

世界規模ネットワークに結ばれた マルチメディア電子図書館システム

Multimedia Digital Library System on Global Information Network

藤澤浩道*

Hiromichi Fujisawa

金田玄一****

Gen'ichi Kaneda

三科雄介**

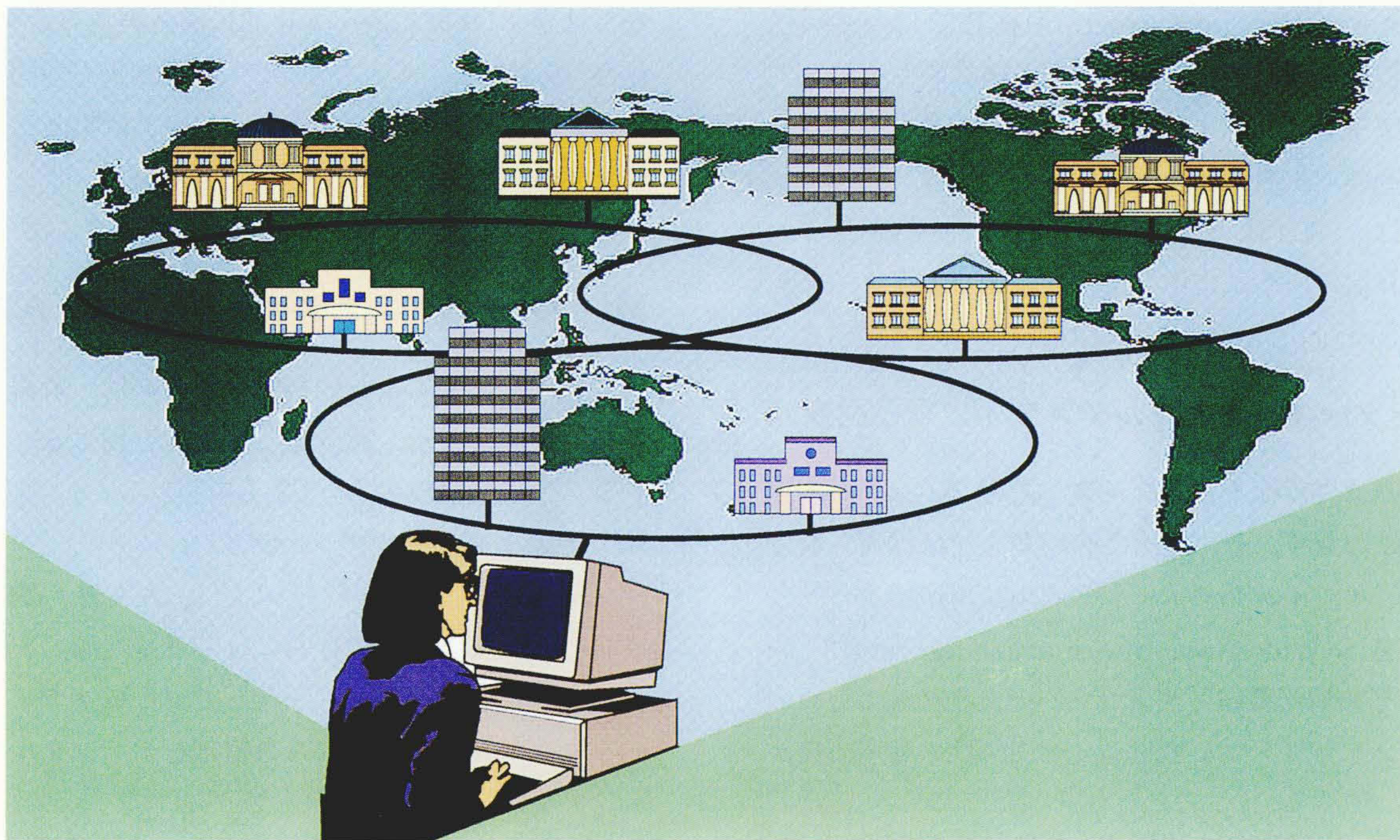
Yūsuke Mishina

菱沼茂樹*****

Shigeki Hishinuma

加藤寛次***

Kanji Katô



マルチメディア図書館のイメージ

ネットワークに接続された自分のマルチメディアパソコン(パーソナルコンピュータ)から世界中の情報に、あたかもそれが手もとにあるがごとくアクセスできるようになる。

マルチメディア電子図書館の構想が、日米欧で検討されている。これは世界規模の計算機ネットワークに接続され、文献や資料などの膨大な知的財産を蓄積する情報蓄積・検索サービスシステムである。これにより、ネットワークに接続された自分のマルチメディアパソコンから、居ながらにして世界中の欲しい情報に即座にアクセスすることが可能になる。すでにインターネット上には多数のWWW(World-Wide Web)^{*1)}と呼ばれる情報蓄積サーバが構築されており、そのためのクライアントソフトウェアMosaic^{*2)}を用いれば、それらサーバの情報に

アクセスして世界中から情報を入手することが可能になりつつある。

現在は紙やフィルムなどの媒体に記録されている膨大な情報を逐次電子化(デジタル化)し、テラバイト級の巨大なマルチメディアデータベースを構築することも現実のものとなってきた。容易な検索を実現するためには、高度な検索技術が必要となり、これらを目指したプロジェクトが日米欧で進行しつつある。これにより、ネットワーク全体が実質的に巨大な「マルチメディア情報の蓄積貯蔵庫」となる。

* 日立製作所 中央研究所 工学博士

** 日立製作所 中央研究所

*** 日立製作所 情報・通信開発本部

**** 日立製作所 システム事業部

***** 日立製作所 公共情報事業部

1 はじめに

