

「漢籍研究資料ハイパーメディアシステムの構築について」

武田 亘明
札幌ソフトウェア専門学校

各機関などで文献目録等の文字情報中心のデータベース化作業が進んでいるが、漢籍研究資料情報には、文字情報に限らず画像情報、音声情報、映像情報の4つの形式があり、これらの形式の異なる情報も、同様に調査研究の対象としなければならない。

本稿では、研究資料情報の管理と電算化について概観した上で、大規模にデータベース化された研究資料を研究に有効利用する一形態として、研究者による研究資料情報パーソナルハイパーメディアシステムの構築について提案する。また、「杜甫研究資料ハイパーメディアシステム」(プロトタイプ)を例示し、その概要と今後の課題について述べる。

A construction of a hyper media system
on Old-Chinese-Literature resources research

Takeda Nobuaki
Sapporo Software School

We know that some laboratories are advancing characters data-base work for report lists. I think we get intelligence from characters, pictures and sounds and I know these type of communication. So intelligence communicated in different ways is should be researched in the same way.

In this report I will show you how to computerlize management of resources and talk about it's outline. I will propose the way of utilizing computerlized data-base resources and propose how to construct a personal hyper madia system for oneseif.

I will give a prototype, what I call, "Dü-Hü resources research hyper media system" to explain it's image and make clear problems to slove.

1 はじめに

各機関で文献目録等の文字情報中心の大規模なデータベース化作業が進んでいるが、漢籍の研究では、文書等の文字情報に限らず、謡曲等の音声情報や、水墨画や拓本等の画像情報、現地調査ビデオ等の映像情報なども同様に調査研究の対象となる。

文字資料情報は整然と書誌分類されているが音声・画像・映像情報の分類は未整備状態と言わざるを得ない。また、データの標準化についても、漢字コードをはじめ、画像、音声や映像データの標準化も各方面で行なわれはじめている。

一方、コンピュータやその周辺機器の性能は上がり、日本語環境も徐々に整い、価格は低下の傾向にある。この状況を反映して、論文の作成やカードなどの資料管理に、ワードプロセッサを用いる研究者が増えているし、学会などへの論文の提出もワードプロセッサによる文書とすることが増えている。

こうした中で、研究者は、その研究を進める上で大規模にデータベース化された研究資料情報の有効な利用方法を検討し、しかも文字情報以外の形式の研究資料情報を統合管理する研究環境について検討していかなくてはならない。

本稿では、研究資料情報の管理と電算化について概観した上で、大規模にデータベース化された研究資料の研究への有効利用の一形態として、研究者による研究資料情報パーソナルハイパーメディアシステムの構築について提案する。また、「杜甫研究資料ハイパーメディアシステム」(プロトタイプ)を例示し、その概要と今後の課題について述べる。

2 研究資料情報について

研究資料情報を、情報のレベル、情報形式、管理機関、管理体制の4つの視点から整理し、研究資料情報管理形態について検討する。

2.1 研究資料情報の情報レベル

漢籍研究上取り扱う研究資料情報は、その資料の現物への加工や価値付加の度合いによって、垂直的に4段階に分類できる。

(1) 資料の現物で発掘された遺物遺跡や、甲骨・金石・木簡・竹簡・帛、石碑、書物、衣服、建築物、水墨画、壁画などの一次研究資料情報レベル。(2) 一次研究資料情報の写真、スケッチ、ビデオ、図画、拓本などの二次研究資料情報レベル。雅楽などの録音もこれにあたる。(3) 二次研究資料情報をまとめた調査報告、索引や目

録などの三次研究資料情報レベル。(4) 一次・二次研究資料情報を踏まえた注釈、辞書や研究論文、その目録類などの四次研究資料情報レベルである。

2.2 研究資料情報の形式について

漢籍研究上取り扱う研究資料情報の形式は、質的に4つに分類できる。

(1) 甲骨文、金文、漢字などの書籍など文字資料情報形式。(2) 地図、写真、スケッチ、水墨画などの画像資料情報形式。(3) 謡曲などの録音テープやCDなどの音声資料情報形式。(4) 現地調査などで撮影されたビデオなどの映像資料情報形式である。

