博士論文

歴史文献のための電子スクラップブックシステムに関する研究

石川 正敏

2004年2月6日

奈良先端科学技術大学院大学
情報科学研究科

本論文は奈良先端科学技術大学院大学情報科学研究科に
博士 (工学) 授与の要件として提出した博士論文である.

石川 正敏

審査委員： 植村 俊亮 教授
伊藤 実 教授
宝珍 輝尚 教授 (大阪府立大学)
吉川 正俊 教授 (名古屋大学)

歴史文献のための電子スクラップブックシステムに関する研究*

石川 正敏

内容梗概

本論文では，木簡や古文書などの歴史文献と注釈を共有するためのデータのモデルを提案し，モデルに従って注釈の編集などを処理する電子スクラップブックシステムの構築について述べ，提案システムを用いて作成した文献データを WWW 環境で閲覧する方法について論じる．

電子スクラップブックシステムを構築するために，まず，歴史文献の特徴について考察し，計算機環境での歴史文献を利用する研究活動を支援するための要求をまとめる．この要求に従って，歴史文献の画像と注釈の関係を管理する文献データのモデルと，歴史文献の分類を管理する電子スクラップブックデータのモデルを定義する．さらに，歴史文献に関する注釈を効率的に再利用するために，その内容に基づいて注釈のデータの構造を定義する．

次に，電子スクラップブックシステムの設計について述べる．データは，利用者間での文献や注釈の編集や交換のために XML で記述し，効率的に共有するために関係データベースを用いて管理する．文献データに対する操作には，注釈の編集，切り抜き操作および検索がある．文献データの検索では，文字列を用いた検索と文献の記述にある時空間情報を用いた検索について述べる．

さらに，これらの設計に基づいたプロトタイプシステムの実装と実行例を示す．プロトタイプシステムは，歴史文献への注釈の追加や，文献の収集，文献データや注釈の共有を支援する．

* 奈良先端科学技術大学院大学 情報科学研究科 博士論文, 2004 年 2 月 6 日.

最後に，歴史文献の注釈の応用として，WWW 環境での文献データの閲覧方法を提案する．閲覧の支援方法としては，歴史文献のレイアウトに従って記述内容のテキストデータを HTML 文書として表示する方法と，歴史文献と注釈の関係を WWW 環境で表現する方法について述べる．

提案システムによって，人文社会科学や歴史学研究における効率的な情報の収集と共有が可能になり，考古学や歴史学における研究支援だけではなく教育や電子図書館の個人利用などの他の分野への応用も広がると考えられる．

キーワード

東アジア圏の歴史文献，文献データ，電子スクラップブックデータ，XML，時空間情報

A Study on an Electronic Scrapbook System for Historical Documents *

Masatoshi Ishikawa

Abstract

This thesis proposes a model for an electronic scrapbook system that makes it possible to share and edit annotations of historical documents, as well as its application for web publishing.

First, this thesis discusses the properties of historical documents and conditions needed to computer aided research activities using historical documents. And then the thesis designs a document model for management of relationships between historical documents and annotations, and an electronic scrapbook model for management of classification of historical documents. Furthermore, the thesis analyses the structure of annotations of historical documents for efficient reuse of annotations.

Second, this thesis explains implementation of a prototype system following these models. The prototype system has operations of editing annotations of historical documents, collecting document data, sharing document data between users. It uses XML and a relational database for management of data following the proposed model. It also has functions such as editing of annotations, as well as extracting and retrieval of data. And also spatio-temporal search of historical documents are mentioned, after discussion about the property of historical documents.

* Doctoral Thesis, Graduate School of Information Science, Nara Institute of Science and Technology, February 6, 2004.

Third, an application of the proposed model is shown. It is web publishing of historical documents data based on the proposed model. The web publishing methods include the creation of HTML documents displaying the original layout of historical documents, and the showing the relationship between images and annotations of historical documents.

The proposed model for an electronic scrapbook system would be expected to enable humanity researchers to efficiently collect and share historical documents and annotations related these documents.

Keywords:

East Asian Historical Documents, Document Model, Electronic Scrapbook Model, XML, Spatio-Temporal Information

目 次

第 1 章　はじめに **1**

1.1　背景と目的 1

1.2　構成 4

第 2 章　関連研究 **7**

2.1　関連研究 7

2.2　XML 10

第 3 章　歴史文献に関する考察 **15**

3.1　歴史文献の特徴 15

3.2　外字情報の扱い 18

3.3　歴史文献の収集 18

第 4 章　文献データ，電子スクラップブックデータのモデル **19**

4.1　文献データのモデル 21

4.2　注釈文のデータ構造 26

4.3　電子スクラップブックデータのモデル 34

4.4　考察 36

第 5 章　設計 **39**

5.1　文献データと電子スクラップブックデータの XML による記述 . 39

5.2　文献データと電子スクラップブックデータに対応した関係表の設計 43

5.3　操作 46

5.4 テキストデータを用いた文献データ，電子スクラップブックデータの検索 48
5.4.1 検索処理 49
5.4.2 問合せ処理 50
5.4.3 異体字検索について 51
5.5 文献データの記述に関する時空間情報を用いた検索 52
5.5.1 歴史文献の時間的，地理的な情報の特徴 52
5.5.2 歴史文献に関する時間および地理情報のデータ構造 . . . 53
5.5.3 時区間関連，領域関連 56
5.5.4 1 次元の時区間関連を用いた 2 次元領域の位相関連と方向関連の判定 62
5.5.5 歴史文献の時空間情報の補完操作 66
5.6 考察 69

第 6 章 電子スクラップブックシステム **71**
6.1 構成 71
6.2 実装環境 73
6.3 実行例 74
6.4 検索の実行例 76
6.5 考察 78

第 7 章 WWW 環境での文献データの閲覧 **79**
7.1 文献画像のレイアウトに基づいた釈文の表示 79
7.2 注釈と関連付けた歴史文献画像の WWW 環境での表示 83
7.3 考察 86

第 8 章 結論 **89**

謝辞 **91**

参考文献 **96**

付録 **101**

A Relax スキーマによる文献データのモデルの記述 101

B Relax スキーマによる電子スクラップブックデータのモデルの記述 105

図 目 次

2.1 XML 文書の記述例 11
2.2 XML 文書の構造を木構造で記述した例 12
2.3 名前空間を用いた XML 文書の記述例 13

3.1 特殊なレイアウトの歴史的文献の例 16
3.2 木簡画像の例 17

4.1 文献データと電子スクラップブックデータの関係 20

5.1 時区間 54
5.2 領域 55
5.3 二つの時間区間の位相関連 57
5.4 二つの時間区間の順序関連 58
5.5 二つの領域間に関する位相関連 60
5.6 二つの領域間に関する方向関連と距離関連 61
5.7 最小領域矩形を用いた領域の簡略化 63
5.8 順序関連を用いた方向関連の検査対象の絞り込み 65

6.1 プロトタイプシステムの構成 72
6.2 文献データの閲覧例 74
6.3 電子スクラップブックデータの閲覧例 75
6.4 電子スクラップブックデータ内の文献データの詳細表示例 . . . 76
6.5 検索インターフェースの実行例 77
6.6 検索結果の表示例 77

7.1 文献画像と注釈から生成した HTML 文書 83

7.2　注釈を関連づけた歴史文献の WWW ブラウザの表示例 87

第1章　はじめに

1.1　背景と目的

考古学，歴史学，人文科学分野の研究資料である古文書や木簡などの歴史文献は，長期に渡り図書館や博物館，公文書館，大学などの様々な研究機関で収集され，蓄積されてきた．歴史文献は，保存が困難である上に所蔵機関の展示場所の限界があることから閲覧が困難であるので，研究機関の中には，歴史文献の利用効率の向上を目指して，歴史文献のテキスト化を行っている [1] [2]．また，スキャナなどの入力機器の性能向上に伴い，研究機関の多くは，原資料を画像として保存する方法としてマイクロフィルムに代わって，計算機を利用するようになってきた [3] [4]．

一方，インターネットの帯域幅と個人所有の計算機の性能の向上によって，利用者間で高精細な画像や動画などのマルチメディアデータの交換が容易になった．そのため図書館などでは，電子化した歴史文献の画像やテキストデータをインターネットで公開するようになってきている [5] [6] [7] [8] [9]．

一般に，公開される歴史文献は公開する組織ごとに互換性のないデータ構造で管理されるので，組織を跨いで歴史文献を効率的に閲覧することは困難であると考えられる．そこで，統一したメタデータを索引に用いることで，組織を越えた歴史文献や研究成果の相互利用が進められている [10] [11]．研究機関などで整備するメタデータの多くは，考古学や歴史学などの研究での利用を想定したものが多いため，一般的な利用者に対しては必ずしも十分な情報を提供しているとは限らない．さらに，歴史文献は，研究者の意見の相違などによってメタデータの内容が統一されているとは限らない．従って，歴史文献の利用効率を向上させるためにメタデータの追加や修正が重要であると考えられるが，

一般に電子図書館やデジタルアーカイブでは，最初に整備したメタデータ以外に，新たなデータを追加することが困難である場合が多い．例えば，歴史文献の内容理解に役立つ注釈は，時間とともに増加するため，すべての情報を公開する組織だけで追加することは，時間的および費用的な制約から困難であると考えられる．

そこで本論文では，インターネットで公開される歴史文献の効率的な利用を支援するために，歴史文献への注釈を利用者が独自に編集し共有するための電子スクラップブックシステムを提案する [12] [13]．利用者による注釈の編集を許可することによって，歴史文献に関連する情報が効率的に蓄積される．また，注釈を共有することで同じ目的を持った他の利用者による文献検索の効率が上がると考えられる．従って，本論文の提案によって歴史文献の価値の向上が期待できる．本論文では日本，中国，韓国などの東アジア圏の古文書や木簡などの歴史文献を対象にする．

電子スクラップブックシステムで歴史文献の注釈や書誌などを管理するためのモデルとして，本論文では，歴史文献と注釈などとの関係を管理するための文献データのモデルと，文献の分類を管理する電子スクラップブックデータのモデルを提案する．一般にインターネットで公開される歴史文献の画像に直接注釈を記述することはできないので，文献データでは対応表を用いて歴史文献の画像と注釈の関係の詳細を記述する．電子スクラップブックデータは，文献を個別に閲覧していたのでは発見が困難な時代背景などの情報を得るために利用する．このデータは，分類を管理する文献データ等の URI を用いてそれらのデータを間接的に管理する．歴史文献の注釈は，利用者の文献の内容理解を支援するだけではなく文献検索や分析への利用が考えられる．そこで，効率的に注釈を再利用するために，ここでは注釈で扱う文字属性や単語の意味などの主な情報を分類し，それぞれについてデータ構造を定義する．

提案モデルに従ったデータの記述には，XML [14] を利用する．XML は，データをシステムに依存しない形式で記述することができるためインターネットでの情報交換に広く利用されている．従って，XML は，電子スクラップブックシステムで扱うデータをインターネット上で交換するのに適していると考え

られる．また，歴史文献は，Unicode などの標準的な符号化文字集合に含まれない文字（外字）を多く含むため文献データの外字情報の記述にも XML を利用する．さらに電子スクラップブックシステムでは，XML で記述された文献データを効率的に検索するために，関係データベースを利用する．そこで，ここでは提案モデルに従った XML スキーマに対応した関係表の設計について述べる．電子スクラップブックシステムでは，利用者による歴史文献への注釈の追加，削除などの編集と，元の文献データの一部を抽出して新たな文献データを作成する切り抜き，および収集したデータを分類し，電子スクラップブックデータを作成する機能を利用者に提供する．本論文では，データの収集支援として，文献データや電子スラップブックデータに対する検索について述べる．検索では，各データの注釈などのテキストに対する文字列検索と歴史文献の分類や分析を支援するために，文献の記述にある地理情報や時間情報に着目した検索処理について提案する．特に歴史文献から得られる時空間情報は，次のような特徴を持つ．

(1) 同じ地域でも時間とともに複数の地名を持つ

(2) 生没年などの期間や領土の範囲があいまいであることが多い

(3) 研究者の意見の相違から同じ地名でも異なる場所を指すことがある

(4) 与えられる情報は，利用者からの注釈であるため地名だけが記述されているなどの値の欠落が多い

本論文では，このような特徴をも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述べる．また，不完全な時空間情報を地名などで分類した結果を用いた値の補完について述べる．

提案モデルの実用性を示すために，これらの設計に基づいた電子スクラップブックシステムのプロトタイプシステムの実装と実行例を示す．プロトタイプシステムでは，歴史文献への注釈の編集，切り抜き操作，文字列検索，電子スクラップブックデータによる文献データの分類操作を実現している．

文献データのモデルに従って集約した歴史文献に関する注釈の再利用方法として，文献データを一般的な WWW ブラウザを用いた閲覧の方法を提案する．ここでは次ようなの方法を提案する．

(1) 歴史文献のレイアウトに則したテキストデータの表示方法

(2) 文献の注釈を WWW 環境で閲覧する方法

文献データを WWW 環境での閲覧を可能とすることで，他の HTML 文書などとの組み合わせや WWW サーチエンジンなどでの検索が可能になる．従って，電子スクラップブックシステムを利用しない利用者に対する研究者の成果の公開や，研究者間の情報交換，一般的な利用者による歴史文献の読解支援が容易になると考えられる．

本論文の提案によって，ネットワーク上に公開される歴史文献の画像を用いて関連する情報を集約することが可能となる．また，提案手法は，歴史文献と関連情報を研究活動以外の教育や電子図書館などに適応できると考えられる．

1.2 構成

本論文の次章以降の構成は次の通りである．2 章では，関連研究とデータの記述に用いる XML について述べる．3 章は，東アジア圏の歴史文献の特徴と歴史文献を用いた研究活動を考察し，インターネットを介して歴史文献を利用するための要求について考察する．4 章では，歴史文献を用いた研究活動を支援するための文献データのモデルおよび，電子スクラップブックデータのモデルについて述べる．さらに，効率的に注釈を再利用するために注釈の種類をあげ，それぞれの注釈のデータ構造を定義する．5 章では，文献データモデルなどを XML 及び 関係モデルで記述するための写像および，それぞれのデータに対する操作について述べる．文献データの検索では，文献データの注釈などのテキストデータに対する文字列一致を用いた検索処理と，歴史文献に関する時空間情報を利用した検索処理について提案する．6 章では，電子スクラップブックシステムのプロトタイプシステムの実装と実行例を示し，提案モデルの

有効性を示す．7 章では，作成した文献データを対象に WWW ブラウザを用いて閲覧するための方法について述べる．8 章では，本論文の結論と今後の課題について述べる．

第2章　関連研究

2.1　関連研究

インターネットを介した電子文書の交換に広く利用されているフォーマットとして Adobe 社の PDF （Portable Document Format） がある [15]．PDF は印刷に適した電子文書の交換を目的にしたフォーマットであり，文書交換の他に，注釈の添付，校正などのためのマークの記入などが可能である．しかし，PDF 文書の切り抜き操作は，ページ単位でしかできず，ページ内の一部分だけを取り出す場合は画像もしくは文字列に変換しなければならない．一方，提案モデルは，PDF とは異なり文献データの任意の部分の切り抜きをすることができる．また，PDF が独自形式のデータモデルであるのに対して，提案モデルに従ったデータは XML を用いて記述するので，HTML などの他の形式への変換が容易であると考えられる．

XML を利用した電子書籍モデルとして Open eBook がある [16]．このモデルは，インターネットを介した書籍の配布と閲覧を支援するためのモデルであり，一般的なビューワなどのアプリケーションは，著作権の保護のために文書の切り抜き操作ができない．一方，本論文は，利用者による情報の収集と整理の支援を目的としている．

