

## 時間情報を持つ人文科学 DB の統合検索のための ユーザインタフェースの検討

安達文夫<sup>†</sup> 原正一郎<sup>‡</sup> 柴山 守<sup>‡</sup>  
<sup>†</sup>国立歴史民俗博物館 <sup>‡</sup>京都大学

**あらまし** 人文科学の様々なデータベースを統合的に検索するシステムの研究が進められている。このデータベースの中で、時間情報は有用であるが、表記が多様であるとともに、本来的に幅が存在する性質を持つ。このような時間情報に対して統合的な検索を実現するには、時間情報を正規化し、区間で記述することが必要となる。このとき幅を持つ時間情報に対しても始点と終点を条件として指定する検索が可能となる。この条件を指定するインタフェースについて検討するとともに、太陽暦を正規化時間とする際のユーザインタフェース上の影響を示した。さらに、幅を持つ情報を含めた検索結果の並べ替えとその表示について検討を加えた。

### A Study on a User Interface for Integrated Retrieval of Humanities Database containing Temporal Information

Fumio ADACHI<sup>†</sup> Shoichiro HARA<sup>‡</sup> Mamoru SHIBAYAMA<sup>‡</sup>  
<sup>†</sup>National Museum of Japanese History <sup>‡</sup>Kyoto University

**Abstract** Temporal information is very important in the humanities. The information is described with various manner and usually has some range. Its normalization and description with start and end points is necessary for integrated retrieval of various Humanities database. A new query that is assigned start and end for data with a temporal range becomes possible on the integrated retrieval system. This paper discusses a user interface to assign temporal information to the system and sort and indication method for data that has temporal range.

#### 1. はじめに

人文科学に関する様々な情報がデータベース化され提供されている。これらのデータベースを、その所在や個別の操作方法を意識することなく統合的に検索できるシステムの検討が進められている[1]。統合検索を実現する上で、共通メタデータに個々のデータベースのデータ項目をどのように対応付けるかが重要な課題となる。比較的汎用的な Dublin Core メタデータを共通メタデータとする際の対応付けやその課題について、これまで幾つか報告されている[2-4]。

人文科学のデータベースにおいて、時間情

報は有用な項目である。Dublin Core も時間情報のエレメントを有し、これを用いて統合検索が可能である。しかし、人文科学の分野で現れる時間情報は、後で詳述するように、表記法が多様であるとともに、記述される時間に幅が存在する性質を持つ。これは、データベースの記述対象やその製作等の時期によって大きく異なる。個別のデータベースを検索する上では、時間情報の内容は比較的均質であり、その記法も個々のデータベースの中では統一されていることから、時間情報を文字列の一致で検索することに大きな支障は生じていない。しかし様々な内容と記法による

複数のデータベースを対象とするには、時間情報を文字列で一致させる検索では有用性を失う。このため、個々の時間情報を標準的な形式に変換した情報を基に検索処理を行うことが必要となる。そして、これにより、全てのあるいは幾つかのデータベースを対象として始点と終点を検索条件として与えるこれまでにない統合的な検索が可能となる。

この統合的な検索を実現する上で、人文科学の分野での時間情報が本来的に幅を持つことから、時間に関する検索条件の分かりやすい指定方法の検討が必要となる。検索結果を表示する上でも、特に時間情報によるソートの際に、幅を持つ情報をどのように並べて表示するかが課題となる。そして、国内外の人文科学に関する情報を対象とするには、システム内では太陽暦による正規化時間で扱うことが適切である。このとき従来と変わらないユーザインタフェースを構成することが求められる。

本稿では、国立歴史民俗博物館（以下、歴博）のデータベースを基に、多様な時間情報の表記について整理を行い、幅を持つ時間情報をも対象とする際の検索条件の指定方法と、太陽暦を正規化時間とする際のユーザインタフェースへの影響、検索結果を並べ替えて表示する際の課題を明らかにし、時間情報の統合検索のためのユーザインタフェースの在り方について提案する。