マルチメディア情報処理技術、マルチメディア通信技術、大容量記憶装置技術などの進展、およびマルチメディアパソコンの高性能化と低価格化により、次世代の魅力的な情報システムが現実のものとなってきた。世界規模の情報ハイウェイの実現を目指す世界情報基盤(GII: Global Information Infrastructure)構想では、遠隔医療、海事情報、環境情報、および電子図書館などのシステムが有望なアプリケーションとして考えられており、日米欧でその取組みが始まっている。

ここでは、「電子図書館」を取り上げて、そのシステム概念と日立製作所が試作中の電子図書館プロトタイプシステム“Cornucopia”(コーヌコピア)について述べる。

2 インターネットとマルチメディア電子図書館

計算機の世界規模ネットワークである「インターネット」は、ここ数年、その規模と利用が急速に拡大し続けている。近年、インターネットの商用利用が始まるとともに、WAIS(Wide-Area Information Servers)^{※1)}やWWWなどの情報蓄積・検索サーバと、それらのためのクライアントソフトウェアMosaicが出現し、利用の輪が

格段に広がった¹⁾。インターネットに接続されたパソコンでMosaicを用いると、世界中に分散して存在しているWAISサーバやWWWサーバから必要な情報を手もとに取り寄せて閲覧したり、印刷することができる。

「マルチメディア電子図書館」とは、このように世界に開かれたネットワークに接続され、世界に情報を発信する新しい情報サービスシステムの概念である。このようなシステムは、歴史的・文化的に価値のある古文書、美術作品、文献、フィルムなどの、人類の膨大な知的財産を電子化して、検索可能なマルチメディア情報として蓄積・管理する。将来に向けては、ビデオやコンピュータグラフィックスで表現された知的財産も蓄積するようになる。

3 マルチメディア電子図書館システムの構成

マルチメディア電子図書館システムは、情報の入力、蓄積、検索、配送、課金などの機能を備える、以下のサブシステムで構成する。(図1参照)。

(1) メディア変換サブシステム

メディア変換サブシステムでは、古文書、書籍、文献、資料などの文書情報、さらには写真や映像、音声(音楽)といった視聴覚情報を電子化してマルチメディアデータベースに登録する。これらの情報をデジタル化したテ

※1) WWWは、スイス・ジュネーブのCERN(欧州共同原子核研究機関)が開発した地球規模ハイパーテキストサーバである。

※2) Mosaicは、米国イリノイ大学のNCSA(The National Center for Supercomputing Applications)が開発し

