

# 歴史文献の時空間情報に関する問合せ処理

石川 正敏<sup>†</sup> 波多野賢治<sup>††</sup> 天笠 俊之<sup>††</sup> 植村 俊亮<sup>††</sup>

<sup>†</sup> 島根県立大学総合政策学部 〒 697-0016 島根県浜田市野原町 2433-2

<sup>††</sup> 奈良先端科学技術大学院大学情報科学研究科 〒 630-0192 奈良県生駒市高山町 8916-5

E-mail: <sup>†</sup>m-ishikawa@u-shimane.ac.jp, <sup>††</sup>{hatano,amagasa,uemura}@is.aist-nara.ac.jp

あらまし 歴史文献に関する時間情報や地理情報は、文献の分類や分析に利用される情報の一つである。しかし、歴史文献から得られる時空間情報は、(1) 期間や範囲があいまい、(2) 同じ地域でも年代によって地名が変わる、(3) 同じ地名でも異なる場所を指すという特徴がある。また、歴史文献の注釈として与えられた地理情報が地名だけのよう不完全であることも多い。本稿では、このような特徴を持つ時空間情報を考慮した時区間関連と領域関連を定義し、それらの関連の判定処理について提案する。また、地名などの分類に基づいた不完全な時空間情報の補完方法を提案する。本研究によって歴史文献の持つ多様な時空間情報に対する検索が可能になると考えられる。

キーワード 歴史文献, 時間情報, 地理情報, 時区間関連, 領域関連, MBR

## A Query Processing for Spatio-Temporal Information of Historical Documents

Masatoshi ISHIKAWA<sup>†</sup>, Kenji HATANO<sup>††</sup>, Toshiyuki AMAGASA<sup>††</sup>, and Shunsuke UEMURA<sup>††</sup>

<sup>†</sup> Department of Policy Studies, the University of Shimane, 2433-2 Nobaracho, Hamada, Shimane 697-0016, Japan

<sup>††</sup> Graduated School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayamacho, Ikoma, Nara, 630-0192, Japan

E-mail: <sup>†</sup>m-ishikawa@u-shimane.ac.jp, <sup>††</sup>{hatano,amagasa,uemura}@is.aist-nara.ac.jp

**Abstract** Temporal and geographical information for historical documents are important information to classify or analyze these documents. Spatio-temporal information of historical documents have some properties, (1) ambiguous scale, (2) variety name about same area, (3) inconsistency. And these information is incomplete such as a geographical information has only a name of place. In this paper, we propose temporal and regional relationships for spatio-temporal information of historical documents and explain a method to decide these relationships between these information. And also we propose a complement method for spatio-temporal information of historical documents, which is based on classify these information.

**Key words** historical document, temporal information, geographical information, temporal relationships, regional relationships, MBR

### 1. はじめに

近年、インターネット上で閲覧可能な木簡や古文書などの歴史文献が増加している。しかし、このような歴史文献の効率的な閲覧や読解を支援するための情報が十分に提供されているとは限らない。そこで、本研究では、歴史文献の注釈の編集と文献の共有を支援する電子スクラップブックシステムを提案している [1]。

歴史文献の注釈は、文字属性、単語の意味、関連文献への参照など様々な種類がある。特に歴史文献の時間および地理に関する注釈は、年表や地図などの情報と組み合わせた歴史文献の分類や分析に利用できると考えられる。そこで本稿では、歴史文献に関する時間的、地理的な情報を用いた文献の分類や分析

を支援するための問合せ処理について考える。

本稿で扱う歴史文献から得られる時間的および地理的な情報は、次のような特徴がある。

- (1) 時区間の期間の長さや領域の大きさがあいまい  
歴史的な時間情報は、人物の生没年が不明などのように期間があいまいであることが多い。同様に歴史的な地理情報の領域も、範囲の大きさを正確に示すことが困難であることが多い。
- (2) 同じ地域でも年代などによって名前が異なる。  
一般に地名は、同じ地域であっても年代によって名前が変わる。例えば、現在の島根県浜田市は、江戸時代以前は石見国と呼ばれていたなどが挙げられる。また、同年代の文献であっても同じ地名を異なる名称で記す場合がある。例えば、姫路城が白鷺

城のような別名を持つことなどが挙げられる。

(3) 同じ地名でも異なる場所を指すことがある。

歴史的な地理情報では、研究者の意見の相違によって関連付けられる領域が二つ以上存在することがある。例えば、邪馬台国の九州説と近畿説が挙げられる。

(4) 歴史文献に関する時間情報や地理情報は不完全であることが多い。

歴史文献から得られる時間情報や地理情報は、検索などで必要な情報がすべて与えられるとは限らない。例えば、歴史文献の地理情報では、地名が分かっても現在のどの場所を指すのかわからない場合が考えられる。また、神話などで扱われる時間には、相対的な関係だけの表現が可能であり具体的な時区間として表現できない場合がある。

