

みんなくりポジトリ

国立民族学博物館 学術情報リポジトリ National Museum of Ethnology

文字画像情報システム整備計画について (資料と情報)

メタデータ	言語: Japanese 出版者: 国立民族学博物館 公開日: 2016-03-08 キーワード (Ja): キーワード (En): 作成者: 柴田, 正美, 久保, 正敏 メールアドレス: 所属:
URL	http://hdl.handle.net/10502/00005853

文字画像情報システム整備計画について

柴田 正美

久保 正敏

一 はじめに

本館では、昭和五九年度から五カ年の予定で、文字画像情報システム整備計画と呼ばれる電子計算機システムの新たな整備計画が進行中である。この計画は、本館所蔵各種資料に対し画像情報としての蓄積、検索を可能とする画像データベースの構築を主眼としている。大量の画像情報を蓄積するための媒体の問題もあって本格的な画像データベースの開発が未だ他所では見られない現在、本館の計画は先駆的なものと言えよう。この計画を概説した本稿によって、読者諸賢の批判、提案を仰ぎ、それらを計画に盛り込んで文字画像情報システムの内容を充実したものにしていきたいと考える。

二 文字情報と画像情報の統合

昭和五三年度から五七年度までの五カ年計画により導入された電子計算機システムでは、本館の所蔵する各種資料（文献図書、標本、映像音響、H R A F等）の文字情報によるデ

ータベース化および検索システムの整備が行なわれた。しかし、文献図書に掲載された図、フィルム等の画像そのものや、現地調査で得られた写真は、文字情報による検索システムにのせることがむずかしく、コンピュータには入力されていない。標本資料の画像については、昭和五七年度に導入した標本画像自動処理装置により、標本資料の写真撮影及び必要な部位の寸法計測作業を自動化し、得られた画像情報の蓄積を行ないつつある。しかし、蓄積媒体としてオフラインの磁気テープを用いているため、蓄積された標本画像情報は文字情報に基づく既存の標本資料データベースと結合されておらず、情報の管理、検索、保存という面で難点が多い。

昭和五九年度から始まった五カ年計画、文字画像情報システム整備計画は、最初の整備計画で得られた研究成果とこれまでの情報の蓄積を基盤として、各種の画像情報をデータベース化し、これを既存の文字情報によるデータベースと統合して、総合的な検索とその結果の画像による提供を行なうシステムの構築をねらいとしている。このシステムによれば、例えば「アメリカ・インディアンのカゴ細工」についての検索を行なった場合、このキーワードに関する文字情報とともに、文献図書のそれに関する記述の含まれたページの画像、カゴ細工標本の写真、それらが作られるいは使われる様子を写した映画フィルム等の画像情報を端末機（ワークステー

区分	年度	59	60	61	62	63	64以降
国立民族学博物館ライブラリーシステムの整備		ライブラリーシステムの基本ソフトウェア開発					
		パイロットシステム					
		データ入力方法の検討				(運用)	
		標本資料データの作成と入力(準備)		標本資料ライブラリーシステム			
		データ入力方法の検討				(運用)	
					映像資料ライブラリーシステム		
						(運用)	
						民族誌資料ライブラリーシステム	
						(運用)	
ローカルエリアネットワークシステムの整備				データハイウェイ及び分岐器の設置			
共用システムの整備				ワークステーションの設置			

