

## 自動演奏生成2000—デモンストレーションとパネル

平賀 瑠美  
文教大学情報学部

片寄 晴弘  
和歌山大学システム工学部

小池 宏幸  
筑波大学工学研究科

鈴木 泰山  
東京工業大学  
情報理工学研究科

野池 賢二  
東京農工大学  
工学研究科

星芝 貴行  
北陸先端科学技術大学院大学  
情報科学研究科

### 要旨

日本における自動演奏生成システム研究の成果を一堂に集め、システムの生成物である演奏のデモンストレーションを行う。これらは、大学の理工学系の研究室で独立に研究が進められてきており、演奏生成のモデル、手法は勿論、生成対象となる楽曲、入力情報の種類や形式、出力情報の種類は全て異なる。デモンストレーションに先立つパネル討論では、デモの各参加グループから一人がパネリストとなり、これまでに進めてきた内容から、演奏生成研究において感じた困難なことや、それに基づいた今後のための提案を発表する。今回のデモンストレーションとパネル討論は演奏生成システムを集める世界初の試みであり、今後のより円滑な情報交換と研究の発展を目指す。

## Performance Rendering 2000—Demonstration and Panel Discussion

Rumi Hiraga  
Bunkyo Univ.

Haruhiro Katayose  
Wakayama Univ.

Hiroyuki Koike  
Univ. of Tsukuba

Taizan Suzuki  
Tokyo Inst. of  
Technology

Kenji Noike  
Tokyo Univ. of  
Agriculture and Technology

Takayuki Hoshishiba  
Japan Advanced Inst. of  
Science and Technology

### Abstract

Researchers who have been working on performance rendering systems gather together to show their demonstration and exchange opinions for a further evolution of the research at a panel discussion. All the systems have been independently developed at engineering/science laboratories in universities. For the mutual understanding and sharing common issues in the research of performance rendering, representatives from each group discuss and present proposals toward the research in the next step. The special demonstration of performance rendering is the world premiere.

## 1 はじめに

演奏生成研究はその結果として演奏が生成される。「百聞は一見に如かず」という諺があるが、演奏生成研究に関しては、まったく当てはまらず、いくら紙上に細かくモデルが記されていてもそこから結果の演奏を予測することは殆んど不可能である。この研究分野の成果は「百見は一聞に如かず」という状況を代表しているものの一つといえよう。メディア技術の進歩に伴い音声、音響、音楽情報がより簡単、高速、安価に PC 上で実現されてきているとはいえ、研究は紙の上に文字で記され、その成果は数値により評価されるという、従来ながらの研究発表形式がまだまだ幅をきかせている。

本デモンストレーションは、演奏生成研究の現状を広く知ってもらうことを第一の目標としている。また、従来の研究発表の枠には入りきれない研究分野の一つとして、その成果の発表のあり方を考えるきっかけとなればと考えている。

デモンストレーションで発表される五件の研究は、その演奏生成のモデル、手法は勿論、研究の目的、生成対象となる楽曲、入力情報の種類や形式、出力情報の種類は全て異なる。これらを念頭におき、以下に紹介するデモンストレーション要旨からのみでは伝わらない研究内容を実際にデモンストレーション会場で演奏を聞いて体験していただきたいと思う。

## 2 デモンストレーション要旨

以下に各デモンストレーションの要旨を記す。この五件のデモンストレーションが演奏生成システム研究に関するもので、各発表グループの中からパネリストが出てパネル討論を行う。

本研究会のデモンストレーションとして、これら五件以外にも、国立音楽大学からマルチメディア・インタラクティブ・システムが出展される。

### 2.1 カスタマイズした重回帰分析による演奏の表情付け

石川修 (大阪大学大学院基礎工学研究科  
システム人間系)

片寄晴弘 (和歌山大学システム工学科)

井口征士 (大阪大学大学院基礎工学研究科  
システム人間系)

我々は人間の演奏時における演奏表現を理解・分析するために音楽解釈モデルを提案してきた。その中で演奏表現を抽出するため、重回帰分析に AND 処理のイタレーション・変数減少法といったアルゴリズムを施し抽出を行っている。重回帰分析の条件となる説明変数は、音符列、明示的に楽譜上に書かれた演奏記号だけではなく、音楽構造解析理論から得られるフレーズ(グループ)表現や音楽心理学の理論を用いた。また、目的変数として実演奏の音量、テンポ、音長、発音位置を入力とした。重回帰分析を用いることで楽譜情報と実演奏との関係(演奏表現ルール)を求めることができ、このルールを未知曲に適用することでその楽譜に演奏の表情を付けることができる。デモとして実演奏から抽出した演奏表現ルールをもとに生成した演奏をお聞かせする。演奏表現ルールの違いにより未知曲の演奏の表情がどのように変化するのかという点に注目してもらいたい。

研究室 HP <http://www.inolab.sys.es.osaka-u.ac.jp>

担当者 [ishikawa@inolab.sys.es.osaka-u.ac.jp](mailto:ishikawa@inolab.sys.es.osaka-u.ac.jp)(石川修)

### 2.2 芸術的演奏の生成における楽曲構造に基づく時間的法則の定量化

小池宏幸 (筑波大学工学研究科 (4月1日より  
ソニー株式会社))

水谷哲也 (筑波大学電子・情報工学系)

田中崇之 (筑波大学理工学研究科)

五十嵐滋 (筑波大学電子・情報工学系)

演奏表情のうち、特に重要な役割を持つと考えられるアゴーギク(局所的演奏速度変化)に着目し、その関数(曲線)的表現に基づいてそれに関する法則を提案し、楽譜にその法則を適用する際の定量化方法を開発することにより、演奏を生成する。実験の題材として、ポーランドの民族舞曲に由来するポロネーズを用いた。クラシック音楽全般に成立つ法則およびポロネーズ特有の法則を精密化し、解析システム DAPHNE から得られた音楽解析情報を基に法則の適用位置を決定し、その上で人間の演奏および楽譜データに基づき種々の方法で法則の定量化を行った。さらに、定量化された法則を他の楽曲に適用することで演奏を生成した実験結果に対し、音楽専門家による評価を求めた。その結果、遺伝的アルゴリズムを用いた定量化手法が最も芸術的な演奏を生成するという結果を得た。これは、演奏表情が不

連続曲線であるという、本研究の本質的な難解さを反映している。

研究室 HP <http://www.psychology.is.tsukuba.ac.jp>

担当者 [peace@psychology.is.tsukuba.ac.jp](mailto:peace@psychology.is.tsukuba.ac.jp) (田中崇之)

### 2.3 Kagurame Phase-I

鈴木泰山 (東京工業大学 情報理工学研究所  
計算工学専攻)

徳永健伸 (東京工業大学 情報理工学研究所  
計算工学専攻)

田中穂積 (東京工業大学 情報理工学研究所  
計算工学専攻)

我々は、事例に基づく演奏表情の生成手法の研究を行なっている。この手法は、対象曲の楽譜情報と与えたい演奏表情の特徴情報を入力すると、あらかじめ用意した人間による演奏データ集の中から対象曲に類似した曲の演奏事例を検索し、それらに見られる演奏表情を対象曲に応用することで対象曲の自然な演奏データを生成する。この手法は、単一の楽曲に対して種類の異なる複数の演奏表情が容易に生成できるという特徴がある。"Kagurame Phase-I"は、この事例に基づく演奏表情生成手法を利用した演奏データ生成システムである。このシステムは、入力として、演奏スタイルなど演奏表情の特徴を指定することで、それに応じた演奏表情の演奏データを生成することが可能である。デモンストレーションでは、同一の楽曲に対して演奏表情の特徴情報を変化させた複数の演奏データの聞き比べなどを行なう予定である。

担当者 [taizan@mori.cs.titech.ac.jp](mailto:taizan@mori.cs.titech.ac.jp) (鈴木 泰山)

### 2.4 表情パラメータ値列を近似した近似関数を用いた演奏表情生成システム

野池賢二 (東京農工大学大学院 工学研究科  
電子情報工学専攻)

野瀬 隆 (東京農工大学大学院 工学研究科  
電子情報工学専攻)

乾 伸雄 (東京農工大学大学院 工学研究科  
電子情報工学専攻)

小谷善行 (東京農工大学大学院 工学研究科  
電子情報工学専攻)

われわれは、「人間が演奏に表情を付けられるのは、楽曲の構造を把握し、楽譜上に記述されている記号を理解しているからである」という考えを基に、表情付加規則を適用することによって表情を付加する、ルールベースの演奏生成システムについて研究している。本システムでは、表情として、次の四つのパラメータを扱う。

- 局所メトロノーム速度値
- アーティキュレーション (打鍵離鍵間時間)
- 打鍵強度
- ダンパーペダル踏度

ダンパーペダル踏度を扱うことにより、アーティキュレーションの信頼性を上げることができる。表情付加規則は、これらの表情パラメータ値列を近似関数として持っている。近似関数として持つことにより、関数の係数値だけで表情を表すことができ、また、表情の分析が容易になる。演奏者の個性は、関数の係数の値や、関数の形として現れる。表情付加規則は、演奏例とそれに対応する楽譜情報の対から獲得する。

研究室 HP <http://shouchan.ei.tuat.ac.jp/>

担当者 [noike@fairy.ei.tuat.ac.jp](mailto:noike@fairy.ei.tuat.ac.jp) (野池賢二)

### 2.5 複数の演奏データを用いた奏者の個性解析と評価

星芝 貴行 (北陸先端科学技術大学院大学  
情報科学研究科情報システム学専攻)