本稿では、このような特徴をもつ歴史文献に関する時間情報および地理情報に対する問合せを処理するための時区間や領域の関連を定義し、領域関連の判定処理を提案する。また、値に欠落のある時空間情報に対する問合せ処理をした場合、選択されるべき情報が選択されないという現象が起きる。さらに、歴史文献に関する利用者の作成する問合せも、時区間、年号、時代名、領域、歴史的な地名、現代の地名など多様な表現を用いた記述が考えられる。そこで、このような多様な問合せの記述を処理するために、本稿では、時代や地名などに基づいた時間情報と地理情報の分類を用いた欠落した値の補完を考える。このような値の補完を行うことによって、時代に関係なく現代の地名などによる検索が可能になる。

## 2. 関連研究

文献 [2] では、1次元軸上の始点と終点の組として表される線分として定義される二つの時区間に対して 13 の関連を定義している。時区間に対する 13 関連は、2 時区間の位相的な関連と順序的な関連を合わせた関係である。本稿では、2次元領域の位相関連の判定に利用するため 2 時区間の関連を位相と順序に分けて考える。

9-intersection model は、領域を外部、内部、境界に分割し、二つの領域のそれぞれの部分の交差の関係から 2 領域の位相関連を定義している [3]。一方、本稿では、時区間関連を用いて領域の位相関連を判定することを考える。

文献 [4] は、あいまいな時空間情報を点の集合として表現したモデルを提案している。本稿で扱う時空間情報は、あいまいな領域をその領域の最大範囲と最小範囲を表す二つの多角形によって表現する。このような多角形を用いることで効率的な関連の検査と領域の入力ができると考えられる。

文献 [5] は、時区間と「までの」のような文書中の時間的表現の組みを管理することで文書中の時間的な文脈に沿った問合せ処理を提案している。本稿の時間情報は、「子の刻」などの歴史的な時間記述と時区間の対応を記述する。また、文献 [5] の時間情報では正確は時区間を用いるのに対して、本稿の時間情報は境界があいまいな時区間を扱う。

## 3. 歴史文献に関する時間および地理情報のデータ構造

本節では、歴史文献の記述などに現れる時間情報と地理情報を記述するためのデータ構造について述べる。本節で定義するデータ構造は、時区間や領域の取りえる長さや大きさの最小値と最大値の両方を記述することで、あいまいな時区間の長さや領域の大きさを記述する。

### 3.1 時間情報

時間情報は、年号と時区間の組として表現する。年号による期間の表現は地域によって様々な記述があるので、それらの違いに依存しない形式で期間を比較するために時区間を用いる。

時区間は、時間軸上の始点と終点の組として表現する。また、時区間が特定の時点を表現する場合は、時区間の始点と終点を同じ値を記述することで表現する。本稿で扱う時区間は、始点や終点があいまいなことがあるので、時区間の期間の取り得る範囲の最大期間と最小期間の併記を許す。このとき、二つの時区間は、最小の期間を表す時区間が最大の期間を表す時区間に内包されていなければならない。例えば、生年が不明で没年が 1420 年である人物の生没年を時区間で記述は、最長の時区間として 1340 年から 1420 年、最短の時区間として 1360 年から 1420 年などの記述が考えられる。次に、年号と時区間の関係について考える。例えば、ある人物の生没年の時区間が、慶応、明治、大正と三つの年号が関係するなどが考えられる。時区間と年号の関連は、1 対 n であると考えられる。そこで、本稿で記述する時間情報は、一つの時区間に対して複数の年号を関係付けることを許す。従って、時間情報は以下のような情報の組として記述される。

定義 1 時間情報

時間情報 t-info は、識別子 TID と、時区間 t-interval、時代 t-period の組として表現する。また、時区間は、時区間のとりうる期間の最大  $ts_{max}$ ,  $te_{max}$  と最小  $ts_{min}$ ,  $te_{min}$  の組で表す。時代は、時区間に関連する年号を、年号  $tn$  と年数  $y$  の組の集合として記述する。

時間情報 t-info = (TID, t-interval, t-period)

時区間 t-interval = ( $ts_{max}$ ,  $ts_{min}$ ,  $te_{max}$ ,  $te_{min}$ )

時代 t-period = ( $tn$ ,  $y$ )

記述例を以下に示す。

例 1

木簡に記述された官位に関する時間情報は、以下のように記述する。

時間情報 = (1, 時区: (649, 649, 664, 664), 時代: {(大化, 5), (大化, 6), ..., (天智, 3)})

この例では、時区間があいまいでないため、期間の最大値と最

小値を同じ値で記述される。

### 3.2 地理情報

本節では、歴史文献の地名と領域の対応を管理するためのデータ構造を定義する。歴史的な地域の領域は、あいまいであることが多いため、領域として考えられる最小範囲と最大範囲の両方を記述する。このとき、最小範囲の領域は、最大範囲の領域に内包されていなければならない。また、地名は、時代とともに変化することが多い、そこで文献に記述された地名以外に関連する地名も列挙する。領域の中心は、二つの領域の距離や方向を表すために用いられる。しかし、領域の範囲があいまいなので、領域の中心を自動的に求めることは困難である。そこで、本稿では、与えられた領域の最小値から得られる重心を、領域の中心とする。従って、地理情報は以下の構造で記述される。

