

歴史学研究支援システムの構築

三浦 崇* 伊東 幸宏* 小西 達裕* 田村 貞雄* 赤石 美奈** 中谷 広正* 阿部 圭一*

静岡大学* 北海道大学**

本研究は、歴史学研究の支援を行うシステムの構築を目的とする。歴史学研究は膨大な史料に基づいて進められる。本稿では、歴史研究に利用される歴史史料の特徴を考察し、それに基づいて歴史学研究において支援を必要とするポイントを検討する。今回は、膨大な史料の中から歴史学者の独自の視点から関心をよせる史料を選び出すこと、選び出した史料を一つの空間に配置して全体を俯瞰しやすく提示すること、の2点を取り上げ、そのような支援を行うシステムの構成について述べる。

Constructing a system for supporting researchers in history

Takashi Miura* Yukihiro Itoh* Tatuhiro Konishi* Sadao Tamura*
Mina Akaishi** Hiromasa Nakatani* Keiichi Abe*

Shizuoka University* Hokkaido University**

In this paper, we describe a method of constructing a system for supporting researchers in history. Historians always handle huge historical materials. Considering characteristics of those historical materials, we present an experimental system. The system can help to retrieve historical materials of their interest from a text database and it can visualize the retrieved data in time-location space. For an example, we take a historical event "Eejanaika" which happened in 1867, and build an "Eejanaika" database.

1. はじめに

近年、パーソナルコンピュータやインターネットの普及により、様々な人が多量のデータを取り扱うことが可能になってきた。人文科学の分野でも研究の新しい方法論として、計算機を活用した研究スタイルが考えられるようになってきている。

本研究では、人文科学の中で、特に歴史学を取り上げ、計算機の利用を前提とした歴史学研究の一つのスタイルを検討し、そのために必要なツールの基本構成について述べる。具体的には、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を題材としてとりあげて検討を進める。

まず、2章で歴史学の研究の進め方と、そこで用いられる歴史史料の特徴について考察し、それに基づいて、計算機で支援すべきポイントを整理する。3章で、支援システムを構築するための基本方針を述べ、4章で試作システムについて述べる。

2. 歴史学研究の特徴と計算機による支援形態

2. 1 歴史学研究の一側面

ここでは、まず、歴史学者へのインタビューなどを通じて観察できた歴史学研究の一側面を整理し、計算機の導入によって支援することができそうな問題の洗い出しを行う。

歴史学では、現存する不完全なデータ(史料)を歴史学者の主観を元に分類し、その分類されたデータを時間的な流れ、地理的な広がりという観点から整理し、独自の仮説を構築し、さらにそれを裏付けるデータを探して検証を行っていく、というのが一つの典型的な研究スタイルである。データを分類する際の着目点は歴史学者によって様々である。このようなことは歴史学以外の科学にも共通したことはあるが、歴史学の場合、取り扱うデータ(史料)は以下のような特徴をもつ。

(1) 史料が各地に分散されている。

史料の多くは各市町村史にまとめられている場合が多い。そのような市町村史は地域的に限定された範囲で史料が収集されている。また、

市町村史は流通性が低く、当該市町村に行かなければ実際の史料を総覧することが難しい。

(2) 史料が多量であり、かつ、必要な部分が不必要な部分に混然と混じり込んでいる。

市町村史などに編纂されている史料の総量は極めて膨大である。一つの市町村で、多いところでは、数百から千ページにも及ぶ書籍が5~7巻も編纂されている。その中で、特定の事象に関する知見を与えうる史料はほんのわずかな分量であることも少なくない。

(3) 史料の形式が一定ではない。

史料の編集の仕方が、各市町村により様々である。時代別に分類してあるところもあれば、各日記ごと整理されているところもある。

(4) 史料は完全ではありえない。

ある対象についての歴史史料が網羅的に現存していることは希であり、多くの場合、記録されなかった事実、記録が散逸してしまった事実があることを前提に、存在する史料を取り扱わなければならない。

(5) テキスト以外の多様なメディアで表現された史料も混在する。

歴史史料は、基本的にはテキスト情報が多いが、写真、絵画などの画像情報や、考古学的史料の計測データ、民俗学的な音声データ、動画画像データなども有効である場合が少なくない。