XLibris [17] は，active reading の支援のために電子文献への下線や注釈，切抜きなどの実際の紙と同様な操作環境を提供している．データ構造などで本研究との類似点も多いが，XLibris は，ユーザインタフェースに着目した研究であるのに対して本研究では，データベースおよび情報検索の手法から文献の有効利用の実現を目指している．

歴史文献の検索システムとして，奈良文化財研究所の木簡データベース [7]

がある．木簡データベースでは木簡画像に併せて木簡研究の成果に基づいたメタデータとテキストデータを併せてデータベース化し，画像の代わりにメタデータやテキストデータを用いた検索によって目的の木簡を探し出す．多くの場合，このようなメタデータを用いた検索は研究者を対象としものであり，専門的な知識を持たない利用者は必ずしも容易に扱えるものではない．一方，提案モデルでは，利用者による注釈の編集と共有が可能であるため．考古学，歴史学以外の分野での資料の利用が可能になると考えられる．

OPALES [34] は，フランス国立アーカイブで電子化された動画に対する利用者の注釈の追加と共有を支援するシステムである．OPALES では，注釈をアーカイブで管理される動画への索引として利用している．このシステムによって利用者は，このような注釈の追加，収集を通して，利用者の目的に合わせたアーカイブを作成する．電子スクラップブックシステムも OPALES と同様に公開後の歴史文献に対する注釈の共有を行っている．さらに本論文で提案するシステムは，利用者間で作成した歴史文献の分類をアーカイブとして利用者間で共有することによって，歴史文献を用いた研究活動の支援を実現している．

木簡と注釈などの関連情報を半構造データとしてモデル化しオブジェクト指向データベースで管理する提案がある [18]．この提案システムでは，公開後の木簡に注釈を追加することが可能であり，本研究と類似点が多い．これに加えて，本論文では，データの記述に XML を利用しているので，インターネットを介した歴史文献と関連情報の共有と WWW 環境での再利用を可能にしている．

電子文献に関する注釈は，既存の文書の校正や捕捉のための意見やコメントの記述に加え，検索などのために文書もしくは一部の記述に対するメタデータとしての利用が考えられる．PDF などの電子文書フォーマットでは，既存の文書に対するコメントとしての利用が中心であり，文書の検索のためのメタデータとしての利用は少ない．一方，OPALES や文献 [18] は，注釈を単にコメントとしてだけではなく文書検索のためのメタデータとして利用している．電子スクラップブックシステムにおける注釈は，これまでの文献検索のためのメタ

データとしての利用だけではなく，歴史文献に関する知識を集約する．そのために本論文では，注釈として集約された文献に関する知識を効率的に辞書データなどに再利用するためのデータ構造の定義を示した．また，電子スクラップブックシステムは，歴史学や人文科学の研究支援のための基本的な機能として注釈の編集だけではなく歴史文献の分類機能がある．この機能によって，本システムは，個々の文献と注釈を閲覧していただけでは発見が困難な文献が記述された時代背景の分析などの支援が可能である．

歴史文献のレイアウトにあわせたテキストデータの表示に SVG (Scalable Vector Graphics) を利用する方法が提案されている [19]．SVG は，XML の書式に従って 2 次元ベクトル画像を記述する規格であり，高品質な歴史文献の再現に適している．しかし，巻物などの長文の歴史文献を再現する場合，表示領域の制約などから SVG での再現は現実的ではないと指摘されている．一方，本論文の提案では，HTML を用いて歴史文献の記述を再現するので，スクロールしつつテキストを閲覧することが容易にできるため，長文の歴史文献の再現にも適していると考えられる．

考古学や歴史地理に関連する地理情報を研究者間で共有するためのプロジェクトとして，TimeMap プロジェクトが挙げられる [11]．このプロジェクトでは，考古学や歴史学に関連するデータの変遷の視覚化に適した時空間データモデルを提案している．しかし，このプロジェクトでは，地理データの修正をそのデータの制作者だけに限っている．一方，本論文のモデルでは，文献データの閲覧者も注釈の追加ができるので，地図を用いた歴史文献の分布も注釈によって動的に変化する．従って，提案モデルでは，歴史文献と地理情報を用いた共同研究支援が柔軟に行えると考えられる．

Allen の時区間関連は，時区間を 1 次元軸上の始点と終点の組で表される線分として定義し，13 の関連として提案されている [20]．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．また，文献 [21] 文献 [22] は，2 時区間の関連を位相関連と順序関連を定義し，その組み合わせから Allen の時区間関連と等価であることを示している．そこで，本論文では，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．また，境界のあいまいな時区間を対象とするため，時区間に関する位相関連に関して 2 時区間の接続を考慮しない．

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 [23]．一方，本論文では，9-intersection model で定義された領域関連のサブセットを利用した上で，これらの位相関連の判定に時区間関連を利用することを考える．

文献 [24] 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 [25] は，時区間と 「までの」 のような文書中の時間的表現を併せて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 [25] の時間情報では正確な時区間を用いるのに対して，本稿の時間情報は期間があいまいな時区間を扱う．

2.2 XML

本節では，提案モデルの記述に利用する XML について述べる [14] [26]．

XML（Extensible Markup Language) は，1998 年に World Wide Web Consortium で仕様が勧告されたデータの階層構造を記述するための言語である．XML は，1986 年に ISO が規格を公開した文書記述言語 SGML (Standerd Generalized Markup Language）に比べ以下のような特徴がある．

(1) XML は SGML に比べ仕様が単純であり，多数の処理系の実装が存在する．

(2) XML は整形式や名前空間などの SGML にない機能や概念が追加されている．

従って，XML は，近年インターネットでのデータ交換に広く利用されている．

```
<?xml version="1.0"?>
<歴史文献>
<メタデータ>
<整理番号>木簡 001</整理番号>
<発掘場所>奈良</発掘場所>
<発掘時期>1980</発掘時期>
</メタデータ>
<釈文>大花下</釈文>
</歴史文献>
```

図 2.1 XML 文書の記述例

図 2.1 に，XML を用いた簡単なデータの記述例を示す．XML に従って記述されるデータや文書は一般に XML 文書と呼ばれる．図 2.1 の 1 行目に示す “<?xml version=”1.0”?>” は，XML 宣言であり，以下に続く文書が XML 文書であることを表している．XML では，データの要素の始まりを，“<” と “>” で囲まれたタグで表現し，データの要素の終わりを “< /” と “>” で囲まれたタグで表現する．前者を開始タグ，後者を終了タグと呼び，タグに挟まれた部分が実際のデータである．XML では，データの意味を表すタグ名をデータの設計者が自由に決めることができる．図 2.1 では，“< 歴史文献 >” から “< /歴史文献 >” で挟まれた部分や，“< 本文 >” から “< /本文 >” で挟まれた部分が一つのデータの要素である．それぞれのタグに囲まれた部分がそれぞれの要素の内容である．

歴史文献データがメタデータやテキストデータなどのような階層を持つ場合，XML は，図 2.1 に示す通り，タグの入れ子を用いてデータの階層を表現する．ただし，XML を用いてデータの階層を記述する場合，“$< A >< B > foo < /A >< /B >$” というようにタグを交差させて記述することはできない．図 2.1 から，歴史文献データはメタデータと釈文データを子の要素として持ち，メタデータは，整理番号データ，発掘場所データ，発掘時期データを子の

要素として持つことが分かる．従って，XML に従って記述したデータは，図 2.2 に示すような木構造で表現することができる．XML に従ったデータの操作は，木構造データのノードに対する操作として表すことができる．

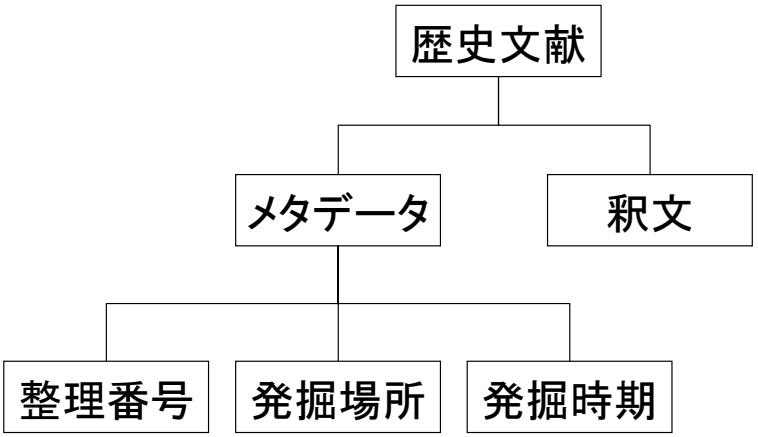

図 2.2 XML 文書の構造を木構造で記述した例

次に，XML の重要な概念の一つである名前空間について述べる．XML では，データ構造を記述するためのタグ名を，データの設計者が自由に定義することができるため，しばしばタグ名の衝突が起きる．例えば，歴史文献の表題と参照文献の表題を記述するためのタグ名を同じ “TITLE” とした場合，XML 文書内ではデータの意味の違いを区別することができない．このようなタグ名の衝突によるデータの意味のあいまいさを解消するために，XML では，名前空間が導入されている．図 2.3 に名前空間を用いた XML 文書の記述例を示す．図 2.3 の “h:参考文献” 要素の開始タグの中の xmlns:h=”http://dtd.ac.jp/history” や xmlns:ref=”http://dtd.ac.jp/refer” が名前空間の宣言であり，http://dtd.ac.jp/history などの URI が名前空間を表し，タグ名の衝突を防ぐ．また，名前空間を用いた XML 文書は，名前空間接頭辞を用いてそれぞれのタグが属する名前空間を識別する．図 2.3 において，“h” や “ref” が名前空間接頭辞である．

```
<?xml version="1.0"?>
<h:参考文献
xmlns:h="http://dtd.ac.jp/history"
xmlns:ref="http://dtd.ac.jp/refer">
<h:title>日本書紀</h:title>
<ref:title>古事記</ref:title>
<ref:title>風土記</ref:title>
</h:参考文献>
```

図 2.3 名前空間を用いた XML 文書の記述例

第3章　歴史文献に関する考察

本章では，東アジア圏の歴史文献の特徴について考察し，計算機を用いた歴史文献の閲覧や共有のための基本的な要求について述べる．

3.1　歴史文献の特徴

本論文では，歴史文献として古文書などの紙に書かれた史料の他に，遺跡などで発掘される木簡などの文書が記述された遺物を対象にする．まず，対象とする歴史文献の例を挙げる．

例 3.1 仏典

「華嚴一乘法界圖」は文書を渦巻き状に記述した歴史文献である (図 3.1) [27]．この文献では，2 次元空間に配置された渦巻き状の字の並びも仏教の精神を表現しているため，文字列としての特徴だけではなく，画像としての特徴も重要であることがわかる．

例 3.2 木簡

木簡は，紙が貴重であった古代において公文書，荷札，習字などで幅広く利用された木の札であり，考古学において貴重な史料であると考えられる (図 3.2) [28]．しかし，遺跡等で発掘される木簡は，腐蝕や墨のにじみなどにより文字の識別が困難であることが多い．また，木簡は，複数の破片を組み合わせることによって一つの木簡に復元される場合がある．

以上の例から，歴史文献には次のような特徴があるため，内容を理解するこ

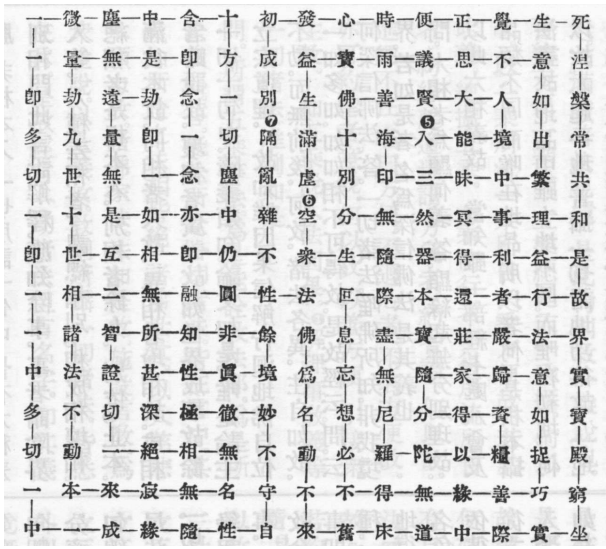

図 3.1 特殊なレイアウトの歴史的文献の例

とが困難であることが多いことがわかる．

(1) 紙や木などの媒体が劣化による変色，虫食いによって文字の判別が困難．

(2) 複雑なレイアウトで記述されていることがある．

(3) 草書体などの崩し字が利用され読むことが困難．

このような文献の特徴は，人文社会学研究にとって年代，地域，人物を特定するために重要な情報である．従って，計算機を用いた歴史文献の閲覧には，原文献の状態の再現に適している画像を用いるべきである．また，画像を用いることによって，複数の破片を組み合わせて文書を再現する木簡を計算機上で処理することが可能となる．本論文では，歴史文献をスキャナ等を用いて作成した画像を文献画像と呼ぶ．

画像だけで歴史文献を管理した場合，文献画像に対する効率的な検索は期待できない．そこで，歴史文献を計算機上で扱うためには，文献画像だけではなく，文献の記述に対応した文字データも併せて管理することが必要である．さらに，歴史文献の記述は木簡のように文字の判別が困難である場合や百人一首のように草書体のため極端に字形が変形している場合があり，文献の注釈を単

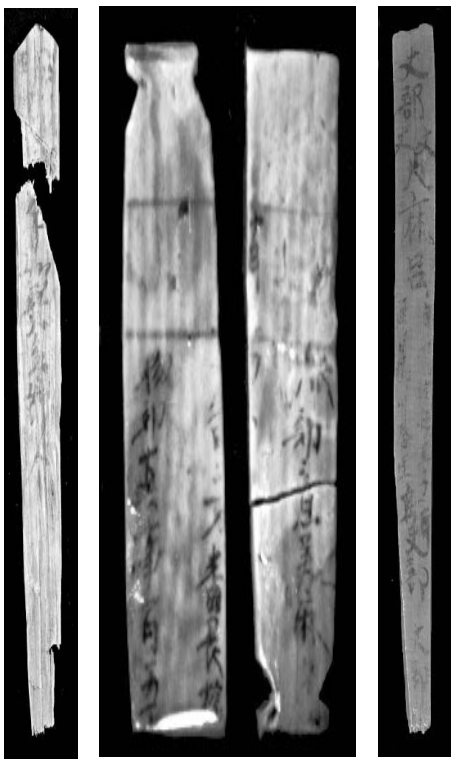

図 3.2 木簡画像の例

に列挙しただけでは歴史文献のどの部分の注釈であるかという判別が困難である．従って，文献に注釈を付ける場合，文献画像の領域を詳細に述べる必要がある．また，注釈を付けるための領域は，文献の文書に限る必要はなく，文献中の絵図や傷などに対しても注釈を付けることができる．

考古学研究では，歴史文献以外に遺跡や土器などの遺物も重要な研究資料である．遺物などは，歴史文献に比べ出土に関する地理情報やある遺物が他の遺物の部品であるなどの複雑なオブジェクトの関連を表現することが必要であると考えられる．つまり歴史文献は遺跡や遺物より下位の概念であり，これらを区別無く管理するためには，新たなデータのモデルを定義する必要があると考えられる．従って，本論文では，遺跡や遺物を対象とせず歴史文献だけを対象にする．