たWWWのためのグラフィカル ユーザー インタフェースである。

※3) WAISは、米国Thinking Machines社が中心になって開発した広域テキスト検索サーバの名称である。

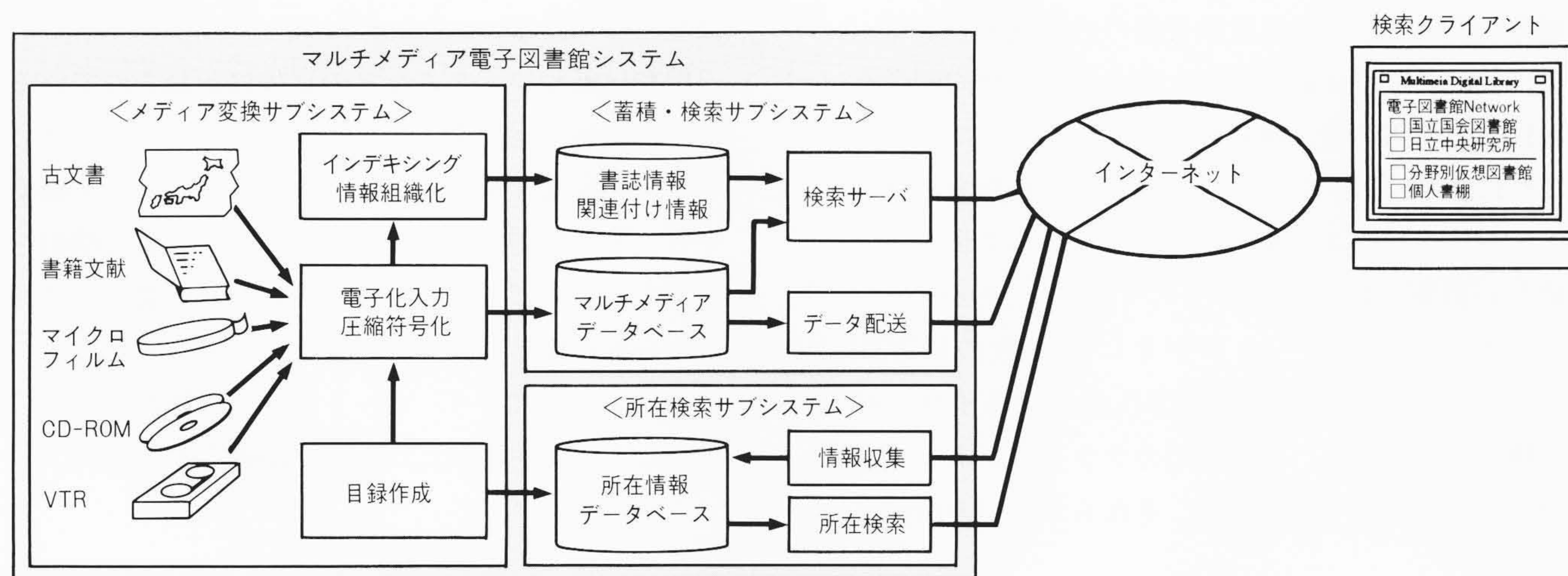


図1 マルチメディア電子図書館システムの構成

ネットワーク上には多数のマルチメディア電子図書館が存在する。ユーザーはいずれかにアクセスして、欲しい情報がどの図書館にあるのかを所在データベースで検索することができる。

キスト、静止画、音声、動画データに変換することをメディア変換という。データ圧縮しながら原本を忠実に再現する変換技術や、古文書の染みなどを回復する「ディエイジング(de-aging)」技術が求められる。

検索を可能にするためには、それぞれの情報に対して、タイトル、主題、作成者、出典などの書誌情報を登録する必要がある、データベース構築上の大きなコストを占めている。従来、人手によって行われるこのインデキシング作業は、パターン認識などの情報処理技術によって自動化することが考えられている²⁾。情報の分類整理や、情報相互間の関連性抽出は蓄積情報の価値を高めるので、このための「情報組織化」技術が重要である。

(2) 蓄積・検索サブシステム

マルチメディアデータベースは動画、静止画、文書などのマルチメディアデータを蓄積・管理する。将来的には、テラバイト級の容量、数億件の超大容量データを蓄積し、同時に1,000クライアント以上の要求に実時間でこたえる必要がある。巨大データの蓄積にはディスクアレー装置や光ディスクを用いた3次記憶装置が、検索には超並列コンピュータがそれぞれ利用される。

検索サーバは、文献などのテキスト情報に対して自由な単語からの高速な全文検索を可能とする。日立製作所の統合文書情報システム“Bibliotheca”(ビブリオテカ)は、このような要求にこたえることができる³⁾。あいまいな情報からの意味的な知的検索は研究課題である⁴⁾。静