2.3 研究資料情報の管理機関について

研究資料の管理機関は、その資料情報をどれだけ多くの研究者が参照することができるかによって、平面的に4段階に分類できる。

(1) 書籍などが目録などでその存在が記録されているが、失われて今に伝わっていないものなどの未認知レベル。(2) 個人・一族などが所有して、公開されていない個人管理レベル。

(3) 公機関の図書館・博物館・研究所・大学などの資料室や研究室に整理管理公開されている公開レベルである。

2.4 研究資料情報の管理体制について

漢籍研究上取り扱う研究資料情報の管理については、その題名や抄録にとどめるか、本文まで管理するのかを情報のレベルと管理機関のレベルの組み合わせにより、立体的に5つの段階に分類できる。

(1) 写真集や、論文集、ビデオリスト、図書目録、出版目録など、形式別の研究資料情報の書誌的目録管理レベル。(2) 形式別の研究資料情報を電子メディアに記録した情報形式別目録データベース管理レベル。但し、このとき本文までは含まない。(3) 研究資料情報の本文にまで検索などを可能にする情報形式別フルテキストデータベース管理レベル。(4) 情報形式別フルテキストデータベースを統合し、すべての情報形式の本文にまで検索などを可能にしたハイパーメディアシステム管理レベル。(5) 研究者や研究機関が分野ごとに開発したハイパーメディアシステムを複数リンクし、どこからでも検索や参照を可能にしたインターネットメディアシステム管理レベルである。

各方面で大規模に構築されている研究資料情報のデータベースは、研究資料情報の管理機関単位が主で、論文目録や論文抄録などその包括範囲が広く、内容は目的的に浅い情報形式別目録データベース管理レベルのものが多い。

これらデータベースは、大型計算機センター等に配置されることが多く、一般の研究者が日常的に利用する状態にはなく、大規模データベースを自由に使いこなせる環境にある研究者は少ない。また、研究をはじめるとに当たり関係資料を膨大な資料群の中から抽出するのに大変有効であるが、それ以降に行なう、資料の現物にあたって資料中の関係部分の抜き出し等には活用しにくいと言わざるを得ない(図1)。

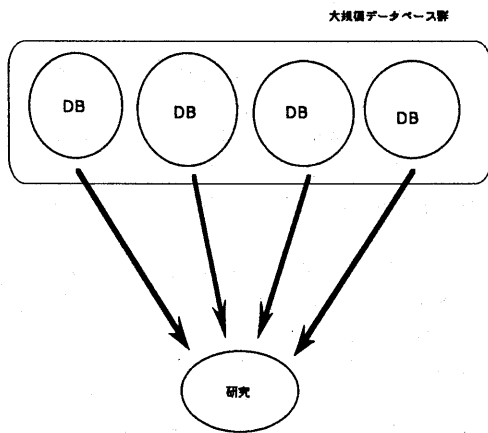


図1 大規模データベースのみによる研究環境

3 研究テーマ別研究資料情報のパーソナルハイパーメディアシステムについて

大規模データベースと研究の間を結ぶ研究環境の一形態として、研究テーマ単位の包括範囲が狭く、内容はフルテキスト的に深い、しかもすべての情報形式に対応したパーソナルハイパーメディアシステム及びそれらをリンクしたインターメディアシステムを構築することを提案する(図2)。このことにより、大規模データベースを研究により有効に活用することができる。と考える。