#### 定義 2 地理情報

地理情報 geo-info は、識別子 GID, 領域 r, 中心 c, 地名 rn の組で表現される。領域は、範囲の最大値 ra\_max と最小値 ra\_min の組として表現される。範囲は、頂点のリストで表現される多角形である。地名は、地域名 n と、関連する地名 rmn の集合の組として表現される。

地理情報 geo-info = (GID, r, c, rn)

領域 r = (ra\_max, ra\_min)

地名 rn = (n, rmn)

記述例を以下に示す。

#### 例 2

石見国に関する地理情報の記述例は、以下の通りである。

地理情報 = (1, 領域: {(130, 140), ..., (200, 210)}, {(120, 130), ..., (190, 200)}, 中心: (155, 165), 地名: {地域名:石見国, 関連地名:(浜田, 島根)})

### 3.3 時空間情報

歴史文献に関する時空間情報は、参照する歴史文献の URI, 関連する時間情報と地理情報の識別子の組として記述する。

#### 定義 3 時空間情報

時空間情報 st-info は、関連する文献の参照 ref, 時間情報への参照 TID, 地理情報への参照 GID の組として表現される。

時空間情報 st-info = (ref, TID, GID)

## 4. 時区間関連, 領域関連

本節では、時間情報や地理情報の検索や分類に利用する時区間や領域の関連を定義する。

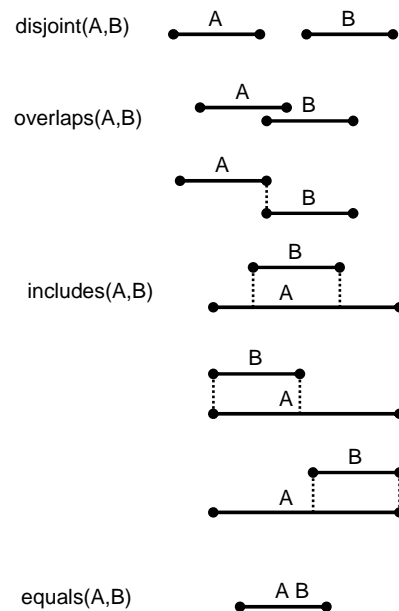


図 1 二つの時間区間の位相関連

Fig. 1 Topological Relationships between Two Temporal Intervals



図 2 二つの時間区間の順序関連

Fig. 2 Order Relationships between Two Temporal Intervals

### 4.1 時区間関連

二つの時区間の関連には、時区間の順序に依存しない関連を表す位相関連と時区間の順序を表現する順序関連がある。

#### 定義 4 時区間に関する位相関連

二つの時区間を表す位相関連として本稿では以下の関連を定義する(図 1)。

##### (1) disjoint 関連

二つの時区間の境界が接していないかつ一方が他方の時区間に内包していない関連を表す。

##### (2) overlaps 関連

一方の時区間の始点もしくは終点のどちらかが他方の時区間に接しているか内包されていること表す関連である。

##### (3) includes 関連

一方の時区間の両端が他方の時区間に内包されている関連である。

##### (4) equals 関連

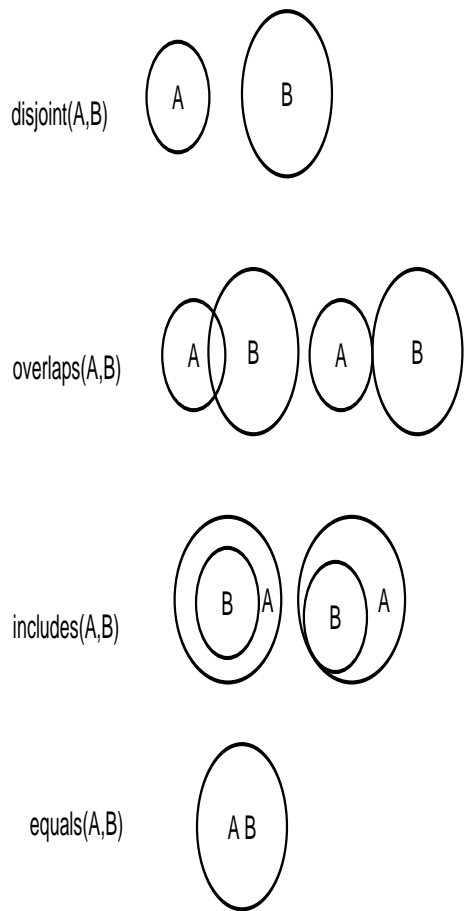


図3 二つの領域間に関する位相関連

Fig. 3 Topological Relationships between Two Regions

二つの時区間が同一であることを表す関連である。

本稿で扱う時区間は境界があいまいであるため境界の接触と内包の関連を区別しない。equals 関連を満たす時区間は、同じ時間情報を比較している場合にだけに限る。

#### 定義 5 時区間に関する順序関連

二つの時区間の順序関連として以下の関連を定義する (図 2)。

##### (1) before 関連

一方の時区間の少なくとも一つの一端が、他方の時区間の前にある関連を表す。

##### (2) after 関連

before 関連の逆の順序である時区間の関連である。

##### (3) disorder 関連

一方の時区間が他方を内包した場合の関連である。

#### 4.2 領域関連

本節では、二つの領域のもつ関連として、位相、方向、距離の 3 関連を定義する。位相関連は、領域の方向や距離に依存しない関連である。方向関連は、二つの領域の方角に関する関連である。方向関連は、二つの領域の距離に関する関連である。

