

特定領域「人文科学とコンピュータ」からの報告

及川 昭文

総合研究大学院大学 教授

科学研究費特定領域研究「人文科学とコンピュータ コンピュータ支援による人文科学研究の推進 -」(以下、じんもんこん)は、1995(平成7)年度より4年間の予定で研究を行ってきた。この特定領域はその副題にあるように『コンピュータ』と『情報科学(的アプローチ)』を核にして、人文科学研究の活性化、新しい研究の視点・手法の確立を目指したものである。この特定領域には実に多くの研究グループが参加しており、活発な研究活動を展開してきた。それぞれの研究グループの成果を問うにはまだ早い、領域全体としてこの4年間に何を目標として、何をやり、何ができたか、そして何が課題として残ったか等について報告する。

1. じんもんこんの目指したもの

特定領域「人文科学とコンピュータ」の背景には何があり、具体的にどのようなことを目標としていたのかについて、領域申請にあたって作成した「領域申請書」(1994年2月に作成)から、抜粋して引用する。

1.1 何が問題なのか 研究の背景

情報化社会、情報化時代という言葉が日常的に使われるようになって久しいが、これを支えているのは、コンピュータや通信技術の発達、広い意味での情報処理技術の進歩である。コンピュータの進歩、とくにパソコン(パーソナル・コンピュータ)の高性能化、高機能化、低価格化はまさに日進月歩で社会の隅々までその普及は著しく、今やコンピュータのない生活は考えられない。研究という世界においても同様で、理工学や医学等の自然科学の分野は言うに及ばず、これまでコンピュータや数理というものからはもっとも縁遠いとされていた人文科学の分野においても、コンピュータを積極的に利用した研究が増えつつある。

ワープロ(ワードプロセッサ)をはじめとして、今やコンピュータは紙や鉛筆と同じように、研究を行う上で必要不可欠な道具として考えるべき時代が始まったと言える。ほんの十年ほど前までは、高価なコンピュータと情報の専門家なしでは到底できなかったデータベースの作成や多変量解析といった高度な数量的分析も、数万円のソフトウェアを購入するだけで自分で容易にできるようになっている。

コンピュータが人文科学研究の発展に大きく寄与し、新し

い研究手法を拓く道具として役立つことは、これまでの先駆的な研究から見て明らかである。しかしながら、問題がまったくないわけではない。

道具としてのコンピュータ

コンピュータは人間が作り出した最も優れた「道具」のひとつであることは万人が認めるところであろう。しかし、便利な道具であり今日の社会生活の必需品である自動車の時として「走る凶器」となるようにコンピュータにも「マイナス」の部分がないわけではない。例えば、ワープロを日常的に使用することによって、「漢字が書けない」「辞書で新しい言葉を探して使うことをしない(漢字変換辞書に含まれている言葉のみを使ってしまう)」というような問題が起きていると指摘されている。また、数量的分析を行った研究の中には、電卓の代わりにコンピュータを使ってこれが新しい研究手法ですといっているような研究もあり、あるいは統計の本質を十分に理解しないまま手法だけを借用し、コンピュータが出した結果はすべて正しいとしているものも少なくない。これは安価なパソコンと初心者でも使える統計ソフトウェアパッケージが容易に手にはいるために、単純な統計解析から「高度な多変量解析まで簡単に行うことができちゃう」ということにも、ひとつの要因があると考えられることでもある。

コンピュータのハード・ソフトウェアの進歩のスピードには驚くべきものがあるが、そのスピードに研究の方が追いついていかない現状がある。新しいハードウェアやソフトウェアが開発されるたびに、それらに対応するために研究手法やスタイルを変えるということが日常的に起こっている。本来人文科学の研究を行うためにコンピュータやデータベースを道具として利用していたのが、その道具の方に研究の視点に移ってしまっている場合が少なくない。確かに新しいハードウェアやソフトウェアを利用した方が、研究の効率も良く、またこれまでにない研究手法を実現できるということはある。しかし、その結果本来の研究の成果が明らかにならないまま、次の試みに取り掛かるということになってしまっている。つまり、ひとつの研究が中途半端なまま、あるいは未成熟のまま、道具に振り回されているということがある。

これらはいずれもコンピュータ側の問題というよりは、それを使う側の人間の問題である。コンピュータというものは、これまで人が作ってきた道具のように、それをどのようなときに、どのように使うかが決められている道具ではなく、その用途は使う側で決めなければならない道具である。したがって、適切な場面で、正しく利用することが、これまでの道具以上に重要になってくる。このことを忘れてただ便利だからといってコンピュータを利用すると、時として大きな失敗につながるということが少なくない。

データベース流通体制の不備と標準化の遅れ

一般的に人文科学系の研究においては、資料(史料)の収集や整理に多大の時間と労力を費やしており、それ自体が研究の目的というような側面もある。これはコンピュータを利

用して省力化や効率化を図ってもあまり変わっていないようである。この結果、例えばデータベースを作っても他の人に使わせたくない、提供したくないということが起こっている。このような研究者の情動的側面だけでなく、流通のための環境が整っていないということが人文科学系のデータベースが広く流通されていないことの大きな理由ではあるが、資源の共有ということを目指して作られるべきデータベースが、個人のものとしてだけ存在し、広く流通されないことは、研究の進歩を阻害することにつながっていくものである。

データベースの流通を容易にするためには、データベースの構造や項目についての標準化を図るということも重要な要素である。ただ、人文科学系の研究では、思弁的作業が大きな比重を占めること、また研究方法が個別的かつ非定型であることから「標準化」ということに対しては大きな抵抗、問題がある。つまりデータベースの標準化ということが研究の標準化、均質化につながる恐れがある。生物界でも、種があまりに均質化してしまうと、突然の環境の変化等に対する抵抗力を欠いて、絶滅してしまうことがあるが、「研究の均質化」というものは研究の質の低下につながるもので、絶対に避けるべきことである。しかし、何らかの標準化が図られなければ、データベースの共有ということはおぼつかない。この二律背反をいかにして解決するかということは大きな課題である。

最先端技術とのギャップ

最近の映画やTVコマーシャル等で使われているCG（コンピュータ・グラフィックス）画像を観ると、以前のそれと比べて格段の差があることがはっきりわかり、いかにCG技術の進歩が著しいかが実感できる。最近では仮想現実感（バーチャル・リアリティ）といわれる新しいCG技術の開発も進められており、その勢いはとどまるところを知らない。また、自然言語処理の分野でも、ひと昔前に比べて飛躍的な進歩を成し遂げており、自動翻訳も決して夢ではないレベルまで達している。このようにコンピュータのハード・ソフトの最先端技術のレベルは一般に考えられているよりはるかに高いレベルに達しているが、人文科学系の研究の分野におけるコンピュータ利用に目を向けてみると、これらの諸技術が十分に活用されているかということ、そうではない状況がある。すなわち、現在コンピュータを利用した人文科学系の研究をコンピュータあるいは情報科学という技術的側面からみても、そこで利用されている技術は最先端からはほど遠く、初歩的なレベルでしかないという事実がある。これは非定型な部分が多く、簡単にコンピュータになじまないところがあるという人文科学系の研究に固有な特徴に起因することでもあるが、その他の要因としては以下のようなことが考えられる。まず研究者側の問題として、

- ・コンピュータに対する無理解、偏見からくるコンピュータ利用に関する無関心。
- ・コンピュータ利用に積極的な少数の先駆的な研究者を除いて、コンピュータ利用の歴史は浅く、また研究者の数も増加の傾向にはあるがその絶対数はまだまだ少数である。

