

## 歴史学研究支援システムの構築

## Constructing a system for supporting researchers in history

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\*

Takashi MIURA\*, Yukihiro ITOH\*, Tatsuhiro KONISHI\*, Sadao TAMURA\*,  
Mina AKAISHI\*\*, Hiromasa NAKATANI\*, Keiichi ABE\*

\* 静岡大学情報学部、浜松市城北

\*\* 北海道大学、札幌市

\*Shizuoka University, Jouhoku 3-5-1, Hamamatsu-City, Shizuoka,

\*\*Hokkaido University,

キーワード: 史料データベース、抜粋、分類、俯瞰

**Keywords:** text database, extraction, classification, indication

あらまし: 本研究では、歴史学研究の支援を行うシステムの構築を目的とする。本稿では、歴史学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し、その具体的な実現方法、その方法に基づいて実装したシステムを紹介する。今回は、対象をテキスト文書に限定し、膨大な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、そして整理され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3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。そして、その方法に基づきシステムを実装する際のポイントやデータ構造の検討を行う。最後に、今回作成したシステムを示し、実際に歴史学研究の仮説検証をし、評価を行う。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仮説を歴史学研究者から提示してもらい、検証を行った。

**Summary:** Our purpose is constructing a system of supporting historians. In this paper, we describe a method of constructing the system, considering characteristics of historical materials and examining how to support historians. In this paper, we describe three points of im-

plementing the system. The first point is how to extract historical materials in which historians are interested. The second point is how to classify extracted materials. The third point is how to display classified materials. We discuss our methods of implementation and structures of data. Finally, we evaluate the effectiveness of our system by verifying the hypothesis which were set up by a historian. The hypothesis is concerned with the historical event "Eejanaika" that happened in 1867.

## 1 はじめに

近年、パソコン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となってきた。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、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章でシステムの構成について述べる。5章で、実際に作成したシステムを紹介し、6章でそのシステムで具体的な歴史学の仮説検証を行った結果を示し、その評価を行う。

## 2 歴史学について

### 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巻も編集されている。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、各日記ごとに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

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。

史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、Intelligent-Pad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[1]。

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。

抽出のプロセスでは、研究者の関心のある史料と関心のない史料を区別するだけではなく、関心のある史料を研究者の主観に基づいて整理する必要がある。

この様に抜粋された史料を分類する時には、

(2-1) 研究者の主観に基づいた史料の分類基準をどのように整理するか、

(2-2) その分類基準に基づいて分類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、

(3-1) 情報を俯瞰するために各史料を配置する空間を定め、

(3-2) 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、

(3-3) 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 (3-3) 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 問題の具体化例

以上の問題を明確にするために、具体的な事例に対して適用してみる。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、

「お札降り」を発端とする臨時の祭典が、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」を発端とする臨時の祭典を、

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の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。

従って、前述の(1-1) から (3-3) の問題は、この事例に関しては以下のように整理される。

1. 史料群から「ええじゃないか」について記述のある史料を抜きだしてくる作業の支援
2. 抜きだされた史料をAからCのタイプの祭りに分類する作業の支援
3. 抜きだ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

以下、この事例を用いながら、各々の解法について説明する。

## 3 解法の検討

ここでは、前節で述べた(1-1) から (3-3) の7つの問題の解法について検討する。

### 3.1 関心のある史料を特定する情報の表現

例えば、2.3で挙げたような例のような発想を歴史学者がもったとき、「ええじゃないか」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「ええじゃないか」が起こったと直接書かれた記述は少ない。過去の史料を元に現代になって再編集された2次史料には、よく見られる表現だが、1次史料にはほとんど書かれていない。そこで、間接的な表現をもとに「ええじゃないか」に関する記述であることを判断する必要がある。そこで、様々な側面からみた「ええじゃないか」の特徴を挙げて、それらの特徴を組み合わせることによって、「ええじゃないか」に関するものと、そうでないものを区別する基準を定めることにする。