3.2 外字情報の扱い

東アジア圏の歴史文献の内容の記述には，外字と呼ばれる Unicode などの標準的な符号化文字集合にない文字が含まれていることが多い．一般に外字は，文献検索で利用される文字列一致の対象にならない．しかし，東アジア圏の歴史文献では，人名などで外字が利用されるため，外字も検索対象として利用できるようにしなければならない．また，注釈や釈文として関連付けられているテキストデータでの外字表示も文献を読む上で重要である．そこで，本論文では，注釈として ekanji[29] や今昔文字鏡 [30] などのインターネット上に公開されている大規模漢字情報への参照と読みなどの文字属性情報を記述する．注釈で外字情報を記述することによって，一般的な文字情報と文献固有の文字情報の併記が可能になる．

3.3 歴史文献の収集

歴史文献の収集では，文献全体を収集する他に，文献の一部だけを収集することがある．例えば，百人一首から “春” に関する記述を抜き出し収集することが挙げられる．このような文献収集の支援には，元の文献画像から任意の部分を切り抜く機能や，利用者が収集したデータの分類を管理する機能が必要である．さらに，文献画像を単に抜き出しただけでは元の文献画像と注釈の関係が失われるので，利用者は内容理解の手がかりを失なってしまう．そこで，文献画像の切り抜きに併せて，その部分に関係付けられた注釈を抜き出す必要がある．

第4章　文献データ，電子スクラップブックデータのモデル

本論文で提案する電子スクラップブックシステムは，前節で述べた歴史文献の研究支援のために，歴史文献の画像と関連するテキストデータを対応付けて管理する．同様に本システムでは利用者の収集した歴史文献と関連情報の組を効率的に利用するためにそれらの分類を管理する．また，本システムで管理される情報は，研究者の意見交換支援のためにインターネットを介して利用者間で共有される．このような様々な要求を満たすために，本論文では，管理する情報を一つのデータモデルだけに従って記述するのではなく，検索や情報交換などの目的に合わせて適した形式で記述する必要があると考えた．そこで，本節では，関係データモデルや XML などの様々なデータモデルに従ってデータを記述するために，まず既存データモデルに依存しない形式でデータの構成を定義する．本論文では，特定のデータモデルに依存せずにデータの記述するための構成や枠組みをモデルと呼ぶ．

本章では，歴史文献と注釈などの関連するテキストデータを記述するためのモデルを文献データのモデル，利用者の収集したデータの分類を記述するためのモデルを電子スクラップブックデータのモデルと呼び，それぞれの構成について述べる．文献データのモデルによって，歴史的文献の文書としての特徴と画像としての特徴を計算機上で同時に扱うことを実現する．また，このモデルは，対応表を用いて歴史文献の画像と注釈の関係を明示する．電子スクラップブックデータのモデルは，文献データへの参照を用いて分類を管理する．文献データと電子スクラップブックデータの概略と関係を図 4.1 に示す．文献データは，文献画像，文献情報，本文，注釈を管理し，対応表で注釈と文献画像の位置関係を管理する (図 4.1 下) ．電子スクラップブックデータは，このデー

タのメタデータである電子スクラップブック情報と文献データの分類を記述するグループからなる (図 4.1 上)．図 4.1 の文献データ A’ は文献データ A からの切り抜きによって生成した文献データである．文献データ A’ も文献データのモデルに従うので，他のデータと区別せずに管理できる．

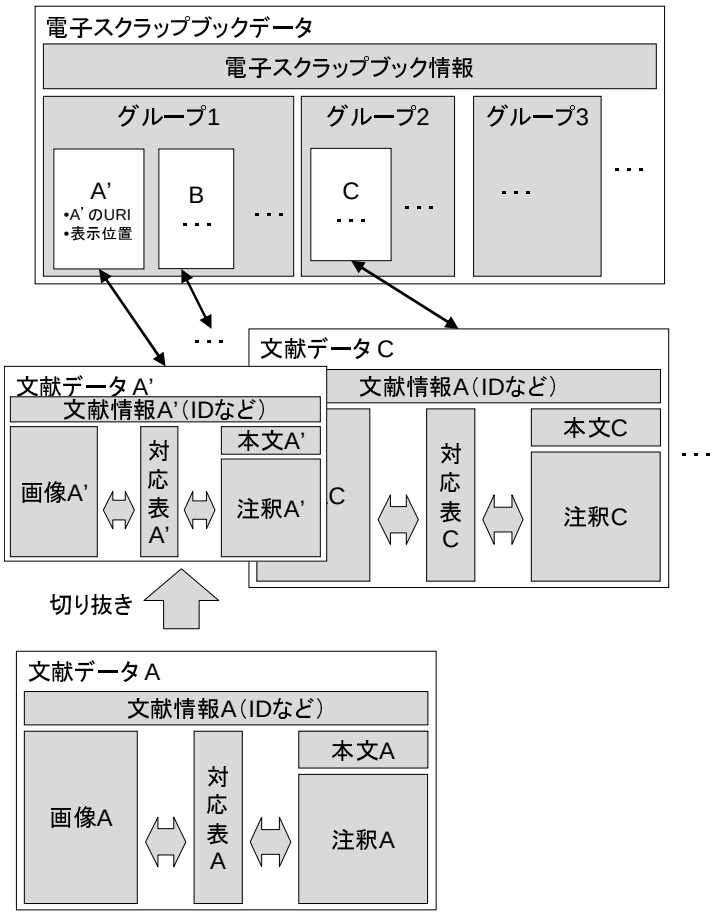

図 4.1 文献データと電子スクラップブックデータの関係

考古学の研究のために遺跡や土器などの遺物のデータベースの利用についての提案がある [31] [32]．遺跡や土器などの遺物をオブジェクトとして考えた場合，木簡や古文書などの歴史文献より一般的な概念と言える．さらに，遺物と遺跡の出土場所などの地理的な内包関係や，遺物間の is-part-of 関係のような

複合オブジェクトの特徴を持つ．従って，遺物を管理するには，注釈の共有だけを考えるのではなく，オブジェクト間の関係を表現できなければならない．本章で提案するモデルは歴史文献間の参照関係などの制約の弱い関係だけを表現し，遺跡や遺物の関係を表現するために必要な制約の強い関係は表現しない．従って，本研究では，歴史文献だけを対象とし，土器などの遺物は扱わない．

4.1 文献データのモデル

文献データのモデルは，インターネット上で公開されている文献画像と注釈などの関係を記述するためのモデルである．このモデルは，歴史文献に関する知識の集約に利用できると考えられる．

まず，このモデルの全体構成について定義し，次にこのモデルの各要素の構成の定義を示す．文献データのモデルは次のように定義する．

定義 4.1 文献データ
文献データは，歴史文献に関するメタデータを表す文献情報 t_1，歴史文献の画像の集合 g, 歴史文献の記述を表す本文 t_2，注釈の集合 an，注釈と注釈を関連付ける文献画像上の領域の関係を表す対応表 r の組として表現される．

文献データ $d = (t_1, g, t_2, an, r)$

(1) 文献情報

文献情報は，表題や分類などの文献に直接には現れないデータをメタデータとして記述する．メタデータとして記述する要素は，Dublin Core Metadata Element Set Ver. 1.1[33] に従った 15 要素を用いる．標準的なメタデータを用いれば，他の図書目録や画像データベースなどのメタデータを利用した情報システムとの情報交換が容易になると考えられる．従って，文献情報のデータ構造は次の通りである．

定義 4.2 文献情報
文献情報 t_1 は，Dublin Core Metadata Elemnt Set Ver. 1.1 の 15 要素の組として表現される．

文献情報 t_1 = (Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights)

文献情報では，他の文献データと識別するための “Identifier” 要素を必須要素とする．“Identifier” 要素に与える値は，文献データの URI を記入する．

例 4.1

図 3.1 を用いた文献情報の記述例を以下に示す．

文献情報 = (Title:華嚴一乘法界圖，Subject: 仏典，
Identifier: http://www.foo.ac.jp/butten001.xml, •••)

(2) 文献画像

文献画像は，歴史文献の画像を管理するための構成要素である．ここで管理する画像を用いて，テキストデータでは表現できないレイアウトや字形などの文献の絵としての特徴を計算機上で再現する．本論文では，インターネット上に公開される歴史文献を対象にする．そこで，この要素では，画像を直接管理するのではなく画像の URI を用いて間接的に管理する．歴史文献の中には，複数の断片から一つの歴史文献が再現される文献もある．従って，複数の画像からなる歴史文献を表示するために文献画像の URI と合わせて配置位置を管理する．配置位置は，画像表示領域の左上を原点としたとき，表示する画像の左上頂点の座標として表現される．表示領域の大きさは，関連する画像の配置後の各頂点の中

で最も大きい x 軸の値と y 軸の値を用いて決定する．従って，複数の画像を用いて一つの歴史文献を表す場合は，配置位置要素と URI 要素の組を画像の数だけ列挙する．ただし，歴史文献が一つの画像で表現される場合，配置位置の値は原点と一致するのでこの属性は省略される．また表示領域の大きさは，画像のサイズと一致する．以下に文献画像の構造を記す.

定義 4.3 文献画像
文献画像 g は，文献画像を表示する情報である配置情報 pos と，画像の URI 組の集合として表現される．また，位置情報 pos は，x 軸，y 軸 それぞれの座標値 pos_x, pos_y の組として表現される．

文献画像 g = (pos, URI)
配置情報 pos = (pos_x, pos_y)

配置情報は，整数値の組である．この要素の記述例は以下の通りである.

例 4.2
図 3.1 が，http://foo.ac.jp/hokai.jpg で公開されている場合，文献画像要素は，その URI を記述する．また，この例では，文献画像が一つだけであるので配置情報の記述を省略している.

文献画像 = http://foo.ac.jp/hokai.jpg

(3) 本文

本文は，このモデルで管理する歴史文献の内容を文字データで記述するための要素である．一般に，歴史文献の記述には，外字と呼ばれる標準的な文字コードにない文字が含まれることが多い．テキストデータに含まれる外字を記述するために，次に述べる注釈への参照を用いる．本文要素の構造を次に示す.

定義 4.4 本文
本文 t_2 は，歴史文献の記述をテキストデータ *value* だけを持つ要素として表現される．

本文 t_2 = *value*

value に含まれる外字を記述するための文字参照は，注釈識別子を用いて記述する．

例 4.3
図 3.1 の文献画像の内容を記述した本文要素の例を示す．

本文 = 佛爲名動 ･･･ 性法

(4) 注釈

注釈は，文献画像の内容に関連する情報を管理するための要素である．注釈は，注釈識別子と注釈文の組の集合である．注釈識別子は，対応表で文献画像の領域と注釈の関係付けと，本文要素のテキストデータの外字の参照に利用される値である．注釈文は，関連情報を記述する要素であり，文字属性，単語の意味，地理情報など様々な情報の記述が考えられる．そこで本論文は，注釈文の値として任意のテキストデータを記述する他に，次節で述べる注釈の内容に基づいて定義したデータ構造に従ったデータを記述する．次に注釈要素の構成の定義を示す．

定義 4.5 注釈
注釈 an は，文献データ内で本文要素などから参照される注釈識別子 aid と，注釈の内容を記述する注釈文 an_obj の組の集合として表現される．

注釈 an = (aid, an_obj)

注釈文 an_obj は，単純なテキストデータの他に，次節で述べる文字属性情報，意味情報，地理情報などの構造を持つデータの記述を許す．

例 4.4
図 3.1 の各文字に対する読みをテキストデータとして記述する例である．

注釈 = {(1，ふつ)，(2，い)，⋯ }

(5) 対応表

一般にインターネット上で公開される歴史文献の画像に直接，注釈を記入することはできない．このモデルでは対応表を用いて注釈を記入する文献画像の位置の詳細を記述する．対応表の構成は，文献画像の注釈位置と注釈の識別子の組の集合である．文献画像の領域は，文献画像の座標系を (1) 画像の左上を原点，(2) 水平方向の左から右の向きを x 軸正の向き，(3) 垂直に上から下の方向を y 軸正の向きとした場合の長方形の原点に最も近い点と最も遠い点の座標の組で表す．注釈と文献画像の領域を明示的に関係付けることで，文献画像の一部を切り出すと同時に，その領域に関連付けられている注釈を抜き出すことができる．その他に文献画像の個々の文字ごとに領域を記述し，関係付ける注釈に文字の出現順を記述すれば，文献画像の読む順序を明示することができる．対応表の構成は次の通りである．

定義 4.6 対応表
対応表 r は，注釈と関連付ける注釈位置 pos_an と注釈識別子 aid の組の集合として表現する．また，注釈位置 pos_an は，長方形で表現される領域の原点に最も近い頂点座標 a と原点から最も遠い頂点座標 b の組として表現する．

対応表 = (pos_an, aid)
注釈位置 pos_an = (a，b)

一般に文献には，n 個の注釈が付く，逆に文献中にある語が m 個あった場合， 文献から注釈への関係は，m 個の関係が存在する．従って，注釈と文献の間には，n 対 m の関係が成り立つ．

例 4.5
図 3.1 の対応表の記述例は，次の通りである．

対応表 = { ((100，100，110，110)， 1)，((100，110，110，120)，2)，$\cdots$ }

例に示す注釈位置では，原点から最も近い頂点の x 座標値，y 座標値，原点から最も遠い頂点の x 座標値，y 座標値の順に記述することで注釈を関連付ける領域を表現している．

4.2 注釈文のデータ構造

本論文では，利用者間で文献画像を効率的に共有するために注釈が重要な役割を果たしている．従って，注釈は，歴史文献の内容理解の他に文献の検索や分析への利用が考えられる．計算機による歴史文献の検索や分類を効率的に処理するには，注釈文の形式を定義することが重要であると考えられる．しか

し，本論文で扱う注釈の内容は，文字属性，意味情報，地理情報，時間情報，外部参照など様々な特徴をもつものが挙げられる．従って，一つの形式ですべての注釈の内容を記述することは困難である．そこで，本論文では，注釈の内容に則した記述形式を定義する．本章では，文字情報，単語情報，地理情報，時間情報，外部リソースへのリンクなどについて注釈文を記述するための形式について述べる．さらに，外部リソースとして利用者が作成する注釈について述べる．

注釈文の記述形式の定義にあたり，次のような方針が挙げられる [34].