## 2. 時間情報の多様性と検索の課題

歴博の公開データベースには、所蔵資料の目録情報を収録した館蔵資料(詳細)DB、研究分野毎の文献や共同研究で調査・収集した成果をまとめた研究成果 DB、主に記録類を対象とした全文 DB がある。時間に関わる情報は重要であり、資料の製作や使用に関する時期、文書や記録類の日付、文献の発行年など何らかの時間に関わるデータ項目を少なくとも一つは有している。この時間情報はデータベース毎、あるいは一つのデータベースのデータ項目毎に様々に表記されている。その例を多様性の内容毎に整理して、表 1 に示す。

時間を表す文字は、同表の(1)のとおり、アラビア数字と漢数字が使われる。アラビア数字では半角と全角の表記がある。漢数字では、原資料の記載どおりにデータベースに記述する場合があることから、元年、正月の表記や廿の文字が出現する。元号も慶応と慶應のような異なる表記が存在する。なお、和暦では閏月があることに留意を要する。

時間情報の表記方法には、(2)のとおり、紀元前後の区分、時代、世紀、西暦年、和暦年と幾つかの形式がある。そして、表現される期間の幅が多様である。年月日による表記も、(3)のとおり、年月日、年月、年、あるいは元号だけと様々で、時間の幅に多様性がある。これは、原資料にどこまで詳しく記されているかによることが多い。なお、元号だけの表記は時代の一種と見ることもできる。

表1 時間情報の各種の表記

項番	表記例
1	安政2年、寛政2年、慶応二年、慶應元年、応仁三年正月廿八日、天和三亥年閏五月
2	BC、弥生時代、近世、17世紀、1810年、文化7年
3	天正10年6月2日、天正10年6月、天正10年、天正、1921年8月10日、1921年8月、1921年
4	鎌倉時代～室町時代、13c～14c、弘化4年～嘉永5年
5	平安前期、鎌倉中期、室町後期、江戸初期、江戸末期 縄文早期、縄文晩期、14c初、18c末、天保後期
6	1190年頃、文政頃、明暦～寛文年間頃
7	室町時代力、天明元年か、明治7年？、927年頃？

一つの時代や世紀で、時期を表現できないときは、(4)のように、時代や世紀の始まりと終わりで表記される。逆に、一つの時代または世紀を区分するため、(5)のように、前期、中期、後期などを付した表記がある。期間が長い時代では初期、末期とさらに細分する場合がある。縄文早期、晩期のように時代に特有な語もある。このように、時期を細分する付加的な語は、一律ではない。

時間情報の記述の中には、(6)のように、時期を確定できない場合に幅を持たせた“頃”を付けた表記がある。また、(7)のように“か”や“?”を付けた表記がある。これば時期を断定できないことを意味する。

以上のように、時間情報の表記は多様であるが、(1)の文字の種別や元号の表記のゆれは、正規化の問題である。ただし、検索結果の表示においては、漢数字は原資料に記述されたとおりとすることが適切とも言える。時間情報が幅を持つことに対し、これを対象として始点と終点を指定して検索するには、時間情報の始まりと終わりをシステム内で記述しておく必要がある。このとき(5)の時代を区分する表現では、その区間を明確に定義できるも

のではないが、何らかの値で記述しておく必要がある。これはシステム側で一方的に与えるのではなく、データベース毎にその作成側で与えることが適切と考えられる。(6)の“頃”も時期を確定できないが、同様に区間による記述が必要である。時期を断定できないことを意味する(7)は、付加的な記号を外した時間情報に対して検索処理を行い、結果を表示する際に断定できないことを示すことが適切である。

