

国文学原文献資料データベース システムの開発

安 永 尚 志

あらまし

本報告は、現在国文学研究資料館において進められている原文献資料データベースの開発報告である。すなわち、開発の目的、概要、研究経緯、研究成果、及び今後の見通し等をまとめたものである。原本等のいわゆる0次情報は、情報処理システムにおいては画像としての情報形態で取り扱うことが、最も適切な方式である。ここではマイクロフィルム資料を原本資料と見なし、これを検討対象とする。本文では、国文学研究に必要な画像データの入力、管理、利用方法について述べる。また、開発の目標は幾つかあるが、主として、オンライン情報検索環境下で所望の本を探し、請求し、かつ入手することを可能とするシステム、即ち原文献資料流通システム（試行版）について報告する。

なお、本研究は主に文部省科学研究費補助金によっている。

キーワード：原文献資料データベース、画像データベース、光ディスク、オンラインデータベース、原文献資料情報、0次情報、国文学学術情報、データベースの横断利用、データベース、マルチメディアデータベース、原資料流通システム

1. まえがき

国文学原文献資料データベースシステム（以下、原資料システムという）は、

0次情報である 原本等を直接電子 ファイル化した データベース システムである。蓄積媒体は、直径 30 cm 光ディスクを用いる。両面で約 2.5 ギガバイトの容量を持ち、本システムでは1枚の光ディスクに約2万枚（またはフィルムコマ数）の原文献資料を蓄積することができる。

原資料システムの開発のねらいは大別して2つある。1つは、貴重な文献資料の新しい蓄積技術の開発である。また1つはその高次活用への対応である。

現在、古文書等の保存はマイクロフィルムを主流とする。国文学研究資料館においても、文献資料は 35 mm 無孔ロールフィルムによる蓄積を柱としている。現在の技術水準では、長期保存に耐え得る蓄積媒体としては、マイクロフィルム等写真による方法が優れていることは論を待たない。

しかし、作成の手間や技術的な問題、化学的条件の設定の困難性、収納スペースの問題、その厳格な保存環境条件の設定、あるいは取扱いの容易性等において、やや利便性に欠く。また、35 mm 無孔ロールフィルムは現時点ではいわゆる標準製品規格ではないため、関連諸技術の進展はなく、さらに既存製品等の保存性にも問題がある。また、いずれ関連機器の欠乏を招くことが予想される。

このような点を考慮して、将来的な文献資料保存技術を各種検討する必要がある。光ディスクによる手法は、その内の1方法に過ぎない。ここでは、この問題は主たる課題とはしないが、有効性の検討を行っておく必要がある。

続いての課題は、国文学研究のための原文献資料の高度活用、利用支援システムの開発である。例えば、国文学研究資料館所蔵“徒然草”“方丈記”“伊勢物語”等の全異本（約 260 点、3 万ページ弱）を標準化し、光ディスクに蓄積した。これは、諸本・異本の比較研究等に際して、その場で即座に比較参照する事を可能にする。また、“松尾芭蕉”及び“井原西鶴”の代表作品等を、異本を含め網羅的に蓄積した（約 180 点）。これから、作家・作品論等を展開するに際して、その場で即座に比較検討する事が容易になる。

国文学原文献資料データベースシステムの開発 (安永)

一方、遠隔地の利用者は、文献資料目録システムから所望の本を知り、本データベースから直接ファクシミリを通して、本の複製を入手することが出来る。このシステムを原資料流通システムと呼んでいる。所蔵されているマイクロフィルム資料を含む文献資料目録データベースと、この原文献資料データベースは、その本の請求記号等のユニークな名前によりリンクされている。すなわち、オンライン情報検索環境下で、本を探し、請求し、かつ入手することを可能としている。この原資料流通システムを構築することを主たる目標としている。

原資料流通システムの先駆的研究は、猪瀬 (1) (2) による研究があり、光ディスクの有効性の実証、及び通信方式等の提案を行っている。本研究は、その成果を継承発展させたもので、いわば実用システムの構築に焦点を当てている。

原資料流通システムに限らないが、一般に画像データベースで問題なのは、そのデータベースの形成法である。現在、画像情報の入力は、原本のマイクロフィルム資料の紙焼きコピーに前処理を施し、直接光ディスクに入力、蓄積している。このための標準作業手順を設定しているが、これは手作業による複雑な工程を必要とする。原文献資料入力手法の標準化について、参考文献 (3) に 1 例をまとめている。

そこで、複合画像システムを研究、開発し、ホストコンピュータとチャンネル接続した。これは、国文学研究資料館独自の 35 mm 無孔ロールフィルムから、直接かつ自動的に入力、蓄積することが出来るシステムである。蓄積は、システム専用の 20 cm 光ディスクに行う。また、A3 版拡大投影映像で、16 本/mm の解像度を持つ。なお、データシートによる高速入力機能も合わせ持っている。

将来的には、入力だけではなく、検索を行いまた伝送を引き受けるシステムとしての機能拡張が考慮されている。

原資料流通システムは、5つのサブシステムから構成されている。すなわち、原文献資料の入力、蓄積、検索・同定、提示及び伝送サブシステムである。以下、これらについて概要をまとめる。なお、詳細なシステムの評価については割愛する。