電子楽器や計算機を結ぶ規格である MIDI と様々な音楽ソフトを用いることにより、計算機上で比較的容易に曲の演奏や作曲ができるようになったが、楽譜通りのデータを入力しても計算機は正確な演奏を行うだけで、人間らしい情緒ある演奏は期待できない。人間らしい情緒ある演奏を行うには、演奏者が楽譜を解釈して演奏する特徴を何らかの方法で抽出し、それを計算機に入力しなければならない。また、奏者の個性も人間らしさの重要な要素と考えられ、これらの抽出も必要となる。我々は、奏者の個性を抽出するための基準となる演奏 (標準的演奏データ) を、複数の奏者によるピアノ演奏データから求める手法と、得られた標準的演奏データを用いて、演奏者の個性の抽出についての研究を行って

いる。標準的演奏データは、同一曲目の複数の奏者による演奏データを1音ずつ相互に対応付け、各音の発音時刻、強弱、長さのそれぞれの平均値を求め設定した。個性の抽出は、このようにして得られたデータと各奏者の演奏データとの差分、各演奏から抽出したルールのパラメータの差分、更に、我々が提案する楽曲構造の階層的な解析手法に基づいた各奏者の演奏の分析から行う。

研究室 HP <http://mitsuko.jaist.ac.jp/>

担当者 hoshisi@jaist.ac.jp (星芝 貴行)

### 3 パネルディスカッション

デモンストレーションで発表される研究は、大学の理工学系の研究室で独立に研究が進められてきたものである。

特に日本では、コンピュータ・サイエンスの研究の一つとして演奏生成研究の論文が紹介されることはあまり多くはなかった<sup>1</sup>。この理由の一つに、第1章で現在演奏生成研究が対峙しなければならないこととして述べたとおり、従来の枠組と評価方法に準拠した発表が求められていたということも挙げてもよいと思う。では、海外<sup>2</sup>ではより多くの演奏生成に関する研究が発表されているが、日本の演奏生成研究は海外でならば発表に値するものであろうか<sup>3</sup>。

研究環境に関して日本と海外の違いとして、思い付くだけでも、以下のようなことがあるのではないだろうか。

1. そのような研究(演奏生成研究)を受け入れる研究土壌がある程度確立されていること。
2. 演奏生成研究に携わる研究者の人口が多少なりとも多いこと。
3. コンピュータ・サイエンスを中心とした研究所でも他分野の専門家との協調がより容易そうであること。
4. 演奏生成に関する文献入手が日本よりは容易であること。

これらは根本的なことではあるが、少人数の研究者で一朝一夕に変えられるような問題でもない。

デモンストレーションに先立つパネルディスカッションでは、日本独特のハンディキャップを甘受す

<sup>1</sup>本段落では、「論文の紹介・発表」で、論文誌への掲載を念頭においている。

<sup>2</sup>英文論文誌。

<sup>3</sup>単純な疑問。

ることなく、世界に伍する研究を発信していくためにどのようなことができるかを考えるきっかけとする。今回のデモンストレーションとパネル討論は演奏生成システムを集める世界初の試みで、今後のより円滑な情報交換と研究の発展を目指しており、何らかの結論を急ぐものではない。

まず、演奏生成研究で課題となっていることを共通に認識することから始まり、さらに、研究の発展のために、研究内容以外のこととして、発表の枠組を変えていく努力をするためのきっかけとなることが期待される。

例えば、本フロンティア領域ジョイント研究会で行われるコンピュータ囲碁大会のように、演奏生成システムによる演奏コンクールを定期的を実施してみるといのはどうかという提案がある。このような催しは研究者のモチベーションにもなり得るし、広く研究の成果を報告することにもなるであろう。実際にコンクールを行うためには、課題曲の選曲、審査員、審査基準など、実際の音楽コンクールにおいても決めなければならないことを演奏生成研究という特殊性を考慮して決定しなければならない。また、入力データの構造や出力楽器、形式といったコンピュータ関係のことがらについても統一する必要が出てくる可能性もある。

このような試みが生み出されるならば、音楽情報処理研究の演奏生成分野に限らず、デモンストレーションが説得力を持つアンサンブルシステムなどでも有効であろう。

### 4 おわりに

本デモンストレーションを機会として、お互いの内容をより良く知るため、また、ステップアップの資料として、各参加者から演奏を提供してもらい、データベースを作成する予定である。著作権に関する問題が明確にわかっていないので、当面はアクセスの制限をするが、なるべく早い時点で一般に公開していきたい。

## Interactive Multimedia Education at Distance (IMED)

池辺 八洲彦\* Maha Ashour-Abdalla\*\* 宮崎 佳典\*\*\* 蔡 東生\*\*\*\*

\*会津大学コンピュータ理工学部 \*\*\*静岡産業大学国際情報学部 \*\*\*\*筑波大学電子・情報工学系

\*\*カリフォルニア大学ロサンゼルス校(UCLA)デジタル革新センター(CDI)兼 物理学科

遠隔教育のためのインターネット線形代数教育プログラムの開発が約 3 年の準備期間を経て 1999 年夏より開始された。UCLA との国際共同計画によるもので、本年夏に初版完成を目指している。役割分担は、第一著者(池辺)の講義ノートに基づき日本側チーム(5 名)が特色あるコンテンツを提供し、Web ベース教材開発に実績のある UCLA チームが情報技術(IT)とプロジェクト管理技術(PMT)を提供している。プロジェクトの概要とこれまでの進捗状況について述べる。

## Interactive Multimedia Education at Distance (IMED)

Yasuhiko Ikebe\* Maha Ashour-Abdalla\*\* Yoshinori Miyazaki\*\*\* DongSheng Cai\*\*\*\*

\*School of Computer Science and Engineering, University of Aizu

\*\*Department of Physics and Astronomy & CDI, University of California, Los Angeles

\*\*\*Faculty of Communications and Informatics, Shizuoka Sangyo University

\*\*\*\*Institute of Information Sciences and Electronics, University of Tsukuba

The development of Internet-based distance learning program on linear algebra started in the summer of 1999, after 3 years' preparation as a joint international project with UCLA. Its first version is aimed to be completed in the summer of 2000.

The Japanese team (5 members) provides unique, quality contents based on the lecture notes by the first author (Ikebe), while the U.S. team (UCLA) with proven track records in the development of Web-based learning materials provides the needed Information Technology (IT) and Project Management Technology (PMT). The outline and the status quo of this project will be discussed.

### 1. はじめに

本稿では、UCLA(University of California at Los Angeles)内の CDI(Center for Digital Innovation)との間で開始した共同研究について報告する。日本側と米国側で構成されるプロジェクトチームを組み、遠隔教育のための線形代数教育プログラムを開発するのがその内容である。線形代数(Linear Algebra)は数学の重要な一分野であり、特に大学の理系学部で学ぶ学生にとっては微分・積分学と並んで必須の数学科目と位置付けられている。なお、本論内ではこのプログラムの略称として IMED-LA(IMED-Linear Algebra, IMED)については後述)を使わせていただく。実は本プロジェクト開始に至るまでに約 3 年必要であった。最初の 1 年は会津大学の研究者たち数名の間で研究の発想と基礎調査を行い、次の 1 年はお礼調査と共同研究者(UCLA/CDI)の発見に、最後の 1 年は CDI との交渉とチーム編成に費やされたのである。

アメリカ側チーム代表 CDI センター長 Maha Ashour-Abdalla<sup>1</sup>教授は会津大学の招聘により、会津大学において 1999 年 10 月 15 日、IMED-LA の概要について講演を行った。講演後日本側チーム代表(池辺)と共同記者会見が行われ、IMED-LA の概要が相当詳しく新聞 5 紙により報道される結果となった(福島民報 10/16、福島民友 10/17、河北新報 10/17、毎日新聞 10/19、読売新聞 11/5。地方版記事を含む)。その第一版の完成予定を 2000 年夏に設定し、現在も継続的に日米間でインターネットによる連携と UCLA 現地共同作業を適当に組み合わせつつ、教材開発に取り組んでいる。現在、予定している約 2/3 の工程が終了した段階である

<sup>1</sup> UCLA デジタル革新センター (CDI) 長 兼 物理学科教授

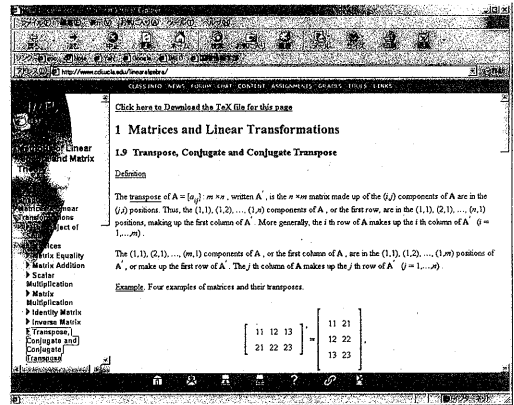
が、これまでの途中経過と、今後のプランについて述べる。

## 2. IMED-LA の構成、特長

IMED (Interactive Multimedia Education at Distance)はUCLAによって開発された、大学レベル講義教材とソフトウェアアプリケーションのウェブベース配信システムであり、既に語学系の応用例としてはスペイン語・イタリア語・ロシア語のバージョンを有し、自然系科目としては宇宙科学学習システム(IMED-Space Science)が開発されている。それでは我々の開発する IMED-LA について議論を始める前に、同じ自然系科目を扱っている IMED-Space Science について簡単に説明しておく。このプログラムは宇宙科学の領域をカバーし、京都大学(リーダー: 美濃導彦教授)と CDI が 1999 年 10 月より 3 ヶ月間にわたり共同で大学講義を中継する遠隔教育プロジェクトとして作成されたものである。京都大学チームは遠隔教育のために必要な情報技術開発(自動追尾システム、教室にどんなカメラを何台どのように配置・制御する、など)に特に重点を置いているのが特徴である。対照的に、IMED-LA ではよりコンテンツ作成に力点が置かれている点が前者と多少性格の違う点である。