### 3. 検索情報入力時のインターフェース

#### 3.1 時間の区間を指定した検索

歴博のデータベースの多くは、時間情報の検索において、年あるいは時代名の文字列の一致により行われている。しかし、時間情報の始点と終点を指定する検索が求められることから、幾つかのデータベースではこれを指定する検索を提供している。その検索画面の例を図1に示す。現状では、年月日または年(月)が確定した時間情報だけが、この検索の対象となっている。これに対して、幅を持つ時間情報を対象にしようとする、利用者がどのような条件を満たすものを一致と見な

---

検索 クリア 全件表示 結果表示件数: 50

表題	
作成	
宛名	
年代	

分類: 分類 a:   分類 b:   分類 c:

西暦年: 自   ~ 至

フリーワード:

分類a一覧
分類b一覧
分類c一覧

---

図1 検索画面の例

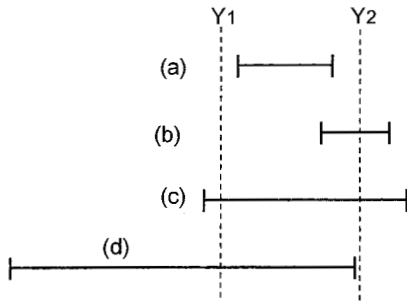


図2 幅を持つ時間情報の検索

しているかが問題となる。この例を図2に示す。ここでは、始点としてY1、終点としてY2が指定されたものとしている。検索対象の持つ時間の幅との関係で、幾つかの検索要求が生ずると考えられる。

(1) 想定される利用者要求

利用者が指定する始点Y1、終点Y2と検索対象の時間情報の幅の関係から、次のような検索要求が考えられる。

- (i) Y1～Y2の区間に収まっているもの(図2(a))だけを一致とする。
- (ii) Y1～Y2の区間に掛かるもの全てを一致とする。(図2(a)～(d))
- (iii) Y1～Y2の区間に掛かるが、その区間に対して外れる部分が大さいもの(例えば図2(d))を除いたものを一致とする。

先に記したとおり、実際のデータベースの時間情報の中には、時期が不確定な“頃”を付した記述がある。これと同様に検索条件の入力においても

- (iv) Y1頃～Y2頃の意味で条件指定をする。要求が考えられる。これは組合せとして、(i)～(iii)それぞれについてあり得る。

以上の検索を実現する上で、システムの処理に大きな問題はない。これらの要求を満たす使いやすいユーザインタフェースを構成できるかが課題となる。

(2) 入力インタフェースと利用者要求の整理  
時間情報に関する検索条件をグラフィカル

に指定する入力インタフェースも考えられるが、ここでは付加的な条件の指定が実用的であるかを検討することが目的であるため、図1にあるような始点と終点を入力するインタフェースを想定して検討する。

先の(i)～(iv)の全ての条件を利用者が選択できるようにするには、検索対象の時間の幅が区間内だけか区間外も含むかの指定、区間外も含む場合にその全てとするか区間外に対して幅を持たせるかの指定、曖昧な“頃”の指定の機能が必要である。これは、入力インタフェース上、3個の選択ボタンの追加で実現できる。しかし、指定の意味が分かりやすいとは言えず、機能を整理することが好ましい。

まず、“頃”の指定は、利用者が始点と終点を広く指定することで代替できる。選択ボタンを少なくする方が、ユーザインタフェースとして簡明になる。歴史分野の研究者からは、条件に一致する可能性のあるものは検索結果として残した方がよいとの意見が多い。このことから(ii)が基本となる。(iii)は関連するもの全てを取り出したいことからすると例外的な要求と言える。また、これは必要であれば、区間を広く取った(i)の結果を、対象となる始点・終点を入力した(ii)による絞り込み検索で得られる。このため、区間から大きく外れるものを一致とするか否かの選択は、入力画面上に置くことはないと言える。そこで、区間に掛かる対象を全て一致とする条件をデフォルトとして、区間内に含まれるものだけの選択機能を設ける入力インタフェースが適切と考えられる。