例えば、「ええじゃないか」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。

- お札が降り、祭典が行われた
- 祭りの日数が何日も続く
- たわけ、狂うなどと評される気分
- 男装したり、女装したり、裸になって踊る
- 金、餅等が蒔かれる

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	表現パターン
お札降り	天降、お札、降札、御降り
継続日数	n 日
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、言語同断
異装	男装、女装、裸、異形、ふんどし
蒔物	米、餅、餅米

### 3.2 史料の内容検索

内容検索は、キーワードによって文字列を拾い出す文字列検索ツールを使えばよく、n 日などという表現に関しては、正規表現によって検索を行うことによって実現できる。検索結果に応じて各々の史料にカテゴリーごと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。このテーブルを史料特徴テーブルと呼ぶ。

現在、先程の「ええじゃないか」のみを例に挙げて検討を行っているが、実際はこれらの史料を参照すべき研究テーマは数多く存在する。従って、各史料には研究テーマに応じて幾つもの表現パターンテーブルによって特徴付けが行われる。よって、各史料特徴テーブルでは、各史料の特徴を各々の表現パターンテーブルで示された基準ごとに整理する。表2に史料特徴テーブルの構造を示す。

表2: 史料特徴テーブル

ええじゃないか	
カテゴリー	表現パターン
継続日数	7 日
気分	狂
主観的評価	前代未聞、筆紙
異装	男之姿、ふんどし

(新居町史より)

### 3.3 抜粋された史料を分類する基準

分類は、抜粋で関心のあるものと、無いものに分けられた史料を、主観に基づいて整理する。2.3で述べた例では、お札降りによる臨時的祭典が起こった史料をA～Cの3タイプに分類している。

歴史学に限らず、一般的に分類結果は木構造であることが多いため、ここでは木構造を用いて表すこととする。木の各ノードには、史料特徴テーブルがどのような値を取った場合に、そのノードに属するかを示すテーブルが必要となる。このテーブルを基準値テーブルと呼ぶ。表3に基準値テーブルの構造を示す。

### 3.4 史料の分類方法

史料の分類は、上で述べた木構造において、各ノードの基準値テーブルと史料特徴テーブルのマッピングを再帰的に行うという方法で、実現する。

表 3: 基準値テーブル

ええじゃないか	
カテゴリー	表現パタン
継続日数	7 < n
気分	狂、ゾメキ
主観的評価	前代未聞、筆紙
異装	ふんどし

### 3.5 史料を俯瞰する空間

史料を俯瞰する空間としては、歴史学を対象としているので、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。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。

### 3.6 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(アボック社)」を利用する。

### 3.7 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。これはパソコン上でも、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

ここでは、支援システムの構成と、その実現方法、システムで扱われるデータ構造について述べる。本研究で構築するシステム全体はHTMLブラウザをインターフェースとしてみち、各機能はJava言語を用いて記述されている。このことには、以下の利点がある。

- 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。

システムは大きく分けて抜粋部、分類部、俯瞰部という3つの部分で構成される。さらに、俯瞰部でデータを扱う準備として、時間・位置属性抽出部が加えられる。また、2.2で史料の電子化が各地で行われるということ的前提としたが、現状ではまだ仮説検証に十分なデータを得られる程ではないため、文献データベース構築支援ツールを用意する。

### 4.1 文献データベース構築支援ツール

市町村に印刷物として保存されている史料から、本システムで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、パソコン上で動作する”読ん de 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HTMLファイルへと自動変換を行う。

### 4.2 時間・位置属性抽出部

3.6で述べた方法で、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。

### 4.3 抜粋部

抜粋部では、研究者の関心のある史料を史料データベースから抜粋する。ここは、表現パタンテーブル構築支援部と、検索部に分れる。

#### 4.3.1 表現パタンテーブル構築支援部

ここでは、ある概念に関する史料を検索するために、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は、基準値テーブルで値のマッチングを行なうために、表現パタンのタイプを宣言する必要がある。現在、表現パタンテーブルが取り扱える表現パタンには以下の2つのタイプを用意している。