IMED-LA はウェブベース遠隔配信システムという IMED の特徴を核として、数学の一専門分野である線形代数を全世界の学生がアクセスできるように学習支援することを目的に開発されている。線形代数の分野で特に重要と思われる証明問題を避けることなく、従来には類を見ない簡潔でわかりやすい証明法、さらに Web 上の双方向性をうまく利用したコンテンツ作成によって学生の理解度をあげることを主眼においている。基本的に対象となるのは理科系を専攻する大学一、二年生で、また、将来的に世界中からのアクセスを想定しているため、使用言語は英語となっている。全体は 9 章から成り、ここに具体的な章立てを述べる(下線付はほぼ作成完了した章を表す):

- 第 1 章…Matrices&Linear Transformations
- 第 2 章…Catalog of Main Facts
- 第 3 章…Equivalence Decomposition and LDU Decomposition
- 第 4 章…Determinants
- 第 5 章…Inner Product and QR Decomposition
- 第 6 章…Schur Decomposition
- 第 7 章…Jordan Decomposition
- 第 8 章…Singular Value Decomposition
- 第 9 章…Further Applications



右の図 1 は本システムの概観を示す、本文内の 1 ページである。

図 1

この構成には従来書に見られない次の特色がある。すなわち、線形代数の方法を 6 大分解定理を中心に整理してあることである。これは「線形代数をどのような形で頭の引出しに整理しておくのがベストか」の立場から、池辺の長年の教授経験に基づいて考えられた整理法である。本プロジェクト IMED-LA 内でもうひとつ力を入れた点は双方向性を活用した練習問題(Exercises)作成にある。双方向性を最大限に活用することによって、学生がアクセスして練習問題を解いたときに、採点結果が即座に教員の手元に送信され、教員は学生の理解度を簡単に把握できる。練習問題も従来の製品に多く見られる選択肢の形式をとるだけでなく、[A]選択肢問題(Multiple Choice type)[B]空所補充問題(Fill-In type)[C]論理ブロック問題(Block

Diagram type)[D]自由記述問題(Free Form type)[E]Mathematica<sup>2</sup>を使った問題の5タイプを用意した。個々の練習問題の性格に応じて、使い分ける幅が広がられたと言えよう。ここでは[B][C]を例示する。

[B]空所補充問題 (Fill-In type)

このタイプは主に計算問題や数理的思考を試すために用いられる。例は行列の乗算に関する問題で、図2は学生が入力をしている画面である。入力後、Submit ボタンを押すとその入力結果がサーバに送信され、採点結果が即座にクライアント側に返送される。図3がその結果例である。括弧内に示された数字が正解である。不正解の箇所は赤く表示されるため、学生側は自分のミスを一目で確認できる。

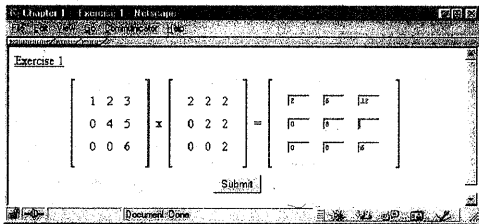


図 2

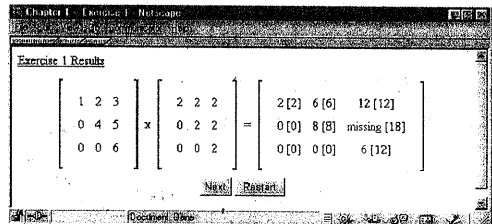


図 3

[C]論理ブロック問題 (Block Diagram type)

この問題は、学生の論理的思考を試すものである。各問題には有限個の論理ブロックが用意され、設問で要求

された解を導くために必要な論理ブロックをその中から選択させる(問題によっては空の論理ブロックも用意し、学生にそのブロックに適切な概念を入力させる)。さらに各論理ブロックがどのように有機的に関連して解に到達するのか、リンクを張ることが要求される。図4がサンプルの問題導入文である。ここで導くべき最終結果(この場合は  $B=C$ (2 行列の同等)、図4 矢印位置)が与えられる。図5が作業場所、この [Start]、[Finish]

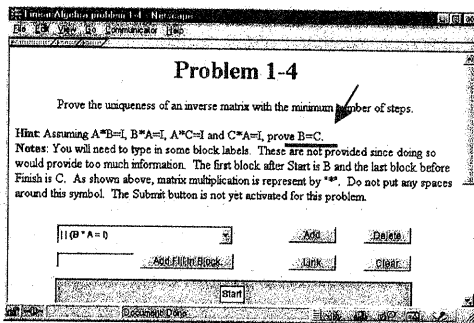


図 4

の間に途中の概念(定理・性質)をコンボボックス( )内の候補より選択・挿入し、リンクをはってゆく。図6がその最終解答である。Submit ボタンを押下すれば、1. どの論理ブロックが選択され、2. 各々の

図 5

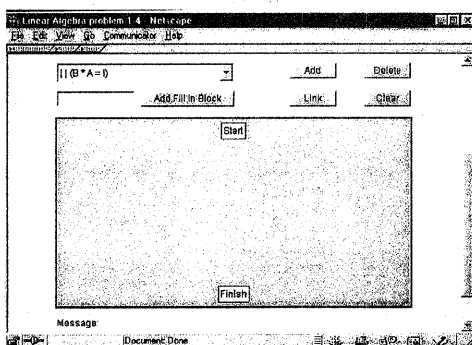
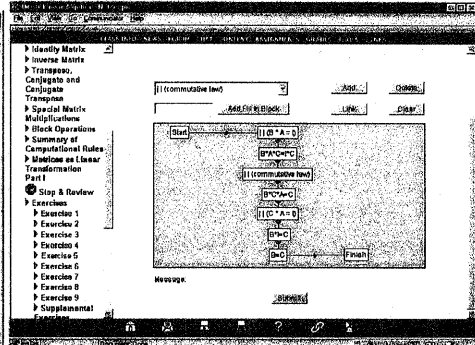


図 6



論理ブロックが解に至るまでに適切な順序で並べられているのか、などの判断基準によって採点が行われる。

<sup>2</sup> Mathematica は、米国 Wolfram Research Inc. の登録商標です。

この解答方法はフローチャート(流れ図)作成に似た作業であり、そういった作業を日常的に要求される理系の学生には自然に受け入れられるものである。

### 3. 教員のニーズに応じたカスタマイズ支援環境

今度は教師側のメリットについて考えてみる。学生が練習問題を解くと、その結果がデータベースに送られる。データはもちろん個々の学生データとして残るから、各学生の理解度や学習の進行状況が把握でき、適当なアドバイスを個人単位で送ることが可能である。逆に、全体としての動向を探ることも統計的に可能であり、どういった概念、定理を学生が理解できないのか、まちがった理解に陥りやすいかなどを調べることができる。図7, 8は、他のIMEDバージョンの統計処理結果の一例であるが、学生が間違えやすい問題の傾向が良く見て取れることが分かる(図7では学生の解答がおおよそ正解と符合しているのに対し、図8では正解ではない箇所を多く選択していることがわかる)。

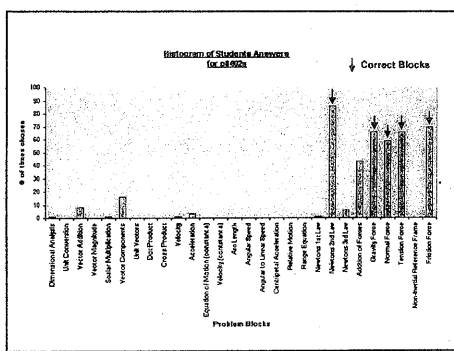


図7

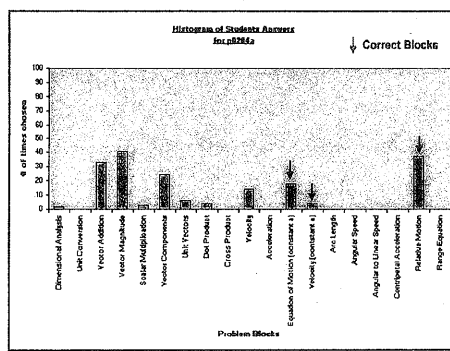


図8

もうひとつのIMED-LAの特長は、教師側のニーズに対応してプログラムをカスタマイズ可能であるということである。例えば学習者のレベルを考えて、新たに選択肢問題の設定を教師が作成したいと考えたとする。しかし一般には、教師側がHTML applet(または servlet)の多くの知識を持っていなければそれは不可能である。それに対し、あらかじめ作成しておいたXML<sup>3</sup>がある特定数の(簡単な)タグを理解し、適当なHTMLのタグや文法に変換されるようにプログラムしておけば、そのタグのみで新たな問題作成が可能となる。つまり、XMLを用いることにより、個々の教師にとって時には大きな障壁となるWeb上言語記述の必要性が大幅に軽減される。この試みは現在実験途中であるが、有力な支援ツールと考えられる。

### 4. おわりに

本稿では、学習者の理解度を向上させることを目指した、遠隔教育のための線形代数教育プログラムについて、その開発の進捗状況を報告した。本年夏の初版完成を目指し、双方向性を重視した、さらにカスタマイズも容易に可能なシステム作りに取り組んでいく。

### 謝辞

本研究は、部分的に文部省科学研究費(国際学術研究、研究代表者 池辺 八洲彦、研究費 ID:10044169)と財団法人福島県学術教育振興財団(研究代表者 池辺 八洲彦)の助成により進められている。

<sup>3</sup> eXtensible Markup Language の略。



## 電子図書館システムを利用した「沖縄の歴史情報」

桶谷猪久夫<sup>1)</sup> 岩崎宏之<sup>2)</sup> 柴山守<sup>3)</sup> 内田智尚<sup>4)</sup>

<sup>1)</sup>大阪国際女子大学 <sup>2)</sup>常磐大学 <sup>3)</sup>大阪市立大学 <sup>4)</sup>日商岩井インフォコム