次に、研究にあたり、研究資料情報を電子メディアと書誌的メディアの両方から入手し、調査分析し、論文作成していく過程を整理する(図3)。

(1) 書誌的・電子メディア(目録等)から必要な研究資料についての情報(存在、管理機関等)を検索し、資料情報リストを作成する(既存データベースの利用)。

(2) リストアップした研究資料情報の現物にあたり、部分または全部を電子メディアに入力変換し、情報スタックを作成する(スタック作成)。

(3) 電子メディア化された研究資料情報スタック等(本文等)を部分とするハイパーメディアシステムを構築する(ハイパーメディアシステムの構築)。

(4) ハイパーメディアシステム中の研究資料情報に対し検索分析処理等を行ない、メモ、図表等の情報を得て、検討を加える。

(5) 研究テーマについて考察する。

(6) メモ・図表をまとめ論文を書く(デスク・トップ・パブリッシング)。

(7) メモ・図表・索引・論文等をハイパーメディアシステムで扱える情報スタックに加える(ハイパーメディアシステムの進化)。

(8) 情報スタックを研究者間で流通・交換しシステムを強化する(システムの強化)。

(9) システムを複数リンクしてインターメディア化をはかる(システムの成長)。

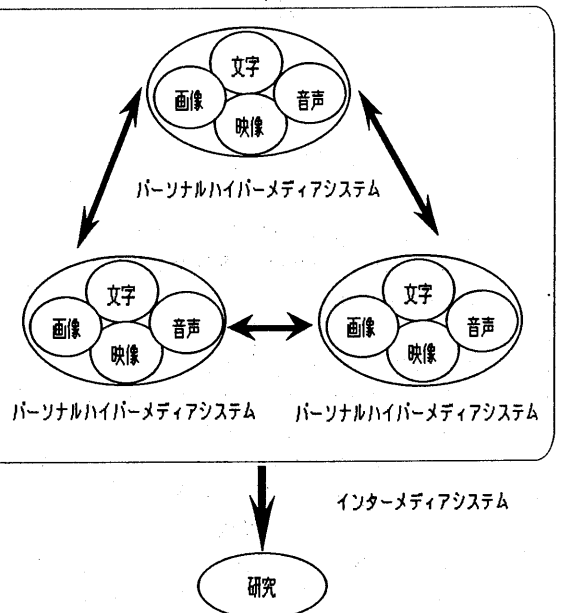
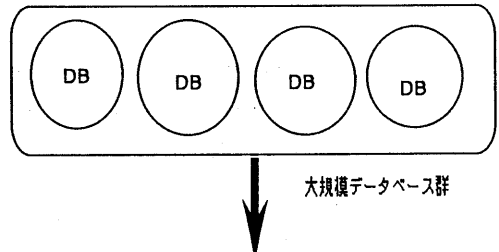


図2 パーソナルハイパーメディアシステムによる研究環境

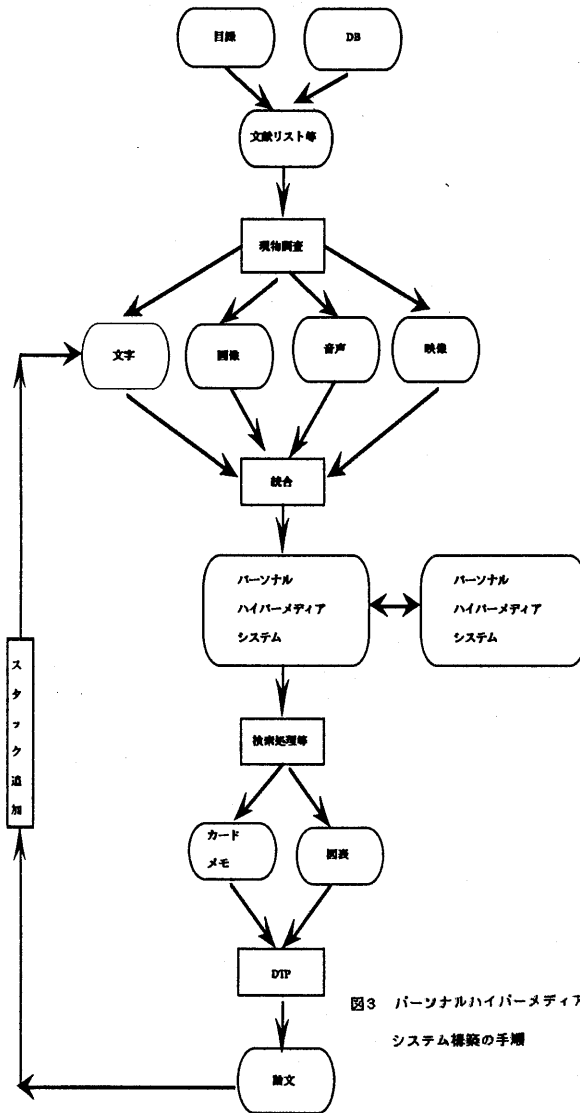


図3 パーソナルハイパーメディアシステム構築の手順

4 研究資料情報の電算管理化について

以下、研究資料情報の電算管理化について、情報形式ごとに検討する。

4.1 文字資料情報の電算管理について

文字資料情報のコンピュータによる管理の方法には、次の4つの方法がある。

(1) 文字資料情報は、書物、木筒、竹筒などのままで、コンピュータはその資料情報の管理場所や所轄等の目録情報を管理する方法。

(2) 文字資料情報を、スキャナなどで画像としてコンピュータの外部記憶装置、例えばハードディスクや、光磁気ディスク等に記憶させておいて、マイクロフィルムのように画像として参照可

