

DTAM / DFRを用いたODA文書の部分検索機能の検討

3 R - 1

中尾 康二 田中 俊昭

国際電信電話株式会社研究所

1. はじめに

ODA(開放型文書体系)に基づいたDTAM(文書転送および操作)/DFR(文書ファイリング検索)の標準化の進捗に伴い、高度なオフィスシステムの構築を目指して、ODA文書の検索およびファイリング機能の実装が開始されている。ここでは、効率的な文書検索機能として、不必要な文書全体の転送を避け、遠隔の文書サーバから必要とする文書の一部(例えば、段落や図表)のみを読み出す通信機能(文書の部分検索機能)の検討が要望されている。このため、CCITT SG VIII およびISO/IEC JTC1/SC18/WG4ではDTAM/DFRを用いた文書の部分検索機能の検討が開始された。これまでの標準化作業では、DTAMが提供する「部分読み出し機能」とDFRの「文書選択機能」をどのように組み合わせて利用するかが焦点となっている。

本稿では、DTAM/DFRを用いた「文書の部分検索機能」の問題点を明確化するとともに、ODA文書を対象とした部分検索機能の実現性について考察したので、以下に報告する。

2. DTAM/DFRの概要

(1)DTAM : DTAMは、CCITT勧告T.430シリーズで標準化されたOSI応用層プロトコルで、テレマテイクス端末間の文書一括転送機能の他に、遠隔のDTAMサーバに蓄積された文書の一部(図表や段落など)をDTAMクライアントが編集操作(挿入、削除、移動など)したり、文書の一部を読み出す機能を提供する。また、編集操作や部分読み出しが可能となるように、ODAのような構造化された文書が必要であり、各構造要素(操作オブジェクト:DMO)は、ひとつの文書内でDMO識別子を用いて一元的に識別される。尚、DTAMの編集操作や部分読み出し機能は、RO(遠隔オペレーショ

ン)サービス上で動作するもので、AGC(オーディオグラフィック会議)[1]やCDH(共同文書処理)などの通信応用にも適用できる。

(2)DFR : DFRは、ISO 10116としてISO/IEC JTC1/SC18/WG4で標準化されたOSI応用層の通信プロトコルで、DFRクライアントがDFRサーバに蓄積された文書群から文書単位で読み出す検索機能、サーバへ文書単位で書き込むファイリング機能、キーワードによる文書探索や文書リスト表示機能を提供する。DFRサーバではすべての文書がUPI(ユニークパーマネント識別子)で一元的に管理される。また、DFRは、分散オフィス応用モデル(DOAM)の通信環境(RO)で動作する。

3. DTAM/DFRを用いた部分検索機能とその問題点

文書の部分検索機能は、一般的にクライアントとサーバの間で提供される通信機能で、サーバに蓄積された文書を選択し、クライアントがその部分情報(例えば、段落や図)を「読み出す機能」である。DTAM/DFRを用いた文書の部分検索機能では、図1で示すように、始めにDFRの「LISTオペレーション」を利用することにより、DFRクライアントがサーバに蓄積された文書群の構成を知る。引き続き、検索した文書群の構成情報(LIST結果)に基づき、DFRの「RESERVEオペレーション(文書選択)」を用いてDFRサーバの文書群より文書d.2を選択する。

その後、DFRサーバの文書d.2がDTAMサーバ上に展開され、DTAMクライアントの「D-GETオペレーション」を用いて、文書d.2の段落P.1を読み出す。従って、DTAMクライアントは文書d.2の全体を読み出すことなく、必要な段落P.1のみを読み出すため、効率的な文書サーバアクセスが可能となる。

