

NAIST電子図書館：図書館におけるデジタルコンテンツ —収集から「知」の創造へ

砂原 秀樹、藤川 和利 (奈良先端科学技術大学院大学情報科学研究科)

1. はじめに

電子図書館の概念の登場は非常に古いが、実用レベルの電子図書館の登場はデジタル情報の展開とインターネットの普及によるところが大きいと考える。国内においても国立情報学研究所（電子図書館サービス開始当初は学術情報センター、1997年4月サービス開始）及び奈良先端科学技術大学院大学（以下「NAIST」という）でのモデル的電子図書館運用（1996年4月）を皮切りに、1997年、1998年の先導的電子図書館プロジェクト、1999年の電子的情報の収集・検索システムを通じて、広く展開が進んでいる。また、国立国会図書館においても2002年10月の関西館開館とともに電子図書館機能の推進が図られ、図書館機能の一つとして電子図書館機能つまりデジタルコンテンツの利用は不可欠なものとなってきている。一方、各出版社においてもデジタルコンテンツとしての出版物提供が促進されており、オンラインジャーナル等図書館で扱う情報も多様化してきている。

ここでは、NAIST附属図書館において蓄積されてきたKnow Howを元に、電子図書館構築の基本的な考え方を示すとともに、オンラインジャーナルとの連携、そして図書館におけるデジタルコンテンツの扱いについて今後の展望を示す。

2. 電子図書館の構成技術

電子図書館の構成技術の核は、デジタル情報とインターネットである。デジタル情報は、文字、画像、動画、3D表現、音声等さまざまな表現を数値化することで、表現形式に依存せず保存、共有、閲覧を可能としている。そして、このデジタル情報を共有する基盤として、インターネットが重要な役割を果たしているのである。特に、Web技術による情報の共有・閲覧機能は電子図書館実

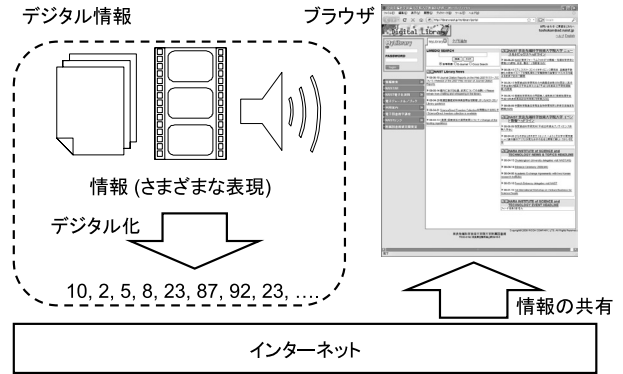


図1 デジタル情報とインターネットによる共有

現に重要な役割を果たしている（図1参照）。

ここで重要なことは、インターネットがオープンスタンダードで構成されていることである。これは電子図書館を構成する技術を、さまざまな部品で構成できることを意味する。標準が公開されているため、各コンポーネントは機能・価格・要求要件などに応じて自由に選択ができる。また、利用者から見ると利用環境を自由に選択できることとなり、電子図書館機能利用の自由度が高まることを意味する。今やWeb機能はPCだけでなく携帯電話などでも利用可能であり、電子図書館の利用形態も大きく変化すると考えられる。

以下では、電子図書館の構成技術を詳細に見ていく。

2.1 デジタル情報の表現形式

まず、デジタル情報の表現形式についてであるが、現在の電子図書館で取り扱われている情報形態として、冊子体情報をデジタル化したものとビデオ情報の2種類が主な情報となっていると考えられる。

NAIST電子図書館においても主な情報は冊子体情報をデジタル化したものである。基本的に、冊子体情報はスキャナによって画像情報としてデ

デジタル化されるが、それだけでは画像情報のみでありそこから先の活用が困難である。そこで、OCR技術（光学文字読み取り技術）を用いて、画像情報から文字情報を取り出すことが可能である。ただし読み取り精度は99%程度であるため完全な文字情報化は不可能である。この誤ったデータを手作業で修正することも可能であるが、膨大な量の冊子体情報を入力することを考慮すると現実的でない。そこで、OCRデータは検索専用とし、画像データを利用者に提示するという方法が多く採用されている。

