

電子図書館におけるマルチメディアデータの 管理モデル

鈴木 麗 川崎 由起子 砂原 秀樹

奈良先端科学技術大学院大学 情報科学センター

〒630-01 生駒市高山町 8916-5

ray@itc.aist-nara.ac.jp yukiko-k@is.aist-nara.ac.jp suna@wide.ad.jp

あらまし 奈良先端科学技術大学院大学では、1996 年 4 月から大学付属図書館を電子図書館として実現し、運用している。ここでは現在、論文等の冊子体情報とデジタルビデオ情報を、それぞれ独立したデータとして管理を行っている。しかし今後、情報は様々なメディアの複合体として提供されることが予想され、それぞれの情報に対する検索機能が要求されると考えられる。本稿では、複合メディアを管理するためモデルを考察し、新たなデータ管理モデルを提案する。

キーワード 電子図書館、データ管理モデル、インターネット

A Note on a model for Data Management on the Digital Library Systems

Rei Suzuki, Yukiko Kawasaki, Hideki Sunahara

Nara Institute of Science and Technology

8916-5 Takayama,

Ikoma, Nara 630-01, JAPAN

ray@itc.aist-nara.ac.jp yukiko-k@is.aist-nara.ac.jp suna@wide.ad.jp

Abstract Since April 1996, Nara Institute of Science and Technology (NAIST) started operation of a digital library system. This system contains the booklet style data and the digital video data. These two type of data are managed in separately on the current system. However, since digital library systems should manage several type of multimedia data, method to manage these data inconsistently is required. In this paper, we discuss a model for management with multiple type of data. Finally, we propose a model for data management on the digital library systems.

Keywords Digital Libraries, A Model for Data Management on the Digital Library Systems, Internet

1. はじめに

奈良先端科学技術大学院大学では、1996年4月から大学附属図書館を電子図書館として実現し、運用している。電子図書館システムは情報を電子化してデータベースを構築し、ネットワークを経由したアクセスを可能とする。利用者は普段利用している計算機から、図書館に格納されている論文、雑誌などを検索、閲覧することができる。情報へのアクセスはWorld Wide Web(WWW)によって実現されているため、特別なクライアントソフトウェアを用意することなく自由に利用することが可能である。また従来の冊子体情報だけでなく、デジタルビデオ情報を取り扱うことを可能とし、マルチメディアに対応した図書館を構築していることを大きな特徴としている。

本学の電子図書館システムは、従来の冊子体による図書館を電子図書館として実現する1つの方法として提供した。今後より進んだ電子図書館を構築するためには、電子化メディア、ネットワーク等の特徴を生かしたシステムの設計が必要となる。また情報のマ

ルチメディア化が進み1つの情報を様々なメディアで表現することが増えているため、これらの各情報の関連付け等の新しい技術も要求されている。

本稿では、現在運用されている電子図書館システムを概観して今後の技術的な問題点を取り上げ、新しいシステムに必要なデータ管理モデルを提案する。

2. NAIST 電子図書館システム

現在運用中の電子図書館システムは図1に示すような構成となっている。システムは、基本的に冊子体情報を取り込む一次情報入力システム、冊子体情報を格納するファイルサーバ、ビデオ情報を取り込むMPEG2エンコーダ装置、ビデオ情報を格納するビデオサーバ、各情報に対するデータベースを構成しWebサーバとして機能するサーチエンジンから構成されている。