止画や映像、または音声などのマルチメディア情報については、将来的には内容からの「マルチメディア内容検索」も求められる²⁾。

(3) 所在検索サブシステム

電子図書館は世界中に多数構築されていくので、自分の欲しい情報がどの電子図書館にあるかを知ることがまず必要である。ネットワークに結合される所在情報データベースは、どのような情報がどの電子図書館に存在するのかを記録・管理する。

将来は、このようにポインタだけを持った「仮想電子図書館」が専門分野ごとに多数登場し、競い合って「権威」ある書誌情報をサービスしていくようになると考えられる。最新の情報のポインタを常に世界中から集めることも「情報組織化」の一つである。

4 電子図書館プロトタイプシステム“Cornucopia”

4.1 システムイメージ

マルチメディア電子図書館構想の一部を実現するプロトタイプとして提案する“Cornucopia”は、マルチメディア文書管理と全文検索にBibliothecaを用いる。その利用イメージを図2示す。

利用者は、Mosaicクライアントから、Bibliothecaで管理されるファイルの内容を高速な全文検索機能で検索できる。検索条件に合致した文献リストから見たいものを選択すると、その本文がネットワークを介して読み出さ

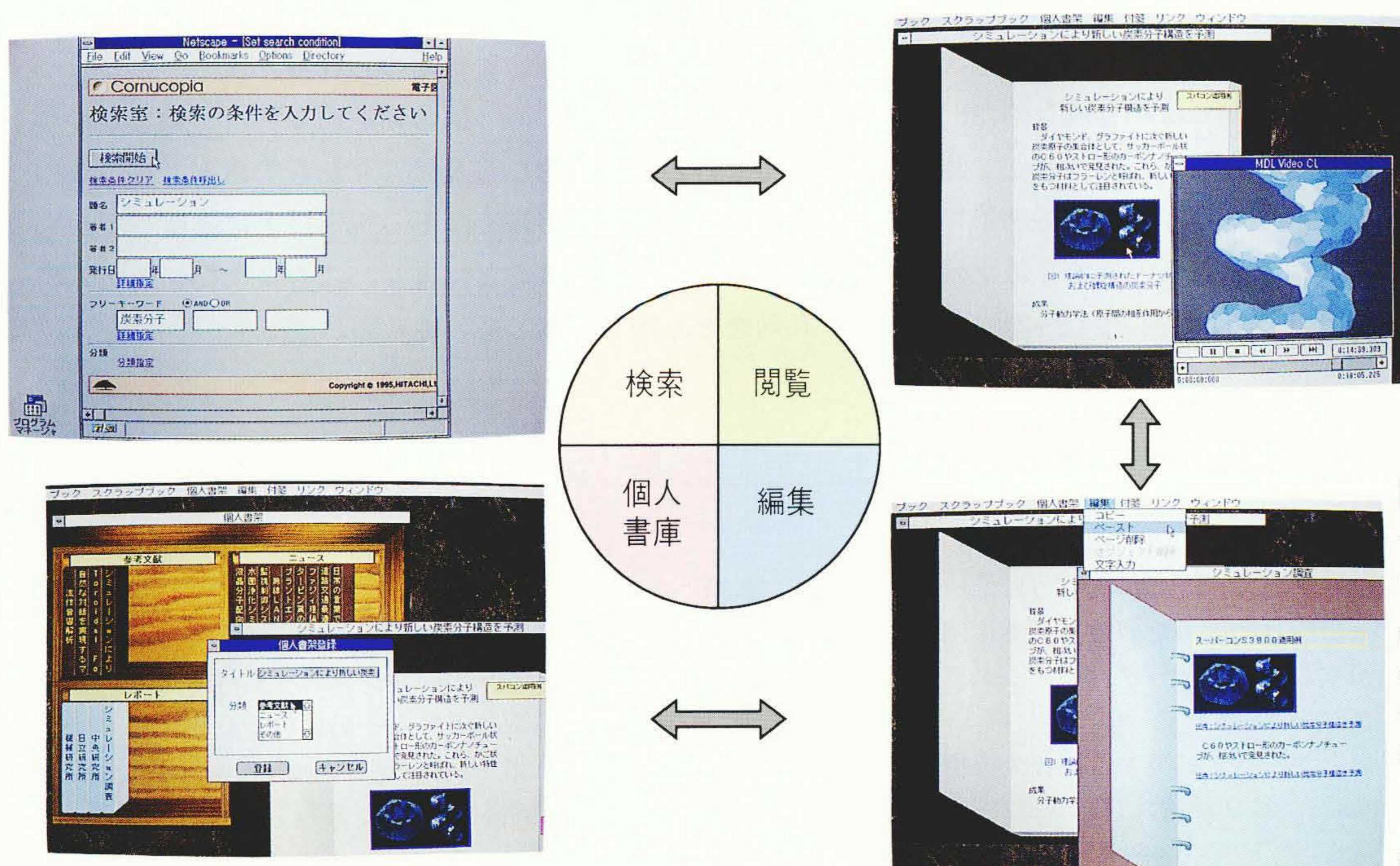


図2 電子図書館システム“Cornucopia”の利用イメージ

MosaicからBibliothecaが管理するマルチメディア文書を検索して本文を表示させることができる。さらに、必要な部分を個人用のスクラップブックに編集したり、これらの文書を個人書庫に格納することができる。

れ、Cornucopiaクライアントによってブックメタファアで表示される。本文中の写真やビデオ映像は、マウスでクリックすることによって別に表示させることができる。