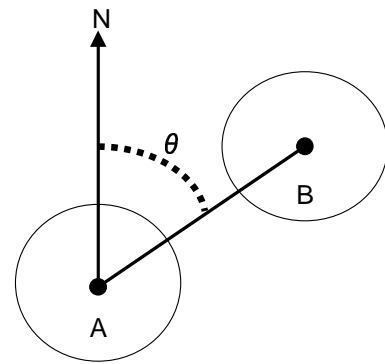


図4 二つの領域間に関する方向関連と距離関連

Fig. 4 Directional and Distance Relationships between Two Regions

#### 定義 6 領域に関する位相関連

二つの領域の位相関連には次の 4 関連がある (図 3)。

##### (1) disjoint 関連

一方の領域が他方の領域を内包せず、それぞれの領域の境界が接していない状態を表す。

##### (2) overlaps 関連

一方の領域の一部が他方の領域と接しているか重なっている状態を表す関連である。

##### (3) includes 関連

一方の領域が他方の領域を内包している場合の関連である。

##### (4) equals 関連

二つの領域が同一であることを表す関連である。

本稿では、領域の境界があいまいであるのでそれぞれの領域の境界の接地と重なりを区別しない。領域の equals 関連が成り立つのは同じ地理情報を比較している場合だけである。

#### 定義 7 領域に関する方向関連

二つの領域の方向関連は、関連を比較する領域の中心から北の向き (xy 平面の場合、y 軸正の向き) に伸ばした直線とそれぞれの領域の中心を結んで表現される線分の角度の大きさとする (図 4)。

本稿では 2 領域間の角度の大きさを表す関数として、以下に示す direction 関数を用いる。

$$\text{direction}(A.c, B.c) = \theta \quad (0 \leq \theta < 360)$$

direction 関数は、引数として比較する二つの領域をとり、前者の領域から後者の領域を見た場合の方向関連を角度の大きさ  $\theta$  で表す。

#### 定義 8 距離関連

二つの領域の距離に関する関連は、それらの領域の中心を結ぶ線分の長さである (図 4)。

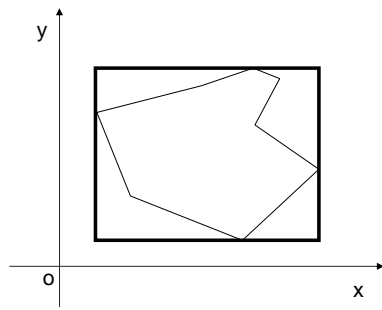


図5 最小領域矩形を用いた領域の簡略化  
Fig. 5 MBR for a Region

本稿では二つの領域間の距離を求める関数として、以下のよ  
うな distance 関数を用いる。distance 関数は引数に二つの領域  
をとる。この関数の返り値は、与えられた二つの領域の中心を  
結んだ線分の長さである。

$$\text{distance}(A, B) = \text{length} \quad (0 \leq \text{length})$$

#### 4.3 時区間関連を用いた領域の位相関連と方向関連の判定

本節では、領域に対する問合せや分類を効率的に処理するた  
めに領域の位相関連と方向関連の調査に時区間関連を利用する  
方法について述べる。一般に多角形で表現される領域の関連を  
判定する場合、それぞれの領域の頂点に対して  $n^2$  回の比較が  
必要である。問合せなどである領域と多数の領域を比較する場  
合、計算量が指数関数的に増加するため、効率的な関連の判定  
を行う必要がある。そこで、本稿では、詳細な関連の判定の前  
に、領域の近似形状を用いた大雑把な関連判定を行い、詳細な  
関連の判定を行う対象の領域の数を減らす。

##### 4.3.1 領域の近似

本稿では、地理情報として与えられた領域の近似として座標  
軸に並行でかつ領域に外接する最小領域矩形 (MBR, Minimum  
Bounding Rectangle) を用いる (図5)。このような MBR を用い  
ることによって、比較する頂点の数が減るので領域の関連の判  
定が容易になると考えられる。

##### 4.3.2 領域に関する位相関連の判定

前節で定義した MRB を直接用いた領域の関連の判定は、各  
頂点間で最大 16 回の比較が必要である。一方、時区間関連を  
1 次元の数直線上の 2 区間の関連の表現に利用し、二つの領域  
の MBR を x 軸、y 軸にそれぞれ射影して得られる区間の位相  
関連の組として判定すれば、8 回の頂点比較で領域の関連が分  
かるため効率的であると考えられる。