2. 原資料システム開発基本計画

2.1 原資料システムの開発目標

原資料システムは、実用化を目指したプロトタイプシステムとする。実用化システムとは、国文学原文献資料（0次情報）の入力、蓄積、検索、提示、転送等が実用レベルで実現できるシステムとする。これを原資料流通システムという。

実用レベルとは、システム全体の運用業務としてルーチンワーク化が可能なこと、作業効率が高く、かつ簡単であること、保守性に優れること等を保障することである。さらに、使い勝手の良いユーザインタフェース、高度利用のための柔軟な処理機能の提供等が不可欠である。一般的には、画像データベースの基礎技術は完成されているが、実用性を達成するためには、実際の運用環境の中で試行を重ね、評価することが必要である。

この意味で、本計画のシステム化はまずプロトタイプシステムを完成し、これを実際に試行運用し、運用実験を通じて、実用システムを完成する工程計画が必要である。

このような要件を考慮して、原資料流通システムは下記の5つのシステムから構成するものとする。図1にシステム構成、関連システム構成、関連を示す。また、図2は全体的なハードウェアシステム構成を示す。

- ① 原文献資料入力システム
- ② 原文献資料蓄積システム
- ③ 原文献資料検索・同定システム

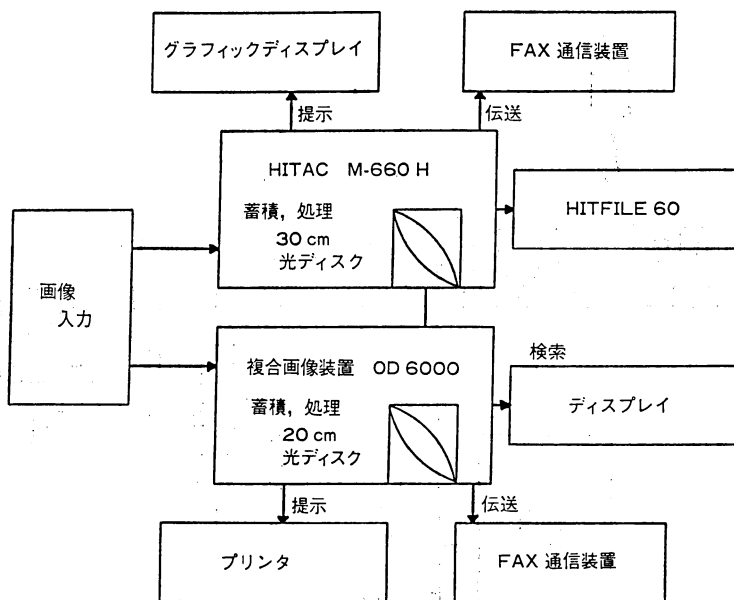


図 1. システム関連図

④ 原典資料提示システム

⑤ 原典資料伝送システム

2.2 前提条件の整備・検討課題

2.2.1 入力データ

入力する原典資料データ (以下、画像データともいう) は、少なくとも数作品について全諸本 (当面、当館蔵) が蓄積されていることが必要である。さらに、特定の作家についての全作品を蓄積する必要がある。これは、実際に国文学者が利用し、評価することを前提にしている。

2.2.2 入力方法・要件

入力方法の検討のためには、まず画像の記録媒体の検討が必要である。

文献資料の形態は、当然ながら原本としての和綴じ本である。本を直接入力

画像処理システム (Graphica M-1006)