能なように管理する方法。甲骨文や金石文等は、この方法で管理するのが現実的である。

(3) 文字資料情報(フルテキスト等)を、コンピュータの外部記憶装置、例えばハードディスクや、光磁気ディスク等に記憶させておいて、そこから目的の資料情報を自動検索し、表示も行なわせる方法。但し、文字と書式を原本と同じではなく、ある標準の文字フォントに置き換えた文字資料情報とする。

(4) (3)の様に外部記憶装置に記憶させておくのだが、文字・書式ともに原本と同様に入力管理する方法。このためには、木版や銅活字、木活字の文字フォントの作成も必要となる。

4.2 画像資料情報の電算管理について

画像資料情報のコンピュータによる管理には次の3つの方法がある。

(1) これまでの管理の方法で、つまり写真のアルバムや、画集、スケッチ集のままで、コンピュータは、その資料情報の管理場所や所轄等の目録情報を管理する方法。この方法だと、画像資料情報を画像資料情報名(文字情報)として扱うことになる。

(2) 画像資料情報そのものを、コンピュータから制御可能な機器、例えばビデオフロッピーディスク等の中に記憶させておいて、そこから目的の資料情報を検索・表示を行なう方法。フロッピーディスクカメラは、2インチのフロッピーに50枚の写真を保存することができ、簡単なプログラムを組むだけで、コンピュータからこのディスクに対して必要な画像を検索することができる。

(3) 画像資料情報をコンピュータの外部記憶装置、例えばハードディスクや、光磁気ディスク等に記憶させておき、そこから目的の資料情報を自動検索し、表示も行なわせる方法である。

また、この画像の記憶装置への取り込み方法は、次の3つの方法がある。(a) スキャナにより、これまで蓄積してきた、写真、図画、写本、スケッチなどを、ハードディスクに取り込む方法。スキャナは、白黒2階調からフルカラーまでスキャン可能である。(b) フロッピーディスクカメラにより撮影された画像をビデオフロッピーディスクドライブを介してハードディスクなどに取り込む方法。(c) 研究資料情報を直接撮影し、その画像をハードディスクなどに取る方法で、ビデオビジュアルライザのステージに資料現物を置き、それをビデオカメラで撮影し、ビデオフロッピーディスクドライブを介して、ハードディスクなどにこの画像を取り込む方法である。

4.3 音声資料情報の電算管理について

音声情報資料のコンピュータ管理の方法には、次の3つの方法がある。

(1) 音声資料情報は、カセットテープ等のままで、コンピュータは、その資料情報の管理場所や所轄等の目録情報を管理する方法である。

(2) 音声資料情報そのものを、コンピュータから操作可能な機器、例えばDAT（デジタル・オーディオ・テープ）やCD（コンパクト・ディスク）等に記憶させておいて、そこから目的の資料情報を自動検索し、再生も行なわせる方法。

(3) FM音源・PCM音源をコンピュータに内蔵させて、オーディオ機器をコンピュータに接続し、専用のアプリケーションソフトウェアを使用して、コンピュータの外部記憶装置、例えばハードディスクや、光磁気ディスク等に音声資料情報をデジタルデータとして記憶し、そこから目的の資料情報を自動検索・再生させる方法である。

4.4 映像資料情報の電算管理について

映像情報資料のコンピュータ管理の方法には、次の3つの方法がある。

(1) 映像資料情報は、ビデオカセットテープや8ミリフィルム等のままで、コンピュータは、その資料情報の管理場所や所轄等目録情報を管理する方法。

(2) 映像資料情報そのものをコンピュータから制御可能なメディア、例えばビデオカセットテープやレーザーディスク等に記憶させておいて、そこから目的の資料情報を自動検索し、再生も行なわせる方法。この制御は専用のアプリケーションソフトウェアを使用すれば簡単に行なうことが可能である。