3.2 時代名による検索

(1) 文字列の一致と区間の一致

検索画面の実例として示した図1に見られる“年代”は、現状のシステムでは文字列によるデータとして格納されており、その検索も入力された文字列との一致を見ている。統合検索として、年代(時代)を始点と終点の情

報でシステム上に持つ場合、入力された時代に該当する区間を条件とする検索が可能となる。一方、データベースの記述どおりに検索する文字列一致も求められる。したがって、時代名入力の際に、利用者が区間一致を求めるものか文字列一致を求めるものかを識別できる入力インタフェースが必要となる。おそらく専門の研究者に対しては、区間の指定は3.1節で記した始点と終点の指定で充分であり、時代名入力は文字列一致を優先とすることが適切である。一方、一般向けには区間検索を優先とすることが有効であろう。時代名入力の主たる利用者がどちらかによって、文字列一致か区間一致かのデフォルトが決まることになる。さらに、区間一致の場合は、始点・終点を入力する場合と同様に、一致の条件が区間内外かの指定も必要となる。

## (2) 異なる解釈がある時代の区間設定

時代の始まりや終わりは、捉え方によって複数の解釈があり得る。例えば、鎌倉時代の始まりを1185年とする場合と1192年とする場合がある。検索条件に一致する可能性のあるものを出力するためには、前後する時代の終点と始点が重なるように値を設定する必要がある。このとき、検索対象の時代の区間と、検索語として入力される時代の区間を、ともに重なるように設定してのでは、例えば先の鎌倉時代を入力したときに、平安時代と記述されたデータが常にヒットする不都合が生ずる。

1185年から1191年までの年(月日)で表記されたデータを平安および鎌倉で一致させるには、検索語を受け付ける側で、重なりがあるよう区間を設定する必要がある。検索対象となる時間情報の始点・終点はデータベース作成側の立場で与えることが適切である。

## 3.3 入力時間情報の解釈

### (1) 入力情報と年月日の精度

ここまでは、時間情報の単位を年として検

討してきた。検索対象の持つ時間の精度は様々であるが、表1に示したように日を単位とする情報が存在する。これを区間を指定して検索するには、それが利用者からY1とY2の年だけが入力されたとしてもY1年1月1日～Y2年12月31日とシステム側で解釈して処理する必要がある。入力された時代名を基にその区間を条件として検索する場合も、時代の始まりと終わりに日までの精度は意味がないが、システム側で始点と終点に年月日を与えて処理する必要がある。

なお、検索対象となる時間情報の始点と終点を、システム内でどのような単位で記述しておくかは、検索のメカニズムとの関係で決まる。

### (2) 正規化時間とユーザインタフェース

国内外の人文科学の情報を幅広く対象とするには、統合化システムの中では、太陽暦による正規化時間で扱うことが適切である。和暦で表記された時間情報も、太陽暦に変換された正規化時間で表現される。

一方、日本歴史の分野では、西暦年として和暦年から形式的に置き換えた年が慣習的に使用されている。このため、始点と終点を指定した検索として、例えば図3に示すように、1600～1610と入力されたときに、システム側で太陽暦の1600年1月1日～1610年12月31日と解釈して検索処理を行うと、利用者が暗黙に期待する慶長5年1月1日～慶長15年12月30日(1600年2月15日～1611年2月11日)ではなく、慶長4年11月15日～慶長15年11月17日に一致するデータが出力される。慶長4年11月15日～慶長4年12月30日のデータが余分に出力されることは許容されても、慶長15年11月18日～慶長15年12月30日のデータが出力されないことは問題となる。このように、検索画面の西暦年の入力欄に、利用者が与えられた時間情報が、太陽暦の意味か和暦年から機械的に置き換えた西暦年の意味かを区別する必要