- ・コンピュータ利用に関する研究情報や技術情報の入手が難しい。等があり、開発者側の問題としては、
- ・開発重視であり、性能や機能の向上にのみ関心を持ち、それらのハードやソフトがどのように使われるかについては十分な考慮が払われていない。
- ・開発者はソフトの開発は行っても、その利用者にはならないため、「利用」ということに対する評価が不十分である。
- ・これまで人文科学系の利用者のニーズに対して十分な関心を持たなかった。

等を挙げることができる。このように人文科学系の研究者とコンピュータとの間には大きな溝が存在しており、それを埋めるには人文科学系および情報処理分野の研究者たちによる幅広い交流、協力の下での研究が推進されなければならない。

まとめ

コンピュータを研究の道具として活用を始めたのは理工系の分野で、ソフトウェアの開発も彼ら自身の手で進められてきた。それは研究の道具であり、また研究の対象でもあるということができる。そのようにして開発されてきたハードウェアやソフトウェアを人文科学系の研究者が道具として利用するには、研究の対象も方法も大きく異なるが故に、改善されつつあるとはいえ、まだまだ未成熟な状態にある。言い換えればこれまで述べてきたような問題を含め『人文科学分野における道具としてのコンピュータ利用のための研究基盤（インフラストラクチャ）の未整備』ということができる。

1.2 何が必要か 研究の目標

これまで「人文科学研究の推進」ということが、学術審議会等を含めいろいろところで議論されているが、具体的な方策となると実効力のある形ではなかなか提案されてこなかった。実際のところ人文科学といってもその対象分野は哲学、文学、芸術、宗教、歴史、考古学、民族学等実に多種多様で、具体的な方策を立てにくい側面があることは事実であり、したがって「人文科学研究の振興につとめましょう」といったスローガンを主張した精神論に終始していた感を否めない。このような中で、特定領域研究「人文科学とコンピュータ」は『コンピュータ』と『情報科学（的アプローチ）』という、いわばカンフル剤を人文科学に注入し、人文科学研究の活性化、支援を目指すものである。

「コンピュータ」「情報科学」がカンフル剤として有効に機能するためには、人文科学研究のコンピュータ利用の現状を把握する必要があるが、人文科学研究のコンピュータ利用の段階を大きく分けると図1のようになる。

既にレベルⅢに達している先駆的な研究もあるが、それはまだまだ少数であり、大部分はレベルⅠからⅡへの移行期にあると考えられる。したがって、その促進を図りレベルⅢへの移行を容易にすることを目標に、研究基盤の整備を推進する必要がある。

具体的には、次頁の図に見られるようにまず研究基盤の整備として「データベースの流通促進」「ソフトウェアの開発」を図り、次に個々の研究の活性化を図るための情報の収集・

レベル I	レベル II	レベル III
<p>これまで手作業で行っていたことを単純にコンピュータで代用させる。</p> <p>ワープロの活用, テキスト処理, 初歩的統計処理, ……</p> <p>⇨データベースの作成</p>	<p>手作業では困難であったことをコンピュータを利用することによって, より効率化, 省力化を図る。</p> <p>パソコンの活用, イメージや音声の処理, 高度な統計処理, ……</p> <p>⇨データベースの活用</p>	<p>従来の方法では不可能であったことを, コンピュータを利用して可能にする。</p> <p>マルチメディア処理, 多変量解析, ネットワークの活用, ……</p> <p>⇨データベースの流通</p>

図1 人文科学研究におけるコンピュータ利用の段階

提供を中心とした研究支援システムの開発, 研究支援活動の推進を図る。

(1) データベースの流通促進

人文科学研究におけるコンピュータ利用は, まずこれまで研究の対象としていた諸々の資料(史料)をコンピュータに入力しデータベースとすることから始まるといってよいだろう。図2からも分かるように, コンピュータ利用はデータベースを軸に行われている。テキスト類のデータベースが中心であるが, 既に多くの研究者によって多種多様なデータベースが, まさに百花繚乱のごとく作られている。ところが, これらのデータベースの流通, いいかえれば情報の共有がスムーズに行われていないという現状がある。その要因としては,

- ・同じテキスト, 例えば「源氏物語」であっても個々の研究者の研究の視点や方法は異なっており, それらに基づいてデータベースが作られており, 他の研究者にとっては使いづらいものになっていることが多い。
- ・膨大な時間と手間をかけて作成したデータベースを他の研究者には提供したくないといった心情的な側面がある。
- ・データベースや情報科学に関する知識や経験が十分でない人文科学系の研究者が作成しているため, データベースの構造や項目の構成等への配慮が不十分であるといった技術的問題がある。
- ・研究者に負担をかけないで流通をスムーズに行うための機構(体制)が整備されていない。

等がある。いずれも単純に解決できることではないが, この問題の解決を図らなければ今後の発展は望めない。

データベースの共有化

個々に作られたデータベースをお互いに共有するためには, 共通のルールが必要である。すなわち標準化というものが必要になってくる。欧米においてはテキストの全文データベースの標準化を目指して, SGML (Standard Generalized Markup Language)^{*1} が提案されたり, TEI (Text Encoding Initiative)^{*2} 等による標準化の活動が進められている。SGML は電子出版を強く意識したもので, テキスト全文データベー

スの物理的・論理的構造の詳細にわたっての規格を定め, その標準化を図ろうとしているものであり, これに準拠したデータベースも作られ始めており, 日本にも数年前から紹介され検討が進められている。SGMLによる標準化やTEI等の活動についての対応は必要であるが, 分かち書きやふりがなといった日本語固有の問題等研究すべき課題も多く存在している。また, 個々の研究者の知的生産物としてのデータベースの標準化が果たして可能かというような問題もあり, 本特定領域ではSGMLのような厳密な標準化ではなく, まず流通できることを目的とした緩やかな標準化を図り, その上でデータベースの共有化を実現すべきであると考えている。

その詳細については, すでに作られている個々のデータベースの調査・検討等をふまえた上で決定されることになるが, 基本的にはレコードの最大サイズ, 各項目の記録形式等の物理的構造についての標準化を中心に, 文字コードの統一, 外字の取り扱いについてのルール設定等, 流通を目的とした必要最小限の規則に基づいたものになると考えている。したがって, ここでは「標準化」という言葉より「共有化」と呼ぶことにする。

データベース流通のための基盤整備

すでに述べたようにデータベースの自由な流通を妨げている最大の障壁は, 流通体制の不備である。理想的にはデータベース流通センターともいえるべき, データベースの収集と集中管理, そして配布等を主たる業務とする機関(組織)の設立が望ましい。しかし, 早期実現が望むべくもない現状では, そのような機関の設立を将来の目標として描きつつ, 現状の中で最適な仕組みを作っていかなければならない。それがどのような形態になるのかは検討すべき事が多く, 今ここで具体的に述べることはできないが, 基本的には,

- ・データベースの所在情報の収集と集中管理, そして提供
- ・データベースの収集, 共有化
- ・総括班をセンターとして本特定領域参加者内での流通
- ・総括班, 計画班のメンバーを中心とした, センター機能を有する組織の設立

*1 「汎用マークアップ言語規約」と訳されており, ドキュメント標準化のためのメタ言語である。ISO (国際標準化機構) で採用されており, 1992年に日本でもJIS (X4151) によって制定された。

*2 1986年以来, ALLC (Association for Literary and Linguistic Computing), ACH (Association for Computers and the Humanities), 及びACL (Association for Computational Linguistics) の3学会が中心になって進められている電子テキストの標準化のための大規模プロジェクト。