2. 2 問題点の整理と計算機による支援形態の検討

以上の特徴より、歴史学研究を難しくする問題点として次の事柄を挙げることができる。

(1) 史料収集の困難さ

史料の分散性により、地域的に広がりのある一つの事象に関する史料を一人の研究者が総覧することが容易ではない。

(2) 収集した史料から関心のある情報を抽出することの困難さ

たとえ全ての市町村史が一箇所に集められていたとしても、その総数が膨大であり、さらに編纂形式が一定でないため、史料の中から自分の関心に基づいて必要な部分のみを抜き出

すには、最終的には史料を通読しつつ内容検索を行わなければならない、膨大な労力を要する。

(3) 抽出された情報を俯瞰し、有用な知見を引き出すことの困難さ

多大な労力をかけて関心ある情報を抽出しても、それを俯瞰することができなければ一般的知見を引き出すことは難しい。俯瞰するためには、抽出された情報が一定の基準のもとで整理されなければならない。しかし、史料の多様さと多量さのために情報をその基準のもとで整理する労力が大きくなりがちである。例えば、従来情報の整理に利用されてきたノートやカード等の紙のメディアでは、限られたタイプの情報しか載せられないし、また、整理されたデータ量がある程度以上になると、全体を俯瞰することが物理的に難しくなる場合がある。

次に、以上の問題に対する、計算機を利用した支援の可能性について検討を行う。

資料収集の困難さの問題に対しては、歴史史料のデジタル化が各所で進められることを前提とすると、計算機ネットワーク上での分散データベース技術を援用することによって支援可能である。ただし、地域的に分散された箇所それぞれ構築されるであろうデータベースのアーキテクチャの不均一性に拘わらずにデータを流通させられるような機構が必要になる。この点に関しては、本稿では議論を省略するが、我々の研究グループでは、IntelligentPadを用いてアーキテクチャの差異を吸収するという方向で、別途検討を進めている [2]。

収集した史料から関心のある情報を抽出することの困難さを解消するためには、

(1-1) 関心のある史料とそうでないものとを区別するための情報をどのように整理するか、
(1-2) その情報に基づいて膨大な史料を対象に内容検索を行う機構をどのように構築するか、という2点が問題となる。

抽出された情報を俯瞰し、有用な知見を引き出すことの困難さを解消するためには、

(2-1) 情報を俯瞰するために各史料を配置する空間を定め、
(2-2) 各史料がの空間上でどのような位置を占めるかをいかにして定めるか、

(2-3) 必要な各史料が配置された空間を人間が認知しやすい形で表示する機構をどうやって構築するか、

を検討する必要がある。

本稿では、(1-1)から(2-3)までの問題について、テキスト形式の史料を対象として検討する。

2. 3 歴史学研究支援の概略

例えば、歴史学者が次のような発想をもったとする。「ええじゃないか」は領主などの体制側が許容する限度を超えて無政府状態に至るまで爆発してしまった現象であるが、無政府状態とまでは至らずに、体制が許容する範囲内で収斂してしまったところもある。「お札降り」を発端とする臨時的祭典を、

A. 無政府状態にまで至ったもの、

B. やや常軌を逸しているものの体制の許容範囲内で収斂したもの、

C. 通常の祭典の延長あるいは単なる臨時的祭典程度で終わったもの、

の3つのタイプに分けて、その分布を調べることが「ええじゃないか」という現象を理解する上で重要なファクターとなりうる。

直感的には、Aは非農村地域、特に、街道筋の宿場町に多いのに対し、Cは農村地帯に多いように見受けられる。これは、その時代における民衆の困窮度、体制の民衆掌握力、民衆の自我意識などにおける地域差と関係している可能性がある。

このことを確かめるためには、お札降りを発端とした祭典について記載された史料をもとに、その祭りがAからCのどのタイプであるかを判定し、その結果を俯瞰して、各タイプの祭りが行われた場所の分布を調べるという作業を行う必要がある。

