みんぱくリポジトリ
国立民族学博物館学術情報リポジトリ National Museum of Ethnolo

民族学とコンピューター

メタデータ	言語: Japanese 出版者: 公開日: 2010-02-26 キーワード (Ja): キーワード (En): 作成者: 杉田, 繁治 メールアドレス: 所属:
URL	https://doi.org/10.15021/00003568

第　I　部

民族学とコンピューター

杉　田　繁　治*

要旨

国立民族学博物館において，コンピューターシステムをどのような経過で導入してきたか，その過去および現状と，近い将来における計画について述べる。また人文科学の分野におけるコンピューター利用について，その特徴を分析し，現在のコンピューターに欠けているものは何か，どのようなことが大切であるかについて述べる。特に利用者の立場からマンマシンインターフェースの問題を論じる。

1　民博の情報システム

民族学の研究に積極的にコンピューターを活用しよう，というのが国立民族学博物館（民博）の設立当初からのスローガンである。1970年代の半ばはまだコンピューターは専門家が扱うもので，プログラムや操作はむずかしいものというのが一般の考えであった。特に人文系においてはコンピューター利用は，統計計算などのごく限られた問題においてだけであった。しかし，コンピューターはやがて研究者にとって万年筆と同じような道具になるであろう，という梅棹忠夫館長の見通しの下に，民博では大規模なコンピューターシステムの導入が進められている。

われわれがめざすものは，民博が所有するすべての資料について，できるだけ現物に近い情報をコンピューターによって検索できるシステムの構築である。書籍ならば，単に書誌的情報が検索できるだけではなく，その内容が例えばページイメージとして，あるいはテキストそのものが文字列として検索されるような，民族誌ライブラリーである。標本資料ならば，いくつかの角度から撮った映像が表示されるものである。スライドの場合は，あたかもプロジェクターで表示しているような映像がコンピューターの端末のところで映し出されることである。また録音された音楽や音声などは単にカセットのあり場所が検索されるだけではなく，音そのものもコンピューターから再生されて，直接音を聞くことができるシステムを構築することである。

われわれはこのようなシステムをマルチメディアデータベースと呼んでいる。それ

* 国立民族学博物館 第5研究部

は様々な情報媒体（メディア）によって表現されている情報をあたかも単一の情報であるかの如く扱うことができるシステムだからである。ここでは文字も画像も音も混在してコンピューターによって管理され，外部に表示される時にはそれぞれの元の形態に近い形に再現されるのである。

このため昭和54年からホストコンピューターとして IBM を導入し，また画像・音響関係の入力，出力のための様々なシステムを導入している。これらは同時期に導入されたものではなく，長期計画の下に毎年その時代の最先端の機器を選択して導入している。従って全体のシステムを構成する機器は同一のメーカーのものではなく，さまざまな会社のものが入っている。そこで異なる機器を統合するために LAN（ローカルエリアネットワーク）を導入し，あたかも同質のシステムであるかの如く見えるようにしている。

1.1　情報システムの整備

情報センターとしてマルチメディアデータベースをいかに構築していくか。これにはしっかりした思想と技術の見通しがなければならない。民博設立準備の段階，1970年代の初め頃は日本におけるコンピューターはまだ専門家が使うという状況で，一般的な応用に耐えるものではなかった。しかし，その当時から創設準備室では，やがてコンピューターは社会のもろもろの分野で重要な役割を演じる道具になるであろうという見通しがあり，それは人文科学研究や博物館においても例外ではありえないという確信があった。そして，技術の開発を見込んで段階的に情報システムを実現し，技術の発達に柔軟に対応し，常に最新の道具を利用できるように，第1期，第2期，第3期など，長期の年次計画を構築した。1期の期間をどれくらいにするか。1970年以降のエレクトロニクスの技術革新は急激である。10年という期間は長いと判断し，5年毎にその計画を見直し，資料の収集具合をも勘案して実現の可能性を見極めて情報システムの構築を行なう方針を立てた。

第1期計画（文字情報システム）——昭和53(1978)年度～昭和57(1982)年度

先ず基本的なデータベースの構築を行なった。それは民博が所有する資料について，その書誌的な事項をコンピューターに入力し，検索することができるシステムの導入である。図書，標本，映像・音響，HRAF（Human Relations Area Files, 人間関係地域ファイル）のいわば台帳データの蓄積・検索である。当時，本格的なデータベースシステムはほとんどなかったが，ユーザーとしての要求を強く表に出して漢字デ

ータを扱える検索システムを導入した。現在，民博が独自に入力したデータは約170万件にのぼっている（平成4年3月現在，表1）。

また博物館として画期的な新しい試みとして標本資料の画像入力と自動計測のシステムを開発した。世界の諸民族から収集された資料を，従来は手作業で計測し写真を撮っていたが，それを自動化した。標本資料の立体計測の可能性も検討したが，当時は自動車のボディの設計などで試みられていた程度であって，民博の資料のように多様な形態と材質に対応できるものはなかった。新しいシステムの開発により，昭和58年以降毎日入力作業をつづけることが可能になり，現在約5万7千点分（約400ギガバイト）のデータが蓄積されている。

第2期計画（文字画像情報システム）——昭和59(1984)年度～平成2(1990)年度

第2期計画は当初5年計画で出発したが，途中種々の事情によって予算編成上7年にわたることになった。

第2期計画の重点はスライド・写真映像の蓄積・検索，それと非図書資料の総合的活用をめざした民族誌ライブラリーの構築である。また第1期で開発されたシステムや第2期で開発された分散システムを統合し，全体が一体となって活用されるようこれらをネットワーク化することであった。

標本資料画像は第1期のシステムで軌道にのせたが，民族学の調査から得られる膨大なスライドや写真をどうするかの問題があった。通常これらは各人が自分の方式で管理活用しているが，もしこれらが共有財産として活用できるようになれば，そこから得られる情報は非常に大きい。民族学では「比較」ということが重要であるが，研究者が自分のフィールドとしている地域とあまりよく知らない地域における物質文化や儀礼などをスライドを通じて比較できるだけでも仮説の設定に役立つところが大きい。これについては，ハイビジョン並の高解像度を持つディジタルカメラを使って入力し，ディジタルデータとして蓄積・検索するシステムとして実現した。

次に民族誌ライブラリーの構築を開始した。従来の図書館では書棚に配架できる形態の図書を中心にシステムが考えられている。雑誌論文や書籍の中の章単位の内容などについては検索の対象にはなっていない場合が普通である。また，本の中の図表や写真，地図などをそのまま画像として検索・表示するシステムはできていない。最近でこそ辞典や百科事典をCD-ROMに入力し，パソコンによって検索できるシステムも登場しているが，これをさらに一般化したシステムの構築を考えている。例えば著名な論文や民族誌などは単に書名や所在場所が分るだけではなく，その文章自体がコ

表1　国立民族学博物館が収集している資料とコンピューター化されているデータ量　　平成4年3月1日現在

資料名	資料の種類		冊(点)数	情報の種類	データベース件数(件)	データベース化率(%)
文献図書資料	単行本	日本語	107,493 冊	単行本・書誌情報（日本語）	102,632	
		外国語	186,845	単行本・書誌情報（外国語）	162,369	
		計	294,338		265,001	90.0
		製本日本語雑誌	18,753 冊			
		製本外国語雑誌	42,324			
		計	61,077			
	雑誌	日本語雑誌	5,496 種			
		外国語雑誌	4,562			
		計	10,058	雑誌・書誌書蔵情報	10,058	100.0
	小冊子類資料		28,756 点	小冊子類・書誌情報	28,756	
	マイクロ写真資料		1,989 種	JP-MARC データ（1977～1992.02.20）	1,015,665	
				US-MARC データ（1969～1991.12.23）	2,836,952	
HRAF 資料	文献情報スリップ		6,353 ソース	民族誌文献書誌情報	6,353	100.0
	テキスト（原典）スリップ		812,290 枚	主題分類コード情報	812,290	100.0
映像音響資料	レコード，CD		35,873 点			
	テープ（言語）		6,107			
	テープ（音楽）		3,039			
	フィルム（16ミリ）		2,803			
	ビデオテープ		1,979			
	計		49,801	管理情報	44,213	89.4
				学術研究情報（日本語）	50,428	
				学術研究情報（外国語）	96,245	
				スライド画像情報*	21,000	
標本資料	海外資料		107,045 点			
	国内資料		82,795			
	計		189,840	管理情報	182,107	95.9
				学術研究情報	117,637	62.0
				標本資料画像情報*	56,885	30.0
国内資料調査報告集	国内資料調査報告集　1～12集		12,910 点	国内資料調査報告集　1～12集	12,958	100.0