以下に、各軸の 2 区間の位相関連と領域の位相関連の関係を  
示す。

##### (a) disjoint 関連の判定条件

x 軸、y 軸のどちらか一方の線分の関連が disjoint 関連の場合、  
二つの MBR は disjoint 関連である。

##### (b) overlaps 関連の判定条件

(a) 以外の状態で x 軸、y 軸のどちらか一方の線分の関連が

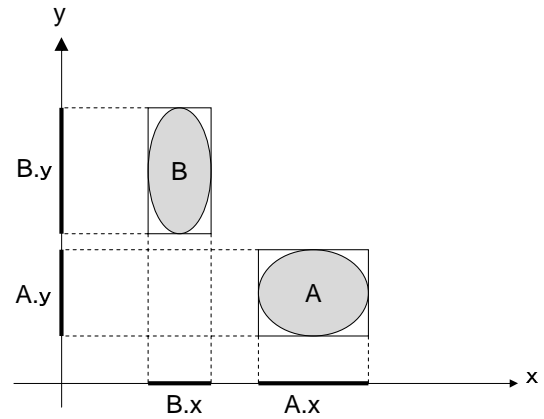


図6 2領域の位相関連の判定例

Fig. 6 An Examination of Topological Relationships between Two Regions.

overlaps 関連である場合、二つの MBR は overlaps 関連である。

##### (c) includes 関連の判定条件

x 軸、y 軸の両方の線分の関連が共に includes 関連の場合、二  
つの MBR は includes 関連である。

##### (d) equals 関連の判定条件

x 軸、y 軸の両方の線分の関連が equals 関連の場合、二つの  
MBR は equals 関連である。

MBR を用いて得られる overlaps 関連や includes 関連は、MBR  
が元の領域を近似した形状であるため実際の領域の形状を用い  
て得られる関連と異なる場合がある。従って、MBR の overlap  
関連や includes 関連を得た場合、二つの元の領域の形状を用い  
た詳細な検査を行う。また、equals 関連は、先の関連の定義か  
ら同一の地理情報を比較している場合にだけ成立する関係である  
ため、本稿では includes 関連として扱う。

#### 例 3

図6に示す領域 A と領域 B の MBR を例に時区間の位相関連の  
組を用いた領域の位相関連の判定例を示す。まず、それぞれの  
MBR を x 軸、y 軸に射影し、それぞれの軸に関する区間  $A.x$ 、  
 $A.y$ 、 $B.x$ 、 $B.y$  を得る。次に、軸ごとに時区間関連を判定する。  
その結果、軸ごとの関連が、すべて disjoint 関連であることが  
分かる。このような時区間関連の組と先に述べた条件から例の  
領域 A、B の領域関連は、disjoint 関連であることがわかる。

##### 4.3.3 時区間関連を用いた方向関連の判定

本節では、多数の領域から目的の方角にある領域を選択する  
処理について述べる。方向関連は、定義から比較元の領域の中  
心と比較対象の領域の中心を結ぶ線分と比較元の領域の中心か  
ら y 軸正の方向に延ばした線分の角度の大きさと表現する。一  
般に多くの領域との方向関連の判定する場合、条件を満たさな  
い領域の方が多いと考えられるため、すべての領域に対して角  
度を計算していたのでは、効率的な領域の選択ができない。そ  
こで、本稿では、多数の領域の方向関連を効率的に判定するた  
めに時区間の順序関連を用いて検査対象の領域を絞り込むこと  
を考える。この判定を行うために、まず各領域の中心を x 軸と

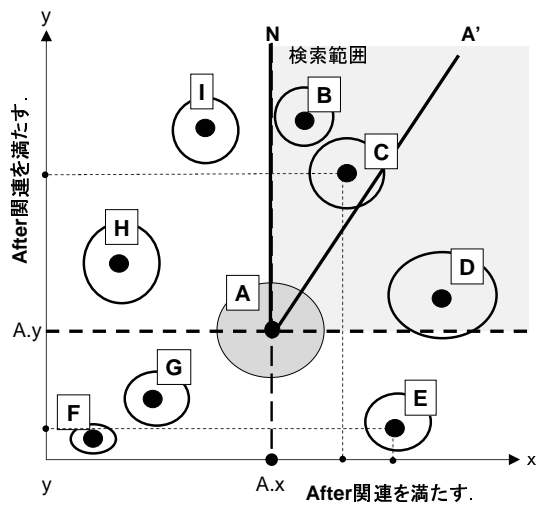


図7 順序関連を用いた方向関連の判定  
Fig.7 Directional Relationships Test Using Order Relationships

y 軸に射影して得られる点を、長さ 0 の線分として扱う。次に、軸ごとに順序関連を判定し、各軸の順序関連の組み合わせから、領域の方向関連を検査する対象を絞り込む。例として、領域 A から見て、0 度から 30 度の方向にある領域を取り出す場合において、方向関連の検査対象の絞りこみを考える (図 7)。条件を満たす領域は、図 7 の線分 AN と線分 AA' に挟まれる範囲として表現される。従って、領域 A を原点とした座標系の第 1 象限にない領域は明らかに方向関連を検査する必要はないことが分かる。つまり、領域 A の中心と他の領域中心が x 軸、y 軸に関してともに after 関連である領域だけを検査対象にすればよいことが分かる。図 7 では、A.x 以上かつ A.y 以上の範囲の領域が例の条件を満たす可能性がある領域を表すので、その領域に含まれる領域 B, C, D が選択される。さらに選択された領域に対して、詳細な方向関連を検査することで、条件を満たす領域 B, C が選択されることがわかる。図 7 から順序関連を用いた方向関連の検査対象の絞り込みは、各領域の中心の座標値の比較で処理できるので、個々に角度を計算するより効率的に方向関連の調査ができる。