本稿では、その過程において、以下のような面から研究支援を行うシステムを検討する。

(1) AからCの各タイプの祭りに関する記述を史料から抜き出してくる作業の支援

(2) 抜き出してきた史料の分布を直感的に把握しやすく提示することによる理解の支援

3. 解法の検討

ここでは、前節でのべた(1-1)から(2-3)の5つの問題の解法について検討する。

(1) 関心のある史料を特定する情報の表現

例えば、上で挙げた例のような発想を歴史学者がもったとき、「無秩序状態に至った祭り」という概念が形成され、史料の中からその実在を見つけ出すという作業に移ることになる。

ところが史料には「無秩序状態に至った祭り」と直接書いた記述はない。そこで、間接的表現をもとに「無秩序状態に至った祭り」に関する記述であることを判定する必要がある。これは、様々な側面からみた「無秩序状態に至った祭り」の特徴を挙げて、それらの特徴を組み合わせて行われる。

例えば、「無秩序状態に至った祭り」がある地域で生じたという事実は、その地域の出来事を記した史料中に以下のような記述が見られることによって確認できる。

- ・ 祭典が行われ、その継続日数が7日以上
- ・ 男装の女性、裸の女性の祭典への参加
- ・ 「たわけ」「狂う」などと評される気分
- ・ 「踊りこみ」といわれる無法行為

そこで、史料中に直接現れることがあり、歴史学者が関心をもった概念に関する史料であることを示す表現を予め整理する必要がある。

ここでは、そのような表現を表1のように表現内容のカテゴリーに基づいて分類して整理することにする。このような表を表現パターンテーブルと呼ぶ。

表1. 表現パターンテーブル

カテゴリー	表現パタン
継続日数	n日 (n>7)
異装	男装、裸、異形
かけごえ	いいじゃないか、ええじゃないか
気分	のぼせ、うかれ、くるう、狂、ゾメキ、忘却
主観的評価	前代未聞、筆舌、筆紙、奇々怪々、言語道断
出来事	踊りこみ

(2) 史料の内容検索

内容検索は、例えばUnixのgrepのような文字列検索ツールを使えばよい。検索結果に応じて各々の史料にカテゴリー毎に見つかった表現パターンを記載したテーブルを作成することにする。このテーブルを史料特徴テーブルと呼ぶ。

(3) 史料を俯瞰する空間

史料を俯瞰する空間としては、歴史学を対象とするのであれば、当面時間と場所からなる3次元空間を考えるのが最も自然である。そこで、本研究では、当面、時間—場所空間における俯瞰のみを考えることとする。

(4) 史料の時間属性、位置属性の決定

時間属性は史料中に現れる時間を表す表現を拾い出すことによって自動的に設定することができる。ただし、歴史史料中の時間表現は、和暦や十干十二支による表現が多く、このままでは取り扱いにくい。そこで、そのような表現から西暦に変換するための変換テーブルを用意し、西暦表現で時間軸を構成する。この際、うるうの取り扱い等に注意を払う必要がある。位置属性についても同様に史料中に出現する地名を拾い出す。次いで、地名から緯度—経度に変換する。このための基礎データとして、「日本地名ファイル(アポック社)」を利用する。

(5) 空間表示機構

時間—場所の3次元空間に視覚的に史料の分布を表示する機構を用意する。これはパーソナルコンピュータ上でも、グラフィック操作機能を持つプログラミング環境(Java, Visual C++等)を利用して比較的容易に作成できる。

4. 支援システムの構成

本研究で構築するシステム全体はHTMLブラウザをインタフェースとして持ち、各機能はC言語及びJava Scriptを用いて記述されている。このことには以下のような利点がある。

- ・ 図形とテキストの組み合わせ、画面内のスクロール、画面内のオブジェクトに対する操作等、ユーザインタフェースに必要な機能の実現が比較的容易である。

- ・ 使用する計算機の機種やOSなどを選ばない。
- ・ ユーザが通常用いている環境でそのまま動作する可能性が高い。そのためユーザが操作を新たに学ぶ必要がないことが多い。またユーザが使用している他のアプリケーションとのデータの授受も容易に行える。