コンピュータ支援による人文科学研究の推進

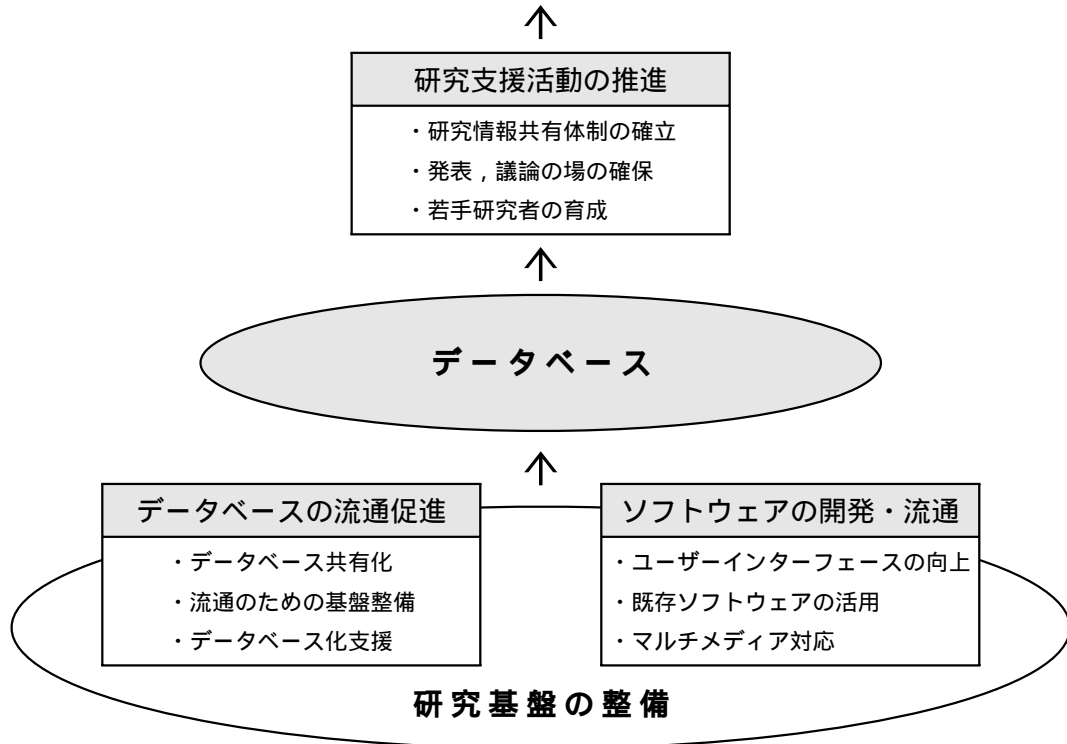


図2 特定領域「人文科学とコンピュータ」の研究目標

・上記組織を中心としたデータベースの収集，管理，提供のようなことを計画している。また，すべての研究者が『知的生産物であるデータベースは人類共通の財産であり，自由な流通が保証されなければならない』ということを理解し，お互いに協力し合うようになるように啓蒙活動もあわせて進めていく必要がある。

資料（史料）のデータベース化支援

人文科学系の大部分の研究者は自分でプログラムを開発できないため，少々不満があってもワープロや市販のデータベース・ソフトを利用してデータベースを作成している。また市販のソフトであっても，それなりの知識や経験なしでは使いこなせないところがあり，データベースを作りたくてもなかなか具体的な作業に着手できない研究者が数多く存在することも事実である。このような現状を改善していくためには，

- ・標準的なデータベース作成のためのソフトウェア開発
- ・一般的な索引やリストを作成するデータベース利用ユーティリティソフトウェアの開発
- ・専門分野や資料（史料）の種類に対応したデータベース作成のための手引き書の作成
- ・初心者を対象としたデータベース作成のための講習会やセミナーの開催

等を組織的，計画的に推進し，資料（史料）のデータベース化支援を積極的に，かつ継続して行っていくことが重要である。

(2) ソフトウェア開発

人文科学の分野では，その対象として取り扱う情報及びその媒体の種類は，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，モノそのもの等，極めて多岐にわたっている。そして，研究においてはそれらから新しい情報を引き出し命題を構築するのであり，情報処理は最終目的ではない。この点が工学系の研究における情報処理と大きく異なる点である。例えば，工学系においては実験等から大量に集められたデータに対して統計処理等を施して，ある一つの量に集約させるということが多いが，人文科学系では観察されたごく少ない情報から人あるいは社会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。また工学系ではデータの解析を行うのに対して，人文科学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もある。あるいは，プログラムに関していえば実験系ではデータを変えて同じプログラムを何度も使うのに対して，知的生産においてはある仮説を検証すればそのプログラムはもはや必要ではないという，使い捨てに近い場合が少なくない。

このように人文科学系と工学系では，ソフトウェアに対するニーズが大きく異なっているにも関わらず，もともとプログラムが工学系の人々によって開発されているため，人文科学系の人々にとって非常に使いづらいものになってしまっているという現状がある。最近ではマッキントッシュ等ユーザーインターフェースに優れた，コンピュータに不慣れた人であ

っても購入したその日からすぐ使えるようなものも増えつつあり、それなりの改善は進んでいる。しかし、人文科学系の研究者の立場からいえば、まだまだ改善すべき点は数多く存在している。本特定領域研究においては、上記のような問題をふまえた上で、コンピュータを単なる情報処理のための道具としてだけでなく、知的活動、知的生産を支援する道具として使えるためのソフトウェアの開発を目指す。実際の開発にあっては、

- ・人文科学系の研究者のニーズに基づいたソフトウェアの性能・機能、ユーザーインターフェースの向上
- ・既存ソフトウェア（市販ソフトウェアを含む）の活用
- ・マルチメディアを意識したソフトウェアの開発
- ・利用者と開発者の緊密な協力

等を基本方針として具体的な計画を立てる。この時、利用者のニーズの調査・集約、プログラムの設計等は研究者側で行い、プログラムの作成は開発後のメンテナンスを確実にするために原則として外注し、マニュアルの作成は研究者と開発者が共同で行なうことを予定している。また開発されたソフトウェアは、原則として誰でもが自由に使用できるものとし、マニュアルの整備、プログラムの配布等を確実に進めるような体制を作っていく。

(3) 研究者の支援

これまでの研究を概観してみると、知的生産を支援する道具としてコンピュータを利用している研究は、まだほとんどないといってもいいぐらいの数でしかない。コンピュータを思考を支援する道具とするためには、前述のデータベースの流通やソフトウェアの開発といった研究基盤の整備とともに、コンピュータを利用した研究そのものの活性を目指し、研究者コミュニティへの支援を図る必要がある。本特定領域においては、下図のような活動を通じて研究者の支援を図るが、このような支援活動は少数の研究者や小さなグループではなく、まさに『特定領域研究』というようなしっかりした組織が必要であり、また特定領域研究として計画的に進めていくことにより一層の効果が上がるものである。

研究情報共有体制の確立

この種の研究に関する情報は、その対象分野が多岐にわたるため、収集が非常に困難である。結果的にどこで、誰が、どのような研究をやっているのか、まったく分からないまま、同じようなデータベース作りやプログラム開発に苦労している例が少なくない。そこで、どこでどのようなデータベースが作られ、どのような研究手法でどのような研究が行われ、そしてどのような研究成果が出ているのか等についての情報を集め、フロッピー等の磁気媒体やネットワークを通じて研究者へ提供できる体制の確立を図る。また、研究の対象となっているデータやデータベース等の共有を目指した、最近話題になっているグループウェア指向のシステムについても検討を行う。