一次情報入力システムは、書籍や雑誌などの冊子体情報を電子的に取り込むために用いられる。

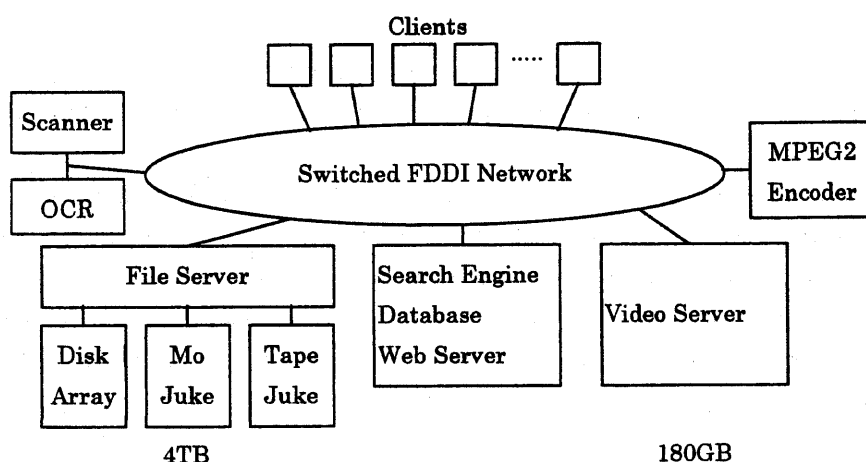
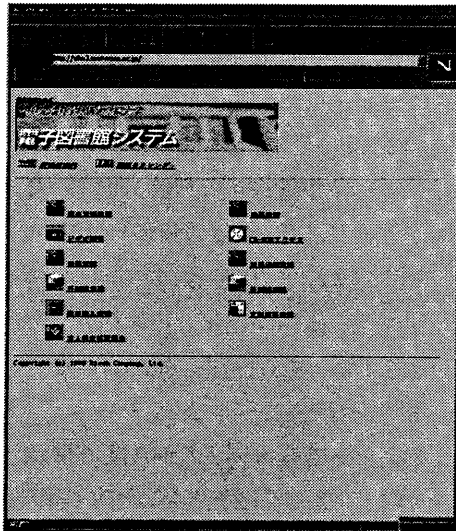
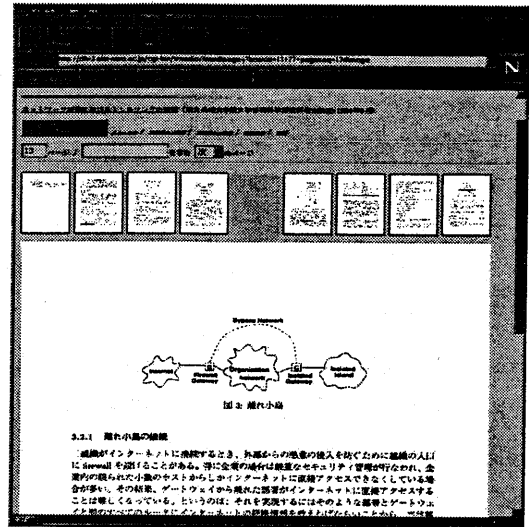


図 1: NAIST 電子図書館システム概要



(a) ホームページ



(b) 冊子体情報の表示例

<http://dlw3.aist-nara.ac.jp>

図 2: Web による電子図書館サービスアクセス例

本システムでは、冊子体紙面をイメージとして入力し、それを OCR によってテキスト情報に変換している。しかし、OCR によって生成されたテキストには誤りが含まれており、これを人手によって修正することは入力作業の効率が低下を招くことになる。そのため本システムでは、イメージ情報と OCR によって生成されたテキスト情報の 2 種類の情報を各冊子体情報に対して有しているが、利用者に表示する情報はイメージ情報だけとなっている。

テキスト情報は全文検索のみに利用されるが、検索の対象となるデータには間違いが含まれていたとしても重要なキーワードは同一文書に複数回登場するであろうことから、いずれかのキーワードが正しく読まれていることが期待され実用となると考えたためである。

一次情報入力システムによって取り込まれ

た情報はファイルサーバに格納される。現在ファイルサーバは、4TB の記憶容量を有している。この容量を実現するために、磁気テープジュークボックス(4TB)、光磁気ディスクジュークボックス(400GB)、ディスクアレイ装置(260GB)の 3 階層構造によって構成されている。ファイルの配置は各装置の残り容量と各ファイルの使用頻度によって自動的に配置されるようになっている。

ビデオ情報は、そのデータの性格を考慮して独立したシステムとして構成されている。ビデオ情報の取り込みは MPEG-2 エンコーダによって行われる。現在はビデオ情報を 4Mbps のビットレートで格納している。

ビデオサーバは、独立したファイルサーバが用意されており、180GB の容量を持つ。これは 4Mbps のビットレートの MPEG-2 ビデオを 100 時間分格納する容量に相当する。