システムは以下のユニットから構成される。

(1) 表現パタンテーブル構築支援部

ある概念に関する資料を検索するための条件を、上述の表現パタンテーブルの形式で定義する作業を支援する。表現パタンテーブルが取り扱える表現パタンには現在のところ以下のタイプを用意している。

- ・連続値をとる属性に対して、その属性を示すキーワードと、範囲を指定する。(例:「祭りの継続日数が7日以上」と指定する際、キーワード“日”と、範囲“>7”を記述する。)
- ・離散値をとる属性に対して値の集合を指定する。(例:かけ声について“いいじゃないか”もしくは“ええじゃないか”というキーワードを指定する。)

ユーザ(歴史学者)は関心をもつ概念を表現するパタンを選択する。システムはパタンのタイプに応じてカテゴリー名とキーワード等を要求し、表形式のデータを生成する。

今後は以下のようなパタンも取り扱えるようシステムを拡張する予定である。

- ・正規表現など、ある程度マッチングに許容性をもつパタン。
- ・予め定義されたクラスによるパタンの指定。例えば、「農具」という指定で、「鋤」「鍬」など全ての農具を表す語を指定できる。

(2) 時間・位置属性抽出部

3の(4)に述べた方法で、史料データベースに登録された各史料ファイルに対して、俯瞰のキーとなる時間属性・位置属性を半自動的に抽出し、ファイルに書き込む。ユーザは抽出結果に対して、適宜修正を施すことができる。

(3) 検索部

ユーザにより作成された表現パタンテーブルをもとに各史料ファイルを検索し、ユーザが関心を持つ概念に関連するファイルを抜き出す。検索は以下のように行う。

- ・時間属性、位置属性については、時間・位置属性抽出部によって各史料ファイルに書き込まれている値とユーザの指定をパタンマッチする。
- ・その他の属性については、表現パタンとして指定されたキーワードを本文とのパタンマッチにより探す。パタンが範囲を指定するタイプである場合には、指定された範囲に該当するか否かを調べる。

(4) 時空間マップ表示部

検索部によって抜き出された史料ファイルを、時間・緯度・経度の3次元空間上に配置して表示する。ユーザが表示させたい時代(年号、期間)を選択すると、その時代に関連する史料が地図上のプロットとして表示される。ユーザはプロットの色・形・大きさ等を、検索に用いた表現パタンテーブル毎に指定することができる。これによりユーザの関心のあることがら時空間上でどのように分布しているかを俯瞰することが容易になる。

(5) 文献データベース構築支援ツール

市町村などに印刷物として保存されている史料から、本システムで史料ファイルとして取り扱えるテキストファイルを作成するツールとして、以下のような環境を用意している。

- ・OCR(Optical Character Reader)として、パーソナルコンピュータ上で動作する“読んデココVer.2.01(AIソフト社)”を用いている。
- ・作成されたテキストファイルをハイパーテキストとして扱うために、HTML文書への変換を自動的に行う。

現在、我々は実験システムを試作中である。実験システムの動作例を図1に示す。現在、静岡県、愛知県内の市町村史24冊(付表)から「ええじゃないか」発生前後の記事を抜粋してテキストデータベースを構築している。全体で約3000文字程度の大きさである。内容的には、静岡県、愛知県内の22地域における「お札降りを発端とした祭典」に関する記述を含んでいる。各史料を、見出しなどの形式的特徴に若干の内容に基づく判断を加えて分割し、結果的に182個のHTMLファイルを構成し、その各々から時間属性、場所属性を抽出している。

図は、お札の種類によって史料を分類して表示した結果である。●が伊勢神宮、□が秋葉神社、○が豊川稲荷、■が津島神宮のお札についての記述がある史料を表しており、その史料に記載されている場所にプロットされている。上段左が1866年9月22日から翌年9月21日、右が同22日から翌年9月20日、下段左が同21日から翌年20日の様子を示している(365日間隔)。また、画面上でプロットをクリックすることにより、そのテキスト史料を開いて見ることもできる(下段、右)。