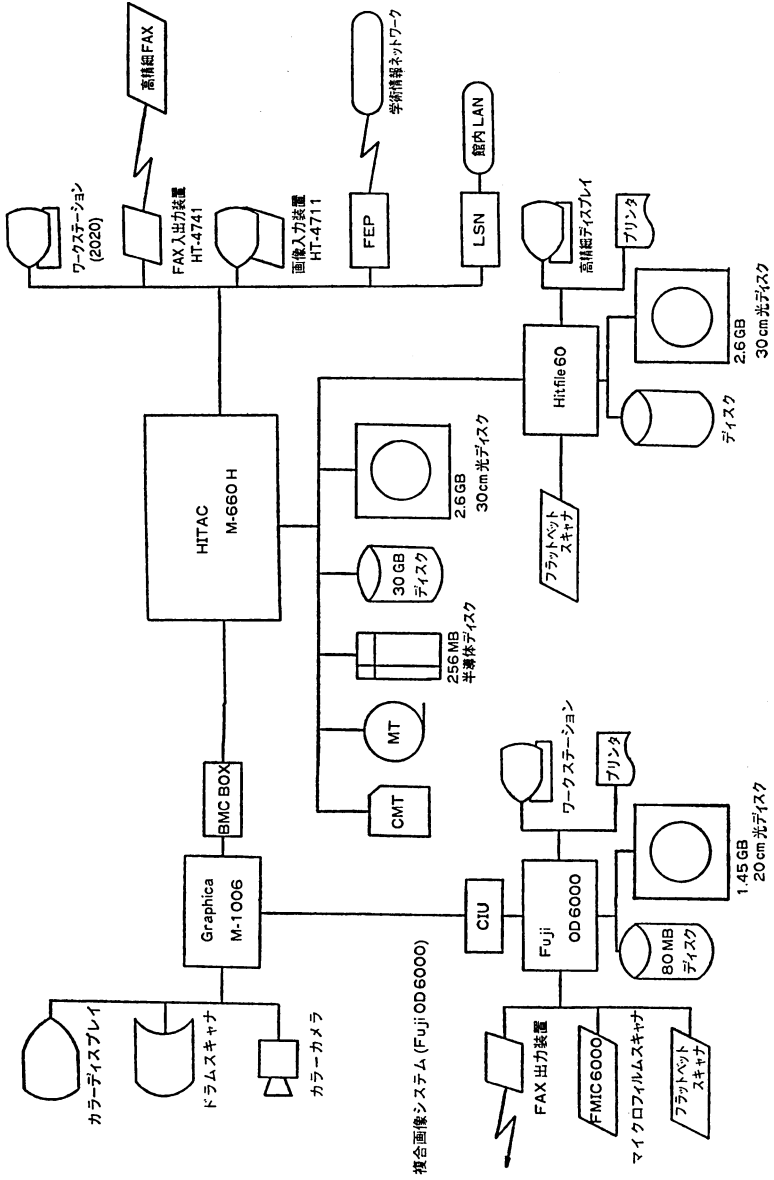


図 2. システム構成 (国文学研究資料館情報処理システム)

国文学原典資料データベースシステムの開発 (安永)

することは出来ないから、ゼロックス等の直接コピーによる。しかし、この方法はあまり意味がない。本の大きさ、形態、あるいは品質等が異なりすぎる。

従って、写真に依らざるを得ない。国文学研究資料館では、マイクロフィルムとして 35 mm 無孔ロールフィルムを用いている。この点を考慮すれば、入力媒体は 35 mm 無孔ロールフィルムを、当面の標準入力媒体として考えるのが現実的である。これにより、入力データの品質が画一化され、標準化、規格化が可能となる。

このとき、2つの方法が考えられる。1つは、フィルムから紙にコピーした(以下、紙焼写真という)による間接的方法である。他の1つは、フィルムから(ネガ、ポジどちらかまたは両方)直接入力する方法である。

紙焼写真は、また2つの画像形態がある。1つは専用のマイクロリーダープリンタによるコピーであり、他の1つはいわゆる写真である。これらは、画像としての品質が異なる。

プロトタイプシステムでは、上記3通りの方法を実験、評価する。なお、品質の設定基準は、当面の間中間調の色情報を除去した完全な白黒画像とする。すなわち、本として読めることを前提とする。また、入力方法は以下の検討を行う必要がある。

- ① マイクロリーダープリンタによるコピーの直接入力 (外注方式も検討)
- ② 紙焼写真の直接入力 (外注方式も検討)
- ③ マイクロフィルムからの直接入力
- ④ 入力用データの標準化、及びその作業手順の規格化
- ⑤ 入力、蓄積工程におけるデータ校正手法・作業 (分析、評価が不可欠)
- ⑥ 入力システムの検討、マンマシンインタフェースの検討
- ⑦ リモート入力方式の検討 (学術情報センター等の高速入力機器の活用実験)
- ⑧ 複合画像システムの設計と制作 (制作は外注)、これによる入力実験

なお、蓄積媒体としての光ディスク関連技術の詳細な検討が必要である。と

くに、光ディスク機器による入力のしやすさ、外注化の可能性等を評価する必要がある。さらに、35 mm 無孔ロールフィルムは、非標準の機器であるため、⑧の複合画像システムを構築することは極めて重要である。

また、入力されたデータは画像校正システムで、校正することが可能でなければならない。これには、標準的なイメージプロセッシングとしての画像拡大、縮小、回転、移動等による画像標準化処理、あるいはノイズ除去、切り出し、埋め込み、回転等による画像編集等の機能も含む。

2.2.3 蓄積法

画像データの効率良い蓄積法の確立、並びにシステム運用のしやすさ等を検討する必要がある。同時アクティブ可能な光ディスクドライブは、多いほど良いが、コンピュータ性能の限界がある。現在、8枚の光ディスク（片面）が実験可能である。将来、光ディスクの自動倉庫システム等の検討が望まれる。

また、蓄積に際してのデータ圧縮化の可否の検討、虫喰い跡の除去等のフィルタリング技術の検討等、情報理論的な研究が必要である。

画像データベースは、1レコード中にアクセス情報を全て含むように設計する場合と、ディレクトリ情報を別データベースやファイルとする設計が考えられる。

本研究では、後者の方法を採用。その理由は、この画像データベースの対象が本である文献資料であり、これを同定するための情報は、いわゆる書誌情報、目録情報、所在情報等多くのデータ項目が必要であるためである。また、将来的には多様なキーワード等を考慮すれば、これらのアクセス情報は、文献資料目録データベースとして別途定義する方がよい。これは、現在稼働中のオンライン目録データベースに必要なリンク情報を加えることで実現できる。一方、画像データベースにもそのリンクに必要な最低限の情報項目を定義する。

この仕様により、蓄積システムの構築が極めて容易となる。また、単純に画像データベースのみの蓄積に専念することが出来る。

国文学原典資料データベースシステムの開発 (安永)

なお、画像データベースの構成には(株)日立製作所製 **FIGURE** を用いる。

以下では、画像データベースは、直接原典資料データベースをいうものとする。

2.2.4 検索法

蓄積法を単純化したため、検索手法は若干複雑となる。しかし、これは、システム内部の問題であって、利用者インタフェースに関わるものではない。すなわち、利用者からみた場合、現在国文研でオンラインサービス中の所蔵文献資料目録データベース利用と、全く同一の利用環境とする。