図1 文字画像情報システム整備年次計画表

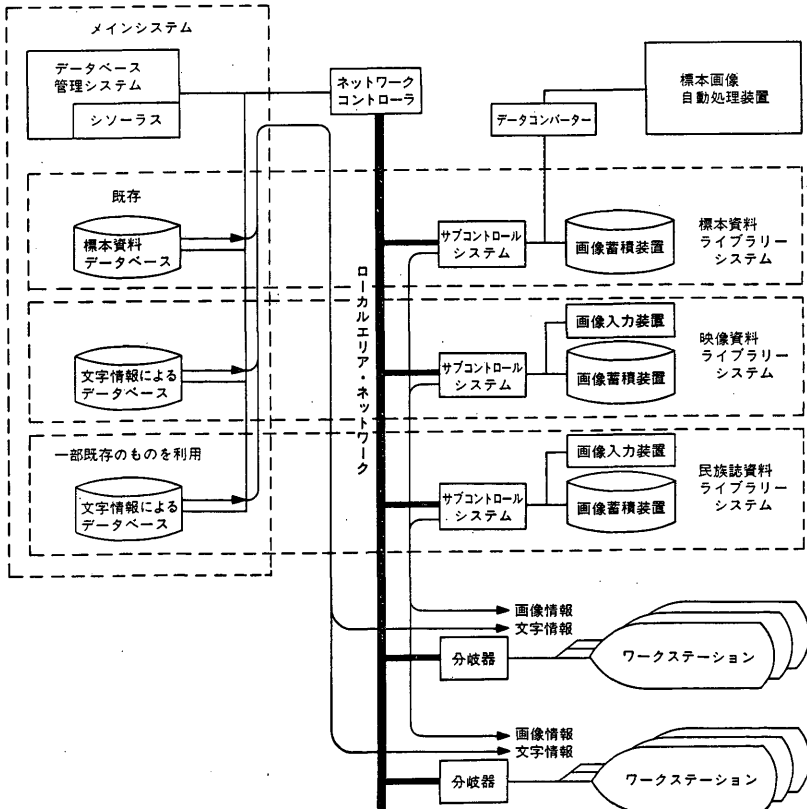


図2 文字画像情報システム概念図

シヨン)の画面上で得ることが可能になる。すなわち、

(a)各資料に対する抽象化された索引として機能するに過ぎない既存の文字情報によるデータベースに比べ、符号化されてはいないが情報の欠落が少ない画像情報によって原資料に肉迫することができると。

(b)特定のテーマについて媒体の異なる資料あるいは情報を画像として同時に一覽、比較することができる。

このようなシステムが実用化されるならば、民族学研究的な広がりや深まりが十分に期待できると考えられる。

また、生活様式の変化により失われてゆく伝統的文化あるいは生活様式に関する映像情報をデジタル情報化することにより体系的な蓄積、保存が可能となるので、映画、写真等におけるカラーフィルム其自然退色といった長期保存上の問題が解消される点もこの計画のねらいである。

三 システムの概要

文字画像情報システム整備計画では、図1に示した年次計画に従って、資料の形態別に設定される三つのライブラリシステムとそれらを統合するデータベース管理システム、システムを構成する情報処理機器を相互に結合し効率的な館内情報処理ネットワークを実現するためのローカルエリア・ネットワークシステム、検索を行なうためのワークステーション

ン等を構築する。システムの概念図を図2に示す。

基本となるライブラリシステムは次の三つである。

① 標本資料ライブラリシステム

既に稼動している標本画像自動処理装置で蓄積されつつある標本資料の画像情報をオンライン蓄積媒体に移してデータベース化するとともに、文字情報に基づく既存の標本資料データベースとの一体化をはかり、利用に供するものである。

② 映像資料ライブラリシステム

研究者が現地調査の際撮影してきた写真や映画などの映像資料が、本館には大量に蓄積されつつある。しかし、現在はこれらの映像資料の有効な検索手段が確立されていないため、他の研究者の利用に供したり情報の交換をはかることが困難な状況にある。このような問題の解決のために作られるのが、映像資料ライブラリシステムである。研究者が撮影してきた写真、スケッチ、フィールドノート等の静止画や映画、ビデオなどの動画に、場所や時期、説明などの文字情報を付加してデータベース化しておく、この文字情報を検索の手がかりとする。これらのデータベース化された映像情報は、前述したように長期保存管理にも役立つものとなる。

③ 民族誌資料ライブラリシステム

本館の所蔵する文献図書資料のうち、民族学研究の基本文献を選択し、本文を文字情報として、また、さし絵や写真、

図表等を含む本文の各ページをそのまま画像情報としてデータベース化し、その保存管理と研究者への提供を行なうシステムである。すなわち、求める言葉をキーワードとして入力すると、関連するページが画像として利用者に提示されるので、文献図書の網羅的な検索が可能となる。

利用者はこれら三つのライブラリシステムに対し、ローカルエリア・ネットワークに接続されたワークステーションを通してアクセスする。地域、民族、年代、文化要素、資料の形態等のカテゴリに基づきキーワードによって検索された資料に関する情報は、文字情報とともに画像としてワークステーションの画面に表示されることになる。

四 システム実現のための技術的課題

このようなシステムを実現するうえで解決していかねばならない技術的課題は数多いが、次の三点に整理できる。

① 画像情報の蓄積・伝送技術

この整備計画が対象とする画像資料は、その形態からみるとフィルムドノートやスケッチ入り手書きカードなどの白黒二値画像、スライドや標本画像などのカラー静止画像、フィルムやビデオテープなどのカラー動画像に分類される。これらをデータベース化し検索された画像を利用者に提供するためには、次のような条件を満たす画像蓄積媒体が必要となる。