## 5. 分類に基づいたデータの補完

歴史文献に関する時間情報や地理情報は、地名と対応する領域が不明などの理由から領域や時区間などの値が欠落した不完全な情報であることが多いと考えられる。そこで、効率的に問合せを処理するために本稿では、関連のある時間情報や地理情報を分類し欠落した情報を補完することを考える。時間情報や地理情報を分類する方法には、時区間や領域に基づいて分類する方法と地名や時代名に基づいて分類する方法がある。また、分類には、年表などの外部データに基づいて分類する方法と、既存のデータだけで分類する方法が考えられる。時間情報や地理情報の分類に必要な外部データは、必ずしも得ることができるとは限らないので、ここでは、まず与えられた時間情報と地理情報だけを用いて値の補完を行う。さらに、本稿では、時間と共に地名の変化する地理情報への問合せを処理するために、

関連する地名の対応付けを行う。地名の対応付けをすることによって、“平安時代の島根県に関する記述のある文献を表示せよ” などのような時代に依存しない地名での問合せが処理できると考えられる。

### 5.1 時間情報の補完

時間情報の補完は以下のような手順で行う。

#### 例 4 時間情報

例として、以下に示す時間情報を用いて補完処理を考える。時間情報 5 は時区間だけが与えられており、時間情報 6 は時代名だけが与えられている。例に示す時間情報は、識別子、時区間の最大期間と最小期間の併記を省略している。

時間情報 1 (時区間: 645 年, 時代: (大化, 1))

時間情報 2 (時区間: 649 年, 時代: (大化, 5))

時間情報 3 (時区間: 662 年, 時代: (天智, 1))

時間情報 4 (時区間: 664 年, 時代: (天智, 3))

時間情報 5 (時区間: 646 年から 672 年, 時代: )

時間情報 6 (時区間: , 時代: (斉明, 2), ..., (天智, 1), (天智, 2))

#### (1) 基準時間情報の作成

この処理では、まず時代要素が一つだけの時間情報だけを抜き出し分類する。次に、分類した時間情報ごとに時区間を取り出し、取り出した時区間をすべて内包する新たな時区間を作成する。作成した時区間と時代の値から新たな時間情報を作成する。この時間情報を基準時間情報と呼ぶ。例 4 の場合、年号が同じである時間情報 1 と 2、時間情報 3 と 4 が取り出され、それぞれの分類から次のような基準時間情報が得られる。

#### 例 4-1 基準時間情報

時間情報 7 (時区間: 645 年から 649 年, 時代: (大化, 1), ..., (大化, 5))

時間情報 8 (時区間: 662 年から 664 年, 時代: (天智, 1), ..., (天智, 3))

#### (2) 時区間だけが与えられている時間情報の補完

この処理では、基準時間情報の時区間と補完対象の時間情報の時区間を比較し、includes 関連もしくは overlaps 関連である時代要素を補完対象の時間情報に与える。時間情報 5 を対象に例 4-1 で作成した時間情報 7, 8 と比較し、条件を満たす時間情報の時代名を時間情報 5 に複写する。

#### 例 4-2 時代の補完

時間情報 5' (時区間: 646 年から 672 年, 時代: (大化, 2), ..., (大化, 5), (天智, 1), ..., (天智, 3))

時間情報 5 と時間情報 7 の時区間は、始点に 1 年の差がある。従って、時間情報 7 から時間情報 5 に写す時代の値は、(大化, 1)ではなく (大化, 2) とする。

### (3) 時代の値だけが与えられている時間情報の補完

この処理では、まず複数の時代要素を持ち時区間が与えられていない時間情報と基準時間情報の時代の値を比較し、時代要素の値が一致する基準時間情報の時区間を集める。次に、集めた時区間の集合から期間が最長になる時区間を作成し、補完先の時間情報に与える。ただし、収集した時区間に中に他のすべての時区間と disjoint 関連となる時区間がある場合、連続した時区間が作成できないため、時区間の値の補完は行わない。例 4 の場合、時間情報 6 と基準時間情報を比較し、年号が一致する時間情報から時区間の値を写す。

#### 例 4-3 時区間の補完

時間情報 6' (時区間: 662 年から 663 年, 時代:(斉明, 2), ..., (天智, 1), (天智, 2))

例 4 は、時間情報 6 にある年号“斉明”と一致する時間情報がないため時間情報 7 だけから関連のある時区間を写す。

## 5.2 地理情報の補完

本節では、次の例を用いて地理情報の補完について述べる。

#### 例 5 地理情報

以下の地理情報を用いて地理情報の補完処理について考える。領域 A, B, C の関連は、A と B, A と C がそれぞれ includes 関係にあり、B と C は disjoint 関連であると仮定する。また、例の記述を簡単にするために識別子と中心の記述は省略する。