若手研究者の育成

本特定領域のような発展途上にある分野がより発展してい

くためには、研究者、とくに若い研究者の数を増やしていくことが必要条件である。入門書や解説書を作成し、それらを通じコンピュータを利用した研究にはどのような可能性があるのか、その方法、新しい研究の視点等の理解を助け、セミナーや講習会を開催しコンピュータ利用への足がかりを提供する。

研究発表・議論の場の提供

研究というものには独りよがりであってはならず、他の研究者からの批判にも十分に耐えうるものでなければならない。そのためには、お互いに研究について議論できる場が必要である。しかしながら、例えばそれぞれの専門分野の学会でコンピュータを利用した研究について発表しても、その人数は少数であったりしてなかなか議論を進展させることが難しいのが現状である。また、人文科学系と情報関連の研究者が議論できる場はそう多くはない。研究会やシンポジウムをできるだけ多く開催し、多くの研究発表の場を作り上げていく。

1.3 研究項目

どのような研究項目を立てるかは、本特定領域が所期の目的を達成することができるかどうかを左右するものである。多すぎても少なすぎても研究遂行上の障害を生む恐れがあり、計画研究との対応等を考慮しながら、研究が効率的に行えるような内容・数としなければならない。また、公募研究は、設定された研究項目のいずれかを選んで申請することになることから、申請者にとって理解しやすく、また申請しやすいということも重要な条件である。

本特定領域はその対象とする専門分野が多岐にわたるため、哲学、文学、美術、歴史、……と研究対象となる分野を基準にして研究項目を立てることは、その数が多くなり適当ではない。そこで、研究手法や情報処理技術の観点を基準にして、どのような手法や技術が応用されているかを、これまでの研究から探った。

まず、それぞれの研究の中で取り扱っているデータの種類を分類すると、おおむね「数値」「テキスト」「イメージ」の3つに分けることが可能である。最近ではこれらのデータをすべて処理対象とする、いわゆるマルチメディアに関する研究も増えつつあるが、それぞれの研究において用いられている研究手法や処理技術は、ほぼこれらのデータの種類に対応していると考えてよい。したがって、まずこれらの3種類のデータに対応した研究項目が考えられる。また、データベースは本特定領域において中心的な研究課題であることから、データベースに関する研究項目も必要である。しかし、データベースだけでは対象が漠然としてしまうことから、横断的な研究等を配慮しながらいくつかの項目への細分化が必要である。これらのことをふまえながら研究項目をリストアップし、それらを分類、整理した。その結果が表1で、大項目として研究項目A「データベース」と研究項目B「応用研究」の2つに分けた。

研究項目A「データベース」には、適用している研究手法や情報処理技術がどのようなものであれ、研究の中心がデー

項目A データベース	項目B 応用研究(ソフトウェア開発を含む)
A01 構築ツール開発 A02 応用(技術開発を含む) A03 共有化	B01 テキスト処理(文字処理を含む) B02 イメージ処理(マルチメディアを含む) B03 数量的分析(シミュレーションを含む)

表1 研究項目一覧

データベースにあるものが含まれる。この研究項目は「A01構築ツールの開発」「A02応用(技術開発を含む)」「A03共有化」の3項目から成るが、これらはいずれも深く関連しているので、それぞれに計画研究班を設けることを行わず、「データベース」としてひとつだけの計画研究とする。なお「A03共有化」については原則として研究の公募を行わず、「共有化ワーキンググループ」が中心になって研究を進める。

研究項目B「応用研究(ソフトウェア開発を含む)」は、情報処理技術の観点からの「B01テキスト処理(文字処理を含む)」「B02イメージ処理(マルチメディアを含む)」、研究手法からの観点の「B03数量的分析(シミュレーションを含む)」の3項目から成る。ここではB01, B02, B03のそれぞれに対応して合計3つの計画研究班を設ける。

1.4 計画研究班の目標

本特定領域では、表1の研究項目に基づいて4つの計画研究班を設定した。計画研究班は、それぞれの研究課題に取り組みとともに、研究項目ごとにシンポジウムを開催したり、また参加している公募研究班の指導や連絡調整を行う役目を担っている。

以下、それぞれの計画研究班の目標について述べる。

データベース班

本計画班の役割は、「データベースの共有化」「データベース構築ツールの開発」および「応用技術の開発」の3項目を基本とする。共有化と構築ツールの開発については、本重点領域全体にかかわる主要課題でもあり、前述のとおり、総括班のもとに組織されるワーキンググループが主体となって研究を推進する。このため、「データベース」班としては、「共有化ワーキンググループ」における研究が円滑に推進されるように、各研究分野の実情を配慮した具体的な提言を行うことを主要な任務とする。このため、さまざまな研究分野における具体的な問題点、とくに共有化におけるユーザ側の要望や技術の動向、および構築ツールについての分野によるニーズの差異などについての調査・研究を行う。

応用技術の開発については、非常に広い裾野をもつ人文科学において、できる限り多くの研究分野を視野にいれながら、研究の活性化を目指した新しい応用技術の可能性を追求する。応用技術の根幹をなす部分は、いわゆる情報科学(工学)的な技術になるが、これを単なる技術の視点から考えるのではなく、人文科学への応用という視点を合わせた複眼的な研究方針に基づいた調査・分析を行い、その結果を考察しながら具体的な開発を行っていく。また、本計画班の研究は、計画研究「テキスト処理」「イメージ処理」および「数量的分析」

とも密接に関係するため、適宜、研究情報の交換を行い研究の円滑な推進を目指す。

諸分野におけるデータベースの将来的な役割を考えた場合、単純なデータ形式を越えたマルチメディア的なデータベースこそが研究の支援効果を発揮できると考えられる。したがって、テキスト・イメージ・数量という3つの情報の側面を同時・並行的に自由に操作できる環境を整備することが主たる課題になると考えられる。このため、上記3つの他の計画研究班の研究成果を総合して将来の研究支援型データベースのあり方に関する見解を集約し、それに基づいた技術的かつ応用的な提言を行う。

テキスト処理班

人文科学の各分野において、研究の基礎となるものに『テキスト』がある。文学、国語学、歴史学、哲学などのように、テキストを研究の一次的素材とする分野は多い。音楽、考古学、民族学などのように、広い意味で“もの”資料を研究の一次的素材とする分野においても、情報交換はテキストによってなされる。従来、テキストは紙という情報メディアの上に文字や記号、あるいは絵などによって表現されてきた。表現の形態は分野により、ジャンルにより、時代により異なる。とくに、歴史的な背景及び国や地域の特殊性を持っている。また、学術研究の場面においても分野毎にテキストのあり方、使い方、取扱い方などは異なる。例えば、文学研究においては研究の対象は文献資料である作品である。これを読むことから研究は始まる。一方、同じ作品を扱う国語学においては語や文単位の認識と規則などが研究対象となる。現在、これら学術研究で使用されるテキストは、大別して以下のような形態がある。

①電子化(機械可読化)されていない、主として紙メディアに書かれたテキスト。通常は印刷物の形式で流通している。例えば、原文献資料としての写本や版本、通常の印刷された書籍や辞書、通常の文書(書式を持つものと持たないもの)、プレプリント、メモなどがある。