- 制約の弱い形式
  制約が弱い注釈の形式は，利用者による記述の容易さを優先させた形式である．例えば，注釈として方角の記述に関する制約が無ければ，同じ方角でもあっても “1 時の方角” や “北方向から時計回りに 30 度” など様々な記述が考えられる．このように，利用者の意図に合わせて注釈を記述することが可能であるため，利用者から多くの注釈の記述が期待できる．しかし，注釈の計算機による注釈の分類や分析，語彙抽出などを効率的に処理することは困難である．

- 制約の強い形式
  注釈の記述に関する制約の強い形式は，計算機による効率的な処理を目的とした形式である．一般に制約が強い場合，利用者が注釈を作成するときの規則が複雑になるため，利用者による注釈の作成数が減少する．

そこで本研究では，一般的な辞書などの形式に基づいて，直観的な構造を定義することで利用者による記述の容易さと，計算機による分析効率の両立を目指す．また，注釈の内容に用いる用語の統一が計算機による情報の分類処理に有効であると考えられる．しかし，考古学による遺物の名称や古文書による解釈が研究者によって異なることが多く用語の統一は容易ではないと考えられるため，本研究では用語の統一は考慮しない．だたし，現代中国語の発音表記方法であるピンインなどの発音の表記や地図上の位置の記述は一般的な表記に基づく．

(1) 文字属性情報

東アジア圏の歴史文献の記述には，漢字，ひらがな，カタカナ，ハングル文字など様々な文字が用いられる．特に漢字は，異体字が多く存在するため，テキストデータとして表現できない外字が多数存在する．そこで，ここでは文字属性情報を外字情報の記述に利用する．従って，この要素は，文献中の文字の読みや画数などの属性情報と外字を表示するための文字画像へのリンクを記述する．文字に関する記述には，国文学研究資料館の外字辞書 [10] や，XML による画像参照交換方式 [35] などの提案がある．前者は読みなどの要素ごとに著者登録の記述があり，注釈で用いるには記述内容の重複が多く制約が強いと考えられる．後者は，文書中に埋め込む外字画像の指定方法を定義した規格であり，文字属性情報の記述に関する制約は弱い．そこで本研究では，国文学研究資料館外字辞書を簡略化した形式を考える．また，ekanji [29] や今昔文字鏡 [30] などのインターネットで公開されている大規模漢字集合は，文字の画像だけなく読みなどの文字属性が利用できるものもある．文字情報は，以下のような項目の組として記述する．

定義 4.7 文字属性情報
文字属性情報 k は，読みの集合 p，漢字の部首などを記述する構造 s，参照 ref，出典 so，解説 m の組として表現される．また，読み p は，読みの種別 pt と，発音 pp の組の集合として表現され，構造は，部首 kk，部首を含む文字の総画数 tc，部首を含まない文字の部分画数 c の組として表現される．

文字属性情報 k = (p，s，ref，so，m)
読み = (pt，pp)
構造 = (kk，tc，c)

読みの種別は，漢字の発音を記述する要素であり，漢音，呉音，訓，ピ

ンインなどの名称を記述する．参照には，外字の表示に用いる文字画像を表す URI を記述する．本論文で利用する文字画像は，eKanji [29] や今昔文字鏡 [30] などで公開されているデータを利用する．出典は，注釈を付ける文字の出典情報を記述する．解説は，文献固有の情報を記述する要素である．外字の属性情報を記述する場合は，読みと参照を必須要素とする．次に文字属性情報の記述例を示す．

例 4.6
以下は，図 3.1 の最初の文字である “佛” に関する文字属性情報の記述例である．

文字属性情報 = ((漢音, ふつ),(呉音，ぶつ)，(にんべん，7，5),
(http://foo.ac.jp/kanji/futsu.gif),(辞書 A)，(仏の旧字体))

(2) 単語情報
この要素は，文献中の人物や言葉に関する意味を記述する．この要素で記述する内容は，一般的な辞書の項目に基づいて以下のように定義する．

定義 4.8 単語情報
単語情報 w は，単語の読み p，記述 des，分類 kind，解説 m，参照 ref の組として表現される．

単語情報 w= (p，des，kind，m，ref)

読みは，注釈を付ける語の発音を記述する．記述は，注釈を付ける文献の内容に対応したテキストデータを記述する要素であり，外字を含む場合，外字情報に関する注釈への参照を記述する．分類は，品詞や用語の分類など，単語の理解を助けるための情報を記述する．解説では，単語の意味を記述する．参照では，出典情報を記述する．次に単語情報の記

述例を示す.

例 4.7
例として図 3.2 の “大花下” の意味の記述を示す.

単語情報=((-), (大花下), (官位), (大化五年二月から天智三年二月の間に使われた第八階の官位), (日本古代木簡選))

“-” は，値が不明であること表している．つまり，上記の例では読みが不明であることを表す.

(3) 地理情報
この要素は，歴史文献に関連する地名などに関する情報を記述する.

定義 4.9 地理情報
地理情報 geoinfo は，領域 area，地名 na，参照 ref の組として表現される．また，個々の領域要素は，緯度 la，経度 lo の組の集合として表現される.

地理情報 geoinfo = (area，na，ref)
領域 area = (la，lo)

領域は，歴史文献から得られる地名に関連する範囲を記述する要素である．領域は，ある大きさを持った範囲として表現されるので，緯度，経度の組を頂点として列挙し，リストの先頭と最後を結び表現される多角形とする．また，地名は歴史文献の記述された地域名を記述する．参照は，出典情報を記述する.

例 4.8
地理情報の記述例を表す．この例で用いる値は仮の値である．

地理情報= ((100，300)，(200，300)，･･･，(出雲)，(http://foo.ac.jp/))

上記の例の地名である “出雲” より先に記述された座標値の集合は，出雲の領域の範囲を多角形で表現している．

(4) 時間情報
歴史文献では，地理情報だけではなく時間情報も文献を分類する上で重要な情報である．歴史文献に関連する年代は，時区間として表現することもできる．本節では以下のような形式で，時間情報を記述する．

定義 4.10 時間情報
時間情報 timeinfo は，時区間 ti，時代 p の集合，参照 ref の組で表現される．また，時区間 ti は開始時間 ts，終了時間 te の組として表現され，時代 p は，年号 ne と年数 ye の組として表現される．

時間情報 timeinfo = (ti，p，ref)
時区間 ti = (ts，te)
時代 p = (ne，ye)

時区間は，開始時間と終了時間の組として表現される時間軸上の線分である．時代には，歴史文献に関連する年号などを記述する．時区間は複数の時代に関連することがあるので，時代は，0 個以上の値を持つ．各時代は，年号と年数の組で表現する．ただし年数が 0 の場合は，その年号に該当する時区間全体を指す．参照では出典情報を記述する．

例 4.9
図 3.2 を例に時間情報の記述例を示す.

時間情報 = ((649, 664)，(大化 5 年，天智 3 年)，(日本古代木簡選))

年号と時区間の多くは，一対一の関係であるが，古い年号や時代は，時区間の範囲があいまいな場合があり，研究者によって期間の範囲が異なる場合があると考えられるため年号と時区間の両方を記述する.

(5) 外部リソースへのリンク
注釈文として，既存の Web ページを参照する場合は，URI を記述する.さらに本稿では，注釈である外部リソースの分類を併せて記述する.

定義 4.11 リンク
リンク link は，分類 div と参照 ref の組として表現される.

リンク link = (div，ref)

分類では，リンク先の外部リソースがどのような目的でリンク付けしているのかを任意のテキストデータとして記述する．参照は，参照する外部リソースの URI を記述する.

例 4.10
ある文献に関連する WWW ページへのリンクの記述例を挙げる.

リンク = ((漢字情報)，(http://foo.ac.jp/infomation.html))

(6) 文献画像と本文データとの対応
文献データの本文のテキストデータと，文献画像の記述の文字ごとの対

応を表現する要素である．この要素によって，切り抜き操作で抽出される元の文献データの部分画像と注釈に併せて本文から関連するテキストデータを取り出すことができる．

定義 4.12 文字位置
文字位置 str_pos は，開始位置 ss と終了位置 se の組として表現される．

文字位置 str_pos = (ss，se)

草書体で記述された文献の記述には一つの文字で 2 文字以上のテキストデータに対応する文字がある．そこで，注釈として記述する文字位置は，文献画像の記述に関係するテキストデータの開始位置と終了位置の組として表現する．また，文献画像中の 1 文字がテキストデータの 1 文字に対応する場合は，開始位置，終了位置ともに同じ値を入れる．

例 4.11
図 3.2 の「花」を例に文字位置の記述例を示す．

文字位置 = (2,2)

この例では文献画像中の一つの文字と本文のテキストデータの一つの文字が 1 対 1 に対応しているので開始位置と終了位置が一致する．ただし，文字位置と対応付ける画像の領域情報は対応表を用いるのでここでは記述しない．

最後に，外部リソースとして利用者が注釈文を生成することを考える．利用者が外部リソースとして注釈を作成した場合，歴史文献間で注釈を共有することが可能であり，注釈を介した文献の効率的な閲覧ができると考えられる．

外部リソースとして注釈を作成する場合，注釈文と作成者などの情報を記録するためのメタデータを記述する．メタデータで記述する項目は，Dublin Core Metadata Element Set Ver. 1.1 で定義された項目を用いる．

4.3 電子スクラップブックデータのモデル

電子スクラップブックデータのモデルは，文献データの分類を管理するモデルである (図 4.1) ．このモデルによって，利用者は個々の文献を閲覧しただけでは発見の困難な文献間の参照関係などを表現することができる．まず，電子スクラップブックデータのモデルの全体の定義を示す．次にそれぞれの構成要素について述べる．

定義 4.13 電子スクラップブックデータ
電子スクラップブックデータ esb は，このデータのメタデータを表す電子スクラップブック情報 esb_meta とグループ group の集合の組として表現される．

電子スクラップブックデータ esb = (tesb_meta, group)

(1) 電子スクラップブック情報

この要素は，電子スクラップブックデータの所有者などのメタデータを記述する．メタデータとして記述する項目は，文献情報と同様に Dublin Core Metadata Element Set Ver.1.1 [33] に従う．このように文献データと属性を統一することで，電子スクラップブックデータと文献データを区別せずにメタデータによる検索が可能である． Identifier 要素は，文献データを識別するための必須要素であり，値にはメタデータを付ける電子スクラップブックデータの URI を記述する．電子スクラップブックデータ情報の構成は次の通りである．

定義 4.14 電子スクラップブック情報
電子スクラップブックデータ esb_meta は，Dublin Core Metadata Element Set Ver. 1.1 に従った 15 項目の組として表現される．

電子スクラップブックデータ情報 esb_meta = (Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights)

また，記述例を以下に示す．

例 4.12

電子スクラップブック情報 = (Title:仏典スクラップ，Subject: 歴史，URI: http://www.foo.ac.jp/scrapbook001.xml，･･･)

(2) グループ

グループ要素は，利用者の収集した文献データや電子スクラップブックデータの分類を管理するための要素である．この要素では，対象のデータの URI を記述することで間接的に管理する．また，グループ要素は，各データの URI と併せて表示位置を持つ．表示位置は，グループとしてまとめた歴史文献を概観するための情報を記述する要素である．表示位置に記述する内容は，表示領域の左上を原点とした場合の歴史文献の画像の左上の頂点の座標である．その他，グループメタデータ要素は，グループに一つだけある要素でありグループの内容を表すキーワードの記述を許す．グループ要素の構成は次の通りである．

定義 4.14 グループ
グループ group は，グループメタデータ group_meta と，グループ要素 ge の組として表現される．また，グループ要素は，グループに属する文献データや電子スクラップブックデータの URI と文献画像を表示するための表示位置 display_pos の組として表現される．

グループ group = (group_meta, ge)
グループ要素 ge= (URI, display_pos)

以下にグループ要素の記述例を示す.

例 4.13
記述例を以下に示す．この例では，一つのグループに二つの文献データが分類されていることを表す.

グループ = (分類 1, (http://www.foo.ac.jp/bunken1.xml, 100,100),
(http://www.foo.ac.jp/bunken2.xml, 300,100))

4.4 考察

歴史文献に対する注釈は，古文書や木簡などの種類や組織によって内容が異なるため，すべての歴史文献に対応した構造を定義することは困難である．そこで，本研究で提案する文献データのモデルでは，どのような種類の文献に対しても変化のない歴史文献の画像と注釈の関係だけを記述する単純な構造を定義した．次に，本章では，提案モデルの拡張として歴史文献に付ける注釈の種類ごとにデータ構造を定義した．このように文献データのモデルは，基本構造と注釈の拡張構造の二つの階層を持つデータのモデルである．特に注釈の種類は，本論文で定義した以外にも考えられる．従って，長期に渡り文献データのモデルを利用するためには，注釈の種類の増加を可能にする必要がある．その方法の一つとして，XML における名前空間と同様の概念の導入が考えられる．

名前空間を導入することで，注釈の作成者自身が注釈の定義を自由に追加できるようになるので柔軟な文献データの拡張が実現できると考えられる．

電子スクラップブックデータのモデルは，利用者の収集した情報を整理し，公開するためのモデルであるので，利用者が保存しているデータの目録として利用できる．従って，注釈を用いた検索以外の利用者間の情報共有に電子スクラップブックデータが利用できると考えられる．長期間に渡り提案モデルの利用を実現するには，このような目録データを利用者間で交換を容易にするための枠組みが必要であると考えられる．

第5章　設計

本章では，提案モデルに従ったデータを XML 文書として管理するための Relax スキーマの作成，検索のための関係表の設計，提案モデルの編集操作及び検索処理について述べる．

5.1　文献データと電子スクラップブックデータの XML による記述

本論文で提案したモデルに従ったデータを，利用者間で交換するために XML [14] を利用する．本節では，本論文の提案モデルに従った XML スキーマの作成について述べる．

主な XML スキーマの定義言語には次のようなものある．

- DTD [26]
  DTD (Document Type Definition) は， XML1.0 の中で規定され要素型宣言，属性リスト宣言，エンティティ宣言，記法宣言を用いて XML 文書の論理構造を定義する．XML の前身である SGML で使用されていた定義言語であるため，DTD は XML 文法に従っていない，名前空間に対応していない，データ記述のためのスキーマを定義するにはデータ型が貧弱などの問題点がある．

- XML Schema [36] [26]
  XML Schema は，World Wide Web Consortium から 2001 年に仕様が公開された スキーマ記述言語であり，DTD と比べ (1) 要素の内容や属性値のデータ型を詳細に指定できる，(2) 名前空間の使用を前提に設計

されている，(3) XML データを使ってスキーマを表現するという特徴がある．しかし，XML Schema は，仕様が複雑であるため，注釈文などのスキーマを追加する場合，利用者の負担が大きいと考えられる．

- Relax [37]
  Relax は，XML Schema と同様に DTD の欠点を解決している．また，XML Schema に比べ数学的基盤を持ちかつ仕様がシンプルであるため，スキーマの記述が容易であると考えられる．

以上から本論文では，スキーマ記述に Relax[37] を用いる．文献データのモデルと電子スクラップブックデータのモデルに従って Relax スキーマを作成する手順を以下に示す．また，文献データのモデルの注釈文の値は，単純なテキストデータだけを扱う．

(1) 文献データ，電子スクラップデータに対応した XML ルート要素を作成し，各構成要素をルート要素の子とする．

例 5.1
文献データのモデルに従ったタグを用いた XML 文書の記述例を以下に示す．

<文献データ>
　　<文献情報> … </文献情報>
　　<文献画像> … </文献画像>
　　<本文> … </本文>
　　<注釈> … </注釈>
　　<対応表> … </対応表>
</文献データ>

(2) 各構成要素の持つ属性は，次のような方針で XML 要素を生成し，対応する構成要素の XML 要素の子とする．

(a) 本文などの属性を持たない構成要素は，その構成要素に対応した XML 要素自体が値を持つ.

例 5.2
文献データの本文要素に従ったタグの記述例を以下に示す.

<本文> ··· </本文>

(b) 文献情報のように記述する要素が決まっているものは，その要素を XML 要素として列挙する.

例 5.3
文献データの文献情報要素の定義に従ったタグの記述例を以下に示す. 文献情報の中で記述する要素は，Dublin Core Metadata Element Set Ver. 1.1 で定義されたものを用いる.

<文献情報>
    <Subject> ··· </Subject>
    <Title> ··· </Title>
    <Identifier> ··· </Identifier>
 ···
</文献情報>

(c) 注釈などの構成要素の中で定義した属性の組が複数回登場するものは，その属性の組を保存する中間 XML 要素を定義し，中間 XML 要素を構成要素に対応した XML 要素の子として追加する.