(3) 映像資料情報をコンピュータの外部記憶装置、例えばハードディスクや、光磁気ディスク等に記憶させておいて、そこから目的の資料情報を自動検索し、再生も行なわせるレベルである。

5 漢籍研究資料情報ハイパーメディアシステム構築について

次に、漢籍研究資料情報ハイパーメディアシステム構築上の課題について検討する。

(1) 開発体制について

研究テーマ別研究資料情報ハイパーメディアシステムの開発は、ある分野・資料情報に詳しい研究者や研究グループ単位で行なう。当面は、情報スタックを研究者間で流通・交換のしくみを整えることが必要である。情報スタックの著作権や編

集権の問題があるので、情報の提供者はその権利を放棄する等の申し合わせ、つまりパブリックドメイン化をはかる等の検討が必要である。研究者は入手した情報スタックを自分のシステムに組み込み、システムを進化させていく。その上で、これらパーソナルに構築されたシステムをネットワーク化し、インターメディアシステム化を目指す。

(2) 資料のコード化について

今日、漢籍の分類はの西洋式分類法によっているが、これを拡張して文字資料情報、画像資料情報、音声資料情報、映像資料情報の総てを包括する分類法を策定する必要がある。学界で広く論議を起し検討をはじめの必要があるだろう。

(3) データ形式の標準化について

日本中国学会で文字種・文字体・文字フォント等について検討作業をはじめたところである（注1）。漢籍研究では、2万字程度の文字種が必要といわれているが、現在JISコードでは6353字しかサポートしていなく版本研究以外の研究でも全く不十分と言わざるを得ない。また、文書の書式・構造に関する標準化については、SGMLなどの標準化の動きがあり、画像情報データは、ハードウェアに依存する部分が多いが、パーソナルコンピュータではTIFF・GIFといった標準データ形式が複数の機種でサポートされるようになってきた。また、音声情報では、MIDIなどの標準化がはかられている。当面は、文字情報の入力に、日本中国学会の動きに合わせてできるだけJIS漢字の範囲にとどめ、必要に応じてグタをはかせるなどして対応していくことが大切であり、画像情報についてはPICTかTIFF形式等でデータ入力しておく、学界などの方向が決まったところで変換するのが現実的と思われる。映像情報については、アプリケーションソフトウェアによってまちまちなので、ソフトウェアをある程度限定して使用する等の検討が必要であろう。

(4) 開発環境について

ハードウェアは、パーソナルに高レベルな機能を活用できる環境が整いつつある。それは、これまで業務用に提供されていた機能が一般向けパーソナルコンピュータやその周辺機器にも付加されるようになってきたことによる。しかし、レーザーディスクやコンパクトディスクが一度しか書き込みできないなど、大記憶容量を必要とする音声情報や映像情報の電算管理は難しい状態である（注2）。

ソフトウェアは、研究者に負担の少ない、でき

るだけプログラミングする必要がない開発ツールを使用することが大切であり、現在、幾つかの一般的なハイパーメディアシステム開発ツールが安価に提供されている(注3)。しかし、これらハイパーメディア開発ツールは、そのほとんどがカラー対応していないので画像情報や映像情報の取り扱い上の問題が残されている。

情報スタックを流通・交換する上で、ハイパーメディアシステム中で一貫性のあるインターフェイスの設計がなされることが大切である。それは交通標識等と同様に、グラフィックユーザーインターフェイスで誰でも同じイメージを持って、使用方法が決められたものにする必要があるだろう。これらについても、学界等で議論して統一をはかることが大切である。

6 「杜甫研究資料ハイパーメディアシステム」について

以下に現在筆者が作成中の「杜甫研究資料ハイパーメディアシステム」の概要について述べる。

6.1 情報スタックについて

(1) 文字情報スタック

杜甫詩集スタックと、注釈スタックを作成する(注4)。文字情報管理レベルは(3)で、フルテキスト入力し、文字はできるだけJIS漢字にとどめ必要に応じてゲタをはかせる。検索は、当初に想定したもの以外の文字をキーとして検索する可能性もあるので、どのような検索文字でも自由にその時々指定できる様にする。各詩から関係注釈の検索をできる様にする。研究資料の現物のイメージを保つためテキストは縦書きとする(図4)。