本デモセッションでは、電子図書館パッケージ“InfoLib”を利用した、「沖縄の歴史情報」のMETA検索システムを紹介する。このシステムでは、平成10年度文部省研究開発予算にて作成された「沖縄の歴史情報」CD-ROM版をコンテンツとし、Dublin Core Metadata Element Setと全文検索を併用した電子図書館システムになっている。METAデータによる画像検索や、情報の多視点・多目的カテゴリ化を実現するコレクション機能、さらに任意なキーワードで目的の研究情報を発見する高速全文検索機能を有し、貴重な研究成果をさらに二次利用できる可用性の高いサービスをご提供できる。また、全てのユーザーインターフェイスをWebブラウザで統一しているため、利用者にとって容易に活用できるシステムである。

### “The Research for Historical Resources of Okinawan Studies” using the Electronic Library

Ikuko OKETANI<sup>1)</sup> Hiroshi IWASAKI<sup>2)</sup> Mamoru SHIBAYAMA<sup>3)</sup> Tomohisa UCHIDA<sup>4)</sup>

<sup>1)</sup>Osaka International University for Women <sup>2)</sup>Tokiwa University

<sup>3)</sup>Osaka City University <sup>4)</sup>Nisho Iwai Infocom

Here we demonstrate the META reference system of “The Research for Historical Resources of Okinawan Studies” using the Electronic Library Package “InfoLib.” This system consists of the full-text search function with the Dublin Core Metabase Element based on the CD-ROM of “The Research for Historical Resources of Okinawan Studies.”

This system provides the picture reference by META data, the collection function to realize the informational multi-viewpoint and multi-purpose categorization, and offers the high-speed, full-text search function enabling the discovery and retrieval of appropriate information with arbitrary key words. Moreover, since all the user interfaces are unified by the Web browser, this is a user-friendly electronic library system.

#### 1. はじめに

貴重な研究成果はもとより、世に「情報」として扱われるものは、それを得た者から他者へ伝達され、さらに利用されることを前提にしている。やがて書き記された「情報」は、知識の宝庫として図書館に集約され、必要とする全ての利用者に対して提供されることとなり、司書という洗練された導き手が

確な案内をしてくれる。

しかし、近年における携帯端末、パソコン、インターネット等の伝達手段の普及・発達により、今や私達の周囲には、あらゆる種類の「情報」がその導き手なしに溢れているのが現実である。

よもや“書物”という形態に留まらなくなった現代の「情報」に対し、それをどこまで有効にやりと

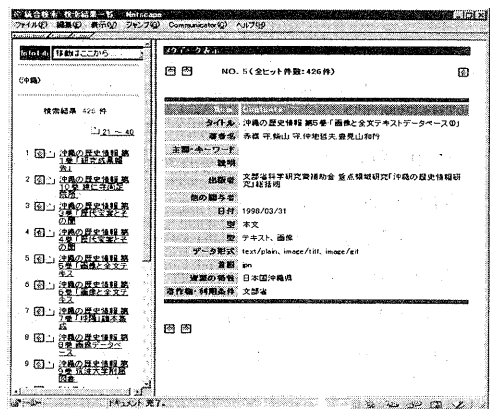
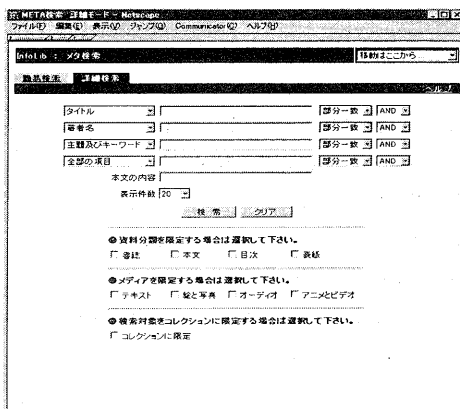
りし、利用していただけるか？ 本デモセッションは、これまで「沖縄」について研究を重ねて来た貴重な成果（テキスト、画像情報による一次情報、二次情報など）を対象に、その試みのひとつとして DC (Dublin Core Metadata Element Set、以下 DC と略す) メタ・データによる「沖縄の歴史情報」統合検索システムを紹介する。本デモンストレーションを機会に、さらに有効な Hyper innovation を実現できればと考えている。

## 2. InfoLib-META (DC メタ・データによる統合検索システム)

DC メタ・データによる統合検索システムのベースは、電子図書館パッケージ OpenLibrary (日商岩井インフォコム (株)) に基づいている。この OpenLibrary は、1997 年に開発したもので、1998 年に Dublin Core Metadata Element Set (DC) を採用したパッケージ InfoLib として図書館 OPAC 等で利用されている。

今回のデモセッションで応用したのは、パッケージ InfoLib のうち、InfoLib-META と呼ぶ部分であ

る。InfoLib-META では、検索対象となるデータについて、そのタイトルや作者などに関する情報を DC に定める内容に集約し、それを検索用情報として扱う。この検索用情報を「DC メタ・データ」と呼び、あらゆる電子情報に対して、この DC メタ・データを付与し、統合的に検索を行うシステムとして実現した。従って、対象となる原情報が文書ファイルであったり、PDF (Portable Document Format) であったり、または動画像情報・音声情報等のマルチメディア情報であっても、DC メタ・データを付与することにより、等しく統合検索を行うことができる。さらに、付与された DC メタ・データ同士を様々な観点からの関連付けそのものに対してもひとつの情報要素として扱えるよう DC メタ・データを付与できるコレクション機能を持つ。今回の「沖縄の歴史情報」では、様々なデータ形式により CD-ROM に格納されているために、各々のデータを有効に扱える本パッケージ InfoLib の機能を活用することとした。



(図-1) InfoLib-META 詳細検索画面 (図-2) InfoLib-META メタデータ表示画面

## 3. 「沖縄の歴史情報」の DC メタ・データ化

「沖縄の歴史情報」は全 10 巻の CD-ROM から構成される、全て電子化された研究成果である。「沖

縄の歴史情報」CD-ROM 版に格納された電子化様式は、以下の通りである。DC メタ・データ化については、プログラム作成と人手の双方による。

(表-1) 「沖縄の歴史情報」 CD-ROM 内のファイル構成

電子化様式	CD-ROM 内での使われ方	ファイルの個数
HTML	GIF 形式、JPEG 形式史料の貼付けなど	728
PDF	研究論文など	154
TIFF	史料など	8,491
GIF	史料など	9,475
JPEG	史料など	11,816
TEXT	CD-ROM 内での目録など	166
※この他にも CD-ROM 内で相互リンクが通れないファイル等が多数あります。		

InfoLib-META パッケージで、標準的に採用して である。  
 いる DC メタ・データの各項目の説明は以下の通り

(表-2) InfoLib-META における標準的な DC メタ・データの各項目説明

No.	項目名	概要
01	Title	名前、題名、タイトル
02	Creator	情報資源の内容に関して責任を持つ人または組織
03	Subject	情報資源に述べられたトピック、主題、キーワード
04	Description	アブストラクトやイメージデータの説明など内容に関する記述
05	Publisher	情報資源を現在の形態にしたもの（出版社、大学など）
06	Contributor	筆者ではないが文書の内容の作成に関わった人または組織（編集者や翻訳者等）
07	Date	現在の形で利用できるようになった日付
08	Type	ホームページ、小説、詩、辞書といった情報資源の型
09	Format	情報資源のデータ形式
10	Identifier	情報資源を一意に識別するための番号あるいは名前
11	Source	情報資源の出所となった情報資源を一意に示す番号もしくは文字列
12	Language	情報資源の内容を記述している言語
13	Relation	他の情報資源との関連づけ
14	Coverage	地理的場所や時間的な内容に関する情報資源の特性
15	Rights	著作権記述などの権利に関する記述や利用条件に関する記述へのリンク（URL もしくは何らかの URI）

上記構造に対し、「沖縄の歴史情報」については以下のようにマッピングした。

(表-6) DC メタ・データ内コレクション（研究単位の場合の例）

情報要素	DC メタ・データ	備考
所属研究テーマ	⇒ 01:Title	CD-ROM 内にて、当該ファイルの格納されている研究テーマを使用
帰属作者	⇒ 02:Creator	CD-ROM 内にて、当該ファイルの格納されている研究テーマの作者を使用
主題等	⇒ 03:Subject	研究論文の記載内容から判断し抽出した語句を使用
PDF のテキスト抽出	⇒ 04: Description	PDF から研究論文本文をテキスト抽出し、全文検索対象とした
出版者	⇒ 05:Publisher	“文部省科学研究費補助金 重点領域研究

寄与者	⇒	06:Contributor	「沖縄の歴史情報研究」総括班”とした 研究論文内に表記された寄与者があれば それを使用
研究成果年度	⇒	07:Date	“1998-03-31”とした
CD-ROM 内役割	⇒	08:Type	テキストは“本文”、画像は“資料”とし た
各ファイルの形式	⇒	09:Format	各ファイルのデータ形式
なし	⇒	10:Identifier	本来は ISBN が入るが、存在しないため空 欄とした
言語記述	⇒	12:Language	“ja”（日本）とした
所属巻号	⇒	13:Relation	CD-ROM 内にて、当該研究テーマが格納 されている CD-ROM の巻号を使用
研究対象地域	⇒	14:Coverage	“日本国沖縄県”とした
研究予算元	⇒	15:Rights	“文部省”とした

#### 4. DC メタ・データ化の効用とコレクション の応用

DC メタ・データを付与することにより、各様であつた情報形式を統合して扱うことが可能になった。また、単に成果としてひとつのディレクトリ下に格納されていた画像史料などが、各々個別の DC メタ・データを持つことにより、単独で意味をもつデータとして扱うことができるようになった。