検索手法は、2次情報から必要な本を知り、0次情報とリンクして、そのコピーをとるというフローを実現する必要がある。すなわち、データベース間の渡りを実現することである。実用性の配慮から、情報検索システムには、(株)日立製作所製 **ORION** を用いる。

また、各種の情報処理機能はイメージプロセッシングを含めて、当面ホスト型としない。一応、ローカルな端末の機能とする。そのためには、大量情報の高速転送機能が必要である。

なお、リモート端末から2次情報検索により、原典資料を同定しその複写を得ることが必要で、このため実用的リモート端末を構成する必要がある。当面、ハードウェアとしては、検索専用の端末(パソコン)とファクシミリに依るものとする。

通信アーキテクチャは、公衆網による無手順系と、大学間コンピュータネットワークによるものとする。

2.2.5 提示法

提示法では、広く情報処理サービスを考慮する。利用者によって、このデータベースの使い方は異なる。個々の利用者が自由にかつ柔軟にデータベースを活用し、得た結果の処理が出来なければならない。しかし、これをシステムとして用意することは不可能に近い。そこで、これを支援すべき範囲を見極め必

要な支援ツールを用意する。

すなわち、サービスの範囲として、まず定型的処理と非定型処理の内容を分析し、その対応法を検討する必要がある。

一般的な提示法として、以下のようなステップが考えられる。

① ステップ 1: 最も単純な場合

- ・離れた端末から、望みの本を知る。望みの本の必要なページを読み、その複写をとる。
- ・複写が手元に届く。

② ステップ 2: 少し高度な場合

- ・ローカル端末で、ある本の異本の探索等の処理を可能とする。同一の画面で、2つの文献資料の比較検討等が出来る。

③ ステップ 3: 最も高度な場合

- ・ローカル端末で、ある本間の必要なページのマージを行ったり、ページから挿絵等の切り出しを行う。一般的には、以下のような要求がある。
- ・画像の拡大、縮小、回転等
- ・画像の重ね合わせ、切り出し、編集機能等
- ・文章の取り出し、挿絵の取り出し、利用処理
- ・虫喰い等のフィルタリング等による除去
- ・うらうつりの除去
- ・コントラストの強調
- ・中間（灰色）調の除去または強調、取り出し
- ・朱書きの処理
- ・1次情報（フルテキスト）との比較による翻刻等。

また、リモート端末においても通信容量が許せば、上記③の問題解決が望まれている。しかし、当面簡易端末で可能な画像情報処理用システムが不可欠である。

2.2.6 伝送法

伝送法は、効率の良い通信アーキテクチャの検討である。また、画像データは大容量となることから、データ圧縮技術の開発が極めて重要である。

伝送では、いわゆる検索システムの通信処理部分と、画像データベースそのものの伝送部分を別途効率よく構成する必要がある。当面、ミックスモード通信は考慮しない。

3. 原典資料流通システムの構成 (プロトタイプシステム)

3.1 原典資料データベースの構成

原典資料流通システムにおけるデータベースは、2.2 で述べたように、原本 (マイクロフィルム資料を含む) である原典資料データを格納している画像データベースと、その画像データベースにアクセスするための書誌情報等を格納した、文献資料目録データベースから構成される。また、前者は0次情報データベースであり、後者は2次情報データベースである。

理論的には、文献資料目録データベースは、画像データベースに対する属性情報を持った、指示あるいは特性データベースである。将来的には、画像の多様な特性情報を持ち、多様なかつ柔軟な検索を可能とする機能を持っている。これを、インデックスデータベースともいう。今回は、単純な書誌情報のみを持つ。

3.1.1 画像データベース

原典資料データベースである画像データベースは、原本単位に蓄積した画像データの本体部分 (ボディという) と、これへのアクセスキー (画像 ID という) を持つデータベースである。

画像データベースのデータ構造は、単純で以下のように定義される。すなわち、画像データベースは1冊単位の本の集合である。個々の本単位のデータベースを、画像データファイルという。さらに、その本の特定のページの1枚単

位の情報を画像データという。マイクロフィルム資料を対象とするときは、1 エレメントである画像データは、原則としてフィルムの1コマ、すなわち本の見開きを1単位とする。これを、ページという。

(1) ページの構成

ページの物理的な構成は、参考文献(3)に譲る。イメージ情報として、ページが持つ情報は、ボディのファイルへのリンク情報、ボディの位置決め情報(位置マーク)、及びボディが持つ属性情報である。リンク情報は、ファイルのIDである(本の請求記号)。位置情報は、イメージ処理に必要なボディの画像としての境界、中央等のドメインを与える。イメージとしての歪も考慮する。

また、属性情報は、裏映り、虫喰い、朱書き、汚れ、挿絵等の有無をマークし、ボディが持つ本来の情報の欠落を示している。すなわち、可能なかぎり潜在情報を示すものとする。これにより、原本が持つ多様な情報を極力保持し、オンラインシステムだけで研究を進めることが出来る。同時に、必要ならば実際に原本にあたることを容易にしている。

(2) データファイルの構成

データファイルの物理的構成は、基本的にはマイクロフィルムの記録方式と同等である。ただし、1冊の本として完結する。なお、標準の書誌情報を第1ページに付し、データファイルの性格を標準化している。データベースのアクセス単位は、データファイル単位である。従って、データファイルにはアクセス情報が必要である。これらは、文字情報により各データファイルにヘッダ的に書き込まれている。