5. まとめ

歴史学の史料の特徴とそれに基づく歴史学研究の困難さを考察し、計算機によって歴史学研究を支援する形態について検討した。また、試作を進めているシステムの概略について紹介した。今後は、ユーザインタフェースを改善すると共に、史料データベースを拡張して、実際の歴史学研究において有用な仮説検証への適応実験を行っていく必要がある。

謝辞

本研究の一部は文部省科学研究費（重点領域07207208）の補助による。

【参考文献】

- [1]田村貞雄：ええじゃないか始まる、青木書店(1987)
- [2]渡辺,中谷,赤石,伊東,阿部,田村：異種メディアからなる歴史史料のデータベース構築、人工知能学会研資、SIG-HIDSN-9603-1, pp.1-6(1997.3)

付表 データベースに格納した市町村史

菊川町史	近現代通史編	pp.33-36
金谷町史	資料編 2	pp.684-696
金谷町史	資料編 3	pp.1-9
舞阪町史	中	pp.14-19
豊田町史	通史編	pp.722-723
新居町史	第 8 巻	pp.884-885
磐田市史	資料編 2	pp.814-817
磐田市史	資料編 3	pp.524
雄踏町史	資料編 1 0	pp.213-214
相良町史	資料編近世(2)	pp.831-832
佐久間町史	資料編 3 下	pp.497-498
豊川市史		pp.447-449
南知多町史		pp.3-6
吉良町史		pp.15-18
中町祭礼帳		pp.409-914
岡崎市史	近世 3	pp.868-877
安城市史		pp.705-709
刈谷市史		pp.646-647
刈谷市史	第 2 巻	pp.899-904
蒲郡市史		pp.527-528
音羽町史		pp.599-602
豊根村史	資料編 2	p.470
一宮市史	資料編 1 0	pp.955-956
小牧市史	資料編 3	pp.547-560

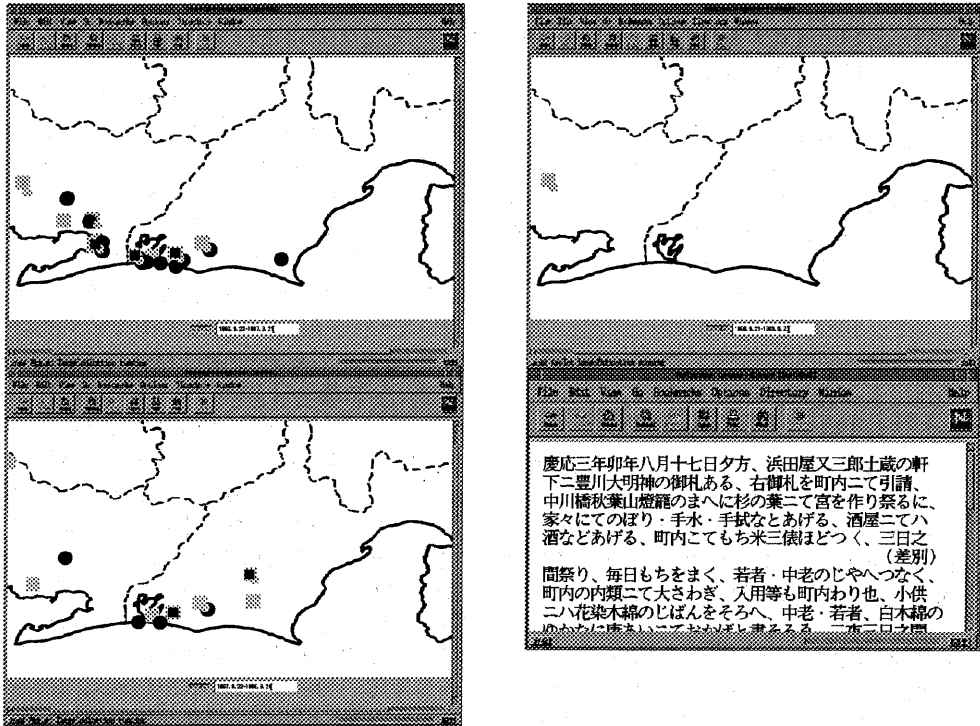


図 1 動作例