- 数値+文字列を取る属性に対しては、n + 文字列を指定する。(例:「祭りの継続日数」に関しては、「n日」という表現を記述する。)
- 離散値をとる属性に対しては、値の集合を指定する。(例:「かけ声」については、「ええじゃないか」、「いいじゃないか」というキーワードを指定する。)

研究者は関心のある概念を表現する表現パタンのタイプを選択し、その概念のカテゴリ、表現パターンを入力し、表現パターンテーブルを作成していく。

今後は、以下のパターンもシステムで取り扱えるように拡張してく予定である。

- 予め用意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。

#### 4.3.2 検索部

ユーザにより作成された表現パターンテーブルを基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。

- 表現パタンのタイプが n + 文字列である場合については、史料中から数字列と文字列になっているパターンを抜き出す。
- その他の属性に関しては、表現パターンとして指定されたキーワードを本文とのパターンマッチにより探す。

抜き出されたパターンについては、史料特徴テーブルに書き込む。

#### 4.3.3 抜粋部におけるデータの構造

抜粋部では、表現パターンテーブルの作成を行い、それに基づいて、文献データベースから抜粋する。ここで、取り扱う文献データベースは膨大な量であるため、抜粋に要する時間を節約するために以下の2つの方法を取る。

##### 1. 時間、場所の範囲指定

史料検索の際に、研究者の概念に明らかに必要のないデータを検索しない様に、時間、場所の範囲を指定して検索範囲を限定できるようにする。例えば、「ええじゃないか」の研究を行う時に、縄文時代の史料は必要のない場合がほとんどである。従って、予め時間、場所の範囲が限定されるような場合は、範囲

を指定して検索を行うと、無駄な時間を省くことができる。

現在、与える時間範囲については年月日で指定でき、場所範囲については県のレベルで指定できる。場所範囲については、関東や中部といった範囲や静岡県西部といった範囲への対応も検討する必要がある。

##### 2. 新たに追加されたデータにのみ検索を行う

研究者が表現パターンテーブルを修正した際に、既に検索したデータに対して再び検索を行う無駄を省くために、スキャン履歴を用意する。

これらの方法を実現するために抜粋部では、以下の3種類のデータを用意する。これら3種類のデータ構造については、後述のシステムの動作を踏まえて検討を行った。

##### ● 表現パターンテーブル

表現パターンテーブルでは、3章で述べたカテゴリ、表現パタンの他に、そのカテゴリのタイプ、表現パターンテーブル名、ユーザが指定した時間、場所範囲、そして、未走査ファイルリストが保持される。

表 4: 表現パターンテーブル

ええじゃないか
1867.01.01 ~ 1868.01.01
愛知, 静岡
file1, file2
祭りの日数: 数値:n 日
服装: キーワード: 女装, 男装, 裸,
気分: キーワード: うかれ, 狂い,

##### ● 史料管理テーブル

史料名、史料の時間属性、位置属性とその位置属性が属する県が書かれる。

表 5: 史料管理テーブル

anjou1.html,1867.09.15,36593458,愛知
arai1.html,1867.09.26,37343441,静岡
iwata1-1.html,1867.9.14,37513444,静岡

##### ● スキャン履歴

スキャン履歴に書かれるのは、過去にどの表現パターンで検索し、どの史料が抜粋されたかという情報である。

表 6: スキャン履歴

eejanaika.ept,anjou1.html
eejanaika.ept,arai1.html
eejanaika.ept,arai2.html

これらのデータを元に、検索部では表現パタンのタイプに基づいて、検索が行われる。検索が行われた史料特徴テーブルは、表7のような構造になる。

表 7: 史料特徴テーブル

時間属性
場所属性
表現パタンテーブル 1
カテゴリー 1 表現パタンのタイプ1 表現パタン
カテゴリー 2 表現パタンのタイプ2 表現パタン
表現パタンテーブル 2
カテゴリー 1 表現パタンのタイプ...