（データベース総件数　5,556,548 件）
（館内作成データ件数　1,703,931 件）

- JP-MARC データ：国立国会図書館で作成されている文献書誌情報。
- US-MARC データ：アメリカの議会図書館で作成されている文献書誌情報。
- 管理情報：資料名・収蔵場所等といった資料を管理するための情報。
- 学術研究情報：たとえば標本資料ならば，その使用法・使用民族名等といった研究者の調査研究によって得られる情報。
- 情報の種類欄で＊が付加されているものは，画像データであることを示す。
- 国内資料調査報告集：民具等標本資料，技術伝承者，民族・民俗関係映像記録，その他の資料，および民族・民俗関係出版物の所在に関する情報。
- 小冊子類・書誌情報：パンフレット，リーフレット，抜刷等の文献書誌情報。

ンピューターの端末で直接見ることができるといったものである。そこには表や図も写真も画像として見ることができる。千冊分くらいのデータを1枚の光ディスクに収納することも可能である。ディスクの自動倉庫を使えば数万冊分のデータをオンラインで扱うことができる。ハイビジョンシステムを応用すればカラーの資料も蓄積できる。このようなシステムを使えば散逸しがちなパンフレットやポスター類，展示図録，建築設計図なども容易に検索することができるようになる。

さらに1期，2期を通じて開発してきたデータベースを統合し，館内のあちこちに分散している端末から，あたかもすべてのデータが一箇所に存在しているかの如くに検索できるよう LAN で結合した。現在このネットワークを学術情報センターのネットワークや大学間ネットワークに接続し，館外からも民博のデータベースが利用できるよう準備を進めている。

第3期計画（総合情報システム）——平成3（1991）年度～平成7（1995）年度

第1期，第2期計画でまだ実現されていない問題がある。その一つは音響データである。民博ではレコード，カセットテープ，CD（コンパクトディスク）など多数収集しているが，それらを活用する場合に書籍とはちょっと違った問題がある。書籍は手に取って誰でもその場で見られる。しかし，レコードやカセットなどはそれぞれの装置を操作しなければその内容を知ることはできない。しかし，もし図書の検索などと同様にコンピューターを使って資料の検索をすれば，その場で音が聞えてくるようになれば音響装置の操作が不要になる。さまざまな民族の音楽や言語を同時に聞いて比較したり，その音の特徴をコンピューターによって解析することもできる。また媒体も劣化せずに大切な資料として保管することができる。

次は標本やスライド以外のイメージデータの問題がある。主なデータとして地図がある。民族学の分野では世界の諸民族を対象にしているので地図は必須のものである。単に地形図だけではなく，言語の分布，交通，産物，その他風俗習慣のマッピングなどもある。これらは大きなサイズから小さなものまであり，また図法もさまざまである。またその検索の仕方も多様である。緯度経度から直接検索する場合もあれば，地名や民族名からその近辺の領域を知りたい場合もあろう。産物名などから検索し，該当地図が目の前に表示されることも必要な機能である。

また展示されている壁面や立体配置の様子など，展示に関わる記録をイメージとしても蓄積し，記述データと連動させて検索・表示するシステムの構築も考えている。これらは民族誌ライブラリーの範疇であると考えることもできる。しかし，高解像度

カラー画像の大量蓄積技術は現在電子機器メーカーにおいても開発中であり，第2期計画の民族誌ライブラリーとしては時期尚早であったが，近い将来新しいメディアが出現するであろうから，第3期の中では実現可能であるという見通しを持っている。

さらに立体データの問題がある。これは物質文化の研究や博物館として展示を行なう場合に必要になる。仮面や器などの概略は第1期で開発した標本画像自動入力計測装置からの映像で見当付けることができる。しかし凹凸の様子や任意の角度から眺めた場合の姿は見る人の想像力に待たねばならない。もし3次元データが入力されておれば，コンピューターによって任意の方向から見た姿を表示させることができる。モノの研究には直接手で触れることが重要であるが，立体形状の計量的分析，特徴比較にはコンピューターが威力を発揮する。このことによって物質文化の研究が従来にはなかった次元への展開も可能となり研究の進展にもつながる。また展示を行なう場合にも実際にモノを配置せずに，コンピューターシミュレーションによっていろいろ試みることができる。このような展示シミュレーションはまだほとんどなされておらず，経験に頼っているにすぎない。立体計測の実現によって新しい展開が可能になるものと思われる。

さて今まで開発してきたシステムは媒体の形態を超越し，関連する資料がコンピューターの端末で統合されるための具体的データの蓄積と検索であった。しかしその検索はある程度資料について知っているという前提が暗黙の了解として存在していた。書物にしても前方一致や後方一致などの機能によって単語の一部から検索することはできた。しかし，対象とする資料のデータに含まれていない語彙での検索は不可能であった。コンピューターによる情報検索がまだ一般に普及しないのは検索者の語彙と内部に入力されている語彙とが必ずしも一致しないために，検索がうまくできない場合がしばしば起こるということに原因がある。そこで検索者が頭に浮べる語彙と内部の語彙を関連付ける仕掛が必要になる。それをわれわれは「シソーラス」と呼んでいる。これは経験的に獲得してきた常識の一部であると考えることができる。すべての事柄に関するシソーラスは到底できないが，せめて民族学に関係する分野に限定した辞書ができないか。民博では，既にデータベース化した標本資料名や HRAF システムが採用しているコード体系と概念関係をもとに，民博シソーラスを編纂し，コンピューターに仲介の役割を持たせることにより，より柔軟な検索を可能にしようとしている。

さて以上の事柄が実現すれば，現在民博が所有している研究資料については一応データベース化が着手されたことになる。データの入力は今後継続的に行なっていか

ねばならないが，これらのデータを生かして，広く一般の観客にもよりよく世界の民族を理解してもらうための仕掛を開発することは民博の役割でもあろう。昭和52年の開館当初からビデオテークというコンピューター制御による映像音響情報自動送出装置によって，観客は数百種類のビデオから選択して見ることができるようになっている。平成元年からはレーザーディスクを使った新システムになっている。やがてハイビジョンシステムを考えねばならないであろうが，さらに性質の異なったシステムによって民族に関わる情報を提供することを考えている。それは簡単にいえば「電子図書館」である。あるいは「学習コーナー」である。モノを見るだけでなく，そのモノに関わるあらゆる種類の情報が検索できるシステムである。民博の展示場にはあまり説明がない。それをカバーし，さらにもう少し深く，民族についての情報を映像，画像，文章，音を通じて知ることができるようにするシステムである。そこにコンピューターを活用しようというわけである。このようなシステムには「ホロテーク」という名称が考えられている。これは今後建設を希望している第6, 7展示場に設置することを想定しているが，その準備としてパイロットシステムの開発を第3期計画で考えている。これにはシソーラスを活用して曖昧な表現からでも知りたい情報に近づけ，関連する情報が媒体の種類に関係なくコンピューターから出てくるようにしようとしている。

このような統合情報システムでは絶えず新しいデータを入力し，全体の統合を行なわなければならない。大量情報の中からデータを迅速に，適確に検索するための研究情報データベースの構築には，様々な工夫を導入しなければならず，その構成はより複雑化し，データの選択，加工，入力など日常的業務も多岐にわたってくる。これを着実にかつ効率的に継続していくためには，より整った組織・体制と予算的裏付けが必要であり，行政面からの強力な支援がなければならない。

