

構造化文書体系(ODA)の処理: 文書プログラミングによる構造化文書自動作成

4C-7

楠本 浩二

富士ゼロックス(株)システム技術研究所

1. はじめに

ODA^[1]文書を対象として既存の文書(ソース文書)を利用することにより、新たな文書(ターゲット文書)を自動的に作成するための手段を考える。その1つとして、ソース文書のスペシフィック論理構造をもとにターゲット文書のスペシフィック論理構造を作成するため、ソース文書の文書クラス(ジェネリック論理構造)とターゲット文書の文書クラスとの間に関係記述を行う。この記述を文書プログラムと呼ぶ。文書クラス間で記述を行うことにより、文書クラスの制約のもとで作成された文書のすべてがプログラムの入力対象となる、文書(インスタンス)と文書プログラムの独立性を保持できる、という利点がある。今回、ODA文書エディタ^[2]上に本方式による文書自動作成実験システムの実装を行った。本論文では、文書プログラムの処理と課題について述べる。

2. 自動作成するターゲット文書

文書プログラムは、ソース文書のオブジェクトクラスからターゲット文書のオブジェクトクラスへのマッピングの記述である。以後、文書プログラムが参照するソース文書のオブジェクトクラスをソース側オブジェクトクラス、ターゲット文書のオブジェクトクラスをターゲット側オブジェクトクラスと呼ぶ。処理系は文書プログラム、ソース文書のジェネリック論理構造とターゲット文書のジェネリック論理構造、そしてソース文書のスペシフィック論理構造を入力としてターゲット文書のスペシフィック論理構造を自動作成する。以下に処理の過程で生成する論理オブジェクト、オブジェクト間に設定されるべき従属関係、生成する文書内容部について説明する。

1) 自動生成する論理オブジェクト

ソース側オブジェクトクラスのインスタンス(ソース側オブジェクト)が存在する時、ターゲット側クラスのインスタンス(ターゲット側オブジェクト)

を生成する。ただし、文書ルート、ターゲット側クラスの従属子生成子の制約上存在すべきオブジェクト、の2つについてはマッピングに関係なく生成する。例えば、ターゲット側オブジェクトクラスAがSEQ(B, C)を持つならば、Aのインスタンスaの生成時にB、Cのインスタンスb、cを生成する。

2) 従属関係の設定

ソース側オブジェクトの従属関係を保存するようにターゲット側オブジェクトの従属子を設定する。ただし、前述したターゲット側オブジェクトクラスの従属子生成子の制約上、必ず生成されるオブジェクトについては一意に設定する。

3) 自動生成する内容部

ソース側オブジェクトの内容部をターゲット側オブジェクトの内容部へ複写する。ただし、ター

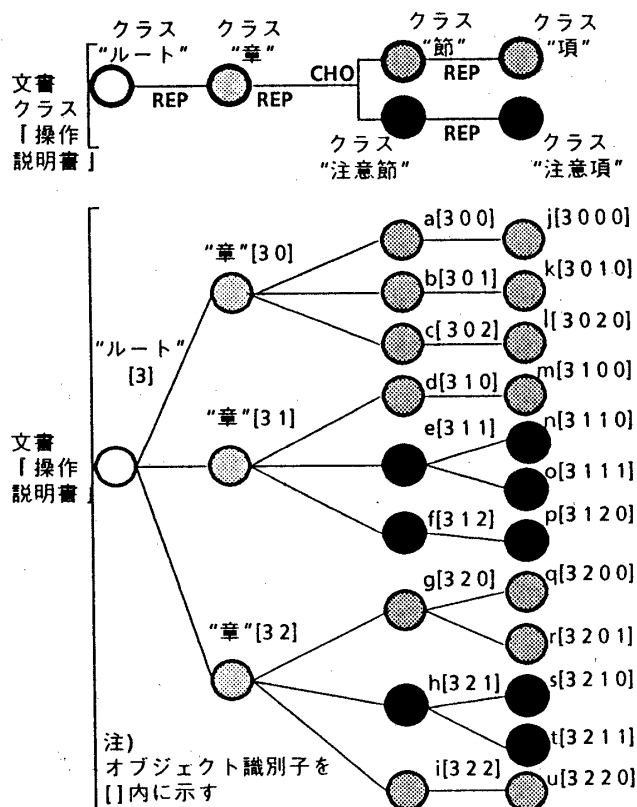


図1 ジェネリック論理構造とスペシフィック論理構造の例

ゲット側オブジェクトクラスが自身で内容を生成する機構(generic content, content generator)を持つ場合は、これらによる内容生成を優先し、マッピングによる内容生成は実行しない。