さらに、コレクション機能の応用により、個別に存在する上記のデータを、各研究の成果単位で統合するとともに、各成果を格納する CD-ROM の巻号単位でもまとめて扱うことができる。これは作者である「研究者という「情報」の帰属属性」と、製本化された状態での「所属巻号という物理的な属性」を表す DC メタ・データの応用方法である。

#### 5. 全文検索の必要性和原情報の提供

DC メタ・データの構成だけで、「沖縄の歴史情報」における研究成果が正しく情報提供できるか？という視点から、研究論文が記載されている PDF 形式のファイルについては、その文中のテキストを全て抽出し、検索対象とすることにした。そして、(少し強引ではあるが) 抽出した全文テキストは、DC メタ・データの「04:Description」に格納し、DC メタ・データ検索と併せて、論文本文に対する全文検索をも可能にした。

上記“全文検索の必要性”とともに、原「情報」

の重要性も考慮している。InfoLib-META では、「情報」に辿り着く一手段として DC メタ・データ検索を行うが、“コンテンツ表示”のメニューで提供される内容は、原情報をそのまま利用者に発信できる。

#### 6. まとめ

1998 年 2 月、W3C により XML に対する正式勧告が公開され、1999 年 2 月には、WWW 文章のためのメタデータ記述方式 (RDF:Resource Description Framework) が承認された。そして、その記述・文法は XML に基づくとされている。また、2000 年 1 月には、同じく W3C から“XHTML1.0”が公開された。

それら新しいマークアップ言語の文法は、“当然 XML を用いる”との前提であるため、今後、文書情報における XML 化が一層進み、データ管理・交換においては RDF 仕様のメタ・データとして応用できるようになると考えられる。

今回の「沖縄の歴史情報」による DC メタ・データによる統合検索システムの試みにおいて、ネット時代における知識管理のあり方を見つめる上で非常に先駆的なものであると考えている。

#### 【参考文献】

[1]岩崎宏之：科学研究費補助金重点領域研究「沖縄の歴史情報研究」CD-ROM 版研究成果報告書、1999.3

## 江戸時代における人口分析システム (DANJURO ver.2.0)

川口 洋・上原邦彦・日置慎治

帝塚山大学経営情報学部

〒631-8501 奈良市帝塚山7-1-1  
tel : 0742-48-8306 fax : 0742-46-4994  
kawag@tezukayama-u.ac.jp

**概要:**「宗門改帳(しゅうもんあらためちょう)」と総称される古文書史料から様々な人口学的指標を算出するために、「江戸時代における人口分析システム(DANJURO ver.2.0)」を構築している。本システムは、「宗門改帳」古文書画像データベースと人口分析プログラム、システムの概要から構成されている。現在、陸奥国大沼郡桑原村、会津郡石伏村、小松川村、鶴巣村、摂津国川辺郡花熊村の5ヶ村における約7万人分の個人情報、約1万世帯分の世帯情報が蓄積されており、60項目の人口学的指標をグラフ表示することができる。インターネットを通じて本システムに接続することにより、江戸時代における人口分析の基礎となる古文書史料と分析方法を研究者間で共有できる研究環境を整備した。

**Demographic analyzing system in Tokugawa Japan using the relational database of Japanese religious investigation registers (DANJURO ver.2.0)**

**Hiroshi KAWAGUCHI, Kunihiko UEHARA, Shinnji HIOK**

**Faculty of business administration, Tezukayama University**

7-1-1, Tezukayama, Nara, 631-8501, Japan

**Abstract:** The most important source data for the historical demography in Tokugawa Japan (1600-1868) is the religious investigation register generally called Shumon-Aratame-Cho (SAC). We have developed a demographic system for analyzing the SAC data. We call this system DANJURO ver.2.0. DANJURO is composed of three parts; the image database of the SAC data, programs for outputting 60 items of demographic statistics and its exposition. Now, we have stored up approximately 70 thousand personal data and 10 thousand household data of five villages. The URL of DANJURO is <http://kawaguchi.tezukayama-u.ac.jp>.

## 1. はじめに

「江戸時代における人口分析システム (DANJURO ver. 2.0)」は、「宗門改帳」と総称される古文書史料から人口学的指標を計算する史料整理の研究過程をできるかぎり自動化するために構築されている [1]。システムの構築目的は、①史料読解から人口学的指標算出に至る作業時間の短縮、②研究過程の再現性の確保、③古文書史料の保存、④研究者間における史料と分析方法の共有の4点である。1999年度には、本システムをインターネット上に公開した。さらに、データ入力作業の効率化を図るため、ニューラルネットワークを利用した古文書文字認識の実験を継続している。

## 2. 「江戸時代における人口分析システム」の概要

本システムの URL は <http://kawaguchi.tezukayama-u.ac.jp> である。Oracle 8.03 を DBMS、Oracle Web Application Server 4.0 を Web Server として構築されている。利用者側コンピュータには、Microsoft Internet Explorer 3.0 以上、Netscape Navigator 3.0 以上などのブラウザと Microsoft Excel 97 以上が必要である。インターネットを通じて、研究者間で史料と分析方法を共有できる研究環境を整備した。

本システムは、①「宗門改帳」古文書画像データベース、②人口分析プログラム、および③システムの概要から構成されている (図1)。現在のところ、「宗門改帳」古文書画像データベースには、陸奥国奥会津地方4ヶ村に攝津国八部郡花熊村を加え、合計5ヶ村、延べ69494人、16272世帯が登録されている (表1)。

表1 「宗門改帳」古文書画像データベースに登録されているデータ

集落名	現在地	史料の年代	延べ人数	延べ世帯数	古文書画像の有無
陸奥国会津郡石伏村	福島県南会津郡只見町	1752-1812	11,593	2,321	無
陸奥国会津郡小松川村	福島県南会津郡下郷町	1792-1868	7,338	1,771	有
陸奥国会津郡鞆巢村	福島県南会津郡南郷町	1790-1859	18,155	4,349	無
陸奥国大沼郡桑原村	福島県大沼郡三島町	1750-1834, 1840-1858	11,486	2,518	無
攝津国八部郡花熊村	兵庫県神戸市	1789-1869	20,722	5,313	無

「宗門改帳」古文書画像データベースは、ア) 個人情報テーブル、イ) 世帯情報テーブル、ウ) 古文書画像情報テーブル、エ) 史料書誌情報テーブルの4テーブルから構成されている。ウ) には入力史料本文の画像データ、エ) には史料表紙、奥付の画像データが登録されている。古文書史料の散逸、劣化が進行する前に史料の現況を画像データとしてコンピュータに蓄積することにより、史料の保存を図った。

ア)、イ)、ウ) の検索結果詳細表示画面には、対応する他のテーブルを検索・表示するためのボタンを付した。利用者が文字データに誤読の疑いを抱いた場合、あるいは史料のレイアウト、文字の配列、筆跡、印形といった文字データとして登録することのできない画像情報を調べたい場合などには、ア)、イ) の文字データとウ)、エ) の画像データを同一画面上で比較対照することができる。このようにして、利用者がシステム構築者の研究過程を再現、検討できる研究環境を構築した。

