

文書情報管理システムBibliotheca/IS(4)
- 文書情報格納方式 -

2W-5

丸山剛男¹

山本洋一²

和歌山哲¹

(株)日立製作所 ソフトウェア開発本部¹ (同)システム開発研究所²

1. はじめに

文書情報管理システムBibliotheca/ISは、ワープロ、スプレッドシート等のPC上の流通ソフトで作成した文書をWSサーバ上で一括管理するクライアント・サーバシステムである¹⁾。Bibliotheca/ISの文書は、文書の書誌情報を保持する文書属性情報と文書本体である文書ファイルから構成される²⁾。前者は、日立独自のオブジェクト指向データベースシステム(以下日立OODB)を利用して管理している。

日立OODBには、上記情報の他に、文書属性情報の定義情報、文書の分類体系情報を保持するビュー情報などを一元管理している。

本論では、Bibliotheca/ISにおける日立OODBを利用した情報管理の概要について述べる。

2. 文書格納管理の要件

Bibliotheca/ISの機能を実現する上で必要な格納管理の要件は以下の通りである。

(1)分類体系と文書間の自由なリンク

文書と独立した複数の分類体系で文書を管理するため³⁾、分類体系を構成するノードと文書属性情報間に自由にリンク付けできるようにしなければならない。

(2)クラスにまたがった可変長文字列属性の高速検索

文書属性情報内の可変長文字列属性を条件とした文書の検索は、文書名等に使用され、文書システムでは利用度が高い。特に文書属性情報を定義した文書クラスが複数存在するため、文書クラス間をまたがった検索を必要とする。

(3)高信頼性の確保

文書情報を共有して使用することを考慮すれば、現状の情報管理システムと同様、情報一貫性の確保、情報回復の必要がある。そのため、通常のDBMSと同様のトランザクション管理、リカバリ等の機能を実現しなくてはならない。

以下、これらの実現方式について述べる。

3. 文書の格納管理

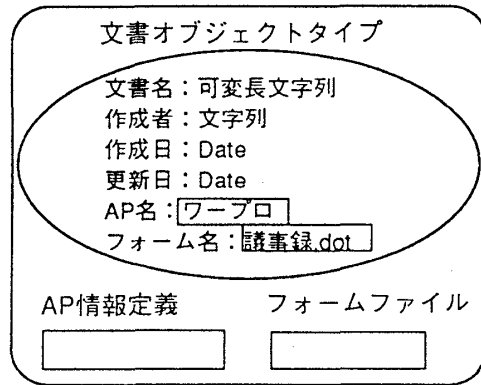
文書情報管理の実現には、上記問題を解決するためにOODBを採用した。以下に、Bibliotheca/ISにお

けるOODBを利用した文書格納方式の特徴を以下に述べる。

(1)文書書誌情報と文書実体に分けられた文書構成管理

同種の文書群は、文書クラスという1分類体系で管理される。文書情報の定義を保持する文書クラスは、OODBによって管理されている。Bibliotheca/ISにおける文書は、文書属性情報と文書ファイルから構成される。図1に文書構成を示す。

文書クラス



文書

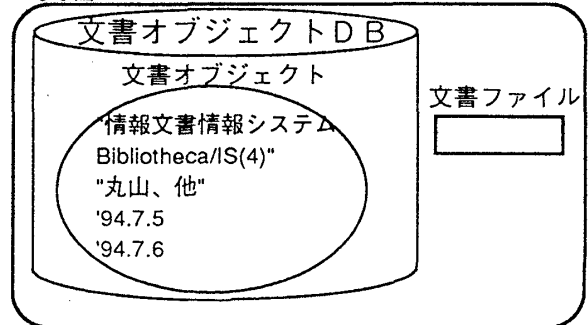


図1 文書構成

文書属性情報は文書オブジェクトとして格納され、その文書オブジェクトの定義情報は文書オブジェクトタイプとして保持される。文書オブジェクトタイプには前述の属性情報定義、同種文書のフォームを定義したファイルのファイル名、文書ファイルを起動するアプリケーション名称などが保持される。文書オブジェクトタイプには前述の属性情報定義、同種文書のフォームを定義したファイルのファイル名、文書実体を起動するアプリケーション名称などが保持される。文書オブジェクトとその定義である

文書オブジェクトタイプがOODBで管理されているため、文書を一意に識別するのに文書名などの文書属性でなく、オブジェクト識別子によって識別することができる。また、文書ファイル、同種文書のフォームファイルやアプリケーションを動作するために必要なAP情報定義は、外部ファイルとして保持している。将来はこれら外部ファイルもOODB内に格納する予定である。