②最初から電子化されているテキスト。例えば、ワープロの文書ファイル、オンラインマニュアル、プログラムなどがある。

②のテキストの利便性は明らかであり、コンピュータを利用することによって、その蓄積、交換、あるいは処理の容易性は、①の形態のテキストの比ではない。そして、従来紙メディア上で行われてきた学術研究が、電子化されたテキストを利用することによって格段に進展することはこれまでの研究から明らかである。例えば、研究はより効率的となり、より生産的な研究が可能となり、考えるという行為が格段に助長されると考えられる。また、従来にない新しい研究のテーマや発想が生起することも十分に考えられる。これらのこと

から、現在多くの研究者個人や研究者グループ、あるいは研究機関によってテキストの電子化が進められている。しかし、ここに大きな問題がある。

例えば、研究者自らがワープロやパソコンを利用してテキストの電子化を行なう場合、これは研究者の個人的な環境においてなされているから、文字コード、外字処理、データの形式や構造等はまったく自由である。これは「研究の自由」ということであるが、自由であるが故に、そこで作成された電子化テキストの流通ということが著しく阻害されてしまっている。この問題を解決しなければ、電子化テキストを利用した研究の発展は望めない。

本計画班は、まずこの電子化テキストの共有化という課題に取り組む。総括班のもとに設置される「共有化ワーキンググループ」を中心に、計画研究「データベース」班と連携をとりながら、現在作成されている電子化テキストの実態調査等を含め諸々の問題の解決に当たる。

共有化と密接に関連してくるが、次の課題としてあるのがテキスト処理のためのソフトウェアである。現状は市販のパッケージソフトウェアを利用したり、研究者個人が苦労しながら開発している。とくに研究者個人が開発したプログラムは、研究者自身のニーズに基づいて開発されているため、他の研究者からはその流通が望まれている。本研究班では、まずテキスト処理の実態を調査し、そこで求められているプログラムの機能を明らかにし、基本的なソフトウェアの開発を行う。これらのソフトウェアは自由な流通を前提とし、マニュアルの整備等を含め流通体制のあり方についての検討も行う。このように、本計画班は「電子化テキストの共有化」「基本ソフトウェアの開発・公開・流通」ということを中心に研究を進める。

イメージ処理班

これまで人文科学分野における本格的なイメージ処理は行われてこなかった。その理由としていろいろなことが考えられるが、最も大きな理由は「イメージ処置が行えるようなハードウェア、ソフトウェアは高価で、人文科学系の研究者が簡単に手に入れることはできなかった」ということであろう。しかし、これは既に何度も述べてきたように昨今のコンピュータの高性能化、高機能化、低価格化によってある程度は解決できていると考えて良いだろう。これから人文科学研究においてもイメージ処理はより活発化していくと予想されるが、まずその必要性和特徴について述べる。

必要性：従来の人文科学におけるコンピュータ利用の形態は、研究用資料の蓄積と効果的利用のためのデータベースの構築が主たるものであった。データベースの内容は、文献情報、数値情報、テキスト情報など、文字や数値で表現されたものが中心である。このため、人文科学におけるコンピュータ利用は、文字コードで表現されたデータベースのみ考慮しておけば十分であるかのような印象を与えてきた。しかしながら、人文科学研究における素材情報は、実は、このようなコード化された情報として存在することはむしろまれで、大部分が、コード化されない情報である。例えば、歴史学や文

学における古文書は現代の印刷物のように標準化された活字で表現されているわけではなく、また、コード化された表現形態で十分ではない。すなわち、文字の字形、紙面への割付けや配置の状況、読者による書き込み、さらには、手垢による汚れなどが、研究上重要な意味を持っている場合が少なくない。したがって、これらの古文書はコード化された形だけでなく、コード化されない生の形でも取り扱えるようになっている必要がある。

特徴：イメージ情報処理が進展した理由は、主にハードウェア資源の性能向上と理論的な研究の進展によるところが大きい。実際に成功しているのは、取り扱う画像、音声などの素データの性質をある程度コントロールすることができる分野である。イメージ情報は、非定型的でばらつきやゆらぎが大きい。対象のカテゴリを定めることができれば、そのばらつきやゆらぎを一定の範囲内に収めることができ、このようなデータを大量に処理することで、高い処理コストを吸収し、十分にペイするシステムとすることができる。

他方、人文科学のような非産業分野でのイメージ情報処理では、ゆらぎやばらつきが一定の範囲に治まる大量のデータを対象とすることはまれで、多くの場合、一個一個のデータがそれぞれまったく違った属性や性質を持つ。したがって、これらを処理するのに必要なアルゴリズムは、個々のデータに対して異なったものになり、手作業で試行錯誤的に解析を進めなければならない。人文科学におけるデータ処理では、むしろ、枠からはずれるような特異なデータを解析する必要があることの方が多い。すなわち、定型的な処理より非定型的な処理が中心となる。したがって、人文科学におけるイメージ情報の取扱いは、産業界や工学におけるアプローチとは大きく異なったものとなり、すでに開発された手法をそのまま適用するだけで目的が達成されるということは、まず考えられない。

以上のようにその必要性は十分にありながら、一般的に数理的な事柄について体系的な教育を受けていない人文科学研究者にとって、コンピュータによるイメージ処理で、何ができて何ができないかということについての認識は不十分である。本計画班では、このようなコンピュータによるイメージ情報処理について、現実の問題として、どのような問題の解決に利用でき、また逆に、なお解決できない問題は何かについて、研究の実例や具体的なソフトウェアの利用を通じて適切な情報を提供する。具体的な研究目標は、以下のようになる。

標準データ形式の検討・変換システムの開発

一口にイメージ情報といっても多種多様なものがあるが、対象とするイメージ情報を以下の3つのカテゴリーに分けて具体的な研究を進める。

- ①グラフィックス・データ 視覚化された表現(図、画像、グラフなど)を作り出すための数値・記号データ、及びその結果のイメージ。
- ②画像データ 静止画像、動画、文書画像等、特にカラー画像。
- ③マルチメディア情報 ①、②を統合した形で取り扱う。すなわ

ち、マルチメディアとして扱うことが特別の意味を持つもの。

ところが、画像データについてみても、白黒画像、カラー画像、動画像等さまざまなデータ形態がある。また、デジタル画像には、それがデジタル化された場合、画像の画素数、一画素あたりのビット数、カラー画像の色表現の手法、データ圧縮の手法等、さまざまな条件があり、これらの情報が適切に保存されていないと、利用時にもとのデータを復元することはできない。一般的にこれらの情報は、ヘッダとよばれるデータ領域に書き込まれ、画像データと一緒にファイルとして保存される。このヘッダの書式、データの保存形式等、いわゆる画像フォーマットには、現在さまざまなものが利用されているが、いわゆる業界標準的なものもいくつか存在している。今後人文科学研究における画像データの流通を考える際にも、このような（業界）標準的なフォーマットを利用するのがよいと考えられる。

グラフィックスデータについては、画像データほど標準化が進んではいないが、傾向としては概ね同様と考えられるので、独自のもの考えるより、このようなフォーマットを採用することがよいであろう。しかし、人文科学者の利用できる機器の環境には、制約が多いであろうから、どのようなフォーマットでも利用できるということにはならない。重要なのは、どれか一つのフォーマットだけを指定して採用することではなく、いくつかのフォーマット間の相互変換を簡単に行えるような仕組みを考えておく必要がある。本計画研究班では、各種データフォーマットの比較検討を行い、これら相互の変換を行うためのシステムの開発とそのソフトウェアの流通を推進していく。