このシステムの一つの特長は、利用者の情報活動の支援機能にある。単なる検索操作の容易性を考えるのではなく、情報の収集、分類、調査、再利用という情報活動全体を支援することをねらっている。利用者は、検索で得られた情報の一部を「ノレッジスクラップブック (Knowledge Scrapbook)」と名づける自分のスクラップブックに仮想的に切り張りすることができる。切り張りは、情報の所在を示すポインタを記録する「トランスクルージョン (Transclusion)」という方法⁵⁾で行うので、見るときに情報課金を行う方式とすることができる。

このスクラップブックは自由に編集することができ、再度開いて見たり、ネットワークを介して別の人に送付することができる。スクラップブックを開くと、システムはポインタから自動的に原情報にアクセスして、あたかもそこに情報があるかのように表示する。また、スクラップブックの一部をクリックして全文情報を表示させることができるので、再度、検索操作する必要がない。

スクラップブックで見ることのできる情報は世界から集めた情報なので、それを格納する書庫は、いわば「仮想パーソナル電子図書館」であると言うことができる。

4.2 文書イメージ情報の検索

Cornucopiaは、文書イメージの全文検索という特徴的な機能も持つ。文書認識によって抽出したテキスト情報を基に全文検索を行い、対応する文書イメージを表示する方法をとる⁶⁾。

文書認識の誤りを完全に無くすことは不可能なので、認識で文字を確定できないときは複数の候補を残して認識結果を格納する。全文検索のときは、この認識結果に残したあいまい性を利用して、認識誤りの影響を最小限にして検索することができる。検索画面例を図3に示す。