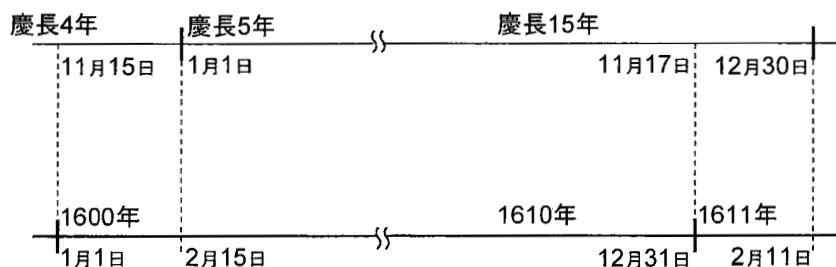


図3 和暦の西暦による検索

がある。

時代名が与えられて区間一致で検索する場合、日の単位まで厳密に考慮する必要はないが、処理する上では、年月日を与えることになる。このとき、利用者の慣習を考えると形式的な西暦年と解釈する方が合理的であろう。一方、世紀と入力された区間一致の場合は、太陽暦か形式的な西暦かを区別することが必要と考えられる。

太陽暦か形式的な西暦かの指定は、利用者の研究分野により異なるため、システム側で機械的にどちらかをデフォルトにすることで解決しない。利用者がある環境で一回設定すれば、その環境において、その設定が有効となるような機能が必要である。

#### 4. 検索結果の時間情報による並べ替え

検索した結果を時間の順に並べ替えて表示する機能は重要であり、歴博の幾つかのデータベースでも提供されている。データベースに記述されている時間情報は、表1に示したように、年月日～年、あるいは元号や時代、世紀などがあって、時間幅が多様である。さらに～前期や後期といった表記もある。ここでは、時間の幅を持つ情報に対して、検索した結果を、時間順序に並べて表示する際に生ずる課題について検討を加える。

##### (1) 従来の並べ替えと要求条件

歴博のデータベースのソートでは、元号、年、月、日の時間の精度に応じて、図4のよ

うな並べ替えをしている。すなわち、元号年月までのデータは、元号年月日を持つデータの該当する月の後ろに配置するというように、不確定な情報はより確定的な情報の後ろに配置している。そして、“近世”といった元号も不明なデータは、最後に配置する。不確定な情報を確定的な情報の前に並べるタイプもある。

これは原理的には、元号年を形式的な西暦年に置き換えた4桁のデータyyyyと、月を表すmm、閏月を1とするフラグl、日を表すddにおいて、不明な部分は9999または99とするyyyymmldddのデータを持って、この順に並べることにより実現している。

以上より、近い時間の情報は集めるが、不確定な情報は離れた位置に並べることが求められていると読み取れる。

文化13年 8月10日  
 文化13年 8月23日  
 文化13年 8月  
 文化13年閏 8月12日  
 文化13年12月  
 文化13年  
 文化14年10月29日  
 文化  
 文政 1年 9月  
 ・  
 近世

図4 精度の異なる時間情報の並べ替え

## (2) 並べ替えのための時間記述

元号年月日の形式で日までの情報を持たない情報は、その月、あるいは年、元号の始点と終点の年月日で記述しておき、その終点(または始点)を基にソートすることで、上述と同じ並べ替えができる。この年月日は、順序の問題であるから、太陽暦による正規化時間で問題ない。世紀で表記されるデータも、この方法で、年(月(日))によるデータと一緒に並べ替えて表示しても不自然ではないと考えられる。

これに対し、時代、特にその終点(または始点)に異なる解釈があるものを年(月(日))で表記されるデータとともに並べて表示することが適切かの問題が生ずる。また、江戸前期、後期といった表記による時間情報は、検索の処理のためには始点と終点を定めなければならないが、確定したものではないため、これを基に並べ替え、年(月日)が確定した情報と合わせて表示することは適切とは言えない。歴史分野の研究者の意見によると、前期等で修飾された情報は、修飾を受けない情報(先の例では江戸時代)とまとめて表示することが適切である。その表示位置は、形式的であっても時代の終了年月日が望まれる場合と、年までが確定した情報と確定しない情報を分けて並べることが望まれる場合があると考えられる。また、“頃”が付く時間情報も、検索用の情報での並べ替えは不適切である。“頃”の付かない情報に続けて並べることが自然である。