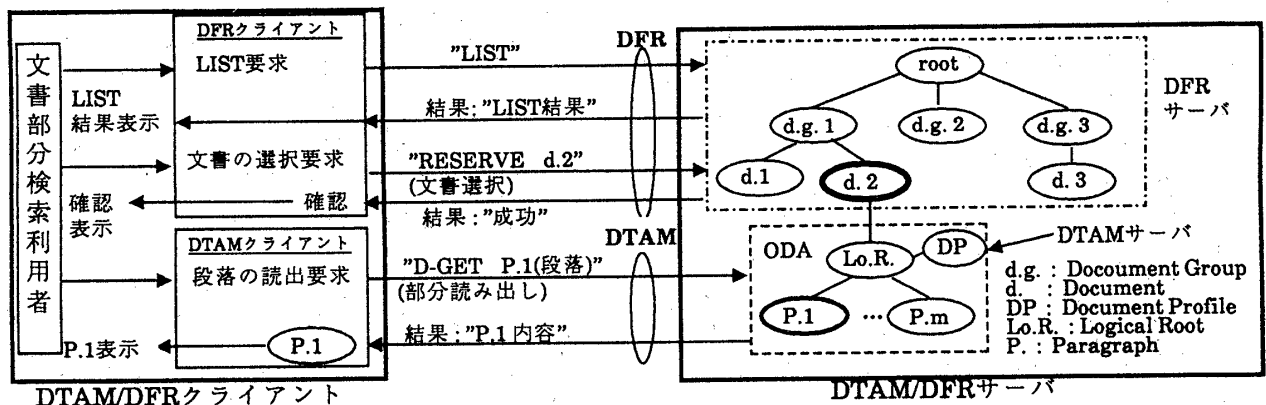


図1. DTAM / DFRを用いた文書の部分検索機能

“Study of Partial Document Read Function for ODA Document by DTAM/DFR”
 Kouji NAKAO, Toshiaki TANAKA
 KDD R & D Laboratories

このような通信機能を実現するためには、以下の問題点が指摘できる。

(1)利用者が選択した文書の内容構成を事前に理解していないと、DTAMの部分読み出しが実行上不可能となるため、クライアントが文書の読み出し単位(操作オブジェクト)とその意味(内容)との対応関係を示す情報(文書構成情報)を管理/作成する手法を明確にする必要がある。

(2)DTAM/DFRは共にROSE(遠隔オペレーションサービス要素)上のOSI応用層ASE(応用サービス要素)であるため、同一の通信プラットフォームで動作するが、それぞれ別々のASEであるため、DFRで選択した文書の識別子(UPI)を、全く別のASEであるDTAMにどのように伝達するかを明確にする必要がある。

4.文書の部分検索機能の問題点の考察

(1)ODA文書のための文書構成情報の管理/生成

文書構成情報(特に読み出し単位)は、使用する文書の構造に依存するため、本稿ではODA文書を取り上げ、文書構成情報の管理方式、およびその生成手法について述べる。

[文書構成情報の管理方式]

ODA文書プロフィール(DP)は、文書の全体的な特性情報(文書作成者、作成日時、文書属性など)を保有するため、文書全体に関わる文書構成情報はDPにて保有する。従って、クライアントが部分読み出しを行う前に、文書プロフィールの文書構成情報をまず読み出す。この文書構成情報は、以下の2つの手法によって生成される論理構造用、およびレイアウト構造用の構成情報から成り立つ。

[文書構成情報の生成手法]

①論理構造の場合

ODAでは、論理オブジェクトに対して章番号やタイトルを示すアプリケーションコメント(APC)を付与することができる。従って、文書構成情報は、アプリケーションコメント(APC)の内容とそのオブジェクト識別子のペアからなるシーケンスとする。図2の例では、APC情報(第1章、タイトル:AAA)と論理オブジェクト識別子(11)のペアからなるシーケンスが文書構成情報としてDPに格納される。従って、利用者(人間)はAPC情報に基づき読み出すべき文書の部分情報を選定し、DTAMクライアントはD-GETオペレーションの引数として、識別した論理オブジェクト識別子を載せる。

