

2.NAIST 電子図書館の状況

2.1 電子図書館システムの現状

本学の電子図書館システムの調達は、1995 年度より 7 ヶ年計画で導入が進行した。この調達において、各年度に調達されるシステムは 4 年間のレンタルを基本としており、2003 年度(2004 年 3 月)に導入されたシステムより、第三世代目の現行のシステムが稼動を開始している。ここでは、現行のシステムの設計方針及びその概要を示す。また、2007 年度に導入されるシステムより展開される第四世代目への展望を示す。

2.1.1 システムの設計方針

大学附属図書館は、あらゆる情報源を利用して教育研究に必要不可欠な情報収集活動を支援するものでなければならない。また、学内で生成される学術情報を外部に対して積極的に提供するための情報発信機能も有していなければならない。このため、本学附属図書館では以下のような目標のもとに、先端的な教育研究活動を支える基盤環境の構築を目指している。

- 国際的な規模での情報流通のグローバル化への対応
- 従来の印刷物中心の図書から音声、画像、映像等を含むマルチメディア情報への対応
- 情報の収集・公開を加速するための情報源に関する情報の収集、一次情報(論文や記事そのもの)の収集、および一次情報保有機関への迅速なアクセスルートの確保
- 利用者に最も近いところでのサービスの提供

さらに、上の 4 つの要件を満たし続けるためには、高度情報化社会における図書館を取り巻く状況の急速な変化(新しいメディアや情報サービス形態の出現、

出版形態の変化、研究発表・情報発信形態の変化など)に柔軟に対応する必要がある。このため、ある時点で可能なサービスを固定的に提供するのではなく、常にサービスの種類・質と図書館の役割を検討しつつシステム構築を進め、時代の変化に柔軟に対応することを目指している。また、附属図書館は利用者サービスの観点からは以下の三つの性格を有することを目標としている。

- **メディアセンター** - 利用者が資料の種別(図書, 雑誌, 音声, ビデオ等)を意識することなく情報提供サービスを受けられることができる。
- **居ながら図書館** - 図書館に出向くことなく、利用者に最も近いところから検索・閲覧を行なうことができる。
- **24時間図書館** - 情報提供の時間が限定されず、開館時間の概念がない。

電子図書館システムは、本学の情報環境(曼陀羅ネットワークおよび曼陀羅システム)を基盤として、上述の理念を具体化するための中核設備である。

2.1.2 電子図書館システムの現状

こうしたことを実現するため、原則としてシステム内に一次情報を蓄積し、それらを利用者が検索して利用する機能を提供することを中核として、これらを支援する機能とともに設計が行われている。具体的には、システムは以下の6つのサブシステムで構築されている。

1. 一次情報入力システム

本システムは書籍情報の一次情報を電子化するために必要不可欠なデータ入力・メディア変換用機器群である。附属図書館が入力の対象とするメディアには、冊子体、CD-ROM 及びネットワーク経由のファイル入力があり、特に冊子体の場合はカラーで表現された写真が含まれている。ここでは、冊子体情報をスキャナで読み込むとともに、OCRによる全文検索用のデータの生成、目次情報の入力、システム内で利用されるデータ形式(基本的に、記事/論文単位のPDF)への変換編集等が行われる。また、CD-ROM やネットワーク経由で入手されたファ

イルについては必要なデータ形式への変換を経て、システム内に取り込まれている。

現在 3 台のモノクロ・カラー兼用スキャナ装置とこれらを支援するための PC 群で構成されている。特に OCR 処理には高い処理能力が必要とされるため、10 台の PC で構成されるクラスタで処理される。これらの処理は自動的に 10 台の PC に負荷分散される。さらに、入力作業用のデータ領域として独立した領域を用意している。

また、ここで生成される情報は、通常の図書館での書誌に相当する。これらの情報が焼失することは、書籍の紛失を意味し、ファイルのバックアップの体制は十分に確立していることが求められる。そのため、第一次のバックアップとして最終的な形式への変換が完了すると書誌単位で DVD-R へのデータのバックアップを作成することとなっている。