1.2　コンピューターシステムの現状

民博における現在のコンピューターシステムの概念図を図1に示す。

ホストコンピューター（写真1）

平成4年3月現在のコンピューターは当初から数えると4代目である。第1代目はIBM370/138であった。主記憶装置は1メガバイトであった。しかし1年後に2メガバイトに増強された。それはいざ検索をしてみると応答速度が予想外に遅いからであった。実はデータはコンピューターの導入よりも3年ほど前から作成しており，その

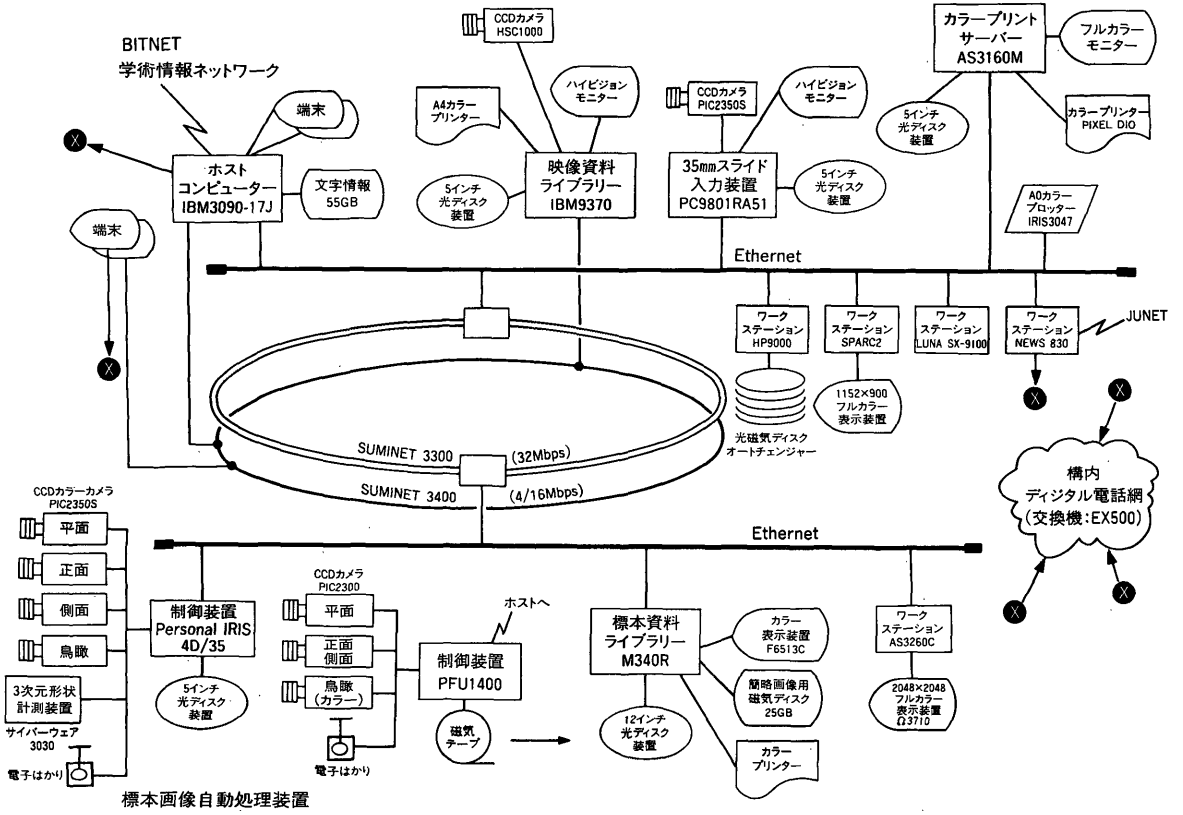

図1　国立民族学博物館のコンピューターシステム概念図（平成4年3月現在）

量の増加が当初の予想より速く，コンピューター導入時に既に予想の3年先の分量になっていたのである。急遽予算を申請してシステムを倍増した。2代目はIBM4341である。主記憶装置は8メガバイトであった。やがてこれは画像関係の実験装置を制御するものとし，ホストコンピューターとして新たに主記憶16メガバイトのIBM4341を導入した。これは2.5代目ということになる。3代目はIBM3090/150，主記憶64メガバイトである。磁気ディスクは45ギガバイトである。そして平成3年3月からは現在のIBM3090/17J，主記憶64メガバイト，磁気ディスク55ギガバイトである。端末は約130台が館内のあちこちに配置されている。

このホストコンピューターのデータベースには，民博が持つすべての資料についてその書誌的情報とやや詳しい研究情報が文字データとして蓄積され，STAIRS，JAIRSなどの検索ソフトウェアによって利用できるようになっている。他のデータベースはこのホストデータベースと連動しており，ワークステーションからの画像検索も，先ず文字情報によって検索し，それをもとに他のシステムに蓄積されている関係画像が取り出されるという仕組になっている。

標本画像自動入力計測装置[1]（写真3）

これは民博が昭和57年度に富士通と池上通信の協力を得て開発したもので，民博では普通「標本画像自動処理装置」と呼んでいる。従来博物館では標本資料のサイズの計測には直接モノを測定するか写真を撮ってその上で計測していた。これをコンピューターによって自動化することを計画した。それと同時に標本の画像もコンピューターに蓄積していくことにした。

この装置は直径2メートルの回転台と3台のCCD（Charge Coupled Device）を用いた電子カメラが入力部にあり，そこで撮られた平面，正面，側面，鳥瞰，の画像がコンピューターへ送られ，画像処理装置によって各面でのサイズが自動的に計測されるものである。カメラの出力は各面1024×1024の点に分解され，濃淡256階調のデータとなる。また鳥瞰画像についてはさらにカラー赤，青，緑の3原色に分解され，各色毎に256階調，全部で1670万色の色が表現できるようになっている。コンピューターのデータとしては標本1点当り7メガバイトの量になる。

物体の大きさは各辺最大1メートル以内最小5センチメートル以上を対象としている。この装置で年間処理できる量は約7千点であるが，民博の標本収集量は年間約1万数千点あり，そのうちこの装置の処理対象となる標本は約85パーセントあるので，

1）本書資料編D「標本画像自動処理装置」参照。

これではいつまでたっても滞貨がなくならない。そこで平成4年3月にやや小型のものを対象にした画像入力装置を導入した。

標本画像蓄積検索装置（写真4）

入力装置から得られた画像は一旦磁気テープに出力され，それを富士通の汎用中型コンピューター M340R を中心とする蓄積検索装置によって，光ディスクと磁気ディスクに蓄積している。光ディスクは1カートリッジ当り2.7ギガバイトの容量があり，1件7メガバイトのデータをそのまま圧縮しないで蓄積して，現在のところこれはオフラインとして保管庫に収納している。

画像データのうちカラーの鳥瞰画像は256×256のサイズに縮小され，磁気ディスクに蓄積されてオンライン検索の対象になっている。磁気ディスクの容量は25ギガバイトであり，カラー画像が10万点以上蓄積できる。

写真・スライド画像蓄積検索装置（写真5）

35 mmのスライドからA4サイズの写真までをハイビジョン対応のカメラによって1030×1920以内の点に分解して入力することができるシステムを導入している。これには池上通信の HSC1000 というカメラと，制御装置として IBM9370 シリーズのミニコンピューターを使っている。蓄積媒体は5.25インチの追記型光ディスクである。ここでも光ディスクのデータは圧縮なしの原データで，オフライン的に使うことにしており，検索には縦・横6分の1に縮小し，約10分の1に圧縮したデータを磁気ディスクに入れてオンライン検索を行なっている。

この装置では変色した画像に対してもモニターを見ながら正常と思われる色に変換して入力することができる。またコンピューターからカラープリンターに出力できるので，写真処理を経ずしてハードコピーを得ることができる。1日に約150枚のスライドを入力している。現在のシステム構成では15万枚分のデータを蓄積することができる。磁気ディスクの容量を増やせば60万件まで扱うことができる。