(a) 画像情報は文字情報と異なり情報量が膨大であり、これを収容できる大きな蓄積容量を持つこと。

(b) 館内でカメラ、スキャナーなどから入力した画像情報を直接その媒体に記録・蓄積できること。

(c) 蓄積媒体から高速に画像情報を読み出せること。

これらを満たすものとして追記型の光ディスクが有望であり、白黒二値画像に対しては、その情報を圧縮しデジタル情報として記録できる、いわゆる文書ファイル装置が市販されている。しかし、カラー画像に対しては、デジタル情報として記録し、しかもオンライン検索に耐える速度で読み出せる光ディスク装置を入手することは残念ながら現時点では困難である。従って、カラー画像については、高速の検索用にはアナログ記録方式の光ディスク装置、前述したような長期保存の目的および高精細画像の検索用にはデジタル記録方式の光ディスク装置、という二種の蓄積媒体を併用する方式が現実的であろう。アナログ記録方式については、現在の商用テレビより画質の優れた高品位テレビに対応するものが望ましい。また、カラー画像の蓄積・伝送情報量をできるだけ減らすために効率的な画像情報圧縮の手法を確立する必要がある。

利用者の検索要求に従ってこれらの蓄積媒体から読み出された画像情報は、ローカルエリア・ネットワークを経由して

ワークステーションに伝送される。ローカルエリア・ネットワークの伝送媒体としては、同軸ケーブルに比べて低損失、広帯域、無誘導、省スペースなどの長所を持つ光ファイバ・ケーブルが最適であり、デジタルデータ伝送やアナログビデオ信号伝送には既に広く利用されている。しかし、この整備計画におけるような、アナログ、デジタルと形式の異なる種々の画像情報を統合・多重し光ファイバを用いて伝送する手法は現在のところ確立しておらず、今後早急に開発していかねばならない技術的課題である。

② データベース管理システム

一般に画像データベースシステムでは、画像そのもの（画像ファイル）とその画像に関して記述された付属情報とを分けて管理する。画像の検索には、この付属情報による方法と画像の特徴による方法とが考えられる。後者は特に内容依存検索と呼ばれる画像データベース特有の検索方法で、例えば「胴がこんな形をした赤いツボの画像を探せ」といった画像そのものの特徴による検索を指すが、これを可能とするには特徴抽出やパターンマッチングなどの画像処理を超高速度で行なう機構が必要であり、現在の技術水準では実現が難しい。従って、本整備計画では、内容依存検索は将来の課題として残し、既存の文字情報によるデータベースシステムを活用し画像を検索する前者の方法をとることとする。

しかしながら、既存のデータベース検索システムには以下に挙げる様々な問題点が存在する。

(a) 各国の文字による検索が不可能である。特に和文のデータベースについては、ヒラガナのヨミをキーワードとする検索のみが許されており、漢字による検索ができない。
(b) 基本的には検索用キーワードとの前方一致によって検索が行なわれ、中間一致、後方一致機能がでない。

(c) 各データベースにわたる同時並行検索ができない。
従来からの懸案事項であったこれら問題点の解消も、本整備計画で各資料データベースと画像データベースを統合するにあたっての目標の一つであり、現行の検索システム STAIRS の大改造が必要となるであろう。

懸案となっているもう一つの重要な問題は、民族学シソーラスの作成である。現行のデータベースシステムでは、自然語によるデータ作成や検索を重視して用語を統制していないために、検索キーワード「くすり」によって同義語「薬品」を検索できないといった不都合が生ずる。この問題を解消するには、まず民族学で用いられる用語間の同義関係を明らかにしたシソーラスを作成する必要があるとの認識から、昭和五七年に「シソーラス研究会」が自発的に生まれ、HRAF の OCM を分類体系の基本とし、データベースに用いられている語彙に OCM コードを割付けるという方針が打ち出され

た。これを受けて、後述の情報化委員会傘下の「シソーラス部会」において、OCMの日本語への翻訳および標本名に現われる語彙の頻度分析の作業が現在進められている。ただ、OCMは文化要素の分類を目的としたものであり、機能の面から標本資料を大分類するには利用できるが、形態、材質などの面からの分類には充分ではないと考えられる。

