下の BodyText
Para	Passage0 直下の新たな Paragraph の始まり その先頭 BodyText
PTitle *	新たな Passage1 の始まり その直下先頭の Title 直下の BodyText
Passage *	新たな Passage1 の始まり その直下先頭の Paragraph の先頭 BodyText
PPara	Passage1 直下の先頭でない新たな Paragraph の始まり その先頭 BodyText
1Num	新たな NumberedSegment1 の始まり その直下先頭の Number および Title 直下の BodyText
1Para	NumberedSegment1 直下の新たな Paragraph の始まり その先頭 BodyText
2Num	新たな NumberedSegment2 の始まり その直下先頭の Number および Title 直下の BodyText
2Para	NumberedSegment2 直下の新たな Paragraph の始まり その先頭 BodyText
PText	任意のレベルの Paragraph の先頭以外の BodyText
(Anchored)	Table もしくは Figure に対応するアンカーフレーム

\*注) これらが 1 つの Passage1 直下に同時に存在することはない

(2) 特定論理構造の生成

処理の初期状態では特定論理構造には文書論理根のみ存在し、その下に生成した論理対象体を順次追加してゆく。追加する論理構造上の位置は、直前に追加した基本論理対象体の上位にある、いずれかの複合論理対象体(もしくは文書論理根)の直下である。これらの候補のうち最も下位の複合論理対象体から順次上位に向かって、追加する論理対象体(の最上位のもの)をその位置に追加することが共通論理構造上で許されているかどうかをチェックし、許されていればそこに追加する。この際に、いずれの位置に追加することも共通論理構造上で許されていない場合、DTP 文書から得られる特定論理構造が、共通論理構造の範囲を越えてしまう場合に相当するので、エラードキュメントとして変換処理を中止する。

(3) 共通論理構造の処理

共通論理構造は処理系でテンプレートとして保持しており、テーブルやフォームを含む文書の場合などを除き、これをそのまま適用する。テーブル、フォームを含む文書に関しては、共通論理構造にテーブルまたはフォームに対応する対象体クラスを追加しながら、特定論理構造の生成を行なう。

(4) 属性の処理

段落属性に関しては、対応する表示属性・割り付け指示属性にマッピングし、表示体裁・割り付け体裁を生成

する。このとき、同一内容の体裁が既に存在する場合には、新たに体裁を生成することは行なわず、体裁識別子により既存の体裁を参照して、生成される ODA 文書データの冗長度を削減している。

表 2: 主な段落属性と表示属性・割り付け指示属性との対応

段落属性	表示属性・割り付け指示属性
行揃え: 左揃え 中央揃え 右揃え	Presentation-style(Alignment) 'start-aligned' 'centered' 'end-aligned'
左インデント	Presentation-style(Indentation)
開始位置: ページの先頭 コラムの先頭	Layout-style(New-layout-object) page frame
アンカーフレームの位置揃え: 左揃え 中央揃え 右揃え	Layout-style(Block-alignment) 'right-hand-aligned' 'centered' 'left-hand-aligned'

4.2 ODA 文書から DTP 文書への変換処理

ODA 文書から DTP 文書への変換を行なう際には、ODA 文書の特定論理構造上を文書論理根から depth-first 探索の手法でたどり、現れる論理対象体を順次処理して行く。

(1) 基本論理対象体の処理

基本論理対象体については、現れる毎に DTP 文書の段落もしくはアンカーフレームを生成する。生成した段落に対しては、論理対象体の種別および論理構造の深さから、適当なパラグラフタグを付与して、表示属性・割り付け指示属性は段落属性にマッピングする。

(2) 複合論理対象体の処理

複合論理対象体については、その属性を下位の基本論理対象体の変換時に使用するために保持し、基本論理対象体に対応して生成される DTP 文書の段落の属性にマッピングする。

5. まとめ

ODA 文書(AE.1136 サブセット)を交換する処理系を既存の DTP システムとのコンバータ方式により実装し、異システムとの文書交換を行ない本方式の有効性を確認した。

今回の方式では、DTP 文書についてパラグラフタグの使用法に制限を設けたため、既存の文書や本処理系を意識せずに作成された文書を、そのまま ODA 文書に変換することができない。

今後はこれらの点を考慮し、パラグラフタグの使用を制限しない、AE.1126 レベルの ODA 文書を対象としたコンバート方式を検討する予定である。

なお、本研究開発の一部は、通商産業省工業技術院大型プロジェクト「電子計算機相互運用データベースシステムの研究開発」の一環として、新エネルギー・産業技術総合開発機構(NEDO)より委託を受けて実施したものである。

参考文献

[1] ISO 8613 : IS Information Processing - Text and Office Systems - Open/Office Document Architecture (ODA) and Interchange Format (1989)  
 [2] JIS X 5003-1987 参考 S 007(V2.0) 文書交換形式(ODA)実装規約(1991)  
 [3] 柴 他: "マルチメディア文書交換形式(ODA)の処理系の実装と評価"、情報処理学会第 44 回全国大会講演論文集(1992)  
 [4] 久保 他: "マルチメディア文書交換形式(ODA)を既存 DTP システムに適用するための割り付け構造変換方式"、情報処理学会第 44 回全国大会講演論文集(1992)