現在のところ、本システムの人口分析プログラムを用いて、60項目の人口学的指標をグラフ表示することができる (表2)。

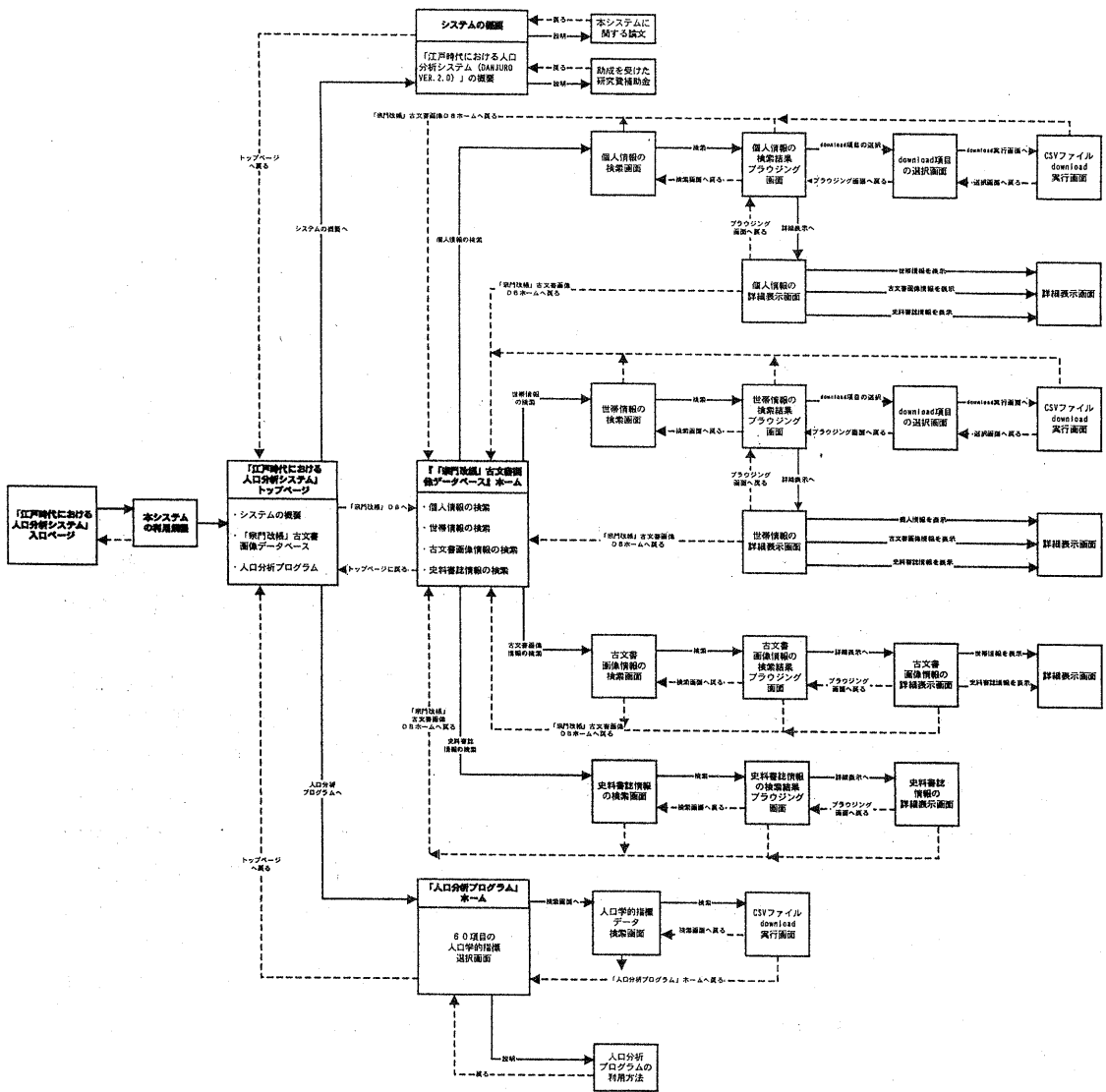


図1 「江戸時代における人口分析システム (DANJURO Ver. 2.0)」の構成

表2 「人口分析プログラム」から出力できる人口学的指標

人口	世帯数	平均世帯規模	世帯規模別世帯数	世帯規模別世帯数の構成比	持高
世帯規模と持高との相関	持高別世帯数	持高別世帯数の構成比	下人・同家人同居世帯数	下人・同家人同居世帯の構成比	同居世代別世帯数
同居世代別世帯数の構成比	家族形態別世帯数	家族形態別世帯数の構成比	平均年齢	年齢階層別人口	年齢構造係数
従属人口指数	性比	年齢階層別性比	宗教・宗派別人口	宗教・宗派別人口の構成比	有配偶率・未婚率
配偶関係別人口(男性)	配偶関係別人口(女性)	出生数	普通出生率	総出生率	出生性比
子供と女性の比率	養子人数	養子年齢	改名者数	改名年齢	改名回数
養子数	養子年齢	流入数	流出数	流入率	流出率
純移動率	婚姻者数	結婚年齢	結婚回数	結婚回数	平均夫婦組数
出産年齢	出産回数	筆頭者となった人数	筆頭者となった年齢	死亡数	普通死亡率
死亡性比	死亡年齢	自然増加率	人口増加率	牛数	馬数

### 3. 古文書文字の自動認識

「宗門改帳」古文書画像データベースを構築する場合、最も長時間の作業を必要とするのが史料読解、文字データ入力である。この研究過程の自動化を目指して、年齢を表記した16種類の古文書文字（ツ、一、二、三、四、五、六、七、八、九、十、壱、弍、年、拾、廿）を対照として、古文書文字の自動認識に関する実験を行った。

まず、「宗門改帳」古文書画像データベースから、「廿」を66個、「廿」以外の15種類の文字を各200個、手作業で採字した。次に、入力層、中間層、出力層の3層から構成されるニューラルネットの範疇でバックプロパゲーション法により、結合係数を出力層各PEの教師付き学習によって求めた。16種類、3066個の古文書文字から各80個（「廿」は33文字）を教師データとして学習させ、残りの未学習文字の認識率を求めた。和泉勇治他の実験によれば、未学習文字1833のなかで正しく認識されたのは1772となり、認識率は約97%に上昇した[2]、[3]。

### 4. おわりに

今後の課題として、第一に、人口学的指標の地域差や婚姻、養子、奉公などによる人口移動を地図上に表示する機能、第二に、家系図をグラフィック表示する機能を開発したい。第三に、研究者のコンピュータをネットワークで結び、相互に比較、利用できるような分散型システムを構築することが望まれる。「宗門改帳」の保存状況が良好な集落は相当数に上るため、全国の史料を個人研究者が収集することは不可能である。古文書史料は、散逸を避けるため、名主や庄屋の子孫、区長の家、教育委員会、図書館、あるいは資料館などで大切に保存されている。民衆生活の具体像を解明するには、研究者や研究期間が地元の「宗門改帳」を分析するだけでなく、人口現象の地域的・時代的特色を相対的に位置付けることが必要となる。そのためにも分散型大規模システムの構築が求められる。第四に、文字データ入力作業の効率化を図るためにも古文書文字の自動認識に関する実験を継続して、認識率の向上を目指したい。今年度は、年齢を表記した16種類3066文字に加え、単位などを表記した16種類（田、畑、高、石、斗、升、合、金、両、分、朱、家、軒、間、馬、疋）、3200文字、合計32種類6266文字を対象として共同研究を続行する計画である。

### 注

- [1] 川口洋・上原邦彦・日置慎治（1999）「江戸時代における人口分析システム（DANJURO ver. 2.0）」の構築、情報処理学会研究報告「人文科学とコンピュータ」、vol. 99, no. 85, pp. 17-24
- [2] 山田奨治・加藤寧・川口洋・原正一郎・石谷康人・柴山守・笠谷和比古・小島正美・梅田三千雄・山本和彦（2000）古文書翻刻支援システム開発プロジェクト報告（1） -プロジェクト概要-、情報処理学会研究報告「人文科学とコンピュータ」、vol. 2000, no. 8, pp. 1-8
- [3] 和泉勇治・加藤寧・根元義章・山田奨治・柴山守・川口洋（2000）ニューラルネットワークを用いた古文書文字認識に関する一検討、情報処理学会研究報告「人文科学とコンピュータ」、vol. 2000, no. 8, pp. 9-15



# GLOBALBASE: 歴史研究における GIS

森 洋久

国際日本文化研究センター  
〒610-1192 京都府京都市西京区御陵大枝山町 3-2  
Tel: 075-335-2165 Fax: 075-335-2090  
E-mail:joshua@nichibun.ac.jp

## あらまし

我々は、歴史研究における空間情報に焦点をあて、地球規模の空間情報の集積を目標とした分散型 GIS システムである GLOBALBASE を開発している。本デモンストレーションでは、この GLOBALBASE の構造を説明するとともに、実際のコンテンツを使った空間情報のブラウジングを試みる。

キーワード GIS、座標変換、マッピング、GLOBALBASE

## GLOBALBASE: a GIS for Historical Studies

Hirohisa Mori

International Research Center for Japanese Studies  
3-2 Oeyama-cho, Goryo, Nishikyo-ku, Kyoto 610-1192, JAPAN  
Tel: 075-335-2165 Fax: 075-335-2090  
E-mail:joshua@nichibun.ac.jp

## Abstract

We develop GLOBALBASE or a distributed GIS system for historical studies which aimed to accumulate global scale spatial information. We illustrate the structures of the server and the client of GLOBALBASE and demonstrate to browse actual spatial information.

**Key words** GIS, Coordinate Transfer, Mapping, GLOBALBASE

### 1. はじめに

近年、GIS(Geography Information System)に代表されるような、大規模な2次元、3次元の空間情報や、また高度な3次元可視化技術の歴史研究への様々な応用が試みられている<sup>1)2)3)</sup>。

我々は特に歴史研究における空間情報に焦点をあて、地球規模で空間情報を集積し得るシステムであるGLOBALBASEを開発している<sup>4)</sup>。

GLOBALBASEのアーキテクチャは、空間情報、およびマッピング情報を集積するサーバと、それをブラウジングし可視化するクライアント、および、これらをつなぐプロトコルからなる。

本デモンストレーションでは、一般発表「歴史研究におけるGIS: GLOBALBASEのための座標変換メカニズムの検討」におけるマッピング・モデルを含めた2次元空間情報のためのGLOBALBASEのアーキテクチャの説明を行う。

### 2.. GLOBALBASE プロトコル(GBP)

GLOBALBASE プロトコルは、GB-LISP と呼ばれるLISPの文法を基本としたLISP言語である。呼び出し側が呼び出される側へ、GB-LISP関数を送ると、呼び出された側がそれを評価し、結果を返すという形で処理が進む。

プロトコルがプログラミング言語であるために、クライアント側が、十分な情報の絞り込みを行う関数を定義し、ネットワーク負荷を下げるなどの最適化が可能である。

### 3. サーバの構造

サーバの構造を図1に示す。サーバはデータベースマネージャ、GB-LISPのインタプリタ、および、ネットワークからの要求の受信処理を行うGBP処理タスク、マッピング・データベースの管理を行うマッピング処理タスクからなる。

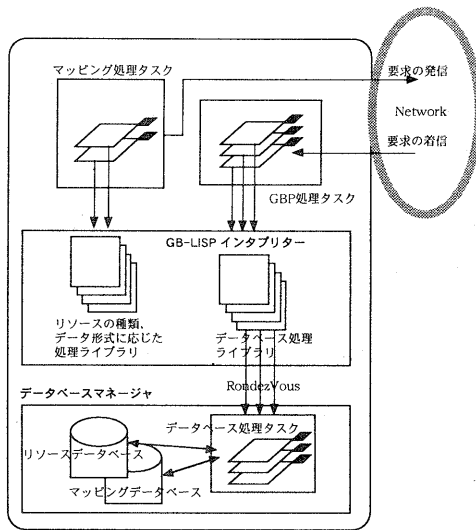


図1 サーバの構造

特に、マッピングに関する情報は、マッピングの始点、終点に登録する必要があるため、マッピング処理タスクは、ネットワークを通じ、双方のデータベースにアクセスする必要がある。