(3) データベースの構成

データファイルが集まってデータベースを作る。将来、文献資料以外の異なるデータベースの蓄積も予想され、また文献資料でも分類の必要等のためから、データベースを区別するアドレス情報が必要である。これを、ボリューム情報ともいう。

国文学原文献資料データベースシステムの開発 (安永)

画像データベースは、現在人手により標準画像データシートを作成し、入力している。また、画像データ管理ソフトウェア (FIGURE) によって管理している。入力された画像データは、MH 圧縮を施し画像データボディとして蓄積する。

蓄積媒体は、直径 30 cm の光学ディスク (両面書き込み可能、片面当たり 1.31 ギガバイトの記憶容量を持つ) を使用している。画像データのデータ量は、ページ当たり (マイクロフィルムで1コマ分) おおよそ 100~150 キロバイトである。したがって、光ディスク片面では、1万枚ないし1万1千枚蓄積することが出来る。

入力精度は、標準入力用画像データサイズを B4 版とし、このサイズで 8 本/mm である。機械精度としては、16 本/mm が可能である。

3.1.2 画像データディレクトリファイル

FIGURE は、画像データを実時間で処理するために、VSAM 形式のディレクトリファイルを使用する。このディレクトリファイルには、各画像データファイルに割り当てられた識別名称 (ファイル ID という) 等を持ち、また画像データファイルをオープンしたときに、そのなかの画像 ID 等を参照するためのディレクトリ等を持つ。

通常、1 つのディレクトリファイルと複数個の画像データファイルによって、画像データベースが構成されている。

3.1.3 文献資料目録データベース

原文献資料 (マイクロフィルム資料等を含む) の書誌情報と、画像データベースへのアクセス情報を持ったデータベースである。システムは、現在オンラインサービス中のマイクロフィルム資料目録データベースの情報検索システムと基本的に同等である。

利用者は、このデータベースを検索することによって、その書誌に対応する画像データへのアクセス情報を取得できる。このデータベースは、ORION 情

報検索システムによって管理している。

なお、文献資料目録データベースは、利用者の便宜を考慮して、書名ソーラスファイルをもつ。すなわち、利用者が入力した任意の書名を、統一書名に変換し、これを検索キイとすることにより、利用者の負担を軽減している。

3.2 原資料流通システムの構成（プロトタイプシステム）

原資料流通システムは、2.1 で述べたように、5 つのサブシステムで構成される。ここでは、プロトタイプシステムの試行版システムとして、2 つのサブシステムに簡単に機能をまとめて述べる。なお、この試行版原資料流通システムは、画像データの対象として、マイクロフィルム資料に限定している。

3.2.1 原資料データベース形成サブシステム

原資料データベース形成サブシステム（形成サブシステムという）は、画像データベースを形成するフェーズ（画像 DB 形成フェーズという）と、文献資料目録データベースを形成するフェーズ（目録 DB 形成フェーズという）から構成される。

通常、画像 DB 形成フェーズを最初に実行し、次に目録 DB 形成フェーズを実行することにより、原文献資料データベースが形成される。各フェーズの機能関連、作業フロー図を、それぞれ図 3(a)、3(b) に示す。

(1) 画像 DB 形成フェーズ

①画像データベース入力機能

マイクロフィルム資料のコピーから、手作業によって作成した入力用標準データシート（画像データ）を、画像処理装置 HITFILE 60（(株)日立製作所製）により、光ディスクに入力・蓄積する。入力方法は、大変簡単で通常の乾式複写機と同じ要領である。