最終的には、この成果を基に、民族学用語間の関係を網羅的に研究し、同義だけではなく反義語、上位概念、下位概念、関連語などの分類構造を表現した民族学シソーラスを作成することが望まれる。しかし、分類構造とはいわば世界観の表現であるから、万人に受け入れられるシソーラスを作成することは難しい。一つの観点からの厳密な用語分類を旨とするよりはデータベース検索の容易さを目的とし、様々な観点からの用語の意味関係をネットワークで表現した事典的なシソーラスを作成するのが良いであろう。このためには、綿密な検討と膨大な作業が必要であることは言うまでもない。このようなシソーラスを検索システムに組み込んでおけば、検索者が発した複数のキーワードがシソーラスを介して翻訳、推論され、最適なキーワード群が対話的に生成されてデータベースを検索することができる。こうした知的検索機能は、民族学の専門家だけでなく入門者にも役立つものとなるろう。

③ マン・マシン・インタフェース技術

文字画像情報システムと利用者との接点はワークステーションであり、良好なマン・マシン・インタフェースの設計が重要な課題である。この問題は、ワークステーションのハードウェア、対話手法などのソフトウェア両面から考える必要がある。前者については、キーボード配置やタッチパネル、マウスなどの操作機器、動画も含む複数画像を合成・編集しマルチウインドウやマルチスクリーンを用いて同時表示する手法、ハードコピー装置、後者については、様々なレベルの利用者にも適応できるコマンド体系のあり方、図的メニューなど、検討を要する事項が多い。また、ワークステーションには、ローカルエリア・ネットワークから受け取った種々の形式の画像データの複合、画像処理などの高度な機能も要求されるので、場合によっては、市販のコンピュータと入出力機器の組合せでは間に合わず、本館独自の設計によるワークステーションの開発が必要になるかも知れない。

これら三点の技術的課題はいずれも利用者の要望と情報処理分野の最先端技術を取入れて解決していかねばならない。

五 計画の具体化に向けて

昭和五九年度予算において、文字画像情報システム整備計画の第一年度の計画であるパイロットシステムの予算が認められた。これを受けて昭和五九年三月、これまでの情報化構

想の集大成と実施に向けての体制作りを目的として、情報システム構想委員会が暫定的に設置された。この委員会は約一カ月の間に集中的な検討を行ない、研究、展示、管理のすべての情報を対象とした全館的な情報化構想をまとめた。その骨子は次のようなものである。

- (1) 将来、標本資料ライブラリーシステム、映像資料ライブラリーシステムを基礎に映像全般のライブラリーシステムを統合してフォトテークと称する研究用システムを構築する。
- (2) ビデオテークは、現行システムの機器更新にあたって全館ネットワークの一部に組み込むが、あくまでも動画の一般向けサービスシステムとして独立して運用できるものとする。
- (3) 文字画像情報システム整備計画終了後に、音響資料に関してフォノテークと称する新たなシステムの整備を考える。
- (4) フォトテークあるいは既存の文字情報による検索システム（ビブリオテークと称する）など研究用システムの成果を基礎として、民族学全般、あるいは展示品解説などの情報を中心とした「目で見える民族学百科事典」とでも言うべき一般向け情報サービスシステムの構築を今後検討していく。
- (5) 最終的には、ビブリオテーク、フォノテーク、ビデオテーク、フォノテーク、及び「目で見える民族学百科事典」をすべて統合し、「ホロテーク」と称するシステムを構築する。

また、これらの構想の実現に向けて新たな体制作りの方針としては次のような方向が示された。

- (1) ビデオテークなどの展示用機器の検討のために設置された展示用機器改善委員会と電子計算機運営委員会とを統合し、全館的な情報システムを検討する組織とする。
- (2) 今後の情報システムの整備に伴う各種情報及びデータのあり方を検討し、あわせてシソーラスの検討を行なうために、新たに情報化委員会を設置する。

このような情報システム構想委員会の結論を得て、文字画像情報システム整備計画は、その実現に向けて実務的な作業に入った。昭和五九年度には、パイロットシステムとして、標本画像自動処理装置で得られた画像の一部を検索・表示する装置、手書きカードや解説書などの白黒二値画像を蓄積・検索・表示する装置を導入し、現在、各ライブラリーシステム構築に向けての基礎実験を進めている。昭和六〇年度には、この成果を基に標本資料ライブラリーシステムの整備に着手することになる。前述したように、この文字画像情報システム整備計画は数々の挑戦課題を含んだ壮大な計画であり、今後、さらに使い易く有効なシステムの実現を目ざして計画を推進していきたい。

（柴田 情報管理施設技術室）

（久保 第五研究部）