#### 4. クライアントの構造

クライアントの構造を図2に示す。クライアントのブラウジングの詳しい説明は一般発表にゆだねる。

クライアントは、ユーザインタフェースから要求される LOD にマッチした座標系をブラウザによってサーバより得る。得られた座標系はマッピングによる座標系同士の関係を維持した状態でキャッシュされる。

可視化タスクによりされた情報はビットマップの座標系フレームに記憶され、重ね合わせられディスプレイに表示される。

#### 5. デモンストレーション

デモンストレーションにおいては、以下に示すオブジェクトの載った4つ座標系をつなぎ合わせ、可視化する。

- (1) 江戸復元図 (縮尺 1:5000) (東京都教育委員会) 約 283Mbyte のオブジェクトを一つの座標系上で計 28 枚つなぎ合わせた。
- (2) 明治9年 1:5000 東京測量図(陸軍参謀本部) 約 328Mbyte のオブジェクト一つを一つの座標系上に配置。

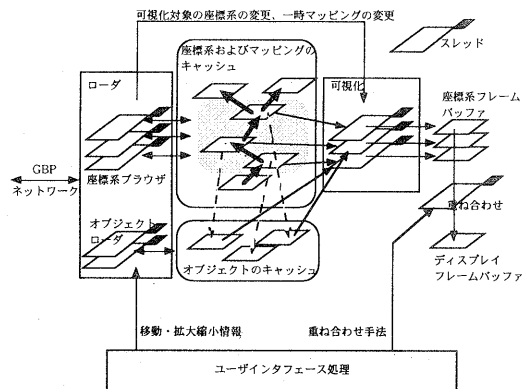


図2 クライアントの構造

- (3) 明治9年 1:1,160,000 大日本全国(陸軍参謀本部) 約 538Mbyte のオブジェクト一つを一つの座標系上に配置。
- (4) ランドサットから撮影された日本全体の写真 約 24Mbyte のオブジェクトを一つの座標系上に配置

この中で(4)の LOD が最小であり、(1)が最大である。これを、拡大縮小しながらブラウジングするデモンストレーションを行う。

#### 6. まとめ

本デモンストレーションでは、GLOBALBASE アーキテクチャのプロトタイプの構造を説明し、実際のコンテンツをブラウズする実験を行った。座標系のブラウジングは、実用範囲のレスポンスで行うことが出来た。

#### 参考文献

- 1) 建設省国土地理院, 「ISO/TC211 地理情報標準化に関する資料」 1996年
- 2) OpenGIS Consortium, *The OpenGIS Abstract Specification Version 4*, June 1999.
- 3) 国際日本文化研究センター, 凸版印刷株式会社, 「洛中洛外散歩」 1999年.
- 4) 森 洋久, 「Globalbase: 歴史研究のための分散型地球規模 GIS」 日本研究とビジュアルデータベース 国際日本文化研究センター 2000年3月

# 国文学と電子資料館

原 正一郎、安永 尚志

国文学研究資料館

国文学研究資料館では電子資料館システムの開発を進めている。本開発では大型計算機からワークステーションを中心とした分散システムへの移行とデータ記述のSGMLが推進されている。プロジェクトのキーワードはデータの標準化、データのハードウェアおよびソフトウェアからの独立およびマルチメディア化である。さらに複数のデータベースを協調させたコラボレーションシステムの実現を目指して、Dublin Core Metadataの導入を進めている。

## Electronic Archival System for Japanese Classical Literature

Shoichiro Hara, Hisashi Yasunaga

National Institute of Japanese Literature

This paper describes the archival system for Japanese classical literature, an ongoing project involving data conversion to SGML and platform migration from a mainframe system to a distributed environment. Project keywords are "standardization of data, system-independent data, and multimedia. This project also promotes collaboration system to link heterogeneous database systems based on Dublin Core Metadata.

### 1. はじめに

国文学研究は、上代の神話から現代に至る我が国の文学作品に関する文学論・作品論・作家論・文学形態論・文学史などを対象とする研究分野である。このうち古典文学は近世以前（明治初期まで）を指す。国文学情報は多種多様であり、その表現形態は文字だけではなく絵や音などを用いたマルチメディアであることが特徴であり、多様な学術情報を多角的に活用することが不可欠である。この国文学研究の対象となる写本・版本などの文献資料は、200万点を越えると言われ、日本国内のみならず世界中に散在しているため、資料の発掘・調査・研究および収集・整理・保存は組織的に行う必要がある。さらに研究者へ情報サービスを行うためには、収集した文献資料をメディアの特性に合わせた各種のデータベースとして組織化しなければならない。つまり国文学研究のための情報システムとは、国文学に関連した各種データベースと、これら複数のデータベースを統合する機構から構成され、研究者の知的発見作業を支援できるものでなければならない。我々はこれを国文学情報システムと呼んでいる。国文学情報システムは目録・原本テキスト・原本画像・典拠など、従来は「モノ」の形態で流通していたデータを、組織的に電子化・蓄積してサービスするものであり、いわゆる電子資料館システムと考えることもできる。

国文学研究資料館は国文学研究に関する資料の調査・収集・整理・保存および研究情報の公開を目的とした人文学系大学共同利用機関であり、主に明治初期までの写本・版本をマイクロフィルム資料として収集し研究者の閲覧に供している。また創設時からコンピュータによる国文学データの組織化に努めてきた。しかし大型コンピュータを中心としたシステムであったため、マルチメディア化やインターネット化への対応が遅れていた。本稿では国文学研究資料館における情報システムの高度化、つまりマル

チメディア化、インターネット化さらにコラボレーション機能の実現を目指した電子資料館システム構築への取り組みについて概説する。

## 2. SGMLを基盤としたデータベースの開発

ハードウェアやソフトウェアの頻繁なバージョンアップ等により、データの維持管理コストが急速に高くなっている。国文学研究資料館のような小規模な組織において、これは大きな問題である。データをハードウェアとソフトウェアから独立させ、同時に多様なサービスを実現するためには、標準規格の導入による資源の効率的運用が必要不可欠である。

目録データにせよ全文テキストにせよ、これらは文字型の不定長フィールドが一定の構造を持ったものとみなすことができる。SGMLは、テキストの構造を記述する能力を持つ国際標準規約である。さらに検索を「テキストデータ中の文字列検索」とみなせば、文字列検索システムを利用したデータベースシステムを考えることもできる。国文学データの検索では、興味の対象を反映した文字列に注目してデータの検索をすることが多いので、この方法は有効である。したがって、SGMLデータを処理できる高速文字列検索ソフトウェアがあれば、これを基盤としたデータベースシステムを構築することは可能である。このような前提に立ち、SGMLを基盤としたデータベースの研究開発を行っている。

## 3. 国文学電子資料館システム

電子資料館システムは、目録データベース、画像データベース、動画データベース、全文データベースなどから構成されている。このうち、目録データベースと画像データベースは関連づけられており、いわゆるマルチメディアデータベースシステムとなっている。

### 3.1 目録データベース

目録データベースとしては、和古書目録（館蔵古書）、マイクロ資料目録（館蔵マイクロフィルム）、論文目録（国文学研究論文）、史料所在およびOPACが公開あるいは準備中である。幾つかの目録データベースは大型計算機上で運用されているため、サービス時間が制限されており、これが海外からの利用の障害となっている。ダウンサイジング化によりサービスの24時間化が可能になるので、この問題は早晩に解消されると考えている。目録データベースの欠点は、所在がわかっても資料そのものにアクセスできないことである。雑誌などとは異なり、国文学資料は稀少でありかつ偏在しているので、これは遠隔地の研究者にとっては大きな問題である。この問題を解消する手段として、次項に述べる画像マルチメディアデータベースの開発を行っている。

### 3.2 国文学研究支援イメージデータベース

画像データベースは、国文学研究資料館蔵古書のマイクロフィルムから作成している。これらの資料は館蔵であるため所蔵権等の問題はない。一方、国文学研究資料館マイクロ資料は、第三者の資料をマイクロフィルム化したものであり、電子的公開を行うためには様々な権利問題を克服する必要がある。

画像データは、白黒2値、解像度をA3換算600DPIでデジタル化を行い、G4圧縮を行った上でTIFF形式でCD-ROMへ蓄積されている。現時点で、約700,000コマ（CD-ROMで約1300枚）のデジタル化が終了している。これは館蔵資料の約70%に相当する。画像データは和古書目録データベースと連携している。利用者は最初に目録データベースを検索して資料の存在を確認し、ついでデータベース間のリンクを辿って画像データへアクセスする。データベース間のリンクにはマイクロフィルムの請求番号を利用している。下図にシステムの例を示す。



図画像データベース

### 3.3 全文データベース

国文学研究資料館が全文データベースの構築に着手した当時、SGMLは普及しておらず、日本語処理の可能なSGML用のツールも存在していなかった。そのため国文学研究資料館では独自のマークアップ規則を作成した。このマークアップ規則をKOKIN規則(KOKubungaku INformation Rules)と呼んでいる。KOKIN規則は、国文学系研究者が利用できるように、明快性と簡潔性を重視して設計されていた。国文学研究資料館には、KOKIN規則あるいはそれに準じた全文データベースとSGMLに基づいたデータベースの、2つの種類の全文データベースが併存している。これまではKOKIN規則による電子化が中心であったが、KOKIN規則用のツールは簡単な語彙解析プログラム程度であり、データを様々に加工することが困難であった。これに対してSGML化されたテキストデータには高度な処理を容易に施すことができる。またKOKIN規則には構文に曖昧性があるなどの問題点もあった。

検索システムにも問題があった。現在KOKINテキストデータのサービスを行っているデータベースシステムは関係データベースモデルに基づいたものである。関係データベースシステムには、SQL(Structured Query Language)やQBE(Query By Example)などエレガントな数理モデルに基づいた、標準的な問い合わせ機能があるが、これらはテキストが有する階層性などの複雑な構造を素直に扱うことが困難である。一方、高速文字列検索装置やプログラムが利用できるようになり、日本語SGMLデータを扱える製品も販売されるようになった。そこで、文字列検索プログラムを利用した全文データベース