次に、研究者が表現パタンテーブルを構築する際に行う操作について検討し、その際のシステムの動作を考察する。

#### 1. 史料の追加

新しく史料が追加された時は、表現パタンテーブルにまだ検索を行っていないファイルとして登録し、次回検索する際に、それらのファイルをどの様に扱うかをユーザの判断に任せる。

#### 2. 史料の削除

史料の削除が行われた場合には、スキャン履歴から対象となる史料と、史料特徴テーブルを削除する。

#### 3. 検索範囲の変更

検索範囲の変更は、増加部分と減少部分に分けられる。検索範囲の増加部分の処理は、増加範囲の史料を検索し、スキャン履歴に追加する。減少部分は、減少範囲の史料を取り出し、各々の史料特徴テーブルから対応する表現パタンテーブルを削除し、スキャン履歴から削除する。

#### 4. 表現パタンテーブル名の変更

表現パタンテーブル名の変更が行われた場合には、もとの表現パタンテーブルの履歴からファイルを取り出し、各史料特徴テーブルに書かれている内容をコピーし、スキャン履歴に追加する。

#### 5. 新しい表現パタンテーブルの作成

新しい表現パタンテーブルが作成された場合には、指定されている検索範囲に当てはまる史料を取り出して検索を行い、マッチした場合には史料特徴テーブルに書き込み、履歴に追加する。

#### 6. 表現パタンテーブルの削除

表現パタンテーブルの削除が行われた場合には、スキャン履歴から、その表現パタンテーブルでスキャンしたファイルを取り出し、各史料特徴テーブルから、その表現パタンテーブルを削除する。そして、スキャン履歴から削除を行う。

#### 7. カテゴリーの追加

新たに、カテゴリーが追加された場合は、指定されている範囲に当てはまる史料を取り出し、検索を行う。見つかった史料のうち、既に史料特徴テーブルにその表現パタンテーブルあるものに関しては、その表現パタンテーブルの部分に新しいカテゴリーを追加する。無いものに関しては、新たに表現パタンテーブル、カテゴリーを追加し、スキャン履歴に追加する。

#### 8. カテゴリーの削除

カテゴリーの削除が行われた場合には、スキャン履歴からその表現パタンテーブルで拾い出されたファイルを取り出し、削除されたカテゴリーを史料特徴テーブルから削除する。その結果、史料特徴テーブルにその表現パタンテーブルの部分空になった時は、スキャン履歴から削除する。

#### 9. 表現パタンの追加

カテゴリーの追加と同様に行う。

#### 10. 表現パタンの削除

カテゴリーの削除と同様に行う。

これらの操作を有効にするために、一度検索が行われると、その表現パタンテーブルのコピーが作られ、次回検索した時にユーザがどのような操作を行ったかを判断する。

### 4.4 分類部

分類部では、研究者が関心のある史料を抜き出した後、主観に基づいて分類を行う。ここは、4つのユニットから構成される。

## 4.4.1 分類木作成部

分類木作成部では、ユーザに分類を行なう木構造を作成してもらう。分類は、各史料が持っている史料特徴テーブルがどのような値を持っているかによって行われる。例えば、2.3の例では、抜粋されたノードにA, B, Cという子を持つ木を考える。その各ノードには、史料特徴テーブルがどのような値を取れば、そのノードに属するかという条件が必要となる。

従って、各ノードには、基準値テーブルが与えられる。但し、基準値テーブルだけでは、単純な単純な条件の羅列しかできず、複雑な条件を指定しきれないので、後述するマッチング条件が必要となる。よって、分類木の各ノードには基準値テーブル、マッチング条件、ファイルリストが保持される。

分類木作成部の動作は以下のようになる。まず、ファイルが木のトップノードに置かれ、子のそれぞれのノードにそのファイルを渡し、子のノードの基準値テーブル、マッチング条件による基準を満たすと、そのノードに追加していく。