このように、検索条件との一致をみるための時間情報と、結果を並べ替えるための時間情報を別に持つ必要があることになる。

## 5. まとめ

人文科学に関するデータベースにおいて、多様に表記され、また本来的に幅を持つ時間情報を対象として、統合的な検索を行うためのユーザインタフェースについて検討を行い、

以下を明らかにした。

- ・時間の区間検索において、検索対象が幅を持つときに、関連する可能性のある情報は全て出力することを原則として、指定区間に掛かるものを一致とする条件検索が基本となる。

- ・検索入力インタフェースを簡明にするため、検索条件の付加的な指定は、上記をデフォルトとし、区間内に収まる情報だけを一致とする条件検索をオプションとする選択に絞るのがよい。

- ・検索語として時代名が入力されるときに、文字列の一致か、時代が表す区間での一致かの指定がユーザインタフェースとして必要である。

- ・区間を検索するための情報は、利用者の指定によらず、始点と終点を年月日の単位で与える必要がある。

- ・システム内の正規化時間は太陽暦が適切であるが、利用者が西暦年として入力する情報が、太陽暦か和暦から機械的に西暦に置き換えた年なのかの識別が必要である。この選択は利用者の慣習によることから、常に指定することなく設定できることが望まれる。

- ・時期に異なる解釈がある時代について、もれなくかつ冗長でなく検索するには、検索対象ではなく、検索条件に終点と始点の重なりを持たせるのがよい。

- ・幅を有する時間情報の並べ替えにおいて、異なる解釈のある時代や、～前期あるいは頃で就職された曖昧な時間情報では、検索用の時間情報による並べ替えでは不都合が生ずる。並べ替えのための時間情報が別に必要となる。

これらの課題は、特定の資源や時代を対象とする個々のデータベースを利用する上では、その分野の慣例として扱えたものが、幅広い分野の情報を統合的に検索することから派生している。システムを具体化するには、利用者の意見をより広く集め、よりよいユーザインタフェースを構成する必要がある。

## 謝辞

ここで検討した課題は、大学共同利用機関法人人間文化研究機構の研究資源共有化の検討の中で見出されたものである。様々な意見と討議を頂いた検討グループの各位に感謝する。また和暦と太陽暦から生ずる入力インタフェースの課題は、国文学研究資料館の相田満助教の指摘が発端となった。本稿での和暦と太陽暦の例示も同氏の対照表によった。ここに深く感謝する。さらに、有益な意見を頂いた国立歴史民俗博物館の各位に感謝する。

## 参考文献

- [1] 山本泰則, 原正一郎, 柴山 守, 安達文夫, 合庭 惇, 安永尚志, “Dublin Core メタデータと Z39.50 にもとづく人文科学系データベースの統合検索に関する実証実験,” 情報処学シンポジウム論文集, vol.2004, No.17. pp.199-205 (2004).
- [2] 原正一郎, 相田満, 入口敦志, 江戸英雄, 五島敏芳, 山田直子, “データベース共有におけるデータマッピングの事例的研究,” 情報処理学会研究報告, 2005-CH-67, pp.31-38 (2005).
- [3] 山本泰則, 中川 隆, “データベース横断検索のための民族学標本資料情報の Dublin Core による記述,” 情報処理学会研究報告, 2005-CH-67, pp.47-54 (2005).
- [4] 安達文夫, 鈴木卓治, “歴史研究データベースの Dublin Core へのマッピングとその課題,” 情報処理学会研究報告, 2006-CH-72, pp.47-54 (2006).