格納されたデータを統合し図書館としての

サービスを提供するのがサーチエンジンである。ここでは、従来の図書館で提供されていた書誌情報検索機能だけでなくOCRによって生成されたテキスト情報を用いた全文検索機能も提供している。また、サーチエンジンがhttpサーバとして機能することにより、利用者はWebブラウザによってサーチエンジンのデータベースにアクセスすることができるようになっている。したがって、図書館サービスの利用は各利用者が利用する計算機に用意されている通常のWebブラウザを用いることで実現できるようになっている。

なお、現在のシステムではビデオ情報を表示するためにハードウェアMPEG-2デコーダを必要とするため、学内に設置された専用のクライアントでのみ表示が可能である。

図2にWebブラウザによる図書館サービス利用の様子を示す。

3. 現在のシステムの問題点

これまで、約1年半にわたって電子図書館システムの運用を行ってきたが、この過程でいくつかの問題点が明らかとなってきた。これらの問題の多くは、電子図書館固有の問題ではなく、Webを中心とするインターネット上の情報システムにおいても重要な課題となるものである。

ここでは、データの管理モデルに視点を置いてこれらの問題についての検討を行う。

3.1 データの管理方式

現在運用中の電子図書館システムでは、冊子体情報とビデオ情報管理はまったく独立に行われている。電子図書館に蓄積されているデータは、冊子体情報かビデオ情報かのどちらか一方であり、現状ではこの管理

方法に問題はない。

ところで、最近の情報はさまざまなメディアの組み合わせとして提供されることが増えている。例えばカラオケのデータは画像と音声、歌詞(テキスト)の組み合わせであるし、音声多重放送のように、1つの画像に複数の音声を組み合わせた番組もある。テキスト情報においても、英語の記事と日本語訳とは内容としては同一であるから、組み合わせて1つの情報と考えることもできる。

このような複合メディアから成る情報を各々のメディアの特徴を生かした形で管理、蓄積するためには、新しいデータ管理モデルが必要である。現状では、カラオケや音声多重放送は1つのビデオデータであり、英語と日本語の記事は独立した別のデータとして管理されている。データ管理モデルは、ビデオデータに含まれる画像、音声、テキストをそれぞれを取り出すことを可能とし、同じ情報はまとめた上で各々にもアクセス可能でなければならない。

したがって、電子図書館を中心としたインターネット上の情報システムで扱われるデータは、情報としてのデータをひとまとめにして管理し、その表現として異なる形式を個々に取り扱うことができるようになっていることが望ましいと考えられる。

3.2 検索機能

情報を電子化することによる大きな利点の1つは、検索が計算機上で行えるようになることである。NAIST電子図書館には全文検索機能が組み込まれており、これが大きな特徴として多方面で活用されている。

全文検索は関連のある情報をすべて抜き出すには確実な方法であり、必要な情報を

探し出せる可能性も高い。しかし効率には問題があり、今後の蓄積データ量の増大に伴い応答時間も長くなることは確実で、現実的にどこまで使用できるかは検討が必要である。関係のない情報が取り出されることも多い。

一次情報の簡易的な検索方法として古くから使われているのは、あらかじめキーワード付与しておく検索方法である。また新しい検索方法として、概念を用いた文書の分類と検索に関する研究も行われている[1]。将来の検索システムは、データの性質や、システムの状態、ユーザの要求などに応じて様々な検索方法を選択、切り替えて使用することにより、最も必要とされるデータを取り出すことが可能でなければならない。

現在の検索システムが対象としているのは冊子体情報の書誌情報と全文情報及びビデオの書誌情報、つまりテキストデータのみである。しかし、最近の画像処理技術、画像検索技術の発達により、近い将来に静止画、動画の検索機能もシステムに組み込めるようになると考えられる。画像の検索機能にも様々な方法があり、組み合わせや切り替えが必要となる。

また 3.1 で述べたような複合メディアを扱う場合は、検索対象となるメディアの種類が混在することになる。複合メディアで目的のデータを探し出すためには、各メディアにあわせた検索方法を選択することが必要である。