また、他の兄弟ノードのどのノードにも属さない場合にそのノードに属するという「その他」という特別なノードを用意する。

## 4.4.2 基準値テーブル作成部

基準値テーブル作成部では、ノードの基準値を作成していく。各ノードでは、表現パタンのタイプによってマッチングが行われる。

- 表現パタンのタイプが数値+文字列である場合は、その数値の範囲を指定する。
- その他の場合は表現パタンの値を選択する。

基準値テーブルが一通り作り終わると、マッチング条件作成部を呼出し、基準値のマッチング条件を作成する。

## 4.4.3 マッチング条件作成部

マッチング条件作成部では、基準値のマッチング条件を作成する。現在システムでは、1段階の積和の形のみを扱うことができる。

## 4.4.4 史料分類部

史料分類部では、木が全て作られた後、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、木を完成させる。

## 4.4.5 分類部におけるデータの構造

## 1. 基準値テーブル

分類部では、複数の表現パターンテーブルによって抜粋された史料を分類することがあるため、基準値テーブルの構造は、史料特徴テーブルと同様に、表現パターンテーブル別にカテゴリー、表現パターンが整理される。

## 2. マッチング条件

マッチング条件では、各ノードの基準値データのマッチングに and や or の条件をつける。この条件のデータは整理しやすいように、常に2項演算の形で保持する。

## 4.5 俯瞰部

## 4.5.1 時空間マップ表示部

俯瞰部では、分類された結果を研究者が俯瞰しやすい様に時間×地図の3次元空間にマッピングを行う。分類された結果は、リストの形で渡され、各リストには表示する、色、形、大きさが含まれる。

## 5 システムの紹介

ここでは、作成した支援システムの各メニュー等の紹介を行う。

## 5.1 抜粋部



図 1: 抜粋

## (1) メニューバー

## (1-1) ファイルメニュー

このメニューはファイルとやり取りを行うメニューで、その内容は以下の5つで構成される。

- 新規作成  
新しく表現パターンテーブルを作成する際のメニュー。このとき、現在表示している内容は破棄する。
- 開く  
既に作成された表現パターンテーブルの読み込みに使われる。このメニューを選択すると、今までに作成された表現パターンテーブルのリストが表示され、選択された表現パターンテーブルが表示される。
- 保存  
「開く」によって読み出されたファイルを上書き保存する。
- 名前を付けて保存  
新しいファイルに名前をつけて保存する。
- 表現パターンテーブルの削除  
現在表示されている表現パターンテーブルを削除する。