ページイメージ蓄積検索装置（写真6）

これは電子ファイルとしてオフィスなどで使われている東芝 TOSFILE 3200 を中心としたシステムである。ここに書類や本のページをそのままのイメージで蓄積し検索することができるが，民博では HRAF 資料の蓄積検索システムとして実験を進めている。HRAF 資料は民族学研究のための資料で，書籍や論文のページ毎にそこに書かれている内容を約600の分類項目に従って分析したものである。各ページは独立

したデータとして分類番号毎にもファイルされている。原データとしては80万ページ分であるが，1枚のページはそこに記された分類記号の数だけコピーが取られ，各々の番号のファイルに入れられているので，全体では400万枚にものぼる資料である。

しかしもしこれをコンピューターを使って検索するとすれば，オリジナルの80万ページ分だけを電子ファイルに入れておけばよく，光ディスクのオートチェンジャーを導入しても5平方メートルもあればよい。現在は約80平方メートルの部屋に100本のキャビネットが必要である。民博では既に HRAF 資料について，誰の論文の何ページ目に何が書かれているかというページ毎の情報は入力されているので，電子ファイルと台帳データベースを連動させるだけでこれが実現する。現在約1万ページ分のイメージを入力して実験を行なっている。

この HRAF が採用しているコード体系には文化項目分類（Outline of Cultural Materials, OCM）と地域・民族分類（Outline of World Cultures, OWC）の2種類がある。OCMは文化要素を約600のカテゴリーで表現したもので，OWC は民族の地域や歴史的区分を表現したものである[2)]。民博ではすべての資料にこのコードを付けることによって，資料間の横の関係を知る手掛りにしようとしている。またこのコード体系を活用してシソーラスの役割も持たせようとしている[3)]。

ワークステーション（写真7）

以上のデータベースは各々独立したコンピューターシステムの上に構築されているが，SUMINET 3300 という LAN によって結合されている。いわば分散データベースになっている。このネットワーク上のデータベースに接近するための端末として，ワークステーションを使う。いくつかの異なったタイプの端末を導入しているが，大型のディスプレイを含め複数個のディスプレイを持って，同時に多数の情報を表示するシステムを構築している。

ビデオテーク

これは直接コンピューターシステムとは関係はないが，動画の映像データベースであると考えることができる。現在は図1のシステムとは別であるが，将来これも新たなネットワークで結合される可能性がある。

民博は開館当初からビデオテークというコンピューター制御の装置によって，世界の民族の生活を動きのある映像で紹介してきた。従来の博物館では単にモノが陳列し

2) 本書「標本資料検索コードとしての HRAF コードの利用について」および資料編B「HRAF/文化項目分類（OCM）コード」，資料編C「HRAF/民族・地域分類（OWC）コード」参照。
3) 本書「民博におけるシソーラスの構想」参照。

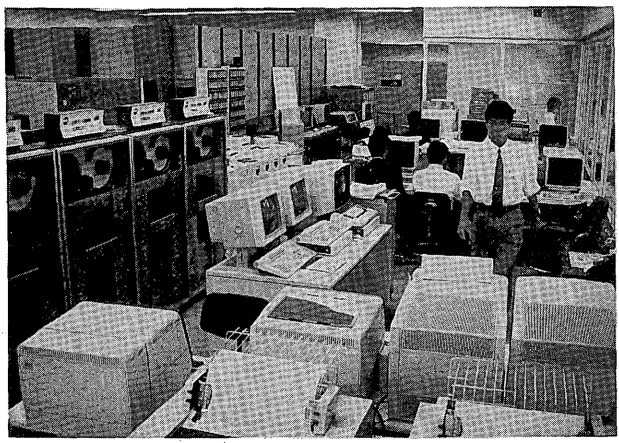

写真 1
主コンピューター(全景)

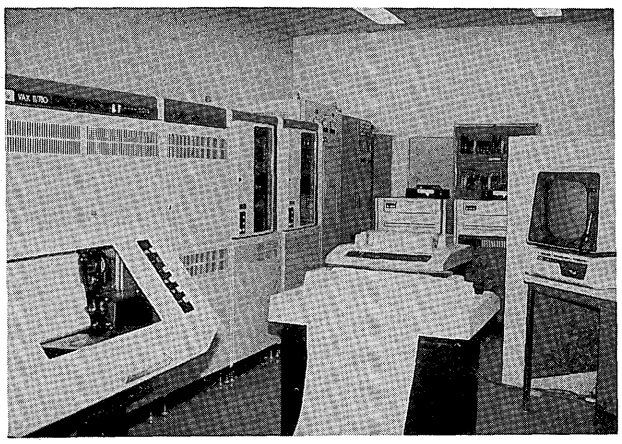

写真 2
音響データ処理実験装置
(a) VAX11/780

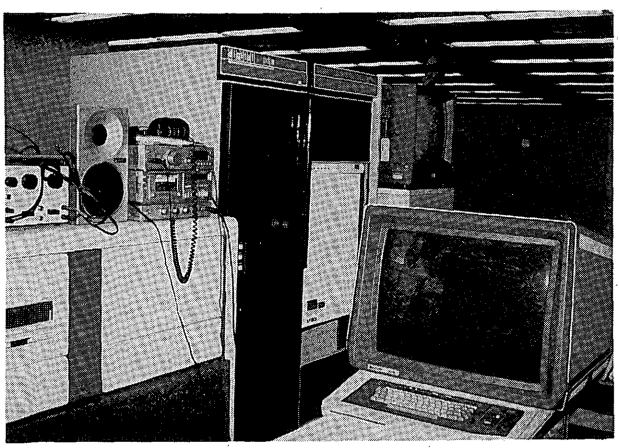

写真 2
(b) PDP11/60

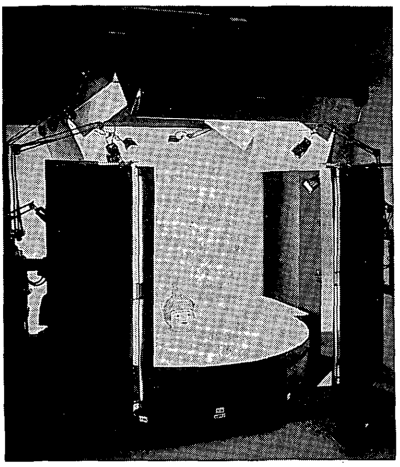

写真 3
標本画像自動入力計測装置
(a) 入力部

写真 3
(b) 操作・制御部

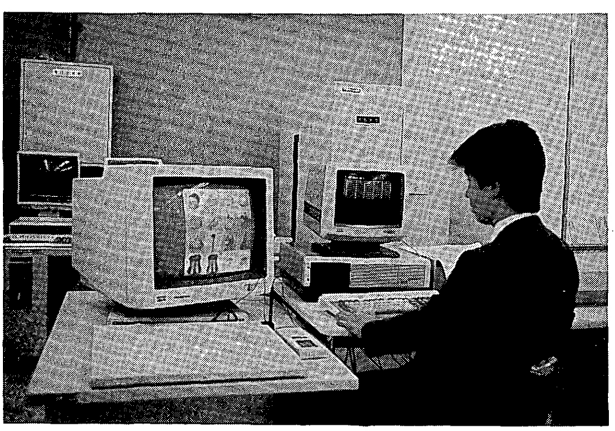

写真 4
標本画像蓄積検索装置
(検索・表示部)