2. デジタルビデオシステム

今日の教育研究活動では、映像情報の重要性が増大している。本システムは、外部から映像情報を収集するとともに、本学において映像情報を製作し、他の図書情報と統合化した形でデジタルメディアに蓄積している。また、それを学内利用するだけでなく、学外への発信をも目指している。システムでは、ビデオ情報を MPEG-2/MPEG-4 形式で保存し、ネットワーク経由での閲覧機能を提供している。さらに、本学において映像情報を作成するためのバーチャルスタジオ、編集装置なども本システムに含まれる。これにより、本学からビデオ情報を発信する環境が構成されている。

現在、8.5TB の容量を有し、4Mbps ストリームの MPEG-2 ビデオストリームに換算すると約 17,000 時間に相当する。

また、授業アーカイブシステムとして各教室に常設のビデオカメラを設置し自動的にすべての授業を収録しアーカイブするシステムの運用を開始している。2005 年度より情報科学研究科での試験運用を行っているが、最終的には全学においてすべての授業の収録を行う予定である。

3. 一次情報蓄積システム

デジタル化された冊子体一次情報を蓄積するための大容量ファイルサーバ

である。最終的に十数 TB の容量を実現することを前提に設計されたため、容量のすべてをディスクアレイで構成するのではなく、磁気テープジュークボックスを用いた階層型の記憶システムによって構成されている。データの配置は、データの利用頻度等によって自動的に行われるように管理されている。ディスクアレイは、磁気テープジュークボックスのキャッシュとして動作し、磁気テープジュークボックスに収容可能な実効容量の 25% の容量を持つ。

現在、システム全体で 12TB の記憶容量を有している。また、ここでも 2 次的バックアップ機能として、磁気テープジュークボックス内において、各テープの複製を自動的に作成し、メディアトラブル発生時には自動的に代替メディアを利用するとともに、オペレータに障害を通知するようになっている。

また、2004 年度(2005 年 3 月)に導入されたシステムでは、第四世代目を睨み階層型記憶システムではなくディスクアレイのみで構成された 12TB のシステムを導入している。このシステムには 16TB の磁気テープジュークボックスシステムをバックアップ装置として用いており、デジタルビデオシステムのサーバとしても共用している。

4. 検索システム

従来の図書館が提供する二次情報を用いた検索機能及び、OCR によって生成された一次情報を用いた全文検索機能を提供する。電子図書館システムにおいてデータは単調に増加することになる。その結果として生じる検索性能の低下を防ぐため、ワークステーションワーククラスタ構成による検索サーバアーキテクチャを採用し、必要に応じてエンジンを追加していくことで全体の検索性能を強化できるようにしている。また、このような構成を採用することでこれまで夜間に行ってきたハウスキューピング処理をクラスタ中の一台のエンジンに割り当てバックグラウンドで行うことができるようになる。これによりハウスキューピング処理による、サービス性能の低下は最小限に留めることができるようになる。

また現行システムでは、検索システムとして全文検索サービスを提供している。ここでは単に文字列を含む文書を提示するだけでなく検索した文字列が該当文書のどこにあるのかを表示する機能を提供している。これは、新しい PDF の機能によって OCR によって認識した文字列を透明なテキストとしてイメージデータの前に貼り付け機能を用いて実現している。これにより、PDF の検索機能

を用いて画面上で検索文字列の存在場所を示している。

現在、検索サーバは9台で構成されるワークステーションクラスタとなっている。

5. ネットワーク接続装置

電子図書館のシステムは、これらの装置単独で動作するわけではなく、これらが有機的に協調することによって全体のサービスを提供している。また、学内の情報システムである曼陀羅システム及び曼陀羅ネットワークとの協調も不可欠である。本システムは、これらシステム間の協調を支援するための高速ネットワークである。現在、Gigabit Ethernet を基本とするスイッチベースのネットワークで構成されている。

6. 業務支援システム

電子図書館といえども、従来型の図書館機能を失っているわけではない。これは、著作権の許諾の関係で電子化できない書籍等に対しての従来型サービスの提供だけでなく、電子化作業の工程管理など電子図書館としても業務を支援する機能を必要としている。これらの機能を提供するため従来型の図書館業務支援システムを拡張する形での実現を行っている。