#### (1-2) 編集メニュー

編集メニューは、カテゴリの削除を行う時に用いる。

- 削除  
このメニューが選択されると、追加するカテゴリに書かれているカテゴリを削除する。

#### (2) 表現パターンテーブル名入力フィールド

表現パターンテーブルの名前を入力するフィールド

#### (3) 時間範囲入力フィールド

検索する時間範囲を入力するフィールド

#### (4) 場所範囲入力フィールド

検索する場所範囲を入力するフィールド

#### (5) 最大化ボタン

このボタンが押されると、現在表示されている表現パターンテーブルがブラウザいっぱいに表示される。

#### (6) 表現パターンテーブル

表現パターンテーブルを表示する部分で、ここで、カテゴリが選択されると、追加するカテゴリに値が表示され、編集を行うことができる。

#### (7) 追加するカテゴリ

新しくカテゴリを追加する場合にこのフ

ールドに書き込む。タイプに関しては、選択式になっており、表現パタンのタイプを選択する。

#### (8) 追加ボタン

追加ボタンが押されると、追加するカテゴリに書かれた内容が、表現パターンテーブルに追加される。

#### (9) 抜粋ボタン

抜粋ボタンが押されると、現在表示されている表現パターンテーブルで、4章で説明したやり方で抜粋が行われる。

## 5.2 分類部

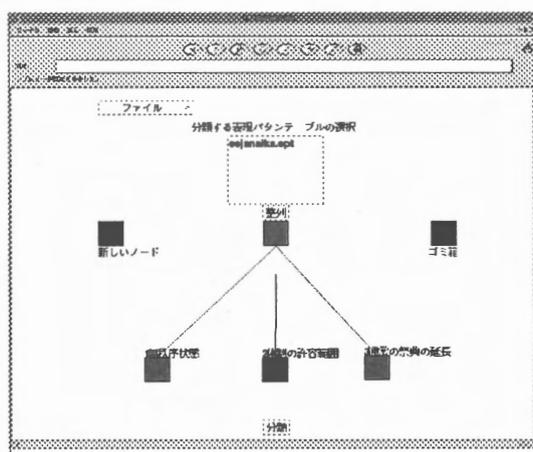


図 2: 分類

#### (1) メニューバー

##### (1-1) ファイルメニュー

このメニューは、抜粋と同様にファイルとやり取りを行うメニューで、以下の4つで構成される。

- 新規作成  
新しく分類木を作成する時に用いる。これが選択されると、現在表示している木は破棄される。
- 開く  
既に作成された分類木を読み出す際に用いられる。
- 保存  
現在表示されている木を上書き保存する。
- 名前を付けて保存  
現在表示されている木を新たなファイルに保存する。

- (2) 基準となる表現パターンテーブルの選択  
分類したい史料の基準となる表現パターンテーブルの選択を行う。複数の表現パターンテーブルを選択し、基準とすることが可能である。
- (3) 整列ボタン  
このボタンが押されると、現在表示されている木を整頓する。
- (4) 木作成エリア  
ここでは、自由に木の作成を行えるインタフェースを用いている。木を作成する際の個々の操作について説明する。
- (4-1) 新しいノードの作成  
新しくノードを作成する時は、左上の「新しいノード」をドラッグし、このエリア内で離すと、新しいノードがワウスが離された位置に作られる。
- (4-2) 新しい枝の作成  
枝を作る時は親ノードをドラッグし、子ノードで離すとそのノード間が結ばれる。離された位置にノードが無い場合には、ドラッグされたノードが移動する。分類木は完全な木構造なので、複数の親ノードは作ることができない。
- (4-3) 枝の削除  
既にある枝を削除したい場合には、既にある枝を枝の作成の時と同じようにドラッグを行うと、枝は削除される。
- (4-4) ノードのコピー  
ノードをコピーしたい場合には、コピーしたいノードをドラッグし、「新しいノード」で離すことにより、内容の複製されたノードが、元のノードのとなりを作成される。
- (4-5) ノードの削除  
ノードを削除する場合には、削除したいノードをドラッグし、右上の削除の位置で離すと、ノードは削除される。この時、削除するノードに子ノードが存在する場合には、削除するノードの親ノードの下に追加される。
- (5) 分類ボタン  
このボタンが押されると、選択されている表現パターンテーブルで抜粋された史料が、木のトップノードに追加され、それらのフェイルを子ノードに渡していき、分類が行われる。