例 5.4
文献データの注釈要素の定義に従ったタグの記述例を以下に示す. この例では注釈要素の属性の組を保存するための中間要素として注釈項目を用いる.

```
<注釈>
        <注釈項目>
            <注釈識別子> … </注釈識別子>
            <注釈文> … </注釈文>
        </注釈項目>
        <注釈項目>
            <注釈識別子> … </注釈識別子>
            <注釈文> … </注釈文>
        </注釈項目>
 …
</注釈>
```

次章で述べる電子スクラップブックシステムで使用している文献データのモデルの Relax スキーマを付録 A に，電子スクラップブックデータのモデルの Relax スキーマを付録 B に示す.

先の例に示すとおり文献データのモデルの Relax スキーマは，文献情報，文献画像，本文，注釈，対応表の５つの部分からなる．文献情報の XML 要素は，Title などの 15 の XML 要素を子として持つ．文献画像や本文の XML 要素は，直接値を管理する．注釈と対応表の XML 要素は，属性の組が複数個出現するので，それらの組を保存する XML 要素を持つ.

電子スクラップブックデータのモデルの Relax スキーマは，ルート要素の子に電子スクラップブック情報とグループを持つ．グループは電子スクラップブックデータの中で一つ以上登場する．電子スクラップブック情報は，Title などの 15 の属性に対応した XML 要素を持つ．グループは，文献データへの URI と表示情報の組が複数登場するので，それらの組を保存する XML 要素を子として持つ.

5.2 文献データと電子スクラップブックデータに対応した関係表の設計

文献データのモデルと電子スクラップブックデータのモデルに従ったデータを検索する場合，それらのデータを XML 文書の形式のまま直接，検索するより，検索に適した形式に変換した方がよい．そこで，本論文では，検索のために XML 文書であるデータから必要な情報を抜き出し，関係データベースで管理することを考える．また，関係データベースを利用することによって既存の歴史や考古学のデータと関連付けた検索や，注釈等の編集権限の管理，文献データへの同時書き込みの制御などにデータベース技術が利用できると考えられる．

本節における文献データに関する検索は，注釈や文献情報要素などに格納されたテキストデータに対する文字列一致を用いる．そこでデータベースには，検索対象となる要素名とその値だけを格納すればよいと考えられる．要素名の記述には，XPath [38] を利用する．また，この検索では，XML 文書の一部分を取り出すのではなく文献データなどの XML 文書自身を取り出すことを考える．提案モデルに従ったデータを格納するための関係表の作成には，以下のような方針が考えられる．

(方針 1) 属性に XML 文書の各要素までのパスとその要素の値を持つ関係表を作成する．

(方針 2) 表名に XML 文書の各要素までのパスを用い属性にその要素の値を格納する関係表を作成する．

検索の対象となる XML 文書は，文献データのモデルと電子スクラップブックのデータの構造は一定であり，検索対象となる要素は，本文，注釈，対応表などのテキストデータだけであるので方針 2 に従って関係表を作成する．ただし，文献データのモデルの注釈文要素は，内容によってデータ構造が異なる．さらに，注釈文要素で扱う内容の種類は，本論文で定義した以外にも注釈の作成者によって様々な種類が考えられるため方針 2 に従って関係表を設計した場

合，注釈の内容の種類が増えるたびに関係表を新たに追加しなければならない．そこで，注釈文は，方針 1 の考えも取り入れて設計する必要がある．従って，本論文では，提案モデルに対する関係表を以下のように設計する．

(a) 要素ごとの関係表の設計

文献データのモデルおよび電子スクラップブックデータのモデルに従ったデータは XML 文書であるので，各データは木構造を用いて表現できる．従って木構造の各リーフ要素ごとに関係表を設計する．

例 5.5

文献データモデルの文献情報に含まれる Title 要素に対応する関係表を作ることを考える．このとき，ルート要素から Title 要素までのパスは/文献データモデル/文献情報/Title と一意であることから，関係表の表名として利用できる．検索結果は文献データ全体であることから，この関係表は Title 要素の値の他に格納する値が属する文献データの URI を外部キーとして管理する．従って，Title 要素の関係表は以下の様に定義できる．

関係表：/文献データモデル/文献情報/Title (value, URI)

属性 “value” は Title 要素の値である文字列を管理する．属性 “URI” は，文献データの URI を記述する．

(b) 関係表の統合

文献データの検索は，一つの要素だけを対象にするのではなく，複数の要素を対象にすることの方が多いと考えられる．従って，要素ごとに関係表を作成した場合，検索条件によっては要素ごとに JOIN 操作が必要となり，効率的な検索処理が期待できない．そこで，先に定義した表を統合して JOIN 操作を減らすことを考える．ここでは，各表名であるパスの最長一致を行い，その結果である共通部分を用いて新たな表を作成する．生成された表の属性は，パスの非共通部分の要素名と 各データの URI を持つ．

例 5.6
例として，文献データモデルの文献情報に関する関係表の統合を考える．先の例から各文献情報の子要素に関して以下のような関係表が作成できる．

関係表：/文献データモデル/文献情報/Title (value, URI)
関係表：/文献データモデル/文献情報/Subject (value, URI)
関係表：/文献データモデル/文献情報/Identifier (value, URI) ･･･

この場合，文献情報の各要素は共通パスとして “/文献データモデル/文献情報 ” を持ち，非共通分が “ Title ”， “ Subject ” ，“Identifier” であることが分かる．従って，表名を “/文献データモデル/文献情報” とし属性に “Title”や “Subject”などの 15 の属性と文献データの URI を持つ表を作成する．

関係表：/文献データモデル/文献情報/（URI，Title，Subject，Identifier，･･･）

(c) 注釈要素に関する関係表の設計
注釈要素は，注釈識別子と合わせて注釈文を持つ．注釈文は先に述べたとおり内容によって異なる構造を持った XML 文書である．従って，注釈識別子と注釈文の値を管理する関係表名は，それぞれの要素の親である注釈までのパスを用いる．注釈文の値までパスは，属性 “パス” に格納し，値を属性 “注釈文” に格納する．

関係表：/文献データモデル/注釈/（URI，注釈識別子，注釈文，パス）

URI は，管理する注釈の属する文献データの URI を記述する．また，上記の属性 “注釈文” は，構造のないのテキストデータも含めてすべての注釈文に記述される値を格納する．

5.3 操作

本節では，文献データへの注釈付けや電子スクラップブックデータに対する分類操作などの設計について述べる．

(1) 注釈の編集

文献データの注釈の編集として挿入と削除，更新について述べる．

- 挿入
  注釈の挿入では，まず，利用者が文献画像の領域と注釈文を与える．次に，対応表と注釈にそれぞれの値を登録するための属性を追加し値を挿入する．その次に，追加した注釈文に対し識別子を与える，最後に識別子を対応表に追加することによって処理する．

- 削除
  注釈の削除は，注釈文だけでなく関係する対応表も削除する操作である．この操作では，入力として削除する注釈識別子を与える．与えられた注釈識別子に従って，注釈および対応表の領域を削除する．

- 更新
  注釈文の内容を更新する場合は，入力として，更新対象の注釈識別子と注釈文，もしくは領域を与える．入力された注釈識別子に従って，更新対象の注釈と対応表を探しだし値を置き換える．

(2) 文献データの切り抜き

切り抜き操作は，元の文献画像の部分画像と関連する注釈を取り出し新たな文献データを生成する操作である．この操作は，歴史文献から利用者が注目する部分だけを収集するための操作であり，得られるデータを切り抜きデータと呼ぶ．切り抜きデータは，元の文献データの URI を持つので，二つのデータ間の参照関係は記録される．この処理は，文献データから指定された情報を抽出する処理と，その情報に基づいて新たな文献データを生成する処理からなる．

(a) 情報の抽出
  入力として文献画像から切り抜く領域を指定する．この領域に従って文

献データから以下の情報の抽出を行う．

(i) 文献画像から取り出した部分画像

(ii) 対応表から指定された領域に内包される領域情報と注釈の識別子

(iii) 先に取り出された識別子を持つ注釈文

(b) 文献データの生成

空の文献データを作成し，先の操作で得た情報を文献データに登録する．切り抜き元の文献データの URI を切り抜きデータの文献情報の属性 Reference に登録する．

この操作には，切り抜き後の元の文献データの編集結果を切り抜きデータに反映させないか反映させるかの違いから，静的な切り抜きと動的な切り抜きが考えられる．これらの処理は，それぞれ切り抜きデータを物理的に別のデータとして作成する処理と，データの閲覧の度に新たな切り抜きデータを生成する処理によって実現できる．

(3) 分類操作

ここでは，電子スクラップブックデータの分類操作として文献データの分類を管理するグループの分割処理と結合処理について述べる．

(a) グループの分割

この処理では，入力として分割対象のグループと，移動させる文献データの URI の集合が与えられる．分割処理は，空のグループを電子スクラップブックデータに追加した上で，指定された文献データの URI の集合を移動させる処理である．

(b) グループの結合

この処理は，入力として与えられた結合元のグループと結合先のグループの組に対して，(i) 結合元のグループに登録されている文献データをすべて結合先のグループに移動させ，(ii) 空になった結合元のグループを削除する処理である．

(4) 文献データ，電子スクラップブックデータの共有操作

利用者が作成する文献データや電子スクラップブックデータは，他の利用者への歴史文献の読解支援や研究成果の公開に利用される．従って，これらのデータを利用者間で共有するための機能が電子スクラップブックシステムには必要である．文献データなどは，XML で記述されたテキストデータであるため利用者間で容易に交換し閲覧および編集することができる．電子スクラップブックシステムは，文献データなどを交換を支援するために，サーバによるデータの一時的な保存機能と，利用者側へのデータの送信機能を持つ．そのための操作は，次の通りである．

(a) データのサーバへの登録操作
この操作は，電子スクラップブックシステムのサーバに利用者が作成した文献データもしくは電子スクラップブックデータを登録する操作である．サーバは，受け取ったデータをファイルとして保存すると同時に，検索などのために注釈等のテキストデータをデータベースに登録する．

(b) サーバからのデータの読みとり操作
この操作によって利用者はサーバから目的のデータを取り出す．取り出すデータの指定は，電子スクラップブックサーバと対象データのファイル名から成る URI を用いる．

この操作は，文献データや電子スクラップブックデータをサーバに登録した時点で，データの制作者以外の利用者による閲覧と注釈の追加を許す．

5.4 テキストデータを用いた文献データ，電子スクラップブックデータの検索

提案モデルに従ったデータに対する検索処理について述べる．文献データの検索は，文献データの注釈などのテキストデータを対象にした文字列一致である．

5.4.1 検索処理

本節では，まず，外字を含まないテキストデータを対象にした文献データ検索の処理について述べる．

(1) 検索条件として検索対象の要素と検索キーである文字列の組の集合を与える．

(2) 各組で指定された属性の値と検索キーが一致するかどうかを検査する．この検査は，データベースへの問合せという形式で処理する．問合せ処理は次節で述べる．一致するデータがあれば，そのデータの URI を一時的な検索結果として取り出す．

(3) すべての組に対して (2) を繰り返し，各検査結果で得られる URI を集計する．

(4) 集計結果で上位になった URI を文献データ検索の結果として利用者に返す．

次に，外字を含むテキストデータの文字列一致について述べる．ここでは文献データの本文を対象に考える．この処理で与える文字列は，直接，読みなどの外字情報を記述したものを考える．

(1) 与えられた文字列から外字情報部分を抜き取る．

(2) 残った部分による部分一致を行い一致する文献データを取り出す．また，本文の値から一致した部分を抜き出す．

(3) 取り出した文献データの注釈要素に対して，(1) で抜き出した外字情報を用いた検査を行う．

(4) 外字情報と一致したものがあった場合，一致した注釈の識別子を取り出す．

(5) 最後に，(4) で取り出した識別子が，(2) で取り出した文字列に含まれるかを検査し，含まれているものを検索結果として取り出す．

最後に，電子スクラップブックデータの検索について述べる．電子スクラップブックデータは，文献データを URI によって間接的に管理しているため，直接，文献データを含む電子スクラップブックデータを検索することはできない．そこで，電子スクラップブックデータの検索は，次のように処理する．

(1) 文献データの検索を行い，該当する文献データの URI を取り出す．

(2) 取り出した文献データの URI を検索キーとして，電子スクラップブックデータのグループの検索を行う．

(3) 検索条件を満たした電子スクラップブックデータを結果として取り出す．

5.4.2 問合せ処理

文献データの検索で関係データベースを利用するために検索条件から SQL 文を生成する．検索条件から SQL を生成するために取り出す値は，検索対象の要素名と，その要素の値と文字列一致をするための文字列である．問い合わせの結果は，条件を満たした要素を含むデータの URI の集合である．与えられた要素名は，関係の表名と属性を連結したものである．本論文では，検索条件に要素名から得られる表名が同じであり属性名の異なる条件が複数あれば，AND 検索として，以下のような SQL 文を生成する．

SELECT Identifier
FROM 表名
WHERE 属性名 *1* LIKE ’ 値 *1*’ AND
属性名 *2* LIKE ’ 値 *2*’ AND ...

上記の SQL 文の表名には，与えられた要素のパスとデータベースにある表名を比較し一致した部分を用いる．一方，属性名は一致しなかった部分を用いる．SELECT 句の Identifier は，文献情報 の属性 Identifier の外部キーであり，文献データの URI である．

検索条件に含まれる対象の要素の中に同じ要素を指定したものが複数ある場合は，OR 検索として処理する．本論文の検索では，問合せで得られる URI を集計し，その集計に基づいて利用者に示す検索結果を作成するので，一つの SQL 文にまとめるのではなく，要素の個数だけ SQL 文を生成する．生成する SQL 文の構造は先に示したものを利用する．

5.4.3 異体字検索について

歴史文献には，異体字を用いた同じ語の記述が多数含まれる．また，歴史文献の記述に対応したテキストデータは，処理を容易にするために外字として表現するべき文字を同じ意味を持つ他の符号化文字に置き換えて記述することがある．従って，歴史文献の検索では，同じ語であっても複数の記述が存在する．本節では，このような記述の違いを利用者が意識せずに行う検索を異体字検索と呼ぶ．異体字や漢字の置き換えによる記述の読みは，一般に同じである．一方，同意語は，同じ意味を異なる記述で表現するため，語ごとに読みが異なる．例えば，読みが “いち” である異体字を用いた記述は,「一」や「壱」などが挙げられる．従って，本節で扱う異体字検索では，記述ごとに読みの異なる同意語を検索の対象としない．

異体字検索では，検索条件に従って SQL 文を生成する際に，まず，利用者から与えられた文字列と関連する異体字を用いた文字列を生成する．次に生成された文字列のいずれかと一致する文献データを取り出すように OR 演算子で連結した SQL 文を生成する．例えば，利用者から「一」という文字が検索条件として与えられた場合，異体字検索では「一」の異体字である「壱」も検索文字列であるため，SQL 文は,「SELECT * FROM table_a WHERE value Like ’%一%’ OR value Like ’%壱%’ 」と生成される．このような検索には文字の異体字関連を表現した辞書データが必要であり，辞書データの作成では手作業の他に ekanji や 今昔文字鏡などのインターネットで公開されている大規模漢字集合の文字属性情報の利用が考えられる．

5.5 文献データの記述に関する時空間情報を用いた検索

本論文では，歴史文献の注釈を研究者によるコメントの他に文字属性，単語の意味，関連文献への参照など様々な種類の情報の集約に利用している．本章では，特に文献の分類や分析に利用される情報の一つである歴史文献の時間および地理に関する注釈の利用について考察する．まず歴史文献の記述にある時間的および地理的な情報の特徴について考察する．次に考察に基づいて管理する時間情報，地理情報のデータ構造と，それぞれの情報の集合から目的の情報を取り出すための時間関連，空間関連を定義し，関連の判定処理について述べる．さらに，一般に注釈として得られる情報は，すべての項目が与えられるとは限らないため，関連する情報から欠落している値の補完について考察する．

5.5.1 歴史文献の時間的，地理的な情報の特徴

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) 同一地域でも年代などによって地名が異なる．
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ことなどが挙げられる．

(2) 期間や領域の範囲があいまい
歴史的な時間情報は，人物の生没年が不明など，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，境界があいまいであることが多い．

(3) 意見の異なる情報の存在
歴史的な地理情報では，研究者の意見の相違によって地名と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) 注釈として与えられる情報が不完全.