システムの開発に着手した。

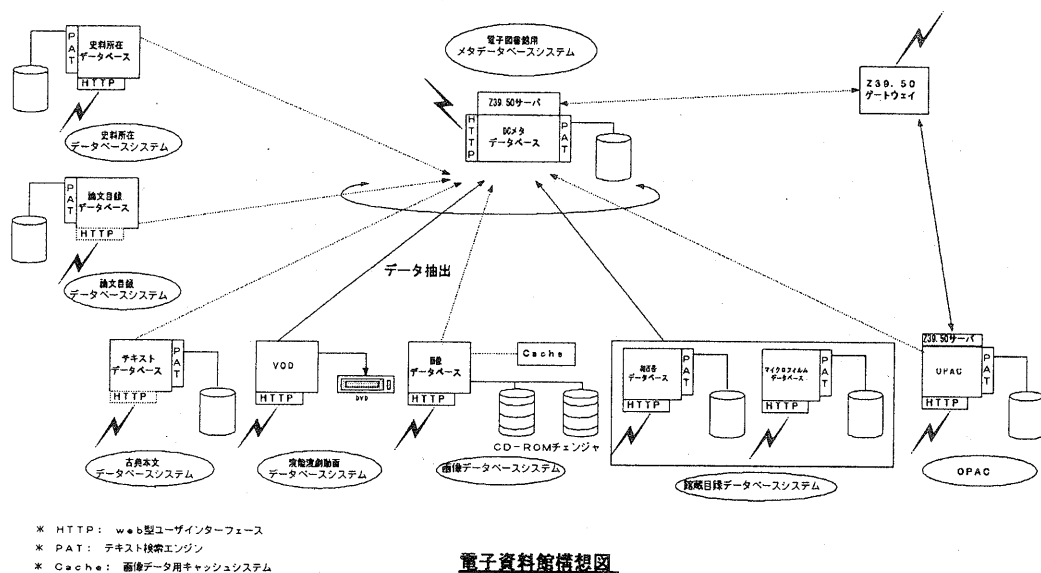
### 3. 4 演能演劇動画データベース

能などのいわゆる演芸も国文学の研究対象であるが、従来は場面の一部を静止画として蓄積して利用するにすぎなかった。演能演劇動画データベースは動画データベースの実験システムであり、DVDに蓄積された動画データをビデオ・オン・デマンド方式で配信するシステムである。

### 4. コラボレーションシステム

複数の情報システムをネットワークで結合し、あたかも単一のシステムであるかのように協調してサービスを行う形態は電子的協調(コラボレーション:Collaboration)と呼ばれている。国文学研究資料館のデータベースシステムは、開発の歴史的経緯から、個別のデータ構造や検索方法を採用していた。そのため、同一組織の史料・資料でありながら網羅的検索ができないという問題を抱えている。

この問題を解決するためにコラボレーションシステムの開発に着手した。目指すシナリオは、例えば、国文学研究資料館の史料所在データベースから「伊能家」を検索すると、やはり国文学研究資料館のマイクロ資料目録データベースから伊能忠敬の「日本経緯度実測」の所在情報、さらに画像データベースからその画像情報、また国文学研究資料館以外の国内外のデータベースからは伊能家に関する研究成果など、関連するあらゆる情報を、単一かつ簡単な操作で、しかも高い精度で検索できることである。これを実現するために、メタデータと検索プロトコルの標準化を考えている。メタデータは検索に必要最小限の共通データエレメントの集合、検索プロトコルは異なるメタデータベースを同じ検索法でアクセスする手段を与える。具体的にはインターネット検索用のメタデータ記述である Dublin Core Metadata Element Set と、書誌情報の検索問い合わせの標準である Z39.50 に基づいたシステムの構築を行っている(下図)。



# テキスト史料の抜粋・分類・俯瞰を支援する歴史学研究支援システム

成瀬聡, 伊東幸宏, 小西達裕, 田村貞雄, 中谷広正  
静岡大学情報学部

本報告では, 対象をテキスト文書に限定し, 史料を電子化して保存すること, 史料の中から歴史学者の独自の視点から関心のある史料を選び出すこと, 選び出した史料を歴史学者の主観に基づいて整理すること, 整理された史料を1つの空間に配置して全体を俯瞰しやすく提示することを支援するシステムの動作を紹介する. なお, 今回のセッションでは, 構築したシステムのデモンストレーションを行う.

## Historical research support system which supports historians in extracting, classifying, and observing historical materials

Satoshi NARUSE, Yukihiro ITOH, Tatsuhiko KONISHI  
Sadao TAMURA, and Hiromasa NAKATANI  
Computer Science, Shizuoka University

Our purpose is constructing a system of supporting researchers in history. In this paper, we illustrate our system that supports historians in constructing text database, extracting historical materials in which they are interested, classifying extracted materials, and observing classified materials. We evaluate the effectiveness of our system by verifying the hypotheses that were set up by a historian.

テムは Java 言語を用いて記述されている.

### 1. はじめに

本報告では, 史料データベースの検索, 史料の分類, 俯瞰を支援する歴史学研究支援システムを紹介する. 1867 年頃, 伊勢神宮などのお札降りを発端として起こった「ええじゃないか」と呼ばれる事件に関する研究と, ほぼ同時期の「コレラ」の流行に関する研究を題材としてシステムの動作を示す.

### 2. 支援システムの構成 [1]

システムは, 研究者の関心ある史料を抜粋する抜粋部, 抜粋された史料を研究者の主観によって分類する分類部, 分類された結果を研究者が俯瞰しやすいように提示する俯瞰部の3つの部分で構成される. また, 文献データベースの構築を支援するツールも備えている. 本シス

### 2.1 文献データベース構築支援ツール

OCR で読み取った歴史史料テキストから, 時間・場所に関する表現を半自動的に抽出し, 史料の時間・場所属性として属性付けられたテキストを HTML 形式に自動変換して史料データベースを構築する. 時間・場所表現の抽出には, 形態素解析ツール「茶筌」[3] を利用している.

### 2.2 抜粋部

抜粋部では, 研究者の関心のある史料を抜粋する. 関心のある史料を抜粋するためには, 歴史学者の関心を何らかの形で表現しておく必要がある. 本システムでは研究者が関心のある史料に多く見うけられる表現のパターンを指定

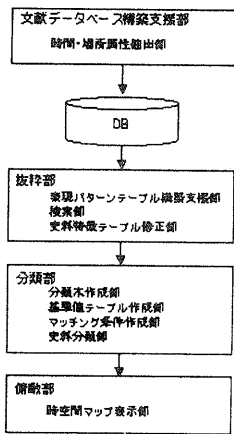


図1.システム構成

することで自分の関心領域を指定する(図2)。ユーザにより作成された表現パターンテーブル中のパターンを各史料ファイルから検索する際にも、「茶筌」を利用している。

### 2.3 分類部

分類木作成部では、ユーザが分類の階層構造を表す木構造を作成するのを支援する。次に、この木の各ノードが表すグループに史料が属するか否かを判定する基準を作成する。判定基準も抜粋基準と同様に、キーワードの出現に基づく。木がすべて作られた後、選択された表現パターンテーブルで、抜粋されたファイルをトップノードから流し、各ノードに該当する史料ファイル郡が対応付けられた分類木を完成させる。

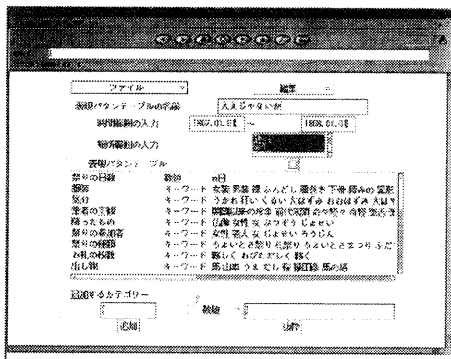


図2. 表現パターンテーブル構築支援部

### 2.4 俯瞰部

俯瞰部では、分類結果を俯瞰しやすいように時間×地図の3次元空間にマッピングする(図3)。

### 3. 実験結果と評価

本システムの有効性を評価するために歴史学上意義ある仮説検証を行う。本研究では、「ええじゃないか」の広がりに関する研究[2]と、幕末期に猛威を振った「コレラ」に関する研究に適用を試みた。その結果、歴史学者から仮説の論拠がより明確かつ、客観的に見渡せるようになったとの評価を得た。

### 4. まとめ

歴史学研究を支援するシステムを紹介した。これによって、今まで見えにくかった全体像や例外事象などの発見が容易になることが期待される。今後、更に多様な歴史学研究課題に対して本システムを適用してみる必要がある。

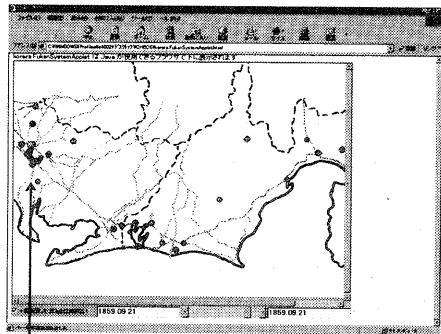


図3. 時空間マップ表示部

### 参考文献

- [1] 伊東幸宏他:テキスト史料の抜粋・分類機能と分類結果の俯瞰機能による歴史学研究支援, 情報処理学会論文誌, Vol.40, No3 (1999)
- [2] 田村貞雄: ええじゃないか始まる, 青木書店 (1987)
- [3] 松本裕治他:日本語形態素解析システム『茶筌』version1.0 使用説明書, NAIST Technical Report, NAIST-IS-TR97007 (1997)