当初、NAISTでも同様の方法を採用してきたが、画像データがページ単位となってしまうこと、文字情報が独立しているため検索したキーワードがページ中のどこにあるかわからないなどの理由から、これらをまとめて一つのファイルで表現できる形式としてPDF（Portable Document Format）を採用している。PDFはISO32000というドキュメント形式標準として認められており、インターネット上でも多く利用されている。特にオンラインジャーナル等で各アートを表現するために用いられており、共通の形式を採用することの意味は大きい。

現在、NAISTではスキャンした画像データを基本としたPDFデータをアールティクル単位で格納しており、OCRデータは透明テキストとして

画像データにオーバレイする形で埋め込んである。そのためPDFファイルに対して文字列検索を行うと該当する文字が含まれている部分に色がつくということが可能となっている（図2参照）。

ビデオ情報については、さまざまなデータ形式があるが、主に現在使われているのはMPEG-2形式とMPEG-4形式であろう。DVDや地上デジタル放送で採用されている方式がMPEG-2であるが、高画質高圧縮が可能であることから、MPEG-4形式の中のH.264形式が多く採用されるようになってきている。NAISTでも、当初MPEG-2形式を用いていたが、新しいシステムではH.264形式とし、ハイビジョン画像にも対応した。なお、YouTube等で利用されているフラッシュビデオ形式も注目されるデータ形式であるが、この形式もH.264形式を利用している。

2.2 情報共有基盤技術

これらの情報を共有する基盤システムとしては、インターネットの技術が活用されている。NAISTのシステムでは、情報はファイルサーバに格納されており、検索サーバが全文検索及び書誌情報検索を含む検索機能を提供している（図3参照）。

利用者は通常のWebブラウザとPDFファイルの表示プログラム（通常はAdobe社のAcrobat

Reader）を持っていれば良いように設計されている。これらのプログラムは、通常標準でインストールされているものであるため特別な準備することなく利用が可能である。また、ビデオの表示にはマイクロソフト社のウィンドウズメディアプレーヤー（WMP）またはリアルネットワークスのリアルプレーヤー（RealPlayer）が一般的に用いられているが、NAISTの現行システムではリアルプレーヤーを採用している。いずれも再生用プログラムは無償で提供されているため容易に再生環境を準備することが可能となる。