3. 自動作成例

図1のようなジェネリック論理構造とスペシフィック論理構造を持つソース文書「操作説明書」から“注意項”を抜き出し、ターゲット文書「注意事項集」を作成する例を用いて本処理を説明する。図2で示すようなソース文書クラスとターゲット文書クラスとの間に文書プログラムを定義する。ソース文書の“注意節”オブジェクトをもとに“節”オブジェクトを作成するために、クラス“注意節”とクラス“節”のマッピングの記述を行う。“注意項”オブジェクトをもとに“段落”オブジェクトを作成し、文書内容を作成するために、クラス“注意項”とクラス“段落”のマッピングの記述を行う。以下にソース文書と文書プログラムからターゲット文書のスペシフィック論理構造を作成する処理を説明する。

1) ターゲット側オブジェクトの生成

ソース側オブジェクトクラスに属するオブジェクトを論理順序に従ってリストし、対応するターゲット側オブジェクトクラスのオブジェクトを作成する。本例では、“注意節”オブジェクトのリスト[e, f, h]と“注意項”オブジェクトのリスト[n, o, p, s, t]を得る。これにより[e, f, h]に対応して、クラス“節”はインスタンスとして3個の“節”オブジェクト[A, B, C]を作成する。同様に[n, o, p, s, t]に対応して、クラス“段落”は5個の“段落”オブジェクト[D, E, F, G, H]を作成する。

2) ターゲット側オブジェクトの従属関係の設定

ソース側オブジェクトの従属関係は、次のように得られる。図1のようにオブジェクト識別子を木構造に従って規則的(例えば左深さ優先)に付与しておくことにより、オブジェクト識別子が最下位の桁を除いて等しいオブジェクトの親は同一であると判断できる。例えば、“注意節”オブジェクトについて

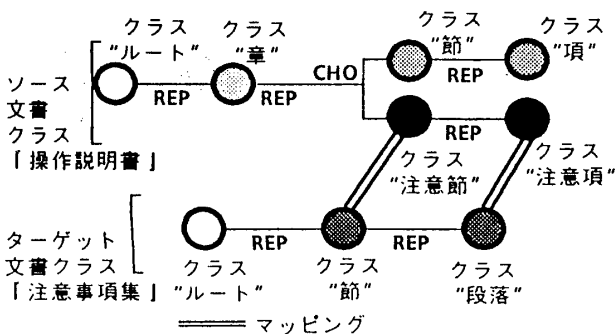


図2 ソースとターゲット文書クラス間のマッピング例

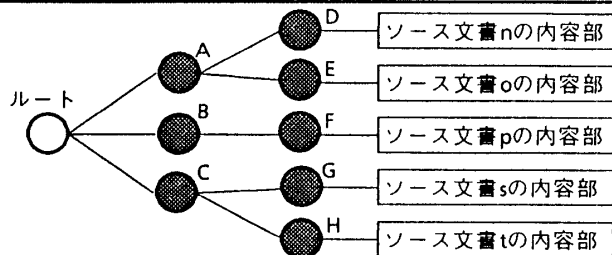


図3 「注意事項集」のスペシフィック論理構造

は、e[3 1 1]、f[3 1 2]、h[3 2 1]であるから、eとfは兄弟である。リスト内で兄弟関係にあるオブジェクトを()内に表すと“注意節”オブジェクトは[(e, f), h]、“注意項”オブジェクトは[(n, o), (p), (s, t)]である。ターゲット文書の“段落”オブジェクト[D, E, F, G, H]はソース文書の情報[(n, o), (p), (s, t)]に対応して[(D, E), (F), (G, H)]となる。これによりDとEは[A, B, C]の1番目のオブジェクトであるA、FはB、GとHはCをそれぞれ親とする従属関係を設定できる。

3) ターゲット側オブジェクトの内容部作成

ソース文書[n, o, p, s, t]の内容を複製してそれぞれターゲット文書の[D, E, F, G, H]の内容とする。こうして作成される「注意事項集」のスペシフィック論理構造を図3に示す。

4. 考察

本論文では、文書クラスの制約を満たす文書を対象として、文書クラス間の対応(マッピング)の記述にしたがった構造化文書自動作成方式を示した。しかし、以下に示す文書プログラムの課題(1)とODAのジェネリック論理構造の課題(2)がある。

- (1) ソース側とターゲット側のオブジェクトクラスの対応を記述するので、ジェネリック論理構造間で対応可能なオブジェクトクラスが陽に存在する必要がある。
- (2) ジェネリック論理構造は、割付けの構造に依存して作成される^[3]ので文書プログラムの都合だけで規定できない。

今後、文書クラスや文書割付けに制限されない記述能力を持つ文書プログラムを定義し、実用的オフィスシステムの構築を目的とした研究を行っていく予定である。

参考文献

- [1] ISO(edt.): ISO8613: Information Processing - Text and Office Systems - Open Document Architecture(ODA) and Interchange Format (1987).
- [2] ODAベースシステムの試作(1)(2)(3), 林他, 情報処理学会第37回全国大会(1988).
- [3] Some ODA Layout Control Problems revealed by a prototype editor, Hayashi et. al, Proceedings of WOODMAN'89 Workshop on Object-Oriented Document Manipulation, pp. 79-90 (1989).