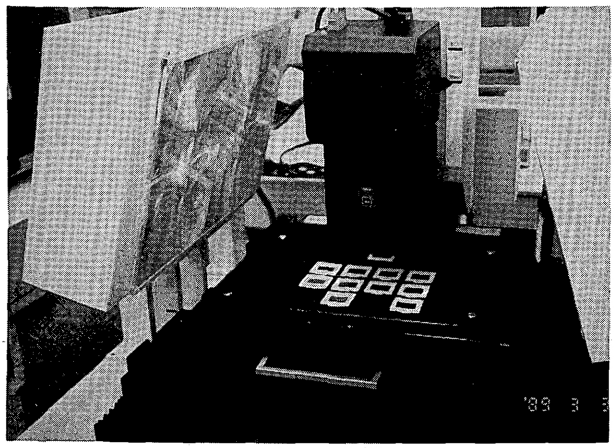

写真 5
写真・スライド画像蓄積
検索装置
(a) 入力部

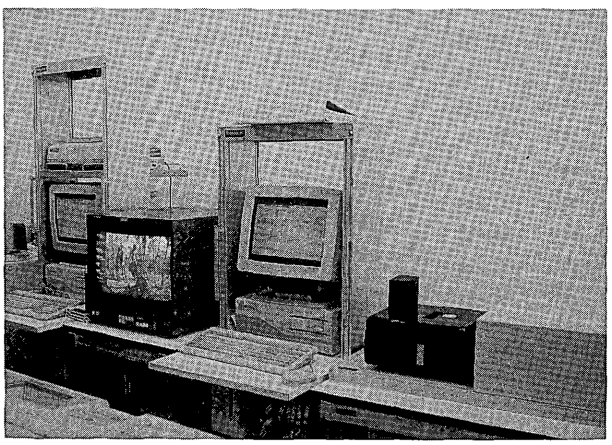

写真 5
(b) 第 2 入力部

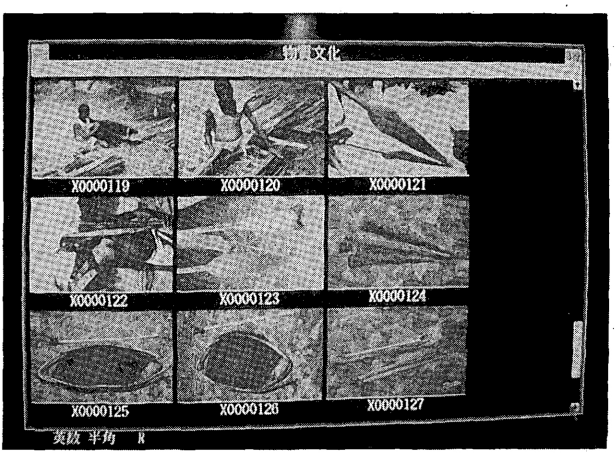

写真 5
(c) 検索・表示部

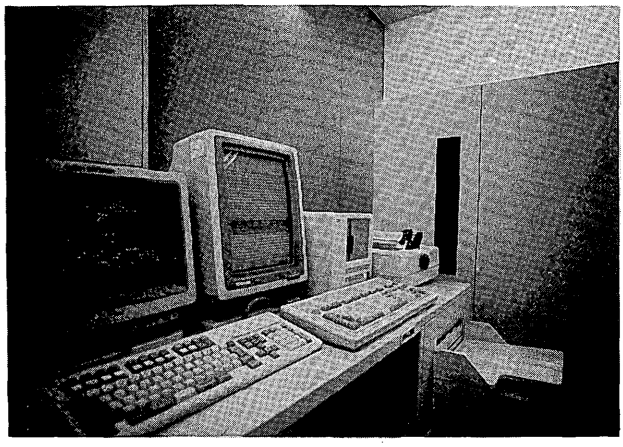

写真 6
ページイメージ蓄積検索装置

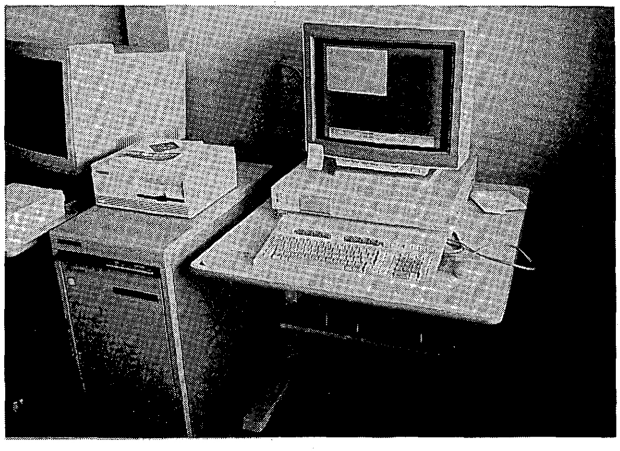

写真 7
ワークステーション
(a)

写真 7
(b)

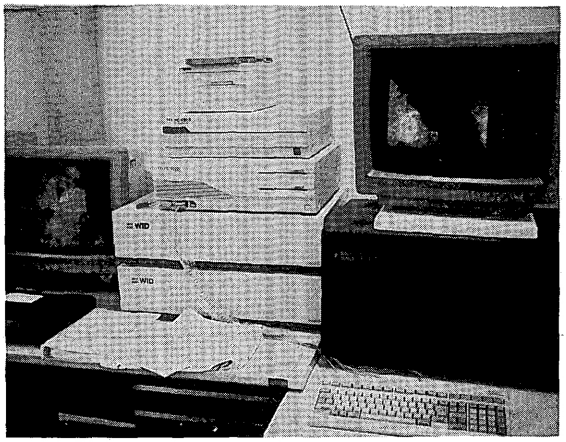

写真 8
画像・テキスト高速検索装置
(a)

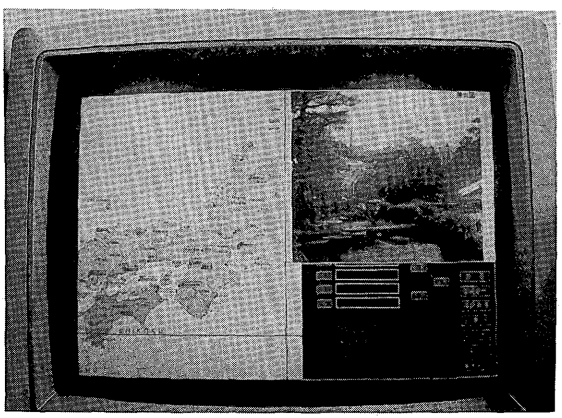

写真 8
(b)

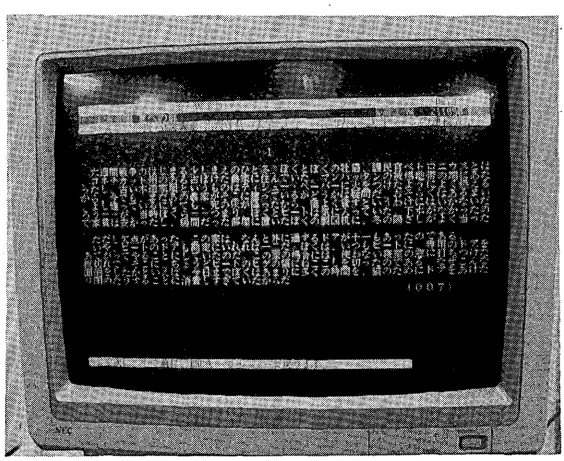

写真 8
(c)

てあったり，簡単な解説TV画面を使って表示されたりするだけであったが，民博ではそれらのモノが実際に生活でどのように使われているのか，自然環境との関わりなども一目で分るように映像を提供している。しかも来観者が自身で好きなものを選択する仕掛を用意し，多様な目的に合うようなシステムを開発した。開館以来11年間はビデオテープを利用していたが，平成元年3月からはビデオディスクのシステムに更新された。

この新システムではキーワードによって番組の中のシーン単位で検索してそこだけを見ることができるソフトウェアも用意されている。

以上コンピューターシステムの概略について簡単に説明を施したが，これは主としてハードウェアの面からであった。実際にこれらが民族学の研究に役に立つかどうかは，使う側からの要求にどれほどマッチしたソフトウェアが存在しているかどうかで決まる。そこで次に人文科学の分野におけるコンピューターの利用の面からの問題点についての考察を試みることにする。

2　利用者から見たコンピューター

「人文科学とコンピュータ」という研究会が平成元年4月に情報処理学会において誕生した。学会ができてからほぼ30年になり，会員も3万人近い大学会ではあるが，今まではどちらかというと工学色が強かった。情報処理というよりもコンピューター技術を推進することが中心であった。ハードウェアにしろソフトウェアにしろ，それを作る側からの問題提起が多かったのではなかろうか。