(2) 画像情報スタック

詩意画スタック、肖像画スタック、地図スタックを作成する。画像情報は、スキャナーからハードディスクに読み込み、ハイパーカードのスタックとして保存する。検索については、文字資料情報に関する画像情報がある場合は、その存在を示すボタンを画面に表示し、必要に応じてボタンをクリックすると関係画像情報を画面に表示する。逆に、画面に表示された画像情報に関する文字情報がある場合は、その存在を示すボタンを画面に表示し、必要に応じてボタンをクリックすると関係文字情報を画面に表示する。

(3) 音声情報スタック

音読スタック、謡曲スタックを作成する(注5)。音声情報は、大記憶容量を必要とするので、本システムでは、音読スタックのみハード

ディスクに記憶させ、謡曲スタックには、謡曲に関する情報を文字で入力することにする。検索については、文字資料情報に関する音声情報がある場合は、その存在を示すボタンを画面に表示し、必要に応じてボタンをクリックするとスピーカから音声を再生するか、謡曲に関する情報を画面に表示する。

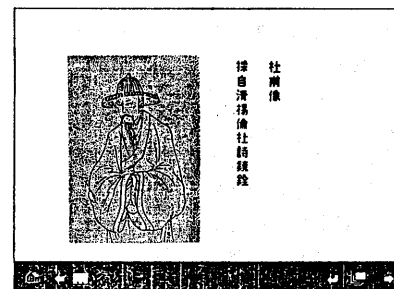
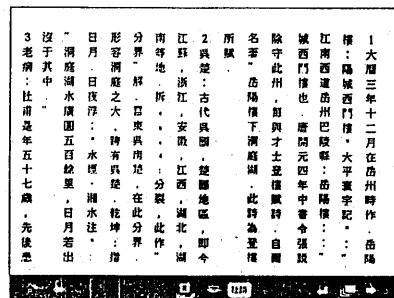
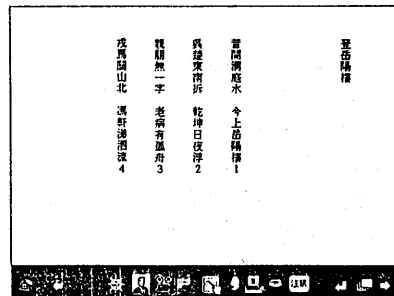
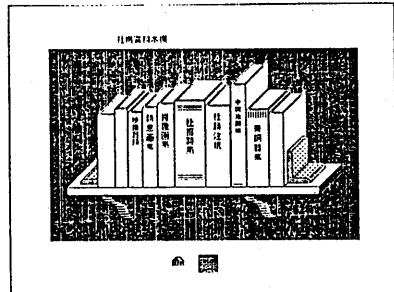


図4 スタック画面

(4) 映像情報スタック

文字資料情報に係る映像情報がある場合は、その存在を示すボタンを画面に表示し、必要に応じてボタンをクリックすると映像情報名、管理場所などのリストを表示する。映像情報は音声情報と同様に大記憶容量を必要とするので、本システムでは、映像情報に関する文字情報を画面に表示するにとどめる。

6.2 システム構成について

(1) 情報スタック構成 (図5)

文字情報 (杜甫詩集スタック、注釈スタック)

画像情報 (詩意画スタック、肖像スタック、地図スタック)

音声情報 (音声スタック、音声情報スタック)

映像情報 (映像情報スタック)

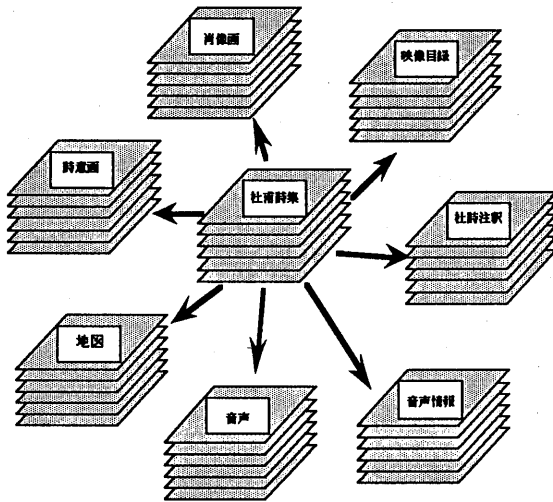


図5 スタック構成

(2) ソフトウェア構成

System v6.07/Hyper Card v2.0/Hyper Scanner・Apple Scan/Hyper Sound

(3) ハードウェア構成 (図6)

