

## 論理構造に基づく多重文書処理

4L-3

野中 士郎<sup>†</sup>、富田 泰志<sup>†</sup>、松本 秀和<sup>†</sup>、森 庸輔<sup>‡</sup>

(株)日立製作所 日立研究所<sup>†</sup>、大みか工場<sup>‡</sup>

### 1. はじめに

文書作成の効率化の観点から、類似文書の作成に対する効率的な手段の提供やシステムティックな対応が文書処理に強く求められるようになってきている。

一方、文書の論理的な構造を利用した文書処理の試みが種々なされている。

本報告では、この文書の論理構造を多重化し、複数の類似文書の一括編集管理を可能とする多重文書処理を提案する。

### 2. 問題点と解決方針

#### 2.1 類似文書の文書処理の問題点

類似文書の作成に対して、定型フォーマットの共通化や、共通的文章図表等の部品化による支援が行われている。

しかし、実際に作成されている文書を考えてみると、例えば、図1に示すように文書の種類としては見積書で一種類であるが、A社向け、B社向け…と内容が一部分のみ異なる文書群が多数存在し、しかも一括して管理したい場合も多い。

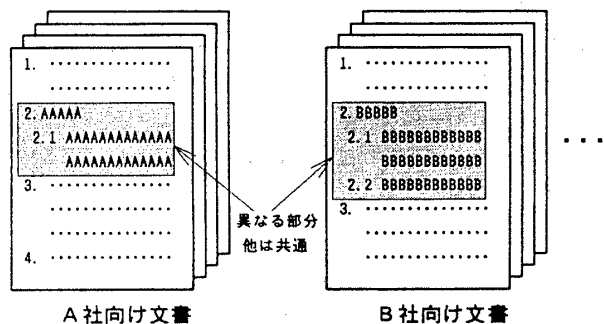


図1 文書の例

この場合にはフォーマットが同じという訳ではなく、また異なる部分のごく一部であることから部品化するほどの必要は

なく、さらに管理上も部品化すると面倒であるといった問題があり、効率的な支援が難しかった。

また、試行錯誤を伴う作成途中の文書にも同様の問題があった。すなわち、作成途中の段階では部分的に複数の文章の案があり、途中段階ではそのまま複数の案を共存させる状態で管理したいというニーズも高いが、このケースも支援が難しかった。

#### 2.2 問題点の解決方針

従来の文書処理は単一の文書のモデルに対してアクセスを行うシステムであると言える。しかし、その枠組みでは、前節の問題点の解決は難しい。なぜならば本質的には複数の文書であるが、文書間で共通な部分と異なる部分を効率よく管理し、しかもそれを一括して扱える仕掛けを必要とするからである。

そこで、図2に示す複数の文書より構成されるモデルに対してアクセスを行う文書処理方式を提案する。図に示すように、文書モデルには種々のフィルターを介してアクセスすることになる。この多重文書モデルとフィルターを介した文書アクセスにより問題点の解決を図る。

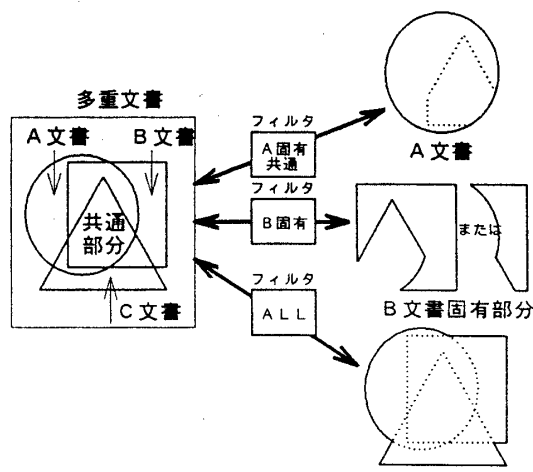


図2 多重文書モデル

Multiple Document Processing  
Using Logical Document Structure

Shiro NONAKA, Yasushi TOMITA, Hidekazu MATSUMOTO  
and Yousuke MORI

HITACHI, Ltd.