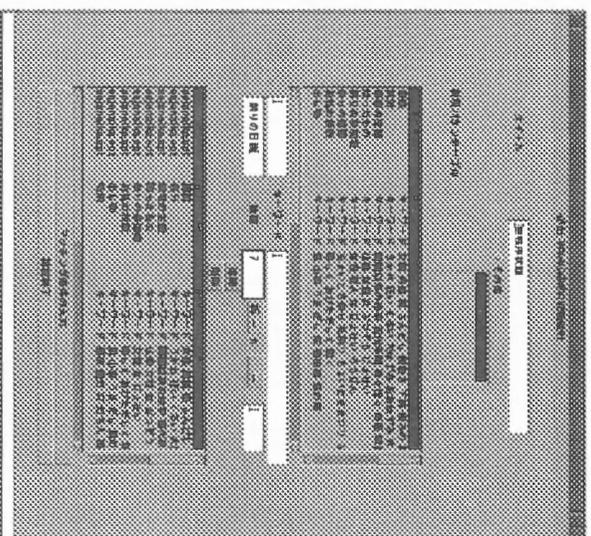


図 3: ノードの設定

- (6) ノードの値設定ウインドウ  
各ノードでダブルクリックされると、そのノードの基準値テーブルの作成用のウインドウが開く。
- (6-1) 表現パターンテーブル選択  
ここでは、(2)で選択したフェイルが表示される。そのうちの一つを選択すると、その内容が(6-2)に表示される。
- (6-2) 表現パターン表示  
(6-1)で選択された表現パターンテーブルの内容の表示を行う。ここで、カテゴリーが選択されると、(6-3)に表現パタンのタイプ別に表示される。
- (6-3) 値入力フィールド  
ここでは、基準値テーブルに追加する値を設定する。タイプが数値の場合は、条件の部分は選択式になっている。
- (6-4) 追加ボタン  
このボタンが押されると、値入力フィールドに書かれている内容が基準値テーブルに追加される。
- (6-5) 削除ボタン  
このボタンが押されると、値入力フィールドに書かれている内容が、基準値テーブルから削除される。

**(6-6) 基準値テーブル表示**

ここでは、基準値テーブルを表示する。ここで、カテゴリーを選択すると、値入力フィールドに表示され、編集することができる。

**(6-7) マッチング条件入力ボタン**

このボタンが押されると、(7)のマッチング条件入力ウィンドウが開かれる。

**(6-8) 設定終了ボタン**

このボタンが押されると、ノードの値設定ウィンドウを閉じる。

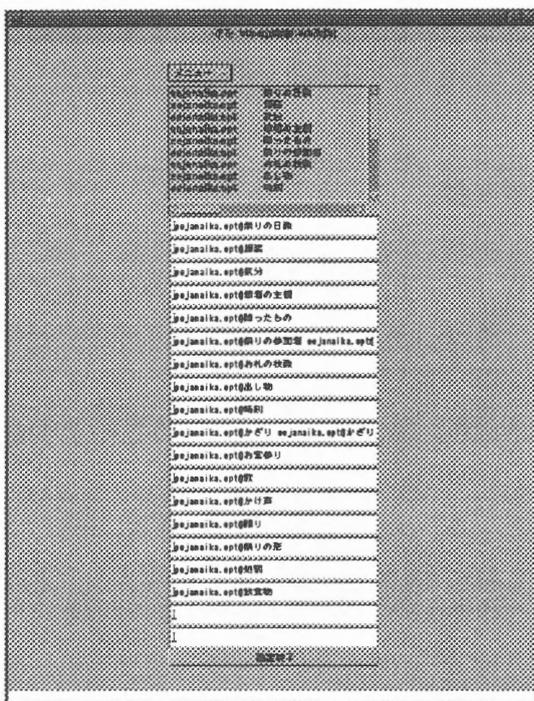


図 4: マッチング条件の入力

**(7) マッチング条件入力ウィンドウ**

ここでは、マッチングする際の条件を設定する。

**(7-1) メニューバー**

ここでは、マッチング条件を入力する際に用いる2つのメニューで構成される。

- コピー  
現在選択されているカテゴリーの内容を保持する。
- 張り付け  
保持されているデータを、現在選択されている条件フィールドに追加する。

**(7-2) 基準値テーブル表示**

(6)で作成された基準値テーブルの値を表示します。条件フィールドに追加する場合には、基準値テーブル表示でカテゴリーを選択し、次にメニューのコピーを選択する。そして、追加したいフィールドを選択して、メニューの張り付けを選ぶ。