地理情報 1 (領域: A, 地名: (地域名: 島根, 関連地名: ))  
地理情報 2 (領域: B, 地名: (地域名: 出雲, 関連地名: ))  
地理情報 3 (領域: C, 地名: (地域名: , 関連地名: ))  
地理情報 4 (領域: , 地名: (地域名: 石見, 関連地名: 島根))

#### (1) 基準地理情報の生成

この処理では、まず与えられた地理情報から領域と地域名の両方の値があるものを抜き出し地域名で分類する。次に、分類された地理情報の領域を互いに比較し、disjoint 関連にある領域があれば、それらを異なる地理情報としてさらに分類する。最後に、分類された地理情報ごとに領域の論理和をとり、新たに作成した領域と地域名を関連付けた地理情報を作成する。このように作成した地理情報を基準地理情報と呼ぶ。例 5 では、領域情報と地域名の両方が登録されている地理情報は、地理情報 1, 地理情報 2 だけである。また、それらの地理情報の地域名は一致しないので、この例では、それらを基準地理情報として利用する。

#### 例 5-1 基準地理情報の生成

基準地理情報 1 (領域: A, 地名: (地域名: 島根, 関連地名: ))  
基準地理情報 2 (領域: B, 地名: (地域名: 出雲, 関連地名: ))

#### (2) 領域に基づく地名の関連付け

この処理では、基準地理情報と補完対象の地理情報の領域を比

較し、overlaps 関連もしくは、includes 関連を満たすものがあれば、それらは関連のある地理情報であるので対象の地理情報の関連地名に値を与える。例 5 の場合、地理情報 1, 地理情報 2 の領域は includes 関連であるため、次に示すようにそれぞれの地域名を関連地名として登録する。

#### 例 5-2 領域に基づく地名の関連付け

基準地理情報 1' (領域: A, 地名: (地域名: 島根, 関連地名: 出雲))

基準地理情報 2' (領域: B, 地名: (地域名: 出雲, 関連地名: 島根))

#### (3) 領域だけの地理情報に対する地名の補完

この処理では、補完対象の地理情報と基準地理情報の領域を比較し includes 関連にある基準地理情報の地域名を関連地名として対象となる地理情報に与える。例 5 の場合、地理情報 3, 地理情報 1, 2 とそれぞれの関連を判定し、例 5 の仮定から地理情報 1 の領域が地理情報 3 と includes 関連にあるため、次に示す通り地理情報 1 の地域名を地理情報 3 の関連地名に登録する。

#### 例 5-3 領域だけが与えられている地理情報に対する地名の補完

基準地理情報 3 (領域: C, 地名: (地域名: , 関連地名: 島根))

#### (4) 地域名だけの地理情報に対する領域の補完

この処理では、領域が登録されていない地理情報の地域名と基準地理情報の地域名を比較し、一致するものがあれば基準地理情報の領域を対象の地理情報に複写する。もし地域名に一致するものがなければ、補完対象の関連地名を用いて基準地理情報の地域名と比較し、一致する地名を探す。例 5 の場合、地理情報 4 の地域名もしくは関連地名、地理情報 1, 2 に地域名の比較を行い、地理情報 1 の地域名と一致するので、次に示す通り地理情報 1 の領域を地理情報 4 の領域に写す。

#### 例 5-4 地域名だけの地理情報に対する領域の補完

地理情報 4 (領域: A, 地名: (地域名: 石見, 関連地名: 島根))

## 6. 問合せ処理について

歴史文献に関する時間情報および地理情報に対する問合せの記述と処理について述べる。本稿で定義した時間情報および地理情報を対象とした問合せは、時間情報に関する時区間や時代名に関する条件と、地理情報に関する領域や地名を対象にした条件を論理演算子で連結した形式で表現される。本稿で行う問合せの結果は、取り出された時空間情報を含む歴史文献の URI の集合である。そこで、問合せ処理は、(a) 与えられた条件ごとに検査を行いその条件を満たす時間情報もしくは地理情報の識別子を得る。(b) それぞれの識別子をもつ時空間情報にある歴史文献の URI を取り出し一時的な結果として保存する。(c) すべての条件に対して (a), (b) の処理を行い取り出された文献の URI の集合に対して、検索条件の論理演算子に従った集合演算から問合せ結果を得る。

このような問合せで得られる文献データの集合は、ある期間

における文献データの地理的な分布として表現することができる。文献データの分布の表現には、地理的な分布だけでなく時間的な分布を表現することも重要であると考えられる。そこで、このような問合せにおいて、利用者は、先に述べた問合せの条件と合わせて文献の時間的な分布を示すための時間間隔を指定することを許す。時間的な文献の分布を表す間隔を指定した場合の問合せ処理は、(i) 問合せ条件を満たす文献データ集合を取り出し、(ii) 指定された間隔に従って結果集合を分割するという処理を行う。問合せの結果は、指定された時間間隔に従って複数のグループに分割された歴史文献の URI の集合となる。