Macintosh II (5MB) / 高解像度12inch Color Monitor/Image Scanner/Hard disk (40MB) / Mac Recorder/LaserWriterII

6.3 今後の課題について

(1) 情報スタックの追加について、例えば、ある字句に典拠がある場合は、その出典の文書スタック (「通鑑スタック」等) を追加していく。

(2) 文字情報については、今後日本中国学会の文字等の標準化に合わせて、現在の文字の置き換えを行なっていく。

(3) 画像情報については、モノクロ画像は問題ないが、カラー画像はスキャンはできるが、「ハイパーカード2.0」が、カラー対応していないので、別のアプリケーションを起動させて表示しなくてはならない。ハイパーカードのバージョンアップを持って、画像情報のハイパーカードデータへの置き換えを検討する。

(4) 音声情報と映像情報は、データの圧縮記憶方式、記憶媒体の大記憶化と価格の低下の動向等をみて完全な電子メディア化を検討する。

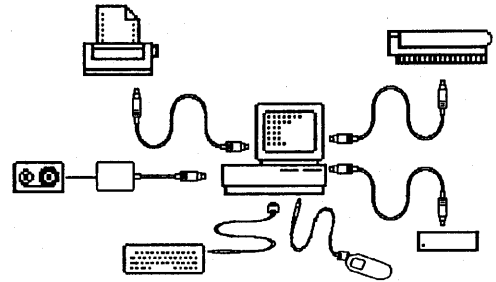


図6 ハードウェア構成

7 おわりに

各方面の機関や研究グループは組織的に研究資料のデータベース化に取り組んでいるところであるが、研究者自身が、情報形式の異なる研究資料情報を統合管理し、大規模データベースを研究活動に有効利用する一形態として、大規模データベースシステムと研究の間をつなぐ中間的システム、すなわち研究資料情報パーソナルハイパーメディアシステムの構築を提案した。

研究資料情報パーソナルハイパーメディアシステムは、今後2つの方向へと発展していくと考えられる。1つは、インターネットシステムとしての大規模化の道であり、もう1つは、教育システムとしてのパッケージ化への道である。

研究者の専門的知識を踏まえた研究資料情報システムのパッケージ化が行なわれれば、教育情報の充実にも寄与することができると考えられる。本提案が人文科学研究者の研究環境の改善のための議論と教育情報システムの発展の一助となれば幸いである。

最後に、本研究に当たり北海道教育大学後藤秋正助教授、フェリス学院大学末岡實助教授にご教示いただいた。また、システムソフト販売株式会社札幌支店蛭名信英氏、札幌ソフトウェア専門学校吉村齋主幹講師、小野哲雄主幹講師には貴重なご助言をいただいた。ここに記して感謝の意を表したい。

(注)

- 1 東洋学文献類目データベース(京都大学)、インド学仏教学論文データベース(日本印度学仏教学会)、国文学研究論文データベース(国文学研究資料館)など研究資料情報のデータベース化は、各方面で行なわれている。
- 2 映像は8ビットカラー環境で約30分で300MB、音声をCDと同じ音質でサンプリングすると1分で約10MBの記憶容量を必要とする。
- 3 "ハイパーカード"(アップルコンピュータ)は、メーカーとライセンス契約することで無償で提供されている。
- 4 注釈は、「讀杜詩説」(清施鴻保著)、「杜甫選集」(森石樞他選注)を入力中である。また「杜甫草堂詩注」(李誼注釈)の入力も検討している。
- 5 音読は、本文の現代中国語の発音を記憶させている。

(参考文献)

- 1 「ハイパーメディア」矢田光治編、日刊工業新聞社(1990)
- 2 「中国書データベース化についての答申」日本中国学会報第四十二集(1990)
- 3 「漢籍の電算処理について—「公孫龍子」一次索引の作成を中心として—」武田亘明、北海道中国哲学会「中国哲学」第17号(1988)
- 4 「Multimedia Tools」宮本学著、HJB出版局(1990)