基本ソフトウェアの開発

従来は難しかったイメージ情報の処理が、比較的手軽に利用できるようハードウェアおよびソフトウェア環境が整ってきている。今後、これらの技術的資源を、人文科学分野の研究者にも提供し、研究の素材となるイメージ情報を広く多くの研究者が利用し、また、コンピュータによる処理技術により、個別データの正確で定量的な処理、また、大量のデータをもとにした統計的取扱いができるように整備していくことが緊急の課題である。本計画研究班では、人文科学研究において「イメージ情報」を取り扱う際の問題点について検討し、それぞれを効率的に管理あるいは加工・処理するための基本ソフトウェアの開発、公開、流通を行う。

新しい研究の視点、新しい研究手法についての検討

イメージ情報のコンピュータ処理が利用でき、また、より原資料に近い形での情報が、たとえ遠隔地にいっても手軽にアクセスでき利用できるようになると、従来、比較的狭い範囲の閉じたグループで行われる傾向のあった研究が、より活性化され、人文科学研究の質的变化をもたらすと期待できる。コンピュータによるイメージ情報処理という客観的解析手段を得ることによって、名人芸的な手法によらなくても、それなりの客観的データが得られるようになれば、研究のすそ野が広がり、それによって、人文科学研究の新たな視点、研究手法が生まれてくることを期待できる。確かに、安易なコン

ピュータ利用は、安易な研究態度を助長し、研究者個人個人が深く考えることを避けるようにしむける危険性があることは否定できない。単に研究の省力化のためにコンピュータを利用するという態度ではなく、常に何が本質的かを見極め、本当にコンピュータ化することが望まれる事柄かについて、十分に議論し検討を進める必要がある。

数量的分析班

コンピュータの著しい進歩・普及は人文科学の諸分野の研究方法にも大きな影響を与えつつある。計量歴史学、数理（計量）考古学、数理言語学、計量文献学、コンピュータ民族学等、コンピュータの膨大な記憶能力と計算の高速性を利用した数量的分析に基づく新しい研究領域が開拓されつつあり、それにともない数量的（計量的）研究の基盤となる各種データベースの構築も盛んに試みられている。しかしながら、自然科学の領域の数量的分析とは異なり、人文科学の領域の数量的分析には、その研究の歴史が浅いことに加え、次のような人文科学固有の問題点がある。

量的データと質的データの混在：人文科学の研究に用いるデータでは、量的データと質的データが混在している場合が多く、しかも、いずれもが重要な情報を含んでいるため、この二種類の情報を一緒にして分析する分析法が要求される。例えば、考古学データの場合、遺物の形状寸法、重量などの量的データと、遺物の型、色、模様、材質等の質的データが混在しており、この二種類のデータを同時に利用できるような数量的分析法が要求される。

欠損値が多い：個別に構築されたデータベースは、収録されている情報がまちまちで、その上、例えばあるデータベースでは遺物の色の情報が入っていないというような欠測値も多くみられる。また、データの精度が測定者によって異なるという問題もある。特に、質的データ、例えば遺物の型、色、模様等の認定方法は個人差が相当あると考えられ、精度的にかなり問題があると考えられる。

これらの問題の他に、人文科学の領域でこれまで作られてきたデータベースは個人的に作られたものが多いため、記述フォーマットが一定でなく、したがって、そのまま計量分析を行えるという状態にはないというような問題もあり、すでに構築されている複数のデータベースの統一的な利用には多くの問題がある。これらの問題は、データ形式の標準化等によってある程度解決できる部分もあるが、量的データと質的データの混在といった、人文科学固有の問題は解決しようと思ってもできない部分である。むしろ、そのような人文科学固有の性質を持ったデータを分析することのできる新しい数量的分析法を探ることが重要である。このような観点から本計画研究班は、人文科学分野の諸データの数量的研究に用いることのできる新しい分析手法の開発を中心に研究を進める。具体的な課題としては、以下のようなことを検討している。

新しい数量的分析法

自然科学の領域のデータとは様相が異なる人文科学のデータの数量的分析法を、具体的な研究課題を通じて研究し、人文科学の諸分野の研究の手助けとなるような、新しい数量的

分析法およびプログラムの開発を試みる。とくに精度的にかなり問題のある質的なデータの分析法や、欠測値の処理法などの問題点の解決に重点をおく。

イメージ情報の数量的分析法

計画研究「イメージ処理」と密接に関係するが、人文科学の研究に必要な画像、図形、音声、音響などのイメージ情報の統計処理法の研究を行う。自然科学の領域においてはイメージ情報の処理に関する研究は進んでいるが、人文科学の場合イメージ情報の利用目的が自然科学と異なる場合が多く、したがって、自然科学の領域での研究成果がそのまま利用できるわけではない。そこで、絵画の数量的分析という具体的な問題を通じ、研究が遅れている人文科学の領域のイメージ情報の数量的分析に関する基礎的研究を行う。

シミュレーション

とくに考古学の分野で、今後有望な研究手段になると考えられる時空間上でのシミュレーションの方法を開発する。これまでの考古学データの分析が、地理的にはある一地点、時間的にはある一時点でのデータの分析が中心であったのに対し、本研究では地理的情報、時間的情報をフルに活用し、大型計算機を用いて時空間での大規模なシミュレーションを行い、古代の人口の推定や、遺跡分布の推定等を試みる。考古学の分野でのシミュレーションに関しては、及川、小山による先駆的研究*1があるが、本計画班ではこれを更に発展させることを試みる。

2. じんもんこんで何をやり、何ができたか

壮大な目標を持って研究を1995年に開始して4年が過ぎたが、その具体的な研究成果を問うには今しばらくの時間が必要である。また、個々の研究に関しては領域全体の研究成果報告書、あるいはそれぞれの研究成果報告書に委ねることにし、ここでは領域全体としてどのようなことをやり、何ができたかについて述べる。

2.1 研究者コミュニティの確立

本特定領域研究のひとつの特徴は、下の表にあるように、人文科学系の特定領域としては、公募研究の数が非常に多いということであるが、それは意図的に公募研究をできるだけ多く採択したからである。

その目的は、研究室にこもりがちで、そしてこの種の共同研究がどちらかといえば苦手な人文系の研究者を、できるだけ多く蛸壺的な研究室から飛び出させて、文系・理系の研究者が共同して研究を進めるといふ、新しい研究スタイルを実践させるということであった。すなわち、低迷している人文科学研究を活性化するためには、まず従来の研究スタイルを

		1995	1996	1997	1998
計 画		5	5	5	5
公 募	申 請	214	175	160	108
	採 択	80	66	69	57

変えることから始める必要があるという判断からであった。結果として、実に多数の公募研究への申請があり、その所期の目的は達成できたのではないかと考えている。ちなみに、参加している研究者の数は、分担者を含め200人を超えており、文系と理系に大きく分けると、前者が約6割で後者が約4割となっている。

200人を越すという研究者集団ということは、一堂に会することも大変であるが、研究を協力して進めていくには、まずはお互いのコミュニケーションを確立することが重要になってくる。とくに、このような分類が適切かどうかは別にして、文系と理系の研究者がほぼ半々という状況の中で、お互いの研究を理解することは簡単なことではない。これまでたどってきた道も、環境も、そして見方、考え方も大きく異なる者が理解し合うには、まずそれぞれの意見を聞き、批判し、議論することである。そのためには、研究発表の場、議論の場を数多く作ることが、最善の方法である。