また、このとき画像データへのアクセス情報として、画像データに対応するマイクロフィルム資料の請求記号と、マイクロフィルム資料内のページ番号も入力する。

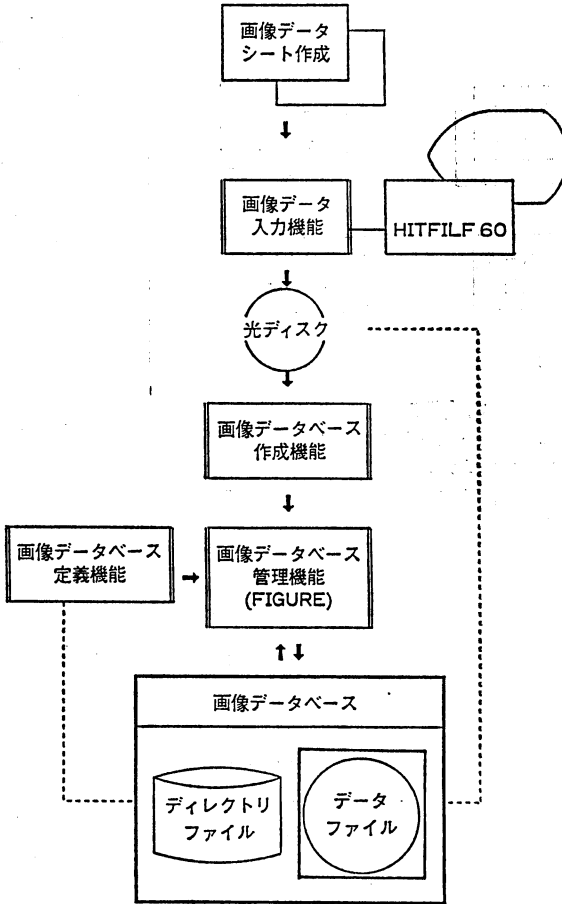


図 3. a 画像 DB 形成フェーズ

実際の入力は外注方式を採っているが、国文学研究資料館の情報処理システムにおいて実施可能である。

なお、入力データシートの形成手順については割愛する。

②画像データベース定義機能

画像データベースのディレクトリファイルの初期的な割り当て、FIGURE

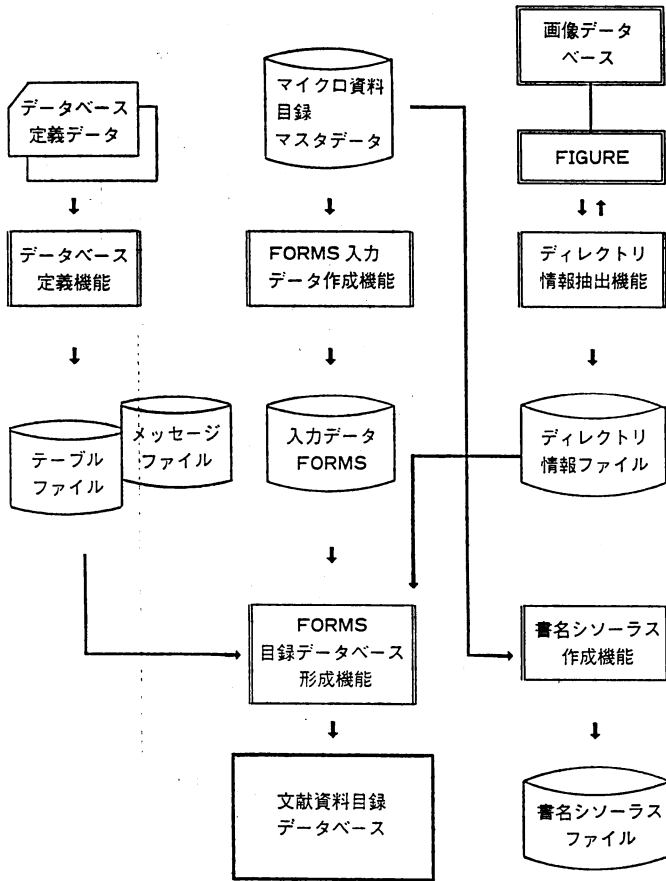


図 3. b 目録 DB 形成フェーズ

ファイル名/ディレクトリファイル名の割り当てを行う。すなわち、FIGURE によって管理される画像データベースの構造を定義する。

③画像データベース作成機能

画像データベース入力機能を使用して作成された画像データファイルに、ファイル名等を割り当て、FIGURE の管理下に置く。画像データファイルに割

り当てた管理情報は、画像データディレクトリファイルに格納される。

(2) 目録 DB 形成フェーズ

①目録データベース定義機能

ORION のデータ定義言語を使用して、マイクロフィルム資料目録データベースの構造を定義する。定義された目録データベースの構造情報は、テーブルファイルやメッセージファイルとして格納され、データベース検索時点で参照される。

②ディレクトリ情報抽出機能

画像データベースから、画像データファイルに関するアクセス情報を抽出する。抽出されるアクセス情報は、画像データとして入力されたマイクロフィルム資料単位に、ページ、画像 ID、ファイル ID、ドメイン名、マイクロフィルム資料請求記号、コマ数等属性情報である。

③情報検索用入力データ (マスタデータファイル) 作成機能

マイクロフィルム資料目録データベースにある各書誌情報に、抽出したアクセス情報を結合して、情報検索システム用入力データファイルを作成する。これを、マスタデータファイルという。情報検索システムは、ORION を用いる。また、ORION への入力データファイルを作成するシステムを、FORMS という。FORMS は、一定形式の入力データから、ORION 形式のデータベースを作成する機能を持つ。

④目録データベース作成機能 (FORMS)

マスタデータファイルから、最終的に利用者が使用する文献資料目録データベースを作成する。利用者インタフェースは、現在のマイクロ資料目録データベースのオンライン検索システムと同等である。

⑤書名ソーラスファイル作成機能

マイクロフィルム資料目録データベースから、記載書名と統一書名の対応関係を抽出して、VSAM 形式の書名ソーラスファイルを作成する。

以上のデータファイル群に、情報検索システムが標準的に使用する各種ファイル群を組織化して、情報検索システムが構成される。これらには、データベースの構造情報を含むテーブルファイル、検索時の対話メッセージを含むメッセージファイル、目録データベースと画像データベースとのインタフェースであるコミュニケーションファイル等がある。

なお、具体的な情報検索システムである原資料データベース検索サブシステムの構成法については割愛する。

3.2.2 原資料データベース検索サブシステム

原資料データベース検索サブシステム（検索サブシステムという）は、文献資料の書誌情報を検索するフェーズ（検索フェーズという）と、検索された書誌情報に対応する画像データファイルを抽出するフェーズ（画像抽出フェーズという）から構成される。

また、画像抽出フェーズは、画像データを画像端末に表示するか、またはファクシミリ装置に転送するかの2つの機能を持つ。前者を、画像表示フェーズ、後者を画像転送フェーズという。図4に、検索サブシステムの構成を示す。