4. 電子図書館におけるデータ管理モデル

電子図書館で扱うデータを管理するために必要な機能は、次の 2 つにまとめられる。

1. 情報として同一な複合メディアを 1 つのデータとして管理する機能

2. 同一情報内の複数の表現(メディア)に対して個々の操作(検索など)を定義し、組み込み、実行する機能

これらの機能を満たすために、データを情報に基づく複合オブジェクトとして管理し、オブジェクト内の各メディアに対する個別の操作(検索機能など)をオブジェクトの一つのメソッドとして実現するデータ管理モデルを提案する。

4.1 複合メディアオブジェクト

電子図書館のデータ管理の単位である。情報に基づいて管理され、それを表現する複数のメディアで構成される。各メディアデータは以下の属性を持つ。

- データの種類(テキスト、動画等)
- データの形式(TIFF, GIF 等の圧縮方法)
- 表示方法(テキスト表示、イメージ表示、動画表示等)
- データの再生時間(連続メディア)
- 検索方法(全文検索、動画検索等)
- 検索のための情報(キーワード、動画の切り替えシーン等)

複合オブジェクトを定義することにより、カラーオケデータ、音声多重番組、英語/日本語の記事等、複数のメディアデータで構成される情報を一つのオブジェクトとして管理できる。MPEG2 形式では画像と音声は多重化された別データであるので、各メディアごとに分解し複合オブジェクトとして再構成するのは比較的容易である。

4.2 オブジェクトに対するメソッド(操作)

複合オブジェクトを管理単位とすることにより、それに含まれるメディアデータにアクセス

するためのメソッド(操作)が必要となる。基本的にこれらのメソッドは拡張可能であるが、各オブジェクトに最低限持つべきメソッドは次のようになる。

- 再生(表示)
- 停止
- データの追加
- データの削除
- 検索
- 属性の取り出し(表示)

各メソッドは、メディアデータの属性にあわせて異なる動作を行うことになる。例えば再生要求がきたとき、テキストデータであれば表示プログラム、MPEG2 データであればデコーダを起動するなどの動作の切り替えを行うことで機能を実現する。

5. まとめと今後の予定

本論文では、さまざまなメディアで構成される情報を一括して管理する仕組みとして、情報に基づくデータの管理方式、複合メディアオブジェクトを提案した。複合メディアオブジェクトでは、従来のオブジェクトとは異なり、組み合わせられる表現メディアに対応する複数のデータを有し、それぞれに対する操作(メソッド)を定義している。そのため、複数メディアで表現される情報を統一された操作によって取り扱うことができ、情報はその意味(単一の書籍など)に基づいて管理することが可能になる。

現在、このモデルに基づき新たな電子図書館システムのプロトタイプを構築中である。このプロトタイプでは、現在の電子図書館に格納された情報を利用し、実際に運用されているシステムに組み込んで評価を行うよう準備を進めている。まずビデオ情報を本モ

デルによって管理することによりビデオ検索機能を組み込み、冊子体情報とともに柔軟に検索を可能とすることを目指している。

また、今回論文中ではふれなかったが、今後電子図書館システムは、さまざまな場所に設置され、それらを組み合わせて利用するという状況が想定される。その際に、各電子図書館システムが統合された情報システムとなるためには、複合メディアオブジェクトモデルを応用し、システム間の連携に基づく情報提供がはかられることが重要であると考ええる。そのため、プロトタイプは、単に一つのシステム内での情報システム構築だけでなく、複数のサーバで構成される情報システムを統合して操作できる環境の構築を目指している。

今後、コーディネートされた情報を提供するシステムとして電子図書館システムは、インターネットでの情報システムの中核をなすと考えられる。そのため、従来情報の管理に関して重要な役割を果たしてきた図書館が変わることによって、新しい情報の形態や情報管理の仕組みが生まれてくるかもしれない。

参考文献

- [1] 堀井千夏, 今井正和, 千原國宏, 嵩忠雄: “デジタル図書館のための概念情報を用いた科学技術論文の検索手法”, 第9回デジタル図書館ワークショップ, pp.85-92, 1997年3月
- [2] Edward A. Fox, Robert M. Akscyn, Richard K. Furuta, and John J. Leggett (ed.), “Special Issue of Digital Library,” Communication of the ACM, Vol.38, No.4, April 1995.