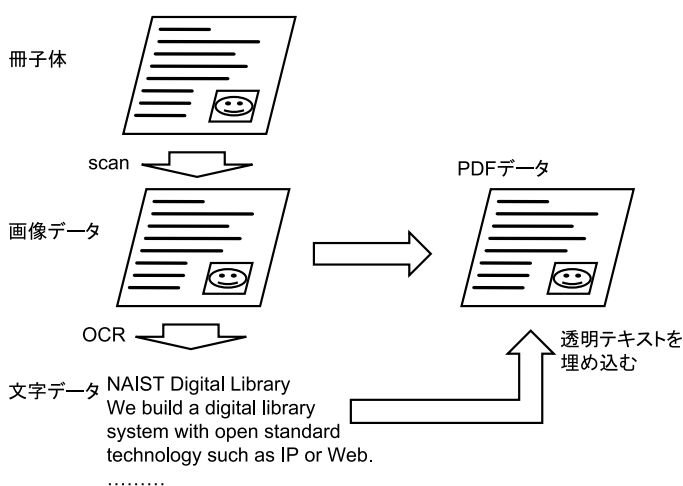


図2 冊子体情報のデジタル形式

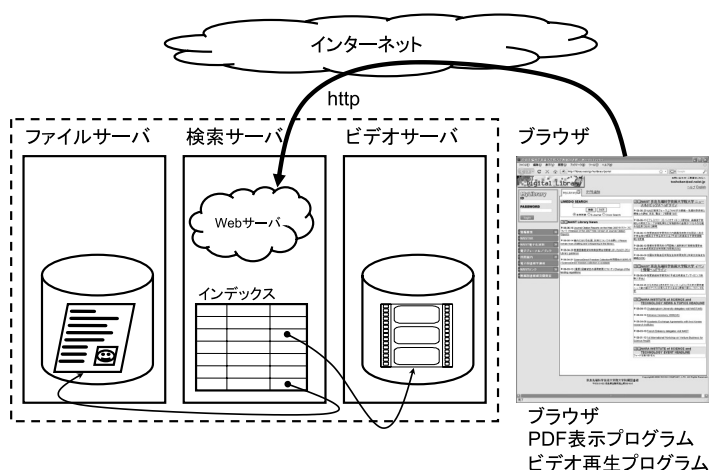


図3 電子図書館システムの基本技術

3. NAIST電子図書館システムの概要

ここでは具体的にNAIST電子図書館システム¹⁾の利用の様子を示す。対象となる利用者は、NAISTに在籍する学生及び教職員である。

利用者は、電子図書館のWebページにアクセスし利用者IDとパスワード²⁾を入力すると、図4に示すようなポータル画面に誘導される。この画面は利用者単位でカスタマイズ可能である。具体的には、利用者画面は「ウィジェット」とよばれるモジュール群から構成されているため、モジュールのレイアウトの変更やモジュールの追加・削除が



図4 NAIST電子図書館 ポータル画面

可能となっている³⁾。モジュールを用いた画面構成は、2008年3月に導入された新電子図書館システムの新しい機能であり、電子図書館サービスの個人化を目的としたMyLibrary (次節参照)を実現するためのものである。現在、NAIST電子図書館システムでは、文献検索、NAIST図書館からのお知らせ、RSSリーダ、貸出手続きなどの図書館業務に関連するものなど20個のモジュールが提供されている。また、モジュールは

すべてRuby on Railsによって構成されており、システム管理者による新たなモジュール作成等の機能追加も容易であり、今後新しい機能を追加していく予定である。

基本的な利用方法は、GoogleやYahooなどの検索エンジンと同様であり、検索用モジュールに検



(a) キーワードの入力



(b) 検索結果

図5 検索と結果の表示

索キーワードを入力することによって行う (図 5 (a))。

ここでは、学内に蓄積された情報のみを検索対象にするか、オンラインジャーナルを検索するか、すべてを検索対象にするかを選択することができるようになっている。当然、検索対象を広くするほど検索速度が遅くなるため、必要に応じて選択することが望ましい。結果は図 5 (b) のように表示され、冊子体情報、ビデオ情報、オンラインジャーナルを区別無く表示するようになっている。それぞれの資料に対応するアイコンをクリックすると、それぞれの資料の書誌情報または資料本体が表示されるようになっている。

ビデオ情報は、図 6 のように表示される。現在 NAIST では、ビデオ情報の共有を特に授業アーカイブとして活用している。そのため図に示したようにビデオだけでなく、講義資料を同時に表示するようにしている。講義資料は、ビデオの進行状況に合わせて更新されるようになっているとともに、資料上のキーワードも検索に用いられるようになっている。

冊子体情報は、図 7 に示すように、PDF ファイル表示プログラムによって表示される。表示では、「Geocrawler」という単語がハイライトされているが、これは検索時に用いたキーワードがハイライトされているものである。多少画像上でのずれがあるのは埋め込まれている透明テキストの場所が微妙に異なっているためであるが、実用上支障となるレベルではない。