本論文で収集する時間情報や地理情報は，電子スクラップブックシステムによって利用者が作成した情報を利用するので，必ずしもすべての値が与えられているとは限らない．つまり，地名だけが与えられ対応する領域情報が与えられない場合などが挙げられる．

このような特徴をもつ時間情報および地理情報に対する問合せとして時区間や領域の比較による検索の他に，“16 世紀の東京に関する歴史文献を取り出せ” のように年代に関係ない地名などの検索や，“江戸時代の出雲に関する記述のある文献を取り出せ” というような年号に依存しない時代での検索など，時代名や地域名の関連に基づいた検索を処理する必要があると考えられる．

5.5.2 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を定義する．

1. 時間情報

   本章では時区間を 1 次元の数直線上の始点と終点の組として表現する．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が同じ値を持つ．年号と時区間の関係は，ある人物の生没年を表す時区間が，江戸，明治，大正と 3 つの時代が関連するように，一つの時区間が複数の時代と関連することが多い．従って，本章で記述する時間情報は，一つの時区間と複数の時代名の組で表現される．また，歴史的な人物の生没を表す時区間の中には，始点や終点が不明であったりあいまいであることがあるので，時区間の始点と終点それぞれに対して取り得る範囲の記述を許す．従って，本章で記述する時間情報は以下のような情報の組として記述する．

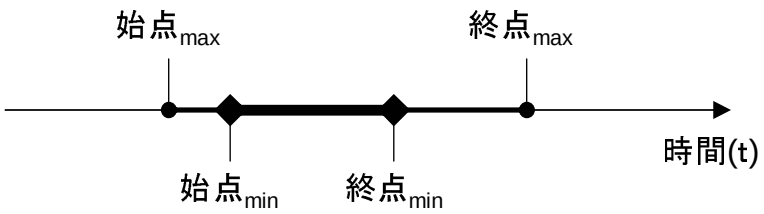

図 5.1 時区間

定義 5.1 時間情報
時間情報は，他の時間情報と区別するための識別子 TID，参照，時区間，時代の集合の 3 要素の組で表現する．時区間は始点と終点の組で表現するが，あいまいな時区間を表現するために取り得る範囲を表す二つの時区間で記述する．時区間の記述について (始点$_{min}$，終点$_{min}$) は，(始点$_{max}$，終点$_{max}$) に内包されていなければならない (図 5.1)．時代名は，時区間に関連のある時代を年号と年 (数値) の組で示す．年の値を 0 とした場合，年号に記された年代全体を指すものとする．

時間情報 = (TID，時区間，時代)
時区間 = (始点$_{max}$，始点$_{min}$，終点$_{max}$，終点$_{min}$)
時代 = (年号，年)

2. 地理情報

歴史文献の地理情報は，地名，領域，中心の組として表現する．歴史文献で示される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の範囲と最大の範囲を記述することを許す．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を列挙することを許す．中心は，二つの領域の距離や

方向の計算に用いられる．従って，本章では以下の構造で歴史文献の地理情報を記述する．

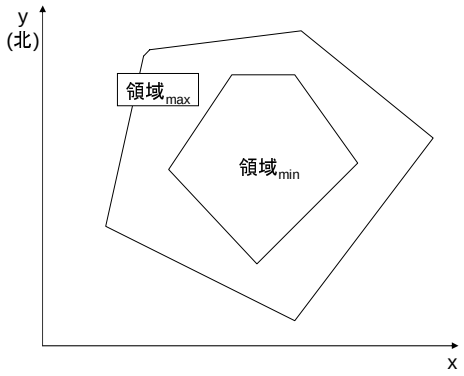

図 5.2 領域

定義 5.2 地理情報
地理情報は，識別子である GID，領域，中心，地名の組として表現される．領域は，領域がとりうる最小の範囲と最大の範囲を記述する．それぞれの範囲は多角形で表現され，領域$_{min}$ は，領域$_{max}$ に内包されていなければならない (図 5.2) ．中心は，与えられた領域の最小範囲から得られる重心とする．地名は，歴史文献に直接関連する地名を記述する文献地名と領域と関連のある関連地名の集合からなる．

地理情報 = (GID，領域，中心，地名)
領域 = (領域$_{max}$，領域$_{min}$)
地名 = (文献地名，関連地名)

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 5.3 時空間情報
時空間情報は，関連のある文献データへの参照，時間情報の識別子 TID，地理情報の識別子 GID の組として表現される．

時空間情報 = (参照，TID，GID)

5.5.3 時区間関連，領域関連

本節では，二つの時区間と領域の関連を定義する．ここで扱う時区間や領域は，期間や大きさがあいまいであるので，2 時区間や 2 領域が接している関連を計ることは困難であると考えられる．そこで，本論文で以下に定義する関連では，2 時区間などの接することを表す関連の定義を省略する．

(1) 時区間関連

二つの時区間 A(s_{max}, s_{min}, e_{max}, e_{min}), B(s_{max}, s_{min}, e_{max}, e_{min}) 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並びを表現する順序関連がある．また，条件から (s_{max}, e_{max}) と (s_{min}, e_{min}) は，以下の関係にある．

時区間 $\mathrm{T} = \{\ s_{max}, s_{min}, e_{max}, e_{min} \mid s_{max} \leq s_{min} \leq e_{min} \leq e_{max}\}$

(1) 位相関連

二つの時区間を表す位相関連として本章では以下の関連を定義する (図 5.3)．

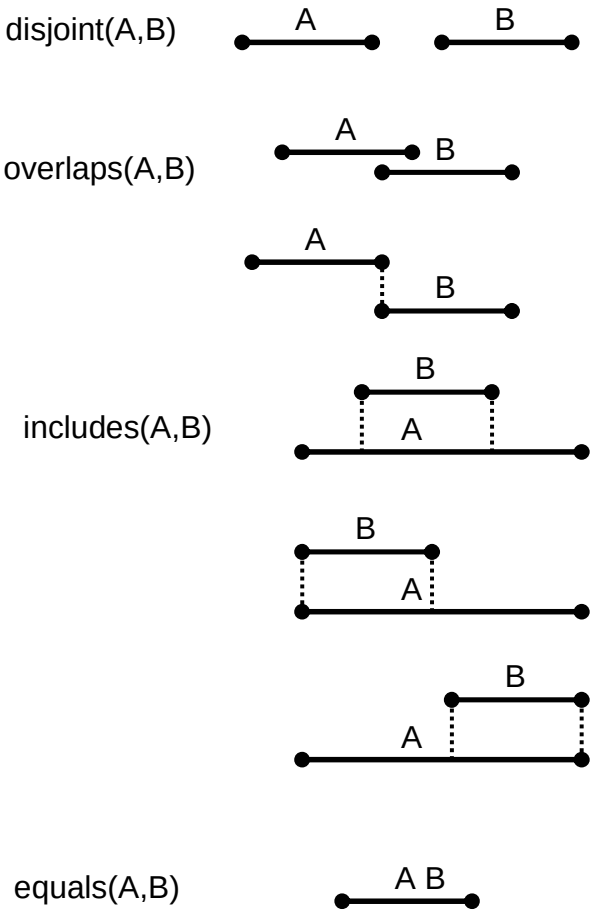

図 5.3 二つの時間区間の位相関連

– **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を内包していない関連を表す.

disjoint(A,B) iff $A.e_{max} < B.s_{max}$ or $B.e_{max} < A.s_{max}$

– **overlaps 関連**
一方の時区間の始点もしくは終点のどちらか一方の点が他方の時区間に接しているか内包されていることを表す関連である．本章で

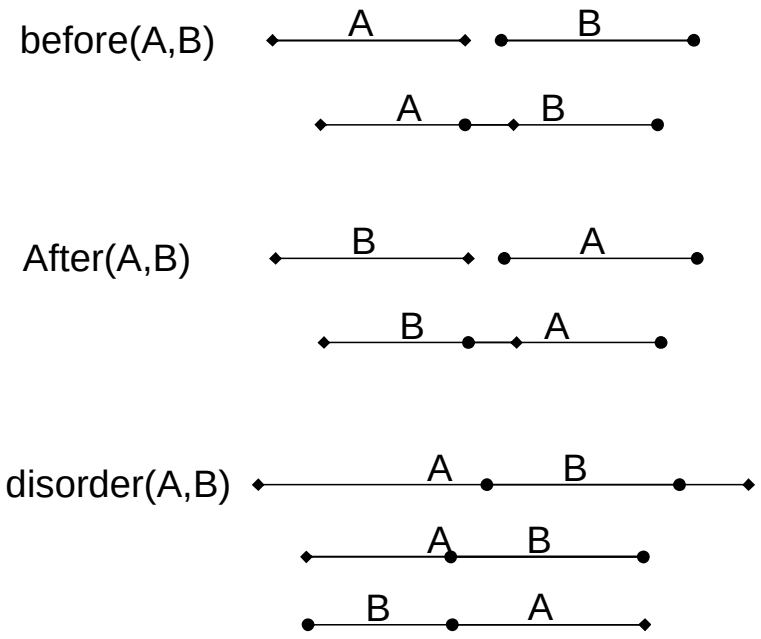

図 5.4 二つの時間区間の順序関連

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を区別しない.

overlaps(A,B) iff $(A.s_{max} \leq B.s_{max} \leq A.e_{max} < B.e_{max})$ or ($B.s_{max} \leq A.s_{max} \leq B.e_{max} < A.e_{max}$)

– **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.

includes(A,B) iff ($A.s_{max} \leq B.s_{max} \leq B.e_{max} \leq A.e_{max}$) or ($B.s_{max} \leq A.s_{max} \leq A.e_{max} \leq B.e_{max}$)

– **equals 関連**
二つの時区間が同一であることを表す関連であり，本章では同じ時

間情報を参照している場合にだけ成り立つ関連である.

equals(A,B) iff $A.s_{max} = B.s_{max}$ and $A.s_{min} = B.s_{min}$ and $A.e_{min} = B.e_{min}$ and $A.e_{max} = B.e_{max}$

(2) 順序関連

二つの時区間 A,B の順序関連として以下の関連を定義する (図 5.4).

- **before 関連**

  時区間 A の少なくとも一つの一端が, 時区間 B の前にある関連を表す.

  before (A, B) iff ($A.e_{max} \leq B.s_{max}$) or ($A.s_{max} < B.s_{max} \leq A.e_{max} < B.e_{max}$)

- **after 関連**

  before 関連の逆の順序である時区間の関連である.

  after (A, B) iff ($B.e_{max} \leq A.s_{max}$) or ($B.s_{max} < A.s_{max} \leq B.e_{max} < A.e_{max}$)

  disorder 関連

  一方の時区間が他方の時区間を内包した場合の関連である.

  disorder (A,B) iff ($A.s_{max} \leq B.s_{max} \leq B.e_{max} \leq A.e_{max}$) or ($B.s_{max} \leq A.s_{max} \leq A.e_{max} \leq B.e_{max}$)

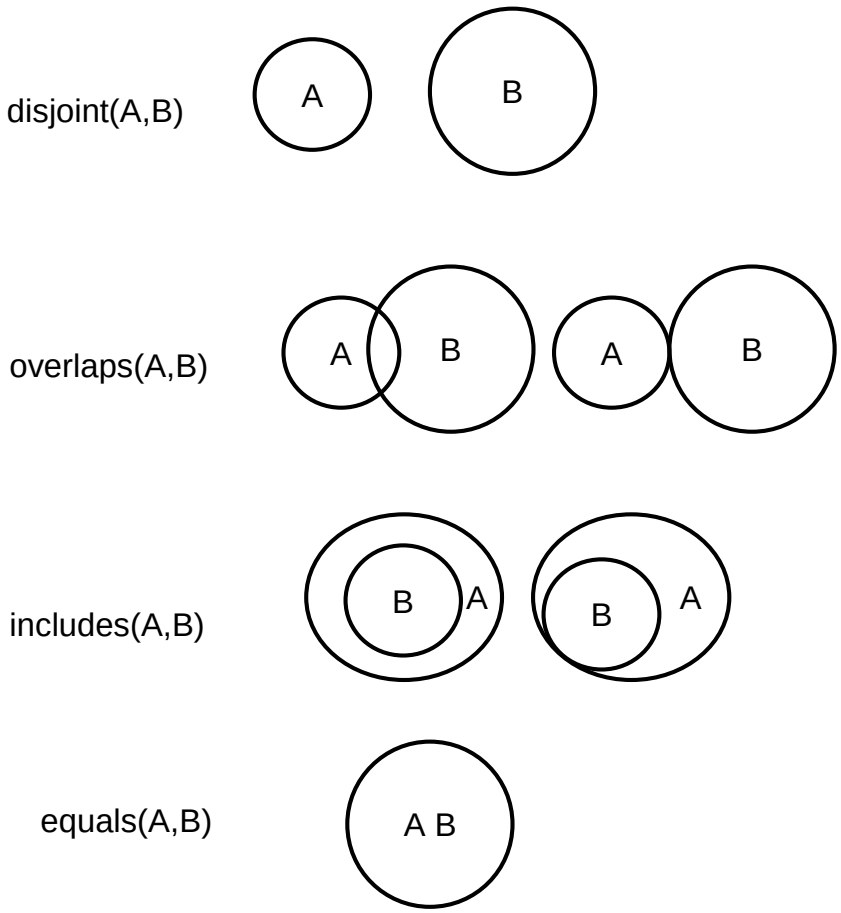

図 5.5 二つの領域間に関する位相関連

(2) 領域関連

二つの領域のもつ関連として本章では，位相，方向，距離の 3 種類の関連を定義する．

(1) 位相関連

位相関連とは，二つの領域の方向や距離に依存しない関連である．本章では以下の 4 関連を定義する (図 5.5).

- **disjoint 関連**

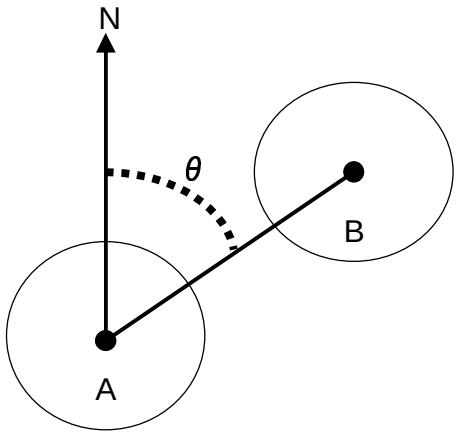

図 5.6 二つの領域間に関する方向関連と距離関連

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.

- **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．本章では，領域情報の定義から境界の接触と領域が重なっていることを区別しない.
- **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を表す関連である.
- **equals 関連**
  二つの領域が同一であることを表す関連である．本章で扱う領域は，境界があいまいなので，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.