このため、本特定領域ではシンポジウムをはじめとして実に多くの研究集会を実施してきた。具体的には、

- ・全体集会（年度始め）
- ・報告会（年度終わり）
- ・各研究項目ごとのシンポジウム
- ・全体シンポジウム（1996、1998年度）

等であるが、この他に本特定領域発足の母体となった情報処理学会「人文科学とコンピュータ」研究会による研究会が年4回開催されてきている。

もちろん、すべての研究者が常に参加するわけではないが、ほぼ毎年10回に近い研究集会がもたれてきたわけで、「発表・議論の場の提供」については、確実に達成できたものと考えている。その結果として次のようなことがいえる。

- ・領域がスタートした頃は、文系と理系の研究者同士の議論がなかなかみ合わなかったが、年を経るごとにかみ合うようになってきた。そして、共同研究等についても真剣に検討されるようになっており、文系・理系間の垣根が非常に低くなった。
- ・どちらかといえば、個人研究に陥りがちであった文系の研究者も、複数の、しかも異分野の研究者との共同研究に積極的に取り組むようになった。

このことは、必然的に研究の活性化につながり、さまざまな研究情報の共有化が促進されることになる。本特定領域は1998年度で終了するが、これまでの活動で作られてきた

* 1 及川昭文「シミュレーションによる遺跡分布の推定」『東アジアの古代文化』69, pp. 52-66, 1991
 小山修三他「縄文時代人口シミュレーション」『国立民族博物館研究報告』9.1 pp. 1-39, 1984

「研究者コミュニティ」を、今後も維持していこうという気運が高まっており、これは「じんもんこん」の大きな成果と考えていいだろう。

2.2 研究成果物

計画班を含めると毎年60以上の研究グループが参加していることになるが、それぞれの研究成果については別の機会に譲り、ここでは領域として作成したCD-ROMや報告書類について紹介する。

データベース (CD-ROM)

参加している研究班の中には、独自に研究用のデータベースを作成している班も多数ある。それらのうち、公開可能なものについてCD-ROM化し、広く研究者に無償で配布している。作成したCD-ROMはVol. 1～3の3枚があり、それぞれVol. 1には2種類の、Vol. 2と3には1種類のデータベースが収録されている。

Vol. 1と2はWindows, Machintoshいずれでも利用が可能であるが、Vol. 3はMachintoshのみでしか利用できない。Vol. 1はインターネットのブラウザから、Vol. 2は市販ソフト「ファイルメーカーPro」から、Vol. 3はマッキントッシュのハイパーカードから利用することができる。

Vol. 1 (Windows95, Machintosh用)

「インタビュー形式による日本語会話データベース」

編集責任者：上村隆一 (福岡工業大学)

1995～96年に行った、日本語の母国語話者と外国人の日本語学習者各50人のインタビュー実験の様子を可能な限り忠実に再現したデータベースである。CD-ROMにはテキストデータと圧縮された音声データ、および全文検索クライアントソフトウェア (Mita-keSearch) のパッケージが収録されている。

「相模集データベース」

編集責任者：近藤美由紀 (千葉大学)

平安時代女流歌人として代表的な人物である相模の歌集である「相模集」の校訂本文と語彙索引とをテキストデータベースとしたものである。

Vol. 2 (Windows95, Machintosh用)

「CD-ROM版音声データベースJCMD大阪」

編集責任者：田原広史 (大阪樟蔭女子大学)

大阪市で生まれ育った5世代の男女各2名ずつ計20名の方言音声を集めたデータベースである。本データベースを利用するには市販されているデータベースソフトウェアの「ファイルメーカーPro」が必要である。

Vol. 3 (Machintosh用)

「MCD 国立民族学博物館所蔵衣服標本データベース」

編集責任者：高橋晴子 (大阪樟蔭女子大学)

国立民族学博物館所蔵の衣服標本資料約1,000点を対象としたマルチメディアデータベースである。検索システムは、マッキントッシュのハイパーカードで作成されている。

これらのCD-ROMについて、残部があるものについては希

望者に配布いたします。入手希望者は、「Vol番号、送付先」を明記して、郵送料 (切手500円分) を同封の上、事務局 (〒240 0193 三浦郡葉山町 総合研究大学院大学 じんもんこん事務局) で申し込んで下さい。

解説書

総括班では、1995年に研究を開始したときから、人文系の研究者や学生を対象とした情報処理に関する解説書を刊行することを検討してきたが、科学研究費の研究成果公開促進費の補助を受けて、講座「人文科学研究のための情報処理」全5巻を1998年に刊行した。これはコンピュータや情報処理に関する入門書や解説書は数多く刊行されているが、人文系の研究を対象としたこの種のものはほとんどないことから企画されたもので、各計画班の代表者が中心になって編集したものである。全5巻の内容を簡単に紹介すると、以下のようになる。

第1巻「入門編」 及川 昭文 編著

- 人文科学とコンピュータ
- コンピュータを知る
- コンピュータを使う
- 資料編

第2巻「データベース編」 小沢 一雅 編著

- 序 - 人文科学の未来に向けて
- データの形態・構造・特性
- データベースシステム基礎編
- データベースの動向と未来
- 終章

第3巻「テキスト処理編」 安永 尚志 編著

- はじめに
- テキストの電子化
- テキストデータの形成と管理
- 情報検索
- テキスト流通とインターネット
- 文字列の操作
- SGML入門
- おわりに

第4巻「イメージ処理編」 八村 広三郎 編著

- 人文科学とイメージ処理
- イメージ処理の基礎技術 画像処理
- イメージ処理の基礎技術
コンピュータ・グラフィックス
- イメージ処理の応用技術
- イメージ処理のシステム
- 人文科学への応用
- 付録 行列の計算方法

第5巻「数量的分析編」 村上 征勝 編著

- データをまとめる
- データをグラフにする
- 違いをはかる
- 違いが認められるか
- 予測する - 回帰分析 -
- どのグループに属するか 判別分析
- 合成変数で分析を簡単にする
- グループに分ける グラスタ分析
- 柔軟なデータ解析法
- 統計解析のソフトウェア

研究成果報告書

毎年度末には、計画班を含めすべての公募班に研究成果報告書の提出を義務づけ、それらを一冊の報告書として作成してきた。初年度は冊子体で作成したが、700頁を越える大部なものになった。印刷費も相当かかったが、それよりも大変だったのは郵送費であり、発送の手間を考えると非常に経費のかかる報告書となった。このため報告書のCD-ROM化を検討し、翌年度から冊子体の報告書はやめてCD-ROMによる報告書に切り替えた。

最初のCD-ROM報告書は、各班に版下原稿を提出してもらい、それをスキャナで読み込みPDFファイルとしてCD-ROM化を行った。PDFファイルを読むプログラムはAdobe社から無償で配布されており、インターネット等を通じて簡単に入手できる。

1997年度からは版下原稿ではなく、報告書をHTMLファイルとして提出してもらい、それらのファイルをまとめてCD-ROM化している。報告書をHTMLファイルとして作成することには、数々の利点がある。具体的には、

- ①作成するCD-ROMの枚数にもよるが、作成経費はほぼ半額になり、郵送費も冊子体に比べ格段に安くなる。
- ②冊子体の場合、A4サイズでカラー印刷すると1頁あたり10万円程度の経費がかかるが、HTMLファイルの場合は、カラー・モノクロの差は全くない。
- ③音声や動画を含むことができる。
- ④関連したホームページへのリンクを設定することが可能で、多彩な報告書を構成することができる。
- ⑤大容量のデータベースであっても報告書の中を含むことができる。
- ⑥著作権の問題は別にして、報告書中の文章や図表、データ等の再利用が容易である。
- ⑦報告書全体の全文検索等の高度な利用が可能である。