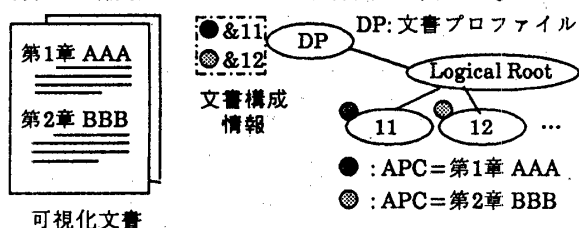


図2 ODA論理構造用の文書構成情報

②レイアウト構造の場合

文書構成情報は、レイアウトオブジェクトのサブプロフィール(SP)情報(例えば、ページ番号)、およびそのオブジェクト識別子のペアのシーケンスとする。図3においては、SP情報(第1ページ)とオブジェクト識別子(21)のペアからなるシーケンスが文書構成情報としてDPに格納される。しかし、

これだけでは各レイアウトオブジェクトの内容(意味)が伝達できないため、文書作成者は第0ページとして「目次情報(SP情報と意味内容の対応)」を文書の始めに必ず挿入する必要がある。従って、利用者は文書構成情報と目次情報を始めに読み出し、選択すべき部分情報を決定する。D-GETオペレーションの引数として、必要とするSP情報に対応するオブジェクト識別子を用いる。

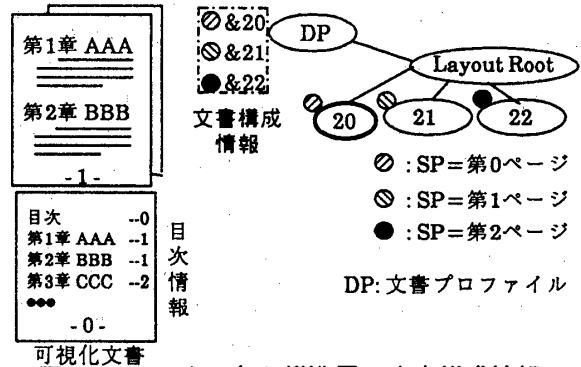


図3 ODAレイアウト構造用の文書構成情報

(2)DFRのUPI情報をDTAMに伝達する方式

以下に、DFRを部分的に拡張する方式1と、将来の高度な相互利用を考慮した方式2を示す。

方式1) DTAMのD-GETオペレーションをDFRがIMPORTして、DFRの応用コンテキストとしてDTAM部分読み出し機能を利用する。

方式2) DFRとDTAMで独立にアソシエーションを設定し、独立に動作させる。但し、DFRで選択したUPIは、DTAM/DFRユーザにおいて内部的にDTAM側へ伝達する。(図4参照)

結論的には、文書の部分検索だけを実現する場合は、上記両方式が適用できるが、検索した部分情報をクライアントが編集し、それを再登録するなど、将来必要となる高度な通信形態を考慮すると、DTAMおよびDFRの全機能が利用できる方式2が有効である。

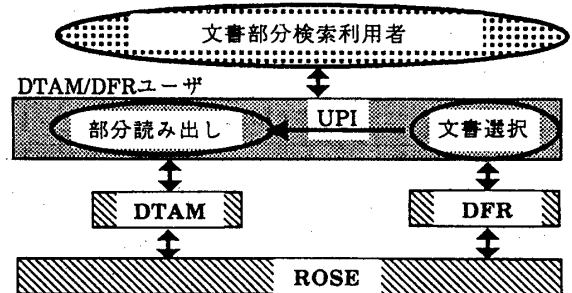


図4 UPI情報をDTAMへ伝達する方式(方式2)

5.おわりに

本稿では、文書部分検索機能の問題点を明確にし、ODA文書を用いた部分検索機能の実現手法を考察した。今後は、他文書フォーマット(EDI文書など)の検討を進める予定である。日頃御指導頂くKDD研究所小野所長、浦野次長、OSI通信グループ鈴木リーダに感謝します。

参考文献

[1]中尾、遠藤:“DTAMをベースとしたAGC会議プロトコルアーキテクチャの提案”情報処理学会研究会DPS-34-4 (1987年10月).