### 3. 多重文書処理の概要

多重文書といっても、その取り扱う単位は種々のものが考えられるが、ユーザと文書処理システム両者にとって認識、操作、管理が容易な、文書の章や節といった論理構造を単位とする多重文書処理を考える。この論理構造は木構造で表現できるが、そのノード毎に与えた文書属性により、フィルターを介した文書を選択が可能となる。

しかし、これだけでは複数の文書間の論理的な関係が不明確となる。そこで、図3に示すように、ノードを多重化することにより論理的な関係の維持管理を行う。本方式は、従来の木構造が2次元構造とすれば、各ノードに奥行（多重化）があることから3次元構造であると考えられる。

### 4. 多重文書処理の評価

#### 4.1 多重文書処理の適用対象と評価

##### (1) 複数の類似文書の統一的管理

類似文書として、章節を単位として異なる内容が複数存在しているような文書（例えば、各種見積書やシリーズものの製品カタログや複数国語のカタログ）のケースでは、共通部分と非共通部分の管理が容易であることから、多重文書処理は有効であると考えられる。但し、フィルターを介して個別の文書としてアクセスする場合は問題ないが、複数の文書全体のアクセスの際にはユーザが混乱しないような仕掛けが必要であり、今後の課題としたい。

##### (2) 試行錯誤を伴う作成途中の文書管理

文書を完成させていく段階を考えると途中で複数の案が存在し、それを最終的に絞り込んでいくことは多いと考えられる。また複数案が存在しないケースでも訂正前の文章を少なくともしばらくは保存してそれを随時参照／復活したりして文章を最終的に練り上げていくことは、プログラムで訂正前の部分をコメントとして残すようなことが多いことを考えても自然である。それらを容易にサポートできるという点で効果は大きいと考えられる。

#### 4.2 問題点

ノードの多重化は子孫のノードを含んだ形で実現されるが、子孫は共通で純粹にそのノードだけを多重化すればよいケースも多い。しかしその機能は論理構造が複雑になることからサポートしていない。従って、本来共通であるものが重複して管理されることになり、管理とデータ量の観点から問題であり、今後検討してゆく必要がある。

### 5. おわりに

文書の論理的な木構造を表現するノードを多重化することにより、複数の文書をひとつの文書の枠組みで取り扱う多重文書処理を提案した。本処理方式により類似文書の一括管理等が容易となる見通しを得た。今後はプロトタイプシステムによる評価を進めるとともに、履歴管理への本方式の適用を検討してゆく予定である。

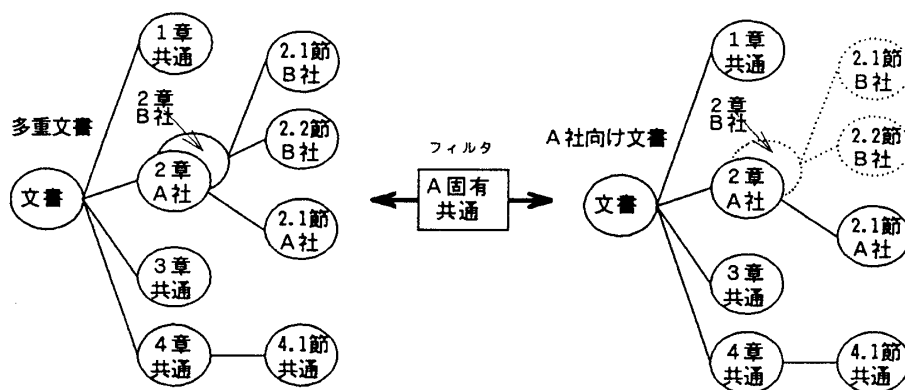


図3 ノードの多重化