検索の実行は、表1に定義したような対話型検索コマンドを用いる。

(1) 検索フェーズ

①対話コマンド解析機能

TSS 端末から入力された対話コマンドを解析して、対応する機能を実行する。コマンド解析は、コマンドのタイプを表わすニーモニック部分と、詳細な機能を指定するパラメタ部分について行う。

コマンドにより、さらにサブコマンドを持つものがある。

②書誌情報検索機能

書名または著者名から対応する書誌情報を、文献資料目録データベースから抽出する。抽出結果は、検索結果集合として番号付けされる。ユニークな名前

国文学原典資料データベースシステムの開発 (安永)

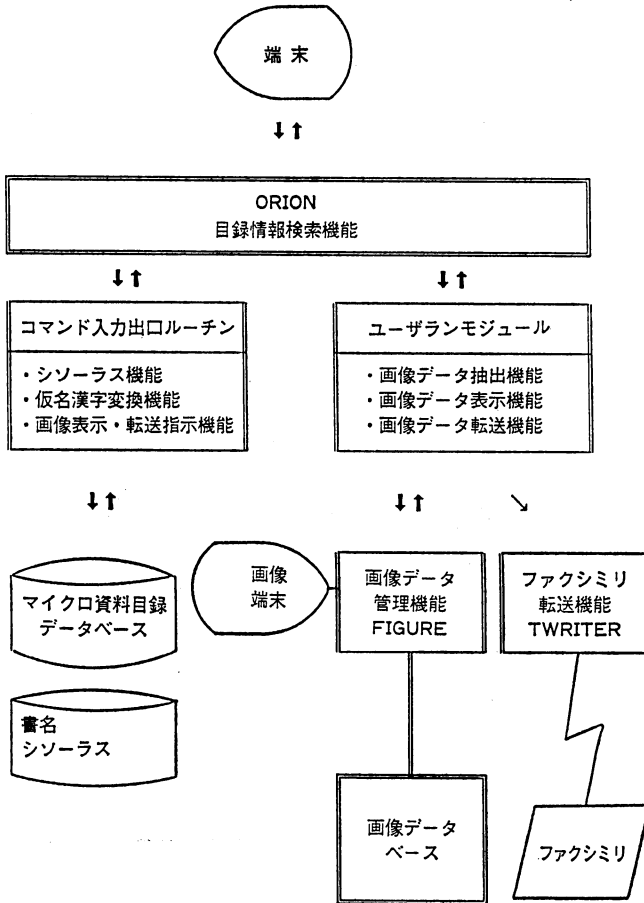


図 4. 原資料データベース検索サブシステム

である。

③検索結果表示機能

書誌情報検索機能によって作成された検索結果集合のうち、指定されたもの
の内容を指定に従って表示する。

④論理検索機能

表 1. 原資料データベース検索サブシステム対話コマンド一覧

コマンド種類	機 能 概 要
B: 書名読み	入力された書名(読み)に該当する統一書名一覧を表示する。一覧表から目的の統一書名を選択し、検索結果集合を作成する。
A: 著者名読み	入力された著者名(読み)に該当する著者名一覧を表示する。一覧表から目的の著者を選択し、検索結果集合を作成する。
DS n	検索結果集合“n”の内容を表示する。
AND n, m	検索結果集合“n”と、検索結果集合“m”の積集合を作成し、新たな検索結果集合とする。
OR n, m	検索結果集合“n”と、検索結果集合“m”の和集合を作成し、新たな検索結果集合とする。
DIF n, m	検索結果集合“n”から、検索結果集合“m”を差し引いた集合を作成し、新たな検索結果集合とする。
LIST	これまでに作成した検索結果集合の一覧を表示する。
IMG n, k	検索結果集合“n”内の書誌番号“k”に対応する画像データを画像データベースから抽出し、第1ページを画像端末に表示する。
PAGE p	IMG コマンドのサブコマンドである。IMG コマンドによって抽出した、画像データの“p”ページを、画像端末に表示する。
FAX n, k, p, q	検索結果集合“n”内の書誌番号“k”に対応する画像データを画像データベースから抽出し、“p”ページから、“q”ページを、ファクシミリに転送する。
FAX n, k, ALL	検索結果集合“n”内の書誌番号“k”に対応する画像データを画像データベースから抽出し、全ページをファクシミリに転送する。
END	検索を終了する。

複数の検索結果集合に対して、論理的な演算を施し、より高度の検索機能を提供する。

論理検索機能は、一般的な論理和、論理差、論理積等を定義している。とくに、この機能によって複数キーワードによる論理検索を行うことができる。

⑤ 検索結果集合一覧機能

検索中に、多くの検索結果集合を作った場合のために、作成した検索結果集合の一覧を表示する。

⑥画像表示指定機能

検索結果集合内の所望の書誌を指定することにより、画像データファイル抽出機能を起動し、所定の画像端末に画像データを表示することを指示する。

⑦画像転送指定機能

検索結果集合内の所望の書誌を指定することにより、画像データファイル抽出機能を起動し、所定のファクシミリに画像データを転送するを指示する。

(2) 画像処理フェーズ

①画像表示機能

実際に、画像表示指定された書誌に対応する、画像データを画像データベースから抽出して、所定の画像端末に表示する。

②画像転送機能

実際に、画像転送指定された書誌に対応する、画像データを画像データベースから抽出して、所定のファクシミリ装置に転送する。

4. 実験結果と評価

原資料流通システムの実用性について、幾つかの実験、評価を行っている。ここでは、その詳細は煩雑であるので、割愛する(別途報告の予定である)が、2, 3の結果だけを述べる。