等を挙げることができる。とくに従来の印刷物の形態では不可能だった音声の記録ができることは、「音」に関する研究を行っている研究者にとっては、実際に音を読者に聞かせることができ、研究の内容をより正確に伝えることができるようになった。

年度末の報告書の他には、次頁にあるように、研究情報誌「じんもんこん」、シンポジウムの報告書等多くの印刷物を発行した。

3. じんもんこんのこれからの課題

4年間の研究活動を通じて、コンピュータが人文科学研究において有効な道具であることは実証できたが、一方取り組まなければならない課題が数多くあることも明らかになった。技術的な課題については、個々の研究者の努力によって解決されるであろうが、長期的かつ組織的に取り組まなければならない課題も少なくない。ここでは、データベースに関する二つの課題について述べる。

共有データベースの構築と流通

人文科学研究においてコンピュータを利用しようとした場

合、まず最初に思いつくのは、これまで手作業で整理、分析していた資料やデータをコンピュータに入力し、コンピュータで処理することである。しかしながら、人文科学では、資料やデータを整理すること自体が研究であるという側面があり、データベース化することは自然科学系のものに比べて、格段に複雑、困難である。

じんもんこんにに参加している公募班でも、古文書やイメージデータの入力、考古学データの構造等、まずデータベースをどのようにして構築するかというところで大変な苦勞をしている班が数多くみられた。また、個々の研究者では対応が難しい「著作権」という大問題もある。この著作権は、自分だけでデータベース化して利用する限りには大きな問題にはならないが、それを公開、流通させようとする、大きな障壁となってくる。

今のところデータベースを作るのに汲々としているというのが現状で、個人的な努力では大きな進展が望めない状況にある。このような現状を改善していくには、大規模かつ組織的なプロジェクトを推進していくことが必要である。主要な古典作品の電子化、公開というようなことが、公的研究機関による、あるいは科学研究費等の公的予算によるプロジェクトとして積極的に展開されていくことが望ましい。

現在筆者を中心とするグループで、SF作家小松左京の全作品の電子化（最終的な容量は約40MBになる予定）を進めているが、この種のプロジェクトが個人レベルではなく、組織的に進められることが重要である。また、もう一つ重要なことは、構築されたデータベースを広く流通させるための組織、機構の確立である。構築するまでは研究者主体で進める必要があるが、流通に関しては研究と切り離れた形で進められることが望ましい。いずれにせよ、膨大な資料を対象とする人文科学研究においては、このデータベースの構築と流通ということが、これからのコンピュータ利用の鍵になるのは確かである。

データベースからの知識獲得

データベースというものは、それ自体は単なるデータの集合体でしかない。データベースを作ったからといって、即座に研究が進展するわけではない。如何にしてこのデータベースを活用するか、データベースの中から有用な情報を取り出せるかということが重要なことである。すなわち、膨大な情報の中から意味のある情報を見つけだし、それをひとつの知識として組み立てる手法を確立することによって、はじめてデータベースとしての利用価値が生まれてくることになる。

最近KDD (Knowledge Discovery in Database) とかデータマイニング (data mining) という言葉が聞かれるようになってきたが、人文系のデータベースを対象にしたデータマイニングの手法の開発が必要であろう。じんもんこんの中にもこの種の研究開発に取り組んだ研究班もあったが、人文科学の中ではまだまだ未成熟の分野であり、今後多くの研究者がこの課題に取り組むことが期待されている。

(1999年1月葉山にて)

「じんもんこん」刊行物一覧

(1999年1月現在)

- #001 「公募研究」申請のてびき第1版(1994年9月発行, 60頁)
- #002 「公募研究」申請のてびき第2版(1995年9月発行, 72頁)
- #003 「公募研究」申請のてびき第3版(1996年9月発行, 88頁) 品切れ!

- #011 研究のてびき1995(1995年5月発行, 50頁)
- #012 研究のてびき1996(1996年6月発行, 120頁)
- #013 研究のてびき1997(1997年6月発行, 114頁)

- #021 特定領域「人文科学とコンピュータ」1995年度研究成果報告書(1996年3月発行, 718頁)
- #022 特定領域「人文科学とコンピュータ」1995, 96年度研究成果報告書CD ROM版(1997年3月発行)
- #023 特定領域「人文科学とコンピュータ」1997年度研究成果報告書CD ROM版(1998年3月発行)
- #031 『じんもんこん』第1号(1995年9月発行, 86頁)
- #032 『じんもんこん』第2号(1996年2月発行, 60頁)
- #033 『じんもんこん』第3号(1997年3月発行, 80頁)
- #034 『じんもんこん』第4号(1997年7月発行, 116頁)
- #035 『じんもんこん』第5号(1998年3月発行, 46頁)
- #036 『じんもんこん』第6号(1998年10月発行, 112頁)

- #051 シンポジウム「人文科学とコンピュータ」1996論文集(1996年9月発行, 164頁)
- #052 シンポジウム「考古学とコンピュータ」論文集(1996年9月発行, 92頁)

- #101 「人文科学とコンピュータ」データベース班1995年度研究成果報告書(1996年3月発行, 142頁)
- #102 「人文科学とコンピュータ」データベース班1996年度研究成果報告書(1997年3月発行, 172頁)
- #103 「人文科学とコンピュータ」データベース班1997年度研究成果報告書(1998年3月発行, 142頁)
- #111 第1回公開シンポジウム「人文科学とデータベース」論文集(1995年12月発行, 106頁)
- #112 第2回公開シンポジウム「人文科学とデータベース」論文集(1996年12月発行, 80頁)
- #113 第3回公開シンポジウム「人文科学とデータベース」論文集(1997年12月発行, 76頁)
- #114 第4回公開シンポジウム「人文科学とデータベース」論文集(1998年12月発行, 84頁)

- #201 「人文科学とコンピュータ」テキスト処理班1995年度研究成果報告書(1996年3月発行, 200頁)
- #202 「人文科学とコンピュータ」テキスト処理班1996年度研究成果報告書(1997年2月発行, 334頁)
- #203 「人文科学とコンピュータ」テキスト処理班1997年度研究成果報告書(1998年2月発行, 142頁)
- #212 シンポジウム「人文科学とコンピュータ - テキスト処理 - 」論文集(1996年10月発行, 144頁)

- #302 「人文科学とコンピュータ」イメージ処理班1996年度研究成果報告書(1997年3月発行, 180頁)
- #303 「人文科学とコンピュータ」イメージ処理班1997年度研究成果報告書(1998年2月発行, 182頁)
- #312 シンポジウム「人文科学とイメージ処理」論文集(1996年12月発行, 92頁)
- #313 シンポジウム「人文科学とイメージ処理」論文集(1997年11月発行, 64頁)
- #314 シンポジウム「人文科学とイメージ処理」論文集(1998年12月発行, 62頁)

- #411 シンポジウム「人文科学における数量的分析」1995論文集(1996年3月発行, 104頁)
- #412 シンポジウム「人文科学における数量的分析」1996論文集(1997年3月発行, 142頁)
- #413 シンポジウム「人文科学における数量的分析」1997論文集(1998年3月発行, 132頁)

残部があるものについては希望者に配布いたします。入手希望者は、「番号, 送付先」を明記して, 郵送料(切手500円分, 5部以上になる場合は1000円分)を同封の上, 事務局(〒240 0193 三浦郡葉山町 総合研究大学院大学 じんもんこん事務局)まで申し込んで下さい。