(2)動的な情報間のリンクの直接的な操作による文書書誌情報の高速化

Bib/ISで実現されている複数分類体系(マルチビュー)は、ビューノードと文書間に自由にリンクが張れ、またそのリンクをたどってビューノードにつながっている文書総てを取得する。これらに関係データベース(以下RDB)で実現するとテーブル間結合(ジョイン)処理が発生し、必要な性能が得られない。OODBでは、オブジェクトの定義であるタイプ定義と独立したオブジェクト間の関連の定義を行えるようにした。よって、オブジェクト間の関連が自由に設定でき、その関連をたどることにより関連先のオブジェクトの集合を取得することが直接的な操作で実現できる。図2にOODBとRDBでのマルチビューの実現例を示す。

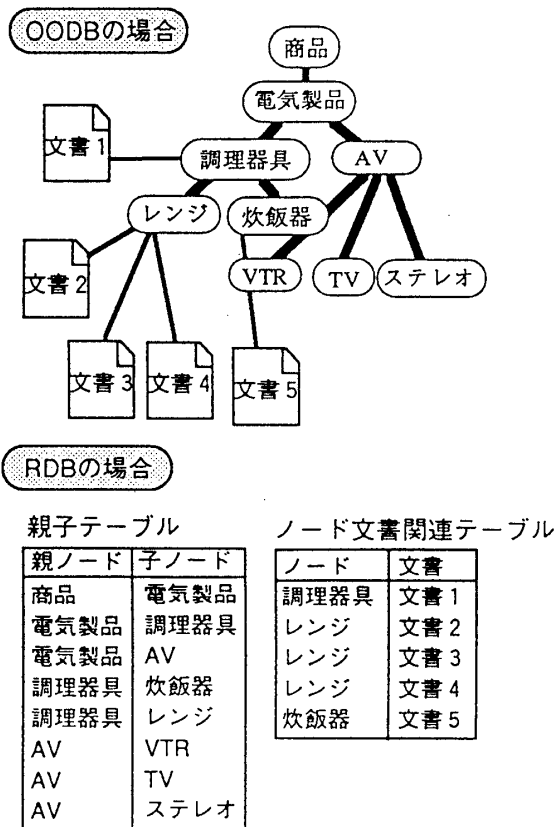


図2 OODBとRDBのマルチビュー実現例

この例で、調理器具に関係する文書を検索したい場合、OODBが親子間リンクとノード・文書間の

リンクをたどることで目的の文書を取得できるのに対し、RDBでは、親子テーブルの親及び子ノード名とノード文書関連テーブルのノード名の結合によって、文書を取得しなくてはならない。

(3)可変長文字列のクラスタリングによる文字列検索時のI/O回数削減

可変長文字列タイプの特定のタイプとして、同一領域にクラスタリングするクラスタドストリングタイプを定義できるようにしたことにより、文書名、作成日など文書属性情報の同一属性をクラスタリングすることで、その可変長文字列の検索時のI/O回数を削減し、それにより高速に検索を実施することができるようにした。図3に可変長文字列クラスタリング検索の例をあげる。

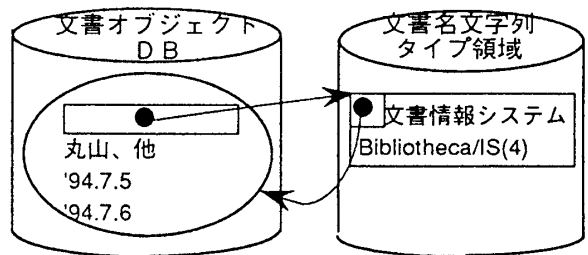


図3 可変長文字列クラスタリング検索例

例えば文書名に対する検索を行うとする。文書名属性に対し、文書名文字列タイプというクラスタドストリングタイプで定義しておけば、文書名文字列タイプ領域をスキャンし、検索にヒットする属性を見つけ、その属性に対する文書オブジェクトのオブジェクト識別子から文書オブジェクトを取得することで、目的の文書を取得することができる。

4. おわりに

自由なリンクの設定の実現による複数分類体系による文書管理やクラスタリングによる文字列検索の高速化など、新しい文書管理システムに必要な機能を取り込むことで、高性能、高信頼の文書格納管理システムを構築できた。

今後は、文書管理システムの大規模化、分散化に対応したシステムの開発を行っていきたい。

参考文献

- [1]松田,他: "文書情報管理システムBibliotheca/IS (1) - 文書情報管理の概要 -",第49回情処全国大会,1994
- [2]三原,他: "文書情報管理システムBibliotheca/IS (2) - 文書情報管理方式 -",第49回情処全国大会,1994
- [3]橋本,他: "文書情報管理システムBibliotheca/IS (3) - 検索機能 -",第49回情処全国大会,1994