コンピューターで何をやりたいかではなく，現在あるコンピューターでは何ができるかという発想が強いものであった。言語の翻訳や文字認識，音声認識などパターン認識の問題も，人工知能としてのコンピューターの開発をねらっていた。

その発想を変えてユーザーの論理からコンピューターに関わるシステムを考えようという動きが出てきたのである。つまり従来の方法がトップダウン的であったのに対してボトムアップ的アプローチへと変化したのである。この変化を生み出したのはTSS（タイムシェアリング）技術，広域ネットワークシステム，そしてマイクロコンピューターの出現であろう。これらによって従来コンピューター専門家のみが扱っていたのが，利用者層の拡大によって非専門家も参加できるようになったのである。

米国においては1960年代の後半からその傾向が見えているが，日本では1970年代の後半からであろう。約10年の差がある。しかし最近10年の間にワードプロセッサーや

パーソナルコンピューターが急激に普及しだして，一般大衆にとってもコンピューターが身近に感じられるようになってきたのである。

そこで利用者からの要求や不満があちこちから出るようになった。一般社会でもコンピューターに対する関心が深まり，それだけ意識も高まってきた。欧米ではタイプライターに親しみがあるからキーボードが自然にコンピューターにつながっていくが，日本では漢字など多数の文字がありそれをどうコンピューターに入力するかで大きな壁があった。しかし，かな漢字変換の方法によって漢字なども容易にコンピューターに入力できるということになって爆発的に広まってきたのである。

コンピューターの利用層の大衆化によって，今までにはなかった問題が発生した。それはコンピューターに関する利用方法について十分な知識を利用者が持っていることが期待できなくなったということである。しかし一般の人に対しては一から使用方法を勉強しなければ使えないというのでは普及はおぼつかない。エンドユーザーにとっては自分のしたいことのみに注意をしていればよいということでなければありがたみがない。テレビは単にチャンネルを押せばよく，電話も番号を押せばよいというようになっている。

ものには階層ということがある。あるいはユニットといえばよいかもしれない。そのユニットは一定の機能を持っているが，その内部がどのようになっているのか，そのユニット以下のレベルに対しては知らなくてもよいというようになっていると分りやすい。テレビがどのような原理でできているのか知らなくてもブラックボックスとしてのテレビがある。電子レンジも時間の設定の仕方のみ知っていればよい。ワープロ専用機は電子手帳などと同様，ある目的にとって必要な機能のみを分りやすくしたキーで表現しているので普及しだしている。

ところがコンピューターでわれわれが日常的に行なっているような知的生産の作業，写真を整理したり，AV（映像音響）資料を見たり聞いたりしていることをコンピューターでやろうとしたら一から開発していかねばならないのが現状である。コンピューターに対する一般的なことはOS（オペレーティングシステム）の発達によって随分楽になった。OSの善し悪しはどのような機能をユニット化しているかによる。CPUの命令体系にしてもRISC（縮小命令セットコンピューター）という概念が出ている。あまり細かいコマンドではかえって使いづらい。かといって大まかなコマンドでは小まわりがきかない。

知的生産のためのコンピューターとしてどのようなコマンドを用意すればよいか，応用的な事柄に関することはこれからの課題である。現在はまだ分りやすいコマンド

レベルでイメージや音響などが手軽に使えるようにはなっていない。パソコンレベルでのマルチメディアシステムの開発が待たれているところである。

そこでいかにして一般大衆，あるいは専門外の人にとっても使いやすいシステムにするか，ということが非常に重要な問題になってきたのである。このようなことは必ずしもコンピューターだけに限らない。自動車やコピー機，家電製品など不特定多数の人が関わる機器については要求される機能である。またエレクトロニクス関係のみならず政治，経済における制度や，都市の構造，建造物，また本や雑誌などすべてについて利用者にとって分りやすく，安全であり，精神的に落着いた気持で対応できることが求められている。このようなものの蓄積，それが文明というものである。従ってフレンドリーを追及することはより高度な文明を指向することにもなる。

2.1　人文系からの発想

コンピューターの利用に関しては，日本の文科系は遅れていた。それは日本の人文科学がどちらかというと論理実証的というよりは思弁的な方法を用いることが主流であったからであろう。しかし最近コンピューターの一般大衆への普及につれて，人文科学の世界においてもコンピューターを使いだす人が増えてきた。大型コンピューターではなくパーソナルコンピューターは個人主義の指向が強い文科系の人にとってなじみやすいものとして受けとられているのではなかろうか。

このように広いユーザー層の出現によって情報処理の分野も新しい変化に対応しなければならなくなってきたのである。しかしコンピューターのように様々な分野に活用されるものに対していちいちその分野特有の問題のみを対象にして議論していても，それはアカデミックな学会活動としてはどうかという面もあろう。確かに個別の問題としては意味が薄いが，かなり大きな集団に共通している問題があるとすればそれを明確にするのは意義のあることである。

情報処理学会を狭い意味に考えるのではなく，例えばそれは総合百貨店であると考えることができよう。いろいろな対象をカバーしていることが望ましいのである。特に情報処理というものは一般論ではなく実際に情報が発生し，それを処理する現場特有の問題を抜きにしては意味がない場合が多い。そういう現場感覚を持ちながら，しかし学問的に面白い問題を発掘していくことが重要であろう。

人文系はどちらかというとコンピューターや数理に弱いという気持の人が多い集団である。その分野の人がどのような目でコンピューターを見ているか，またどのような問題を抱えているかを知ることは作る側にとって有用な情報となるであろう。また

それによって新しいシステムが作り出されるとしたらユーザーにとっても大変ありがたいことである。このような両者を結ぶチャネルが今まで制度として用意されていなかったのが不思議なくらいである。やっとこのたびそれが情報処理学会内に「人文科学とコンピュータ」研究会の発足という形で実現する方向に一歩進んだのである。

2.2　マンマシンシステムの必要性

今一つ人文系でコンピューター利用があまり活発でなかった理由をあげるとすれば，その研究内容が実際問題として，コンピューターで処理できないような性質のものが多いということではなかろうか。しかしそれは問題全体としてすべてをコンピューター化できないというだけで，部分部分には利用可能なのであるが，人と機械がうまく共存して問題を解決していくようなシステムの開発ができていなかったのである。

特に民族学のような総合的な分野では，その対象として取り扱う情報およびその媒体の種類は，極めて多岐にわたっている。数値や文字情報のみならず，写真，スライド，フィルム，音楽，音声，モノなど，しかもそれらから新しい情報を引き出して命題を構築するのであり，情報処理は最終目的ではない。おのずとそこに工学的情報処理の開発目的と，人文系での利用者に対する情報処理技術には異なった観点が出てくる。

2.3　コンピューター利用の比較

工学系と人文系におけるコンピューター利用の特徴を対比させると次のようなことがいえるであろう。

情報処理　vs　知的生産
解析　vs　統合
大量少種類　vs　少量多種類
汎用　vs　個別
多重利用　vs　単発利用

技術系においては実験などから大量に集められたデータに対して統計処理などを施してある一つの量に集約させる，あたかも大量のゴミを処理するようなやり方に対して，人文系では観察された極少ない情報から人間生活の深層構造を明らかにしようとする知的活動が中心であるという違いがある。工学系がデータの解析を行なうのに対

して，人文系では様々な種類のデータを総合的に考えることが重要であるという違いがある。実験系ではデータを変えて同じプログラムを何回も使うのに対して，知的生産においてはある性質を検証すればそのプログラムはもはや用はないという，使い捨てに近い場合が多いという違いもある。

2.4　フレンドリーシステム

コンピューターが様々な分野で活用され，その使用者の層も多様になって，過剰な期待が持たれたり，また誤解されたりしている。いわゆるシステムの開発者のレベルに関係する情報と，エンドユーザーに関係するところが混乱している。どのようなOS がよいかということはエンドユーザーにとってはそれほど重要なことではない。ソフトを開発する立場からは OS の選択がプログラム作成の容易さや効率に関係するから重大関心事ではあろう。エンドユーザーにとってはその上で開発された応用ソフトウェアのインターフェースがいかにうまく工夫されているかが問題である。