**(7-3) 条件フィールド**

この条件フィールドは、基準値テーブルのカテゴリーの個数分作られる。一つのフィールド内のカテゴリーはand条件、それぞれのフィールド間の関係はorとする。

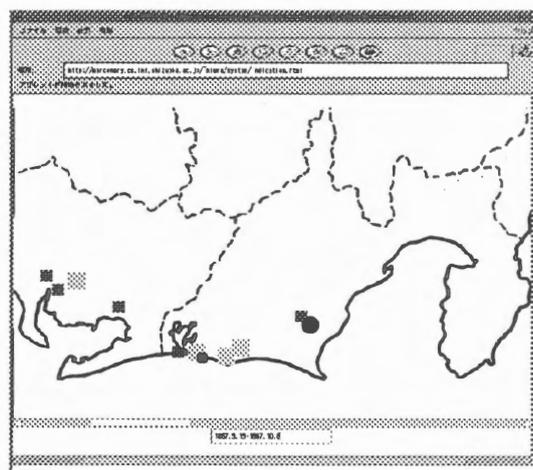
**5.3 俯瞰**

図 5: 俯瞰

**(1) 表示エリア**

ここでは、分類結果を表示する。ここで、プロットがクリックされると、そのプロットが示しているファイルを表示する。

**(2) 時間範囲表示フィールド**

ここでは、現在表示されている時間範囲を表示する。

**(3) 時間範囲変更バー**

ここでは、表示したい時間範囲を変更する。この値を変更すると、(2)も変更され、(1)で表示する内容も変化する。

**6 評価**

本システムの有効性を評価するためには、実際に歴史学上意義のある仮説検証を本システムを利用

して行ってみる必要がある。我々は、現在、以下の4つの課題を取りあげて、仮説検証作業を行っている。

1. 2. 3で述べた仮説を静岡県西部、愛知県東部において検証する。
2. 名古屋などの大都市では、300種類以上のお札が3000枚降っている。これらのお札は、伊勢神宮、秋葉神社、津島大社、豊川稲荷が、ほとんどを占めている。中には、旅行者が置いて行く場合もあるが、これらのお札の種類を調べることにより、その地域の人々の信仰状況が把握できる。
3. 1の仮説を関西地域で分布させると、関西地域では街道筋や、農村地帯というようなきれいな分布にならないと思われる。その理由は、4つめの仮説で説明する。
4. 3で行った結果を政治的な緊迫度で色分けされた地図に表示する。関西のお札降りは、純粋なお札降りによる祭りというだけでなく、政治的な意図を感じとることができる。

第1のステップとして、1の仮説検証に必要な静岡県西部、愛知県東部の江戸時代末期の史料データベースを構築し、仮説検証作業を進めている。現在のところ、史料が不足していることと、表現パターンや基準値などが、最初に直観的に設定したものをそのまま利用しており、リファイン作業が十分に行われていないことなどにより、十分な検証には至っていないが、試用している共著者の歴史学者の印象評価によると、自分の思いつきが比較的簡単な作業で可視化して確認できる点で、有用性を認められるという感触を得ている。

## 7 まとめ

歴史学研究における史料の特徴とそれに基づく歴史学研究の問題点を整理し、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、それに基づいて、実際に仮説検証を行った。今後は、課題1の検証を進めると共に残る3つの課題に関して検証を行って、本システムの有用性を実証する必要がある。そのためには、今までの機能の他に、以下の追加機能が必要となる。

- 街道や宿場町を地図に追加する機能  
現在、白地図にプロットを行なっているが、当時の街道、宿場の位置は記述されていない。

それらの位置を自由に追加できる機能が必要となる。

- 政治の緊迫度の表現

4の仮説では、地図上に政治的な緊迫度を表すプロットが必要となる。ええじゃないかとは違い、緊迫度は連続値であるために補正する必要がある。例えば、「ある日にお札が降った。」という記述があり、次の日にお札ふりの記述が無い場合には、その日はお札ふりは無しと判断することができるが、「ある日ある場所の政治的緊迫度が高かった。」という記述があり、次の日に緊迫度について何も書かれていない場合でもその緊迫度は0ということにはならない。

## 参考文献

- [1] 渡部、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)
- [2] 田村貞雄：ええじゃないか始まる, 青木書店(1987)
- [3] 三浦、伊東、小西、田村、赤石、中谷、阿部：歴史学支援システムの構築, 情報処理学会研究報告 Vol97, No.48, pp.25-30(1997.5)