(2) 方向関連

二つの領域の方向に関する関連は，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) に延ばした直線から時計回りにそ

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で表現する (図 5.6)．本章では，二つの領域の方向関連を表す関数として 以下の関数を定義する．

定義 5.4 direction 関数
direction 関数は，二つの地理情報 A，B について，領域 A の中心から y 軸正の向きに延ばした直線と線分 AB の角度の大きさを返す．また，k は角度の大きさ表し，0 度から 359 度の間の整数値をとる．

direction(A,B) = k ($0 \leq k \leq 359$)

(3) 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表現する (図 5.6)．

定義 5.5 distance 関数
distance 関数は，二つの領域 A,B の中心を結ぶ線分の長さを返す関数である．d は，長さを表す．

distance(A,B) = d

5.5.4 1 次元の時区間関連を用いた 2 次元領域の位相関連と方向関連の判定

本章で扱う地理情報の領域は多角形で表現される．多角形のまま領域の関連を検査する場合，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．

形状の複雑な領域や領域の数が多い場合，計算量が指数関数的に増加するため，効率的な関連の検査は困難である．そこで，本章では，詳細な関連の検査の前に，領域の近似形状を用いて検査対象を絞り込むことを考える．

(1) 領域の近似

本章では，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 5.7).

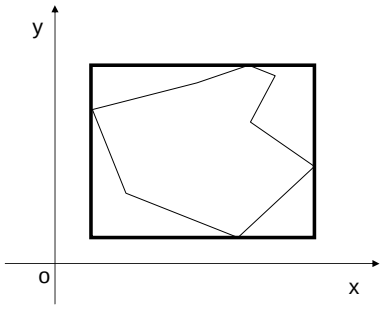

図 5.7 最小領域矩形を用いた領域の簡略化

(2) 位相関連の判定

時区間関連は，1 次元の数直線上の二つの区間の関連の記述に利用できる．そこで，本節では，2MBR の位相関連の判定に MBR を x 軸と y 軸に射影して得られる区間の位相関連の組を用いて判定する．

- **disjoint 関連の判定**
  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．
- **overlaps 関連の判定**
  上記の条件を満たさず x 軸，y 軸のどちらか一方の線分の関連が overlaps 関連である場合，MBR は overlaps 関連である．

- **includes 関連の判定**
  x 軸，y 軸の両方の線分の関連が includes 関連の場合，二つの MBR は includes 関連である．

- **equals 関連の判定**
  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 overlaps 関連や includes 関連の関連を得た場合，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，ここでは includes 関連として扱う．

(3) 方向関連の判定

多数の領域から目的の方角にある領域を選択することを考える．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と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で表現する．一般に方向関連の検査では条件を満たさない領域の方が多いと考えられるため，検索対象のすべての領域に対して角度を計算することは必ずしも効率的ではない．そこで，ここでは，多数の領域の方向関連を効率的に検査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 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の領域を絞り込む．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の検索対象の絞りこみを考える (図 5.8)．方向関連の検査対象は，領域 A の中心から y 軸正の向きの線分からを時計回りに，線分 A A’ までの領域である．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，それぞれの軸に対して，領域 A の中心の x 座標値が after 関連かつ y 座標値も after 関連にある領域だけが検査対象であることを示している．図 5.8 では，A.x の座標値以上かつ A.y の座標値以上の条件をみたす領域が例の方向関連の条件を満たす可能性がある領域を表すので，領域 B，C，D が選択される．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 が選択できることがわかる．図 5.8 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大小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実現できると考えられる．

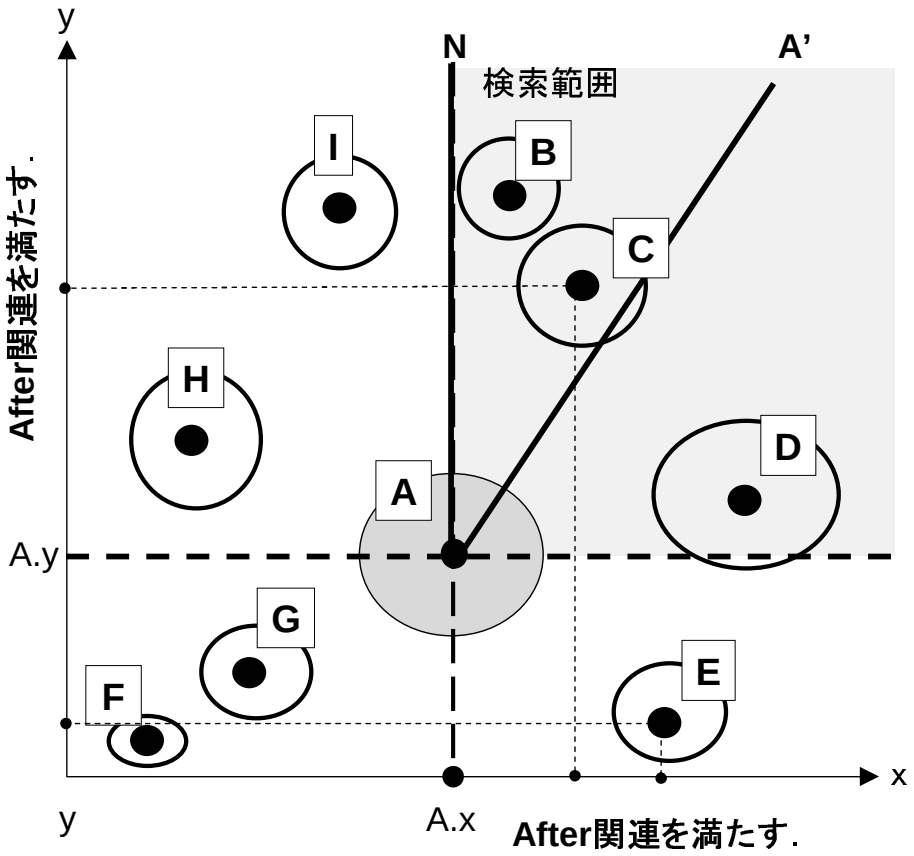

図 5.8 順序関連を用いた方向関連の検査対象の絞り込み

5.5.5 歴史文献の時空間情報の補完操作

歴史文献の注釈で与えられる時間情報や地理情報は，地名や時区間などの値が欠落している不完全な情報であることが多いと考えられる．また，地理情報への問合せでは，時間とともに変化する地名に関係ない問合せが考えられる．例えば，”平安時代の島根県に関する記述のある文献を表示せよ”などが挙げられる．そこで，効率的に問い合わせを処理するための前処理として関連のある時間情報や地理情報を分類し，欠落した情報を互いに補完することを考える．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が，本章では後者の方法だけについて述べる．

(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．

(a) 準備

時代要素が一つだけの時間情報を抜き出し分類する．分類した時間情報ごとに時区間を集約し，集約した時区間のすべてを内包する時区間を新たに作成する．作成した時区間と時代の値から一つの時区間と一つの年号を持つ時間情報を作成する．この時間情報をここでは基準時間情報と呼ぶ．

(b) 時区間だけが与えられている時間情報の補完

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，それぞれの時区間で includes 関連もしくは overlaps 関連がある時代要素の値を補完対象に写す．

(c) 時代だけが与えられている時間情報の補完

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．集め

た時間区間の組み合せの中で最長の時区間を補完先の時間情報に与える．ただし，収集した時区間の中で一つでも他のすべての時区間と disjoint 関連であった場合，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．

(2) 地理情報の補完

地理情報の補完は次の通りである．

(a) 準備

与えられた地理情報から領域と地名の両方の値があるものを抜き出し地名で分類する．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．分類された地理情報域から，ここの分類ごとに領域の論理和を取り領域を新たに作成する．このように作成した領域と地名を関係付けた地理情報を作成する．新規に作成した地理情報を基準地理情報と呼ぶ．また，地名で分類された地理情報に関連付けられている時間情報に基づいて時空間情報も作成する．

(b) 領域に基づく地名の関連付け

基準地理情報と対象の地理情報の領域を比較し，overlaps 関連もしくは，includes 関連にあれば，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基準地理情報の地名を与える．

(c) 領域情報だけの地理情報に対する地名の補完

こ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名を取り出す．地名が複数ある場合は，それぞれの地理情報に関連付けられている時間情報の比較を行い時区間が includes 関連にある地名を補完対象の地理情報の地名とする．ただし，複数の地名がある場合は，利用者が選択する．

(3) 検索

歴史文献に関する時間情報および地理情報に対する検索について述べる．本章で定義した時間情報および地理情報を対象とした検索式は，(1) 時区間関連と領域関連を用いた条件と，(2) 年号や地名を用いた条件を AND / OR 演算子で連結した形式で表現される．このような検索に対する結果は，歴史文献に関する情報を記述した文献データの URI の集合である．従って，検索処理は，与えられた個別の条件ごとにそれを満たした文献データの URI の集合を一時的な検索結果として保存し，各条件を接続する AND / OR 演算子に従って集合演算を行った結果を最終的な検索結果とする．

このような検索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は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，このような検索において利用者は，先に述べた検索条件と合わせて文献の時間的な分布を示すための時間的な間隔を指定する．このような間隔を指定した場合の検索は，(1) 検索条件を満たす文献データ集合を取り出し，(2) 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．従って検索結果は，指定された時間的間隔に従って複数のグループに分割された文献データの URI の集合である．このようなグループに分割された URI の集合の記述には，電子スクラップブックデータの利用が有効であると考えられる．

(4) 分析支援

歴史文献に関する地理上から得られる地理的な分布を視覚的に操作することは，歴史文献から得られる情報の分析に役立つと考えられる．このような分析の支援には，検索処理以外にも以下のような機能が必要であると考えられる．

(a) プレゼンテーション機能
文献データの種類や分布を地図上のアイコンなどを用いて視覚的に表現する機能である．

(b) レイヤー操作機能

この操作は，検索などによって異なる視点や条件に基づいて作成された複数の地図を重ね合わせて新たな地図を生成する操作である．この操作は視覚的な情報の分析に役立つと考えられる．ここでは，それぞれの地図をレイヤーと呼び，次のような論理演算を再帰的に用いることで新たな地図を生成する．

(1) 積

利用者の指定した条件を満たすそれぞれの地図の領域の重なる部分だけを取り出し新たな地図を生成する操作である．

(2) 和

利用者の指定した条件を満たすそれぞれの地図の領域をすべて新たな地図に射影する操作である．

(3) 否定

利用者の指定した条件を満たさない領域を取り出し新たな地図を作製する操作である．

さらに，地図を用いた分析を行うためには，スケール変換などの画像処理機能の提供が必要であると考えられる．

5.6 考察

本節では，まず提案モデルの XML 文書としての記述と操作について考察する．歴史文献に対する注釈は長期に渡り追加や修正されることが多いと考えられるため，歴史文献に関するデータの記述は，システムに依存しない形式での記述が望まれる．文献データモデルでは，タグ付きのテキストとしてデータを記述する XML を用いるのでシステムに依存しない長期的なデータ管理が可能であると考えられる．XML 文書としてこのモデルに従ったデータを記述し

た場合，このデータは，木構造を用いてデータを表現することができる．従って，データの追加や削除，更新は木に対するノードの追加削除操作として実現できる．また，文献データを WWW ブラウザで閲覧する方法には，本論文で提案したもの以外にも XSLT などのデータ変換モデルを用いて HTML 文書を生成する方法などが考えられる．

次に文献データや電子スクラップブックデータの共有について考察する．本論文のデータの共有では，公開された文献データに対して利用者は無制限に修正などを行うことが可能である．歴史学の研究支援への利用を考慮した場合，注釈の作成者以外に修正を禁止するなどの権限の管理や，文献データに対する注釈の同時書き込みの制御などが必要であると考えられる．

また，本章では，提案モデルに従ったデータ管理に関係データベースを利用するために，文献データや電子スクラップブックデータに対応した関係表の作成について述べた．データ管理に関係データベースを利用することによって，文献データベースの効率的な検索の他に，先に述べた文献データなどの編集に関する権限管理やファイルへの同時書き込みの制御に関係データベースの権限管理やロック機能が利用できると考えられる．

さらに文献データの検索について考える．提案モデルのデータは XML 文書であるため，XML 文書に対する問合せ言語として提案されている XQuery [39] などの利用が考えられる．XQuery などの問合せ言語は，XML 文書集合から新たな XML 文書を生成する機能があるが，本論文の問合せは，検索条件にあった XML 文書全体を取り出すことであるため，XQuery より単純な機能で十分であると考えられる．従って，本論文では，検索条件から直接生成する SQL 文を用いて文献データの検索を行う．本章で行う注釈として与えられる歴史文献に関する地理情報の補完は，値の欠落を補うだけでなく予め地名の関連付けを行うため，領域関連を用いずとも時代に依存しない地理情報の検索が可能である．

第6章　電子スクラップブックシステム

本章では，利用者による歴史文献に対する注釈の編集などを支援する電子スクラップブックシステムのプロトタイプの実装と実行例を示す.

6.1　構成

本プロトタイプシステムは，利用者間で歴史文献の注釈や文献の分類を共有するために，一般的なクライアントサーバ形式のシステムとして構築されている．また，注釈を付ける対象はインターネット上の歴史文献であることから，クライアントサーバ間の通信プロトコルは，HTTP を利用する．プロトタイプシステムは，文献データの編集や閲覧を支援するクライアントと，文献データと電子スクラップブックデータを利用者間で共有するためのサーバからなる (図 6.1).

(1) クライアント

クライアントは，文献データ操作インターフェースと電子スクラップブックデータ操作インターフェースからなる (図 6.1).

(a) 文献データ操作インターフェース

このインターフェースは，文献データの閲覧，注釈の編集，切り抜き，文献データのサーバへの登録，サーバからの文献データの読み込み，切り抜き操作によって作成したデータを電子スクラップブックデータへの登録，検索条件の作成と発行操作を利用者に提供する．また，このインターフェースは，文献データのサーバへの保存とローカルディスクへの保存

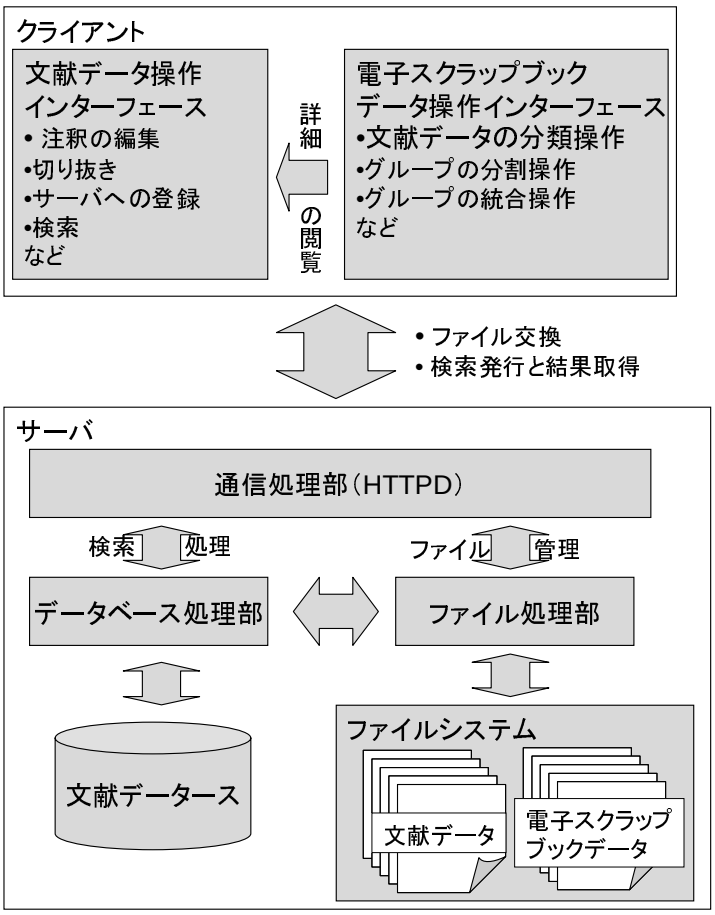

図 6.1 プロトタイプシステムの構成

を許す.