また最近では人工知能が再び脚光をあびているが，文字認識や音声認識などはエキスパートシステムとは異なる問題意識であるし，また OA (Office Automation) や HA (Home Automation) などにおける AI (Artificial Intelligence) は少し違った機能を持っている。にもかかわらず十把ひとからげに論じられている場合が多い。新聞や雑誌では様々な AI システムの広告がなされているが，その多くは特に AI というに値しない内容のものである。AI がブームであるからそれに便乗した表現であり，また開発過程において OS に UNIX を使い，言語として PROLOG や LISP を使うとなんとなく AI といいたくなるようであるが，安易な考え方である。

人文科学研究の分野におけるコンピューター利用については，AI というよりはいかにして研究者の持っている知識を引き出すか，あるいはばらばらな情報を統合させて新たな知識を生み出すか，そのための道具の開発が重要である。コンピューターが出現するまでの長い間，人は手作業でいろいろなことをやってきた。文献を読み，カードを作成し，計算やグラフ化を行ない，写真をファイルしたり，そのようにして仮説を頭のなかに構築したりしてきたのである。このプロセスをコンピューターを活用して行なうことができないかというのが一番の問題である。

文科系の研究者はどちらかといえば機械類を嫌うタイプの人が多い。恐怖感を与えず安心して自然なやり方で使えるようなコンピューターシステムの開発が望まれているのである。機械が勝手に判断して何かをやるようなタイプの AI よりも，何かやりたいことが自然な操作ですぐできるようなフレンドリーシステムが求められているの

である。

その内容は必ずしも高級な事柄ではない。しかし大量のデータから必要なものを探すということは単純ではあっても実際に役に立つものである。特に画像や音響データなどの場合，単に所在情報だけではなくその内容そのものが映像や音響として目や耳で聞くことができれば知的生産にとって大いに役立つ。文字情報だけではなく映像・音響などが総合的に関係付けられたマルチメディアシステムの構築が重要となるのである。

2.5　コンピューターに望むもの

コンピューターが日本に導入されてからもう数十年になるが未だに当初からの形式に引きずられており，新しい展開ができていないように思われる。それは専門家だけのものであったからで，これからは急激な変化を示すに違いない。ちょっと気が付くことだけでも次のようなことがある。

- 面倒なキーボード
- 役に立たないマニュアル
- 不自然な形式での対話
- 一覧性を殺す小さな表示部
- 遅い画像情報処理

コンピューターの知識のない人はいくらそのマニュアルを読んでもどうしたらよいか分らないものである。既に他のコンピューターを使ったことのある人でも他社の製品だと分らないところが多い。本当に素人にも分るマニュアルというのはほとんどない。誰かに説明してもらうと分るのにマニュアルだけでは分らないというのはどこかに欠点があるのであろう。本当はマニュアルなどなしでも使えるようになっていなければいけないのである。

2.6　何が欠けているのか

現在のコンピューターシステムは満足できるものになっているであろうか。ゲームやワープロにおいては多くの人がある程度満足しているが，いざ実際に今まで人間がたやすくやっていたことをコンピューターの下で自動化しようとすると，うまくできない問題があまりに多いのに気付くであろう。

コンピューターは先ず数値や文字を扱う道具として発達してきた。しかしわれわれ

の日常生活では画像や音響情報が普通である。にもかかわらず現在のコンピューターは画像や音響の扱いに向いていない。文字や数値の延長線上から考えているからであろう。むしろ画像から発想して新しいシステムを開発すべきであろう。

コンピューターの出現した1950年頃から既にパターン認識や学習などが研究されたが，結局まだたいしたことはできていない。最近再び AI ということが話題になっているが，いかにハードウェアが進歩したとはいえ人間の知的活動の機構についての知識はまだあまり進展していない。

狭い意味での問題に絞ってみても現在のコンピューターには欠けているものが多い。例えば，

- 文字，画像，音響の自然な入出力，
- 関連情報の同時表示，
- 五感との連動，
- 運動感覚と知覚感覚との連動，
- 予備知識なしで使えるシステム，

など。これらを解決しなければフレンドリーとはいい難い。

3　人文系コンピューターの環境整備

3.1　マルチメディアシステム

人文系では数値化することがむずかしいデータが多い。写真や，考古発掘資料，美術作品，音楽，音声などそのまま対象として扱うことが普通である。現在のコンピューターではそのままでは扱えない場合があり，ましてやパソコンレベルではできない場合が多いので，やむなく文字・記号で代理させているのが現状である。しかし人文系の研究には知的生産の道具として，直感的なヒラメキを刺激するような情報を提供するシステムが必要である。この点からも本来人文系のコンピューターシステムはマルチメディアシステムであるべきなのである。

3.2　自然な入力

従来コンピューターへの情報入力はほとんどがキーボードからであった。しかし人文系のデータはアナログ的なものが多いのでそれらを直接入力できるようになっていないと使いづらい。CCD カメラによる画像の入力や，音響情報の入力も利用者は単

に資料をセットするだけですむようなシステムが開発されていなければならない。その扱いが容易であればまた新しい応用が考え出されるものなのである。必ずしも機械による自動認識でなくてもよい。録音テープやVTR，映画などの資料から，研究者が情報を選択しながら入力する時にやりやすい装置の開発が必要である。

美術館などで光ディスクを用いて収蔵品やカタログの作品を蓄積し検索しているところも出てきている。テレビカメラから直接ビデオディスクに入力する簡便な装置が利用できるようになってきたからである。これをさらに安価にし，かつ画像の質を良くすることに努める必要がある。ただデータとして画像が入ればよいというのではなく，写真と同じくらいの質を持っていないと満足できない。

3.3　シソーラス

コンピューターを使って検索をしたり，言語を扱う場合に一番問題になることは，同類語の扱いである。人間であれば必ずしも同一の表現でなくても，意味的に似た単語であれば一致したと判断する。しかしコンピューターにはそのような常識が与えられておらず，普通の場合，勺子定規な一対一の完全一致でしか検索しないから，漏れてしまう情報が多くなる。これを救うにはシソーラスというものを導入することが最低必要である。シソーラスは一種の語彙集であるが，同義語関連語などを定義した辞書である。特に民族学で使用する用語を中心としたシソーラスを開発する必要がある。これは今はやりの人工知能やエキスパートシステムの問題ともつながるもので，コンピューターにインテリジェンスを持たせる第一歩である。

例えば「湯飲み」を検索しようとする場合，データとして他の表現で記述されているかもしれない。「コップ」とか「茶碗」，「杯」，「グラス」などいろいろな可能性がある。それはデータを作る人が全体のことを考えて名称などを付けているのではなく，自分の分野からの判断のみに基づいているからである。工芸や芸術の分野ではある程度共通の命名法があるようであるが，民族学での対象は日常雑貨に等しく，定まった方式がない。その土地・民族によってばらばらであるのが普通である。

ワープロの普及はかな漢字変換辞書の出現によって可能になった。人文系のコンピューターが普及するには自然言語の扱いが容易になることであり，その為にはシソーラスが用意されることが第一歩である。一般語彙，ある分野毎の辞書を作成することが是非必要である。

3.4　データの相互利用

文科系の場合，図書・雑誌論文について，自分の専門領域の中心的な部分は，各自がよく把握しており，それを検索することはあまり必要でないかもしれない。しかしふだん目にとめないところで出ているものや，別の分野のものまでは手がまわらない。国の内外，広範な分野をカバーして常に全体が把握できる状態にあることが，先端レベルの新しい研究をする上では重要である。

特に民族学研究は人間生活全般にわたっており，その対象とする事柄は非常に広い分野となっている。また一人の研究者が調査できるフィールドは限られている。自分の知らないフィールドのデータをいかにうまく活用するか。そのためにはコンピューターを中心において，各自のデータが相互に活用できるシステムが必要になることは明白である。民族学は単にフィールドの記録（エスノグラフィー，民族誌）を書くことだけではなく，諸民族を比較することによって，その異同を生じさせている要因，自然環境との関係や民族の交流などを明らかにしようとするものである。そのためには自分のフィールドの知識を持ちながら，他の地域・民族のデータを活用することが必要になり，コンピューターによる情報検索技術の活用が重要な役割を演じるのである。