## 7. 考 察

地理情報や時間情報の補完を行った場合の効果について考察する。例えば地理情報の地名について補完処理をした場合、地名による地理情報の検索で地名が与えられていなかった地理情報も検索結果として選択されるようになるので、検索として選ばれる地理情報の割合が増加すると考えられる。しかし、一般に地名は同じ名称を持ち異なる領域を指す地理情報が多数存在するので、検索結果の中に含まれる利用者が必要とする情報が含まれる割合が減る可能性がある。特に地理情報の数が多ければ多いほど前者の割合が増えるが、後者の割合が少なくなる。このような現象は、時間情報の年号など地理情報の領域に関する補完を行った場合も同様であると考えられる。必要のない検索結果の増加を減らす方法としては、地理情報の検索に地名だけではなく領域などに関する条件追加する方法が考えられる。また、地名の分類や検索などの基準に地図などの外部データを利用する方法が挙げられる。ただし、例えば邪馬台国の位置が九州にある地図と近畿にある地図の2種類が考えられるような研究者の意見の相違によって、地理情報などの分類の基準となる地図が複数存在することがある。従って、外部データに選択によって、ある利用者には効率的な検索が可能であっても、他の利用者にとっては有益な検索ができない場合があると考えられる。

次に問い合わせ処理について考察する。本稿で述べた問合せ処理では、時間情報に関する問合せ条件と地理情報に関する問合せ条件を独立させて処理している。そのため個別の条件で一時的な結果として選択された歴史文献であっても、集合演算の結果、問合せ結果として選択されない場合があると考えられる。このような不要な一時的な問合せ結果の数を減らすには、時間情報か地理情報に関する問合せ条件のどちらかを優先的に処理し対象を絞り込んだ後に他方の条件の処理をする方法が有効であると考えられる。どちらの条件を先に処理するかは、(1) 利用者による選択や、(2) 各条件の結果によって得られる一時的な問合せ結果の集合の大きさの見積もりによって決定する方法が考えられる。

本稿で提案する問合せを処理するには、すでに数多く存在する歴史文献に対して時空間情報を追加する必要がある。このような情報の一つの方法として本稿では、文献[1]で提案している電子スクラップブックシステムのような利用者が歴史文献に対して自由に情報を追加できる環境を用いて、地理情報や時間

情報を収集する方法が挙げられる。また、木簡データベース[6]などのように木簡画像と合わせて公開されている木簡のメタデータから時間情報や地理情報を得ることができる。既存の電子図書館などでの情報の記述は統一された形式がないため、情報源ごとに時間情報や地理情報を取り出すには、個別に情報の抽出処理を定義しなければならない。また利用者による歴史文献に地理情報などを追加する方法は、一度に大量に情報を獲得することは困難であるが、歴史学的な発見などによる情報の追加が柔軟にできると考えられる。

## 8. ま と め

本稿では、歴史文献の記述に関する時間情報や地理情報への問合せを処理するための関連の判定について述べた。歴史的な時間情報や地理情報は、領域の境界があいまいであったり、値の欠落があったりするなどの特徴がある。そこで本稿では、歴史文献の時空間情報を考慮したデータ構造と時区間関連と領域関連を定義し、関連の判定方法を提案した。また、問合せを効率的に処理するために既存データからの地名などの値を補完する方法を提案した。さらに、本稿では時間情報と地理情報を用いた問い合わせ処理について述べた。最後に地理情報と時間情報の値の補完による問合せ結果の効果と問合せ処理の効率、歴史文献に関する時間情報や地理情報の作成方法について考察した。本稿の提案する時間情報および地理情報の値の補完を用いることで、歴史文献に対する多様な時間および地理表現を用いた問合せが処理できると考えられる。

今後の課題は、実際の歴史文献を用いて提案手法の実行を示し評価することである。

### 文 献

- [1] 石川正敏, 波多野賢治, 天笠俊之, 植村俊亮, 勝村哲也: “歴史的文献画像のための電子スクラップブックシステム”, 情報処理学会論文誌: データベース, Vol.44, SIG12(TOD19), pp. 110 - 122, 2003年9月.
- [2] J. F. Allen: “Maintaining Knowledge about Temporal Intervals”, Comm. of the ACM, pp832-843, Nov. 1983.
- [3] M. J. Egenhofer, “Point-Set topological relations”, International Journal of Geographical Information Systems, Vol. 5, No.2, pp. 161-174, Taylor&Francis, 1991.
- [4] Erlend Tøssebro, Mads Nygard: “Uncertainty in Spatiotemporal Databases”, ADVIS 2002, LNCS2457, pp. 43-53, 2002.
- [5] 細川宜秀, 清木康: “文脈認識をともなった時空間的関連性評価方式”, 情報処理学会論文誌: データベース, Vol.43, SIG05, pp. 118 - 133, 2002.
- [6] 奈良文化財研究所: “木簡データベース”, <http://acd.nabunken.jp/Open/mokkan/mokkan1.html>.