画像入力方式については、複合画像システムに自動機能を付加することで、省力化が期待できる。しかし、人間が行う程の柔軟なデータ作成と判断を伴うデータ作成には限界がある。これは、例えば絵の輪郭を決める程度の簡単な問題においても、相当の困難な技術的検討が必要である。なお、マイクロフィルム資料からの入力方式は、ほぼ確立したといえる。

蓄積方式には、殆ど問題が無いが、電子図書館を形成するほどの大量データの形成時には、専用の高速大容量光ディスク装置が必要である。現在、約16万コマの文献資料が同時にアクセス可能である。現在のところ、光ディスクに

に対するアクセスにエラーは皆無であった。

伝送方式は、やや遅れが目立つ。通信方式、画像圧縮法等に改良が必要である。不特定多数のファクシミリに対する送信機能はまだ検討していない。

検索方式についても、殆ど問題はない。より高度の要求に対しては、対応するシステム開発が必要である。今回採用した目録データベースと画像データベースとのリンク方式は、方式が柔軟であるので、多様な要求に対応することが極めて容易である。あらためてこの方式の優れていることが確認された。

提示方式については、殆ど今後の検討課題と思われる。実験上 2, 3 のシステムを作成したに過ぎない。

5. 今後の検討課題

以上、原資料流通システムに的を絞って報告した。システム評価については割愛したが、これまでの実験結果からみる限り、このシステムの実用性は実証されたと確信する。現在、鋭意データ蓄積も行われており、またシステム改良も進められている。以下、積み残した検討課題をまとめておく。

2.2 において、本研究の仕様並びに前提条件の整備を行った。これらに加えて、個々のシステムは基よりトータルシステムの運用・管理のしやすさ、使い勝手の良さを追求することは論を待たない。また、3. において、プロトタイプシステムの概要を示した。この研究成果は、トータルシステムとしての原文献資料データベースシステムから見れば、ほんの1部の成果でしかない。検討事項は山積している。

本研究では検索機能は、目録情報から得られるキーワードによる限定検索、すなわち、書誌事項以外にはキイを持たない方式としている。しかも、書名及び著者名から本を探すことに限定している。これは、本研究がシステム技術の研究であることによるが、画像データを探し、同定するキイは多様なものが考えられる。例えば、本の系列や位置づけ、分類やジャンル等である。この様な

国文学原典資料データベースシステムの開発（安永）

キイを考慮すれば、かなり使い勝手の良いシステムを構成することが出来る。今後の展開が望まれる。

一方、画像自身の特徴を指示または特性データベースとしてのインデックスデータベースに持ち、多様な絵の特徴から検索を実施することも考慮すべきである。さらに、研究者のコメント等が、その場で自由に定義でき、それらの知見に基づく利用も望まれる。

なお、システムとしては、当該目録データベース（この意味で、画像データベース指示データベースということが出来る）に、情報項目を追加すれば良い構造を採っているので、柔軟な対応が可能である。すなわち、インデックスデータベースの構成は可能である。

プロトタイプシステムの範囲内では除外した、今後の大きな研究テーマとしては、以下のようなものがある。

①国文学原典資料の高次情報処理、要求仕様をさらに深め、これに対応する情報処理技術を開発すること（4）～（6）。

②各次データベース間の横断的利用とそのオンラインリモートアクセス方式の研究。

③システム内 OCR の可能性、すなわち画像として読み取った情報から、文字や挿絵を切り出し、これを認識すること。

④原典資料とフルテキストを同一画面にて処理できるかどうか。すなわち、翻刻、定本の作成、異本の同定、挿絵や花押索引、及びこれらの比較等に関すること。

⑤色情報の処理に関すること。

⑥電子図書館システムの検討。

また、本来の画像データベースとして、国文学研究上何をデータベース化すべきかの問題が大きい。あるいは、今回は所蔵マイクロフィルム資料に限定したが、直接全国単位での原本のリモート入力技術の検討等の課題がある。

最後に、本報告はすべて技術的検討のみを述べている。また、運用システムについては未検討な部分が多い。さらに、今後ニューメディアといわれる技術革新の社会的環境の中で、文献資料情報等の収集・蓄積・利用方法の制度的な検討が必要であると思われる。

謝辞

本研究は、学術情報センター猪瀬博所長の御指導を得て進められた。また、同センターの安達助教授、橋爪助教授、大山助教授を始め、多勢の方々の御協力を得た。また、国立民族学博物館杉田教授、国立歴史民俗博物館照井教授の御協力を得た。合わせて深謝する。

なお、当館小山館長を始めとして関係各位の御指導御協力を得た。深く御礼申し上げる。

参 考 文 献

- (1) 猪瀬 博：科学研究費報告書，#00522004，1983。3
- (2) 猪瀬 博：科学研究費報告書，#58460234，1985。3
- (3) 小山弘志：科学研究費報告書，#60810009，1988。3
- (4) 安永尚志：大学と科学シンポジウム，1987。12
- (5) 安永尚志：知識情報の世界を拓く，朝日出版，1988。10
- (6) HORI, YASUNAGA: Learning The Space of Word Meanings for Information Retrieval Systems, Proc. COLING '86