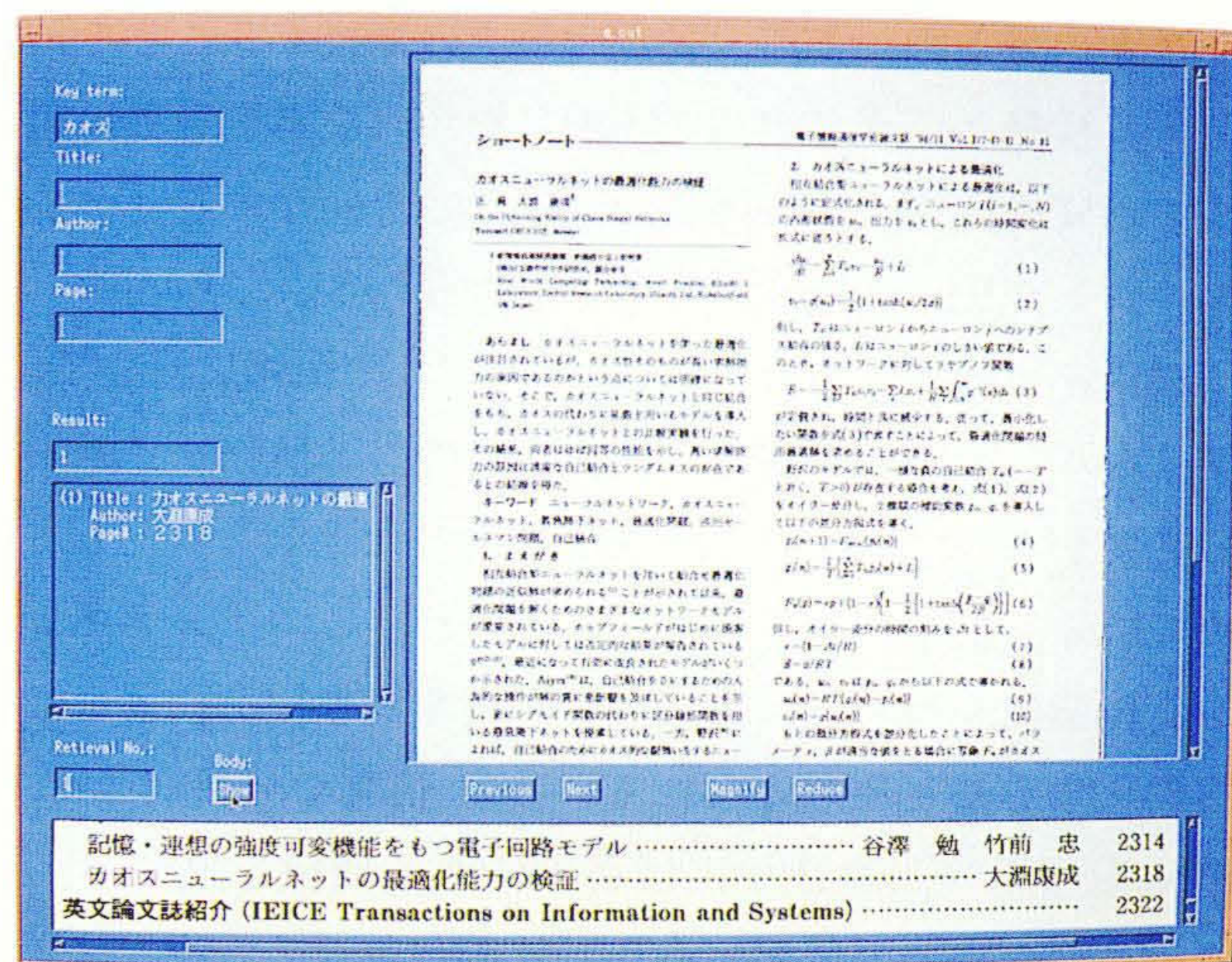


図3 文書イメージ検索の画面例

裏に格納しているテキスト情報を検索することにより、あたかも文書イメージを検索しているかのような感覚で全文検索が行える。

5 おわりに

マルチメディア電子図書館システムは、いわゆる「図書館」という既存の概念を越えた広い意味を持っている。著作権、課金といった課題の解決が求められる一方で、これらにとらわれない情報も多く、企業内利用も含め、世界規模での普及が進展していくものと考えられる。

技術的には、内容からのマルチメディア検索、機械翻訳による多言語検索、エージェントを用いた広域分散情報検索などの検索高度化技術、著作権料決済方式や課金方式、およびプライバシー保護などの体系的な技術課題を解決していくことが求められる。

インターネットでの一部の先進的な利用が始まりつつあるものの、本格的な普及はこれからである。有用かつ価値の高い電子化情報を、ある規模以上に蓄積しながらシステムを構築することにより、このようなシステムの有効性を今後も提案していきたい。

参考文献

- 1) 会津：進化するネットワーク，NTT出版(1994-7)
- 2) 藤澤，外：マルチメディア内容検索のための文書画像属性の自動登録，「電子計算機相互運用データベースシステム」講演予稿集，159-168(平3-11)
- 3) 和歌山，外：クライアントサーバで実現する統合文書情報システム—Bibliotheca—，日立評論，77，5，367～370(平7-5)
- 4) 藤澤：概念知識の体系化のためのメディア・スペース，コンピュータ科学，2，1，39～45(1992-1)
- 5) T. ネルソン：リテラリーマシナーハイパーテキスト原論，竹内・斉藤監訳，アスキー出版(1994-10)
- 6) H. Fujisawa, et al.: Full-Text Search and Document Recognition of Japanese Text, Proc. Symp. Document Analysis and Information Retrieval, 55～80(1995-4)