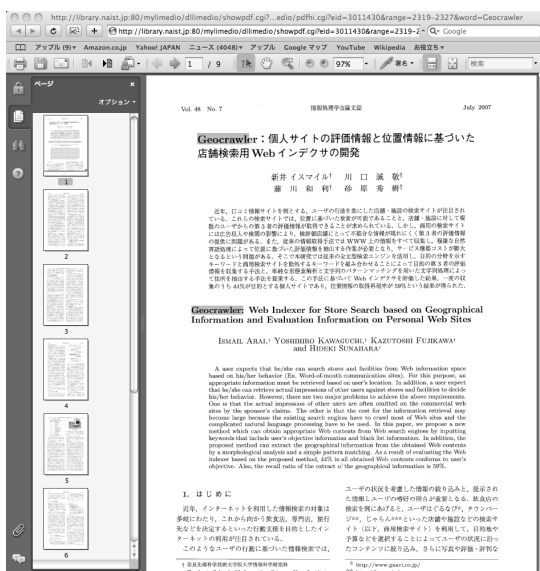


図 7 冊子体情報の表示

これら機能を実現するために、NAIST 電子図書館システムは、一次情報入力システム、デジタルビデオシステム、一次情報蓄積システム、検索システム、業務支援システムから構成される。一次情報入力システムは、書籍情報の一次情報を電子化するために用いられ、データ入力・メディア変換用機器から構成されている。デジタルビデオシステムは、学内で行われる授業や講演会を映像により記録したものを電子図書館システムにおいて、書籍情報と同様に扱えるようにするためのシステムである。このシステムを用いることで、映像のみならず、授業・講演に用いられたスライドとの同期を取った形 (図 6 参照) で利用者に提供することが可能となっている。一次情報蓄積システムは、一次情報入力システム及びデジタルビデオシステムによってデジタル化された情報を蓄積するための大容量ファイルサーバである。具体的には、ホットスタンバイディスクを用いた RAID4 構成及び RAID6 構成で実現されており、実効記憶容量としてそれぞれ 12TB 及び 48TB を有している (図 8、9 参照)。検索システムは、従来からの二次情報を用いた検索のほか、一次情報

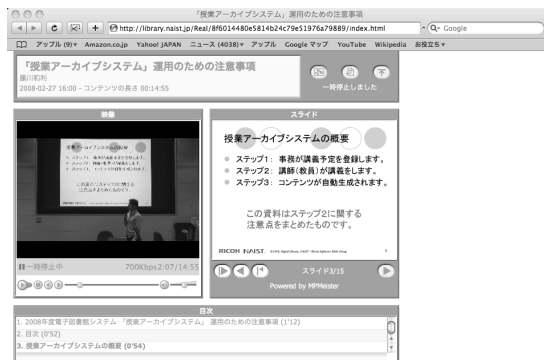


図 6 ビデオ情報の表示



図8 ファイルサーバ (12TB)



図10 検索エンジンと負荷分散装置

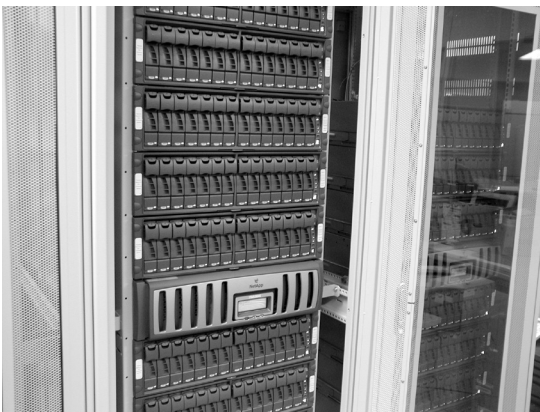


図9 ファイルサーバ (48TB)

入力システムのOCRによって作成された索引情報を用い、全文検索機能を実現している。また、NAIST電子図書館システムに登録されている外部の電子ジャーナルやデータベース⁴⁾に対しても検索を行うことが可能となっている。検索システムは、9台の検索エンジンと1台の負荷分散装置から構成されている(図10参照)。業務支援システムは、従来の図書館業務を支援するものである。これらのシステムは連携して運用されているため円滑に処理が進むように、また利用者の学内外からの多数のアクセスに耐えうるように、1Gbpsのネットワークで相互接続されている。