また近年図書の管理において RF-ID を用いたシステムの導入が進んでいる。本学においても、学生証の IC カード化と並行して導入を進めている。

2.1.3 電子図書館サービス

現行のシステムが提供しているサービスは、基本的に図 1 に示した通りである。利用者は手元にあるシステムより Web ブラウザを用いて、電子図書館システムにアクセスする。通常の場合、キーワードを入力し全文検索機能で該当する文献を探し出すということを行うことになる。また、MPEG-2 のビデオ閲覧は MPEG-2 のデコード機能を有する端末のみで行うことができるようになっている。

また、雑誌リストから巻号にしたがって、雑誌をブラウズすることも可能である。

これ以外に、新着図書の中から利用者が登録したキーワードを含むものがあった場合、それを電子メールで知らせるサービスも提供されている。電子メールには文献名とともにそれにアクセスするための URL が含まれているため、通常はこれをクリックするだけで Web ブラウザが起動され該当する一次情報を参照できるようになっている。



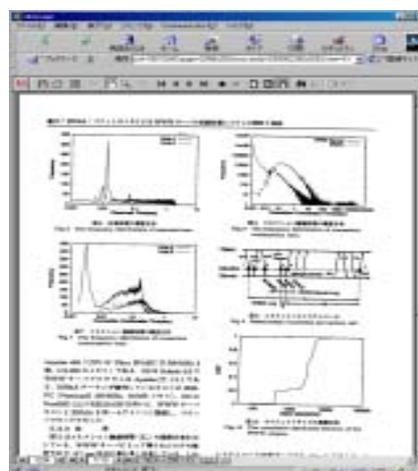
(a) Home Page
<http://dlw3.naist.jp>



(b) Specify "Keywords"



(c) Search Results



(d) Browse Pages

図 1: 電子図書館の利用例

2.1.4 次世代電子図書館サービスへ向けて

現在のシステムは、「情報館」構想に基づき新しい機能の導入を行っている。ここでは、次世代の電子図書館システムへ向けての展望について述べる。

(a) ファイルサーバの構成

すでに述べたが、数十～数百TB級のファイルサーバはディスクアレイのみで構成ことが可能な価格レンジへ収まるようになってきた。データへのアクセスパターンを考慮してもディスクアレイのみでの構成への移行は自然な流れである。第四世代目についてはこうした構成を基礎として計画を進めている。

(b) 電子書齋/司書サービス

利用者が電子図書館に求めるものは、単純に情報を提供することではなく、どのような情報を見れば良いのかを示唆することである。現行のシステムでも、利用者の興味のある単語を登録してもらい、新着図書からその単語を含むものを抽出し、メールで通知する機能を提供している。しかし、利用者の検索履歴、閲覧履歴、ブックマークなどを用いることで、利用者の興味を抽出しより適切な示唆を与えることが可能であると考えている。電子図書館システム内にユーザ毎の書齋のようなスペース(MyLibrary™)を提供し、利用者の情報を蓄積し適切な情報の示唆を行う電子司書の機能の提供を検討している。

こうした機能の基盤として、ユーザ毎の情報保存スペースを用意し、定期購読や検索・閲覧履歴、ブックマークなどを管理する機能を提供する。これらの技術は近年注目されているWeb2.0と呼ばれる仕組みによって実現することが可能であり、RSSによる新着情報の管理や類似利用者の抽出や推薦書誌の提示等の機能を実現するための開発を開始している。

(c) 情報発信機能

電子図書館に求められる機能は、単に情報の閲覧だけではない。論文、テクニカルレポート、テクニカルメモなど、利用者が作成した情報を発信する機能が不可欠であり、これを自由に行えるようにすることでインターネットを経由した情報の広いサーキュレーションを実現することができるようになる。ユーザ毎の書齋スペース(MyLibrary™)では、情報発信のための機能も提供し電子図書館に組み込んでいく。また、従来の書誌形式の情報だけでなく、ソフトウェアや遺伝子配列情報、材料成分等の本学が有する「知」を発信する仕組みとし

での活用を前提に柔軟性に富んだ情報管理システムの開発を進めている。

さらに前述の通り大学が有する重要な情報として「講義」を捉え、講義を収録し公開していく機能を用意しているが、これらを PodCast 等の技術で配信する実験を開始している。