この場合特に検索の手掛りが問題である。自分の分野であればどのような概念が使用されているかが分るが，他の分野になるとどのような表現で検索すればよいのか見当が付かないことがある。データはそれぞれの分野で作成されるので共通のタームが期待できない。このためにもシソーラスが重要である。

3.5　分類をしない

データベースを作成する時に先ず問題となることは，資料に対してどのような分類をするかということであろう。既に体系のできている図書などの場合は一応それに従っておくのもよいが，新しい分野ではここでひっかかる場合が多い。しかし所詮事柄を完全に分類することは不可能である。データにはいわゆる分類コードとか，キーワード，ディスクリプターとか呼ばれるものを付けておくのではなく，あまり分類にこだわらずにデータを作成するのがよい。むしろ自由な記述をしておき，そこに含まれる全ての単語が検索の対象になるようにしておく方が自然である。

今まで分類ということに大きなウエイトが置かれていたのは，情報の利用の観点からではなく，情報を管理する立場からの発想であったように思われる。分類体系がよ

く分っている管理者には分類番号で整理し検索する方が楽であろう。しかしその体系を知らない利用者にとっては，思いつくままに単語で検索できることが望ましい。また実物を配架する場合にはある程度分類されていると，直接棚を探す場合に便利であるから分類が必要になる。しかし徹底的にコンピューターを活用したシステムにするならば分類ということをあまり考えなくてもよいのではなかろうか。

3.6　漏れのない検索

コンピューターを使った情報検索で重要なことは，該当するデータ，あるいは何等かの関係があるデータが漏れなく検索されることである。その場合必ずしも必要最小限のデータのみでなく，余分なものまで検索されることがある。それが少ないことは望ましいことではあるが，しかし必要なものが欠落していることのほうが問題は大きい。対象が分類されており，その分類記号の一致で検索される場合は，曖昧さはないが，文献検索や標本資料検索，映像音響資料検索などにおいて，自然言語で書かれた情報を検索する場合，語の意味の一致に関しては人間の判断に頼らねばならない場合が多い。コンピューター利用では効率や省力を重要視するために，無駄な努力を極力省こうとするが，カードシステムなどで余分な部分でも繰っているうちに，関係のある資料を発見したり，またそのときには役立たないが，思わぬ情報に気が付くことがある。コンピューターによる情報検索においても，ある程度範囲を絞り込んだら，あとは人間が選択するようにし，そのための使いやすいシステムを用意する方式とすべきである。

3.7　融通のある検索

従来のカード方式は，人間の頭と目と手が連動したシステムで，それなりに有効な面も多い。しかし量の増加にどう対応するか，また配列の仕方に限界がある。コンピューターによる情報検索では，従来のカード方式ではできないような検索を可能にするようにしなければ意味がない。例えば図書カードに含まれているすべての語句，あるいはその一部からでも検索可能になっていることが必要である。正確にタイトルや著者名，出版社などを覚えているとは限らないからである。

現在のところ，単語の完全一致のみならず，前方一致や後方一致を可能にするシステムはあるが，中間一致を許すものは小規模のものにしかない。さらに完全一致ではなく，2・3文字異なっていても他に一致するものがなければ候補として検索するような柔軟なシステムは皆無に近い。また自動分ち書き，自動読み振りも最近では辞書

を持つことにより可能になりつつあるが，まだ信頼性に乏しい。むしろ分ち書きをしないで任意の文字列で一致がとれる方式が望ましい。

3.8　画像による画像の検索

シソーラスを併用しても，画像データの場合などの検索には限界がある。インデックスが明確に付けられるものはいいが，画像を印象で記憶している場合には言葉で表現しにくい。サンプルとなる写真や画像を与えて，それとよく似たものを検索せよという指示ができるとよい。これが実現するためには情報工学の分野の研究がもっと進む必要がある。そのとき実際のデータを持っている者が参画しないとうまくいかない。

標本資料をその名称や収集地など記述された文字列から検索するのではなく，その形態や属性，例えば色などからも検索したい場合がある。これには同時に多数の映像を表示して見ている人が自分で選ぶという方法もあるが，できればサンプルを示してそれの類似物が検索できればよりフレンドリーなシステムといえよう。

まだ初歩的な実験の段階にすぎないが，今回の共同研究においても画像の類似検索をする試験システムを構築した。一つは，予め輪郭線を抽出しておき，それをある方法でコード化し，入力例の輪郭線コードとの一致の度合を計算することで，形態の似た画像を検索する試みである。もう一つは色の似た画像を検索する試みである。仮面などは世界中に存在しており，その形態は千差万別である。そこで形態ではなく色で検索したいという場合もある。赤が多く使われている仮面とか，黒っぽいもの，白っぽいものなどの指定で，該当するものがその類似の程度の順に表示されるようにしておき，あとは人が直接表示されたものを見て判断するというシステムも実験した。

3.9　マルチメディア端末

幾種類かのデータベースがある場合に，それらが単独で検索できるだけでは効果は少ない。文字，画像，音響などが連動していることが重要である。それは関係のある情報が同時に検索できることでもある。その場合一つの表示画面ではなく，複数の情報が同時に表示できる新しいワークステーションが必要である。

また文字情報のみならず画像情報に対しても高速に検索され，それを好きなように加工し表示できるシステムになっている必要がある。全体のデータベースから自分の研究に必要なデータのみを抜き出して，個人的なデータベースを構築し，自由に変形加工できるようになっていることも必要であろう。

現在のコンピューターネットワークでは画像データを異機種間で高速に伝送することがむずかしいが，今より100倍くらいのスピードが欲しいものである。そのときには従来のカード方式，台帳方式で直接資料に触れることにより得られる情報，つまり目と運動器官の連動や，目的以外の情報の獲得に相当するものが，コンピューターシステムにおいても実現できるような工夫が必要である。

4　今後の計画

地図データベース

民族学では世界の地図を活用する仕掛が役に立つ。地名や緯度・経度から該当する領域の地図が表示されるシステムである。人工衛星から撮られたデータとも関連させ，巨視的な考察と解析ができ，また様々な文化要素との重ね合せ，いわゆるマッピングの機能も含むような装置の開発も必要ではないかと考えている。

疑似開架図書検索

コンピューターによる情報検索の欠点は，はっきりとは必要な資料が分っておらず，書棚や収蔵庫の棚を直接見ることにより，関係のありそうなものを探すという体験ができないことである。ファイルを開いてこのあたりであったというような，実感的な記憶に基づく検索が不可能である。そこで，このような実物に接しているかの如き状況を作り出すことも重要であろう。

例えば書棚に本が並んでいるその映像を表示するというのはいかがであろう。あたかも書庫の中を歩きまわっているように，次から次へと書棚の映像を動かしていくのである。本の背中を見ているだけであるが，書誌的事項以上の情報を与えるに違いない。必要な本に対しては，図書番号によって，さらに詳しい情報がデータベースの方から得られる。標本資料についても収蔵庫の各棚を順次見てまわっているような映像を見ながら，実際の標本資料の姿を見て判断ができる。一点一点の標本資料の番号は映像では見えないが，棚番号が分れば，そこに収納されているものの情報は取り出すことができるから，対応させて知ることができる。

音響データベース

次のデータベースは音そのものの蓄積と検索である。この場合，検索結果は音で聞けることになる。音と画像とではディジタル化してしまえばそれほど差はない。文字データベースと連動させて，例えば「インドネシアの子守り唄」を検索したとすれば，

その曲がすぐ聞けることになる。これは平成5年3月に実現する予定である。

ネットワーク

現在は館内からであればLANによってマルチメディアデータベースに接近できるが，外部からは利用できない。文字データであれば電話回線を介してアクセスも可能であるが，まだ正式には公開していない。近いうちにディジタル回線網との接続を行ない，他の機関とも接続し，広く全国から利用できるようにしたいと考えている。