4. 電子図書館機能の今後の展望とデジタルコンテンツ

インターネットの普及に伴いデジタル情報が爆

発的に増えたことにより、電子図書館機能は非常に重要な役割を果たすと考える。一方で、検索エンジン等が提供する機能は単にキーワードによる情報検索機能を提供しているだけであり、情報のフラット化を招いている。しかし、従来図書館が提供してきた機能は、情報に熟知した司書の知恵などによる「知」の提供であったと考えられる。そこで今後電子図書館システムに求められる機能は、単に情報を蓄積し、検索機能を提供するだけでなく、「知」を利用者に提供するものであろう。

また、Web2.0という言葉に示されるように、今後のWebサービスのあり方として、利用者による情報の自由な整理や積極的に利用者情報を活用した個人化サービスが求められている。

そこでNAISTではまず電子図書館システムの機能を拡張し、電子司書と呼べるMyLibrary機能の開発を進めている。これは、MyLibrary内に格納される資料の参照履歴や検索履歴を用いて、電子図書館システムが利用者の興味に応じた資料を提示し、より効果的な資料閲覧を可能とする機能である。つまり、利用者専用の司書が存在することと同じ能力の提供を目指したものである。

一方、電子図書館の普及にあわせて、各出版社はオンラインジャーナル等のサービスを展開している。これにより、電子図書館システムの主な役割の中から、「出版されている雑誌や書籍の情報の収蔵」という役割は薄れつつある。このような現状を背景に、NAISTの電子図書館システムに

においても学内で生産される情報の収集と蓄積という役割が重要となってきている。現在、学位論文や授業アーカイブ、講演ビデオ等の情報を収容しているが、今後学内で生産される情報はすべて電子図書館システムに収容するよう検討を進めている。こうして収容される資料には、論文や参考資料、それらを閲覧した際に作成したメモ、プログラム、実験結果など利用者が教育・研究活動を通して生成した情報とその行動履歴を含んでおり、これらの情報を体系化し管理・提示する機能の開発を進めている。これは、「知識・知恵」を構成する技術であり、先端的研究者が集結するNAISTにおいて新しい知識・知恵を集約・体系化することで、世界へ向けた知識発信基地となることを目指すものである。

デジタルコンテンツの扱いは、それらを収集し検索エンジンを構成することから、大量のデータからさまざまな関係、利用者の動向を抽出し、より有益な情報を構成することへと移行しつつある。これは、すべてのデジタル情報が同じ共通の

基盤の上に格納されているというインターネットの特徴を活用したものであり、今後より大きな役割を果たすと考えられる。爆発的に増殖する大量のデータはもはや一箇所に集約することは不可能であり、電子図書館システムやオンラインジャーナルシステム等が連携し機能することが求められている。そういった意味においても、電子図書館システムの今後の注目がされる。

(すなはら ひでき、ふじかわ かずとし)

- 1) <http://library.naist.jp/>
- 2) NAISTでは、「曼陀羅システム」とよばれる全学情報環境システムにより利用者IDとパスワードを一元的に管理している。
- 3) NAISTの学生・教職員でなくてもNAIST電子図書館システムを利用できるが、画面レイアウト変更やモジュール追加・削除は保存されない。
- 4) 例えば、NDL-OPACやGoogle Scholar Advancedなど。

NAIST Digital Library: Digital Contents on Libraries

-- Toward Creating Knowledge with Digital Contents

(By Hideki Sunahara and Kazutoshi Fujikawa, Graduate School of Information Science, Nara Institute of Science and Technology)

In 1996, the library of Nara Institute of Science and Technology started operating as a model project of digital library in Japan. We have an operating and management experience over 12 years. Based on the experience, we renew the digital library in 2008 in terms of system architecture and user service model. In this paper, we describe the basic structure of our digital library. Moreover, we discuss the management model of digital contents.