(b) 電子スクラップブックデータ操作インターフェース
このインターフェースは，電子スクラップブックデータの閲覧，分類，サーバへの作成した電子スクラップブックデータの登録と読み込み操作を利用者に提供する．このインターフェースは，文献データと同様に電子スクラップブックデータもサーバへの保存の他にローカルディスクへの保存を許す.

(2) サーバ

サーバは構成は次の通りである (図 6.1).

(a) 通信処理部

通信処理部は，クライアントからの要求に合わせてデータベース処理部やファイル処理部を起動し，各処理部からの結果をクライアントに送信する処理を行う．クライアントとの通信には HTTP を使用するので，一般的な WWW サーバが利用できる．

(b) データベース処理部

データベース処理部は，文献データ，電子スクラップブックデータのデータベースへの登録と検索条件から文献データベースへの問合せの生成と発行を処理する．

(c) ファイル処理部

ファイル処理部は，クライアントから受信した文献データと電子スクラップブックデータをファイルとして管理する．利用者からの各データの閲覧要求があった場合，文献データベースからデータを再構成するのではなくファイル処理部で管理しているファイルをクライアントへ返す．

(d) 文献データベース

文献データベースは，検索を効率的に処理するためにデータを管理する．本論文では，一般に広く利用されている関係データベースを用いる．

6.2 実装環境

プロトタイプシステムは，Java1.4 を使用して実装した．クライアントとサーバ間の通信処理は，Apache2.0 と Tomcat4.0 を利用し，関係データベースには MySQL4.0.4 を使用した．データベース処理部やファイル処理部は Servlet として実装し必要に応じて通信処理部の Tomcat4.0 から実行される．

プロトタイプシステムのクライアントとサーバは，同一の PC（OS：WindowsXP，CPU：Intel Pentium III 844MHz，メモリ：512Mbyte，ハードディ

スク：30Gbyte）上で実装している．ただし，クライアントとサーバは別のプロセスで実行し HTTP による通信をしているので，クライアントとサーバを異なる計算機上で実行することも可能である．プロトタイプシステムは，文献画像として玉川大学図書館で公開されている百人一首を利用した [40].

6.3 実行例

本プロトタイプシステムの実行例を示す.

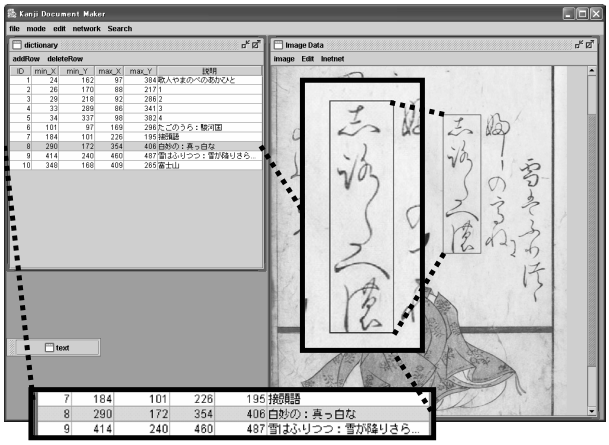

図 6.2 文献データの閲覧例

図 6.2 は文献データ操作インターフェースを用いた百人一首に関する文献データの表示例である．このインターフェースは，文献画像，書誌情報などを表示するウィンドウ群からなる．対応表と注釈は，注釈の識別子に基づいて一つの表にまとめている．このインターフェースは，文献閲覧モードと注釈編集モードがある．前者は，文献データの注釈編集を禁止し，切り抜き操作を許可するモードである．後者は，前者の逆のモードである．図 6.2 は，文献データの対応表の利用例として選択した注釈に関連付けられている文献画像の領域を視覚化している.

図 6.3 は，電子スクラップブック操作インターフェースの実行例として，百

図 6.3 電子スクラップブックデータの閲覧例

人一首から切り抜いた著者名を並べて表示している．図に示す電子スクラップブックデータは，一つのグループに三つの文献データが登録されており．インターフェースの下部にそのデータの URI を示し，上部に対応する各文献画像を表示している．このインターフェースでは，グループの追加，削除，文献データの移動，削除の機能を実装している．電子スクラップブックデータに登録された文献データの詳細の閲覧には，文献データ操作インターフェースを利用する (図 6.4) ．図 6.4 で示している文献データは，ある百人一首からの切り抜きデータである．図 6.4 の表部分は，文字の出現を記した注釈とそれに対応する文献画像の領域を示している．

利用者はこれらのインターフェースを用いて作成した文献データや電子スクラップブックデータをクライアントもしくはサーバのどちらかで保存する．サーバ側で保存した場合は，ネットワーク上にデータが公開されたものとして，文献データベースに登録する．

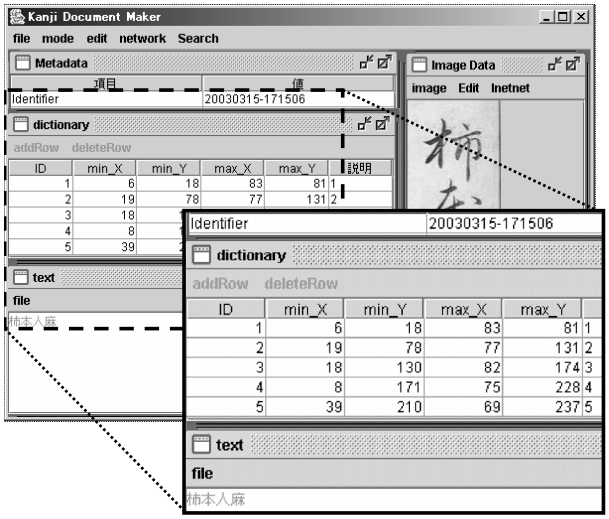

図 6.4 電子スクラップブックデータ内の文献データの詳細表示例

6.4 検索の実行例

本節では，プロトタイプシステムに実装した文献データの検索例を示す．図 6.5 は検索インターフェースであり，文献データ操作インターフェースの一部として実装した．このインターフェースを用いて，利用者は，検索対象の要素と検索キーとなる文字列を指定する．図 6.5 は，“歌人 ”もしくは “鳥 ”に関する注釈を含む文献データの検索例である．文字列に含まれる “% ”は 0 文字以上の任意の文字列と一致することを表す．作成した検索条件は，文献データモデルに従った XML 文書としてサーバに送信される．このインターフェースによって，利用者は，文献画像の注釈編集と同様の操作で検索条件を記述できる．

図 6.5 で与えられた検索条件は，同じ要素に対して二つの条件が示されているので，サーバでは次のような二つの SQL 文を生成する．問合せの結果は，与えられた文字列を注釈に含む文献データの URI の集合である．

(1) SELECT Identifier FROM ipan WHERE body LIKE ’歌人%’;

(2) SELECT Identifier FROM ipan WHERE body LIKE ’%鳥%’;

図 6.5 検索インターフェースの実行例

FROM 句の表名は，ルート要素からのパスを記述するが，ここでは簡略化のため付録に示した要素名を用いている．

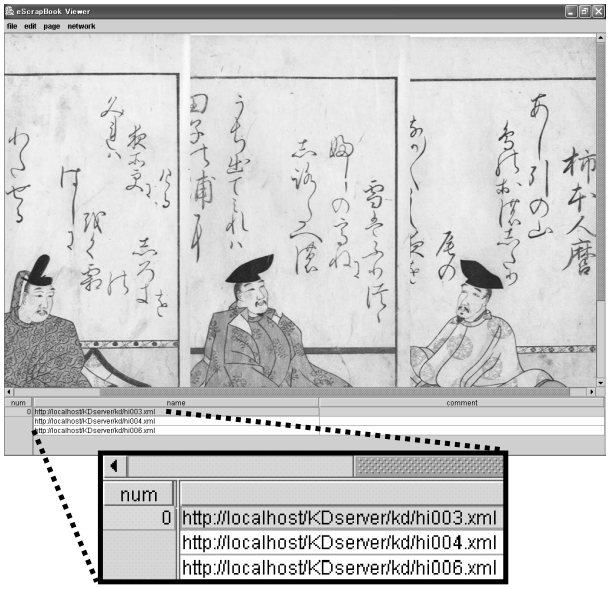

図 6.6 検索結果の表示例

サーバからクライアントに送信する検索結果の形式は，電子スクラップブックデータのモデルに従う．図 6.6 は，図 6.5 の検索結果を電子スクラップブック操作インターフェースで表示した例であり，検索結果である文献データ の URI が一つのグループに登録されている．検索結果は，新たな電子スクラップ

ブックデータであるので，利用者による分類や文献データの詳細の閲覧が可能である．文献データの詳細の閲覧は，文献データ操作インターフェースを用いるため，さらに注釈の追加や切り抜きなどの編集が可能である．

6.5 考察

プロトタイプシステムのクライアントは専用のインターフェースであるが，通信プロトコルに HTTP を用いるため，Web ページなどの外部リソースと文献データの統合が可能である．また，このインターフェースによって，利用者は提案モデルの書式を知らなくとも自由に注釈の編集や分類ができる．このような直観的なデータ操作は，資料収集や分類作業の効率化に役立つと考えられる．

奈良文化財研究所の木簡データベースは，既存の学術情報をインターネットで公開することが目的であるため，利用者による情報の追加などの機能はない．一方，本論文では，利用者間の知識共有が目的であるため，データの公開後の動的な情報の追加が可能である．また，プロトタイプシステムのクライアントは WWW 環境を利用していないが，Java で実装しているため，アプレットなどの WWW ブラウザのプラグインとして構築することも容易にできる．

利用者が検索条件として指定できる要素は，提案モデルで定義された要素だけである．従って，検索条件も文献データであり，検索処理は条件として与えた文献データと既存の文献データがどの程度一致するのかを検査する方法であると考えられる．また，文献データでは，文献画像中の文字の出現位置を注釈として記述できるため，注釈を用いた文字や語句の配置に着目した検索が可能であると考えられる．

第7章　WWW環境での文献データの閲覧

7.1　文献画像のレイアウトに基づいた釈文の表示

文献データの注釈の応用として，本章では歴史文献の記述に対応したテキスト，文献画像，文字位置の注釈と対応表を用いて，文献画像のレイアウトに基づいた本文データの表示について述べる．

木簡の記述には，文字の記述された位置が意味を持つことがある．例えば，荷札であれば，荷物の送り先と送り元の違いは，木簡上の記述位置で決まると考えられる．また，仏典には，仏教の思想を表すために文書を渦巻き状に記述しているものがある．従って，歴史文献の文書の2次元的な配置が重要な資料になることが多い．さらに，木簡などの歴史文献には墨のにじみなどにより文字の判別困難な文献も多いため，文献の画像としての表示に併せて，歴史文献の閲覧支援のために文献画像のレイアウトに基づいた釈文の表示は有効であると考えられる．さらに，草書体などの記述を楷書体で表示することで，一般的な利用者による文献の内容理解の支援に役立つと考えられる．

WWW 環境の上で歴史文献のレイアウトに従ってテキストデータを示すことができれば，一般的な利用者による歴史文献の利用効率の向上が期待できる．また，研究者間の研究成果などを効率的に公開する手段として，WWW 環境での表示は重要であると考えられる．

原資料のレイアウトに従った本文データを WWW で表示する方法として，SVG を用いる方法と，HTML 文書として再現する方法が挙げられる．本論文では，閲覧のために特別なソフトウェアの追加が必要ではなく，手作業による修正が容易な HTML 文書として表示することを考える．

そこで本研究では，文献画像のレイアウトに従って本文データの表示方法として以下の二つの方法が挙げられる．ここで記す釈文とは，歴史文献画像の記述をテキスト化したデータであり，文献データの本文の値である．

(1) 文献画像の文字とフォント画像の置き換え．
文献データの対応表に記述された領域と対応付けられた釈文のフォントに基づいて文字画像を作成し，文献画像に文字画像を上書きする方法である．この方法では，釈文を画像として表現するため，テキストの順序を表現することとはできないが，文献データとして再利用可能である．

(2) HTML 文書としての表現.
文献データの対応表の領域情報に従って文字ごとの位置の記述に CSS を利用することで文献のレイアウトを HTML 文書として再現する．文献のテキストを HTML 文書として表現するので，SVG のような表示領域の制限なしに画面サイズ以上の文献の再現が可能になると考えられる．

ここでは， WWW ブラウザでの閲覧に適した後者の方法について述べる．以下に処理の手順を示す.
(1) 準備
歴史文献のレイアウトに従った HTML 文書を作成するために文献データの (a) 文字位置情報に関する注釈，(b) 対応表，(c) 本文を用いる．

例 7.1 文献データの例
本文:(大花下)
注釈 (文字位置):
(注釈識別子：1, 開始位置:1，終了位置:1),
(注釈識別子：2，開始位置:3，終了位置:3),
(注釈識別子：3，開始位置:2，終了位置:2)
対応表:
(位置:(10,10,100,110), 注釈識別子:1),

(位置:(10,120,110,240), 注釈識別子:2),
(位置:(15,260,95,300), 注釈識別子:3)

(2) 文字情報の並べ替えと釈文，対応表との結合
一般に注釈に順序は存在しないため，注釈から文字位置情報を単に抜き出しただけでは釈文の文字の並びに従って HTML 文書を作成できるとは限らない．従って，注釈の文字位置情報の開始位置に関して並べ替えを行う．さらに注釈識別子を使い対応表と注釈を結合する．

例 7.2 文字情報と対応表，釈文の結合
文字 1:(位置:(10,10,100,110), 大)
文字 2:(位置:(10,120,110,240), 花)
文字 3:(位置:(15,260,95,300), 下)

(3) レイアウト情報の生成
歴史文献の画像のレイアウトに従った HTML 文書を作成するには，各文字を表示する位置と文字の大きさの情報が必要である．HTML 文書中の文字の位置は，対応表にある領域情報に示される長方形の左上の頂点の座標を用いる．歴史文献中の文字の大きさは一般に幅と高さが異なるが HTML ではフォントのサイズを幅，高さが異なる大きさの値を指定できない．そこで，ここでは，対応表の表示領域から得られる文字の高さと幅のうち短い方を選択する．ただし，注釈の文字位置情報が 2 文字以上の釈文に関係づけされている場合は，領域の高さを文字数で割った値を用いて比較する．

例 7.3 文字情報と対応表，釈文の結合
文字 1:(表示位置:(10,10), 文字サイズ:100, 大)
文字 2:(表示位置:(10,120), 文字サイズ:120, 花)
文字 3:(表示位置:(15,260), 文字サイズ:40, 下)

(4) HTML 文書の生成
HTML の DIV 要素と，CSS を用いて，文献データに含まれる文字位置情報ごとに先の操作で求めた文字の表示位置と文字の大きさを指定する .

例 7.4 HTML 文書としての記述

```
<div style="top:10;left:10;
font-size:100;position:absolute">大</div>
<div style="top:10;left:120;
font-size:120;position:absolute">花</div>
<div style="top:15;left:260;
font-size:40;position:absolute">下</div>
```

木簡の文献画像とその画像のレイアウトに従った釈文の表示例を図 7.1 に示す.

テキストデータに外字が含まれる場合は，以下のように文字画像を HTML 文書に埋め込む.

(1) 釈文で参照している文字属性情報を含む注釈から文字画像への参照を取り出す.

(2) IMG タグを用いて HTML 文書中に文字画像への参照を記述する.

また，このときの文字画像のサイズは，対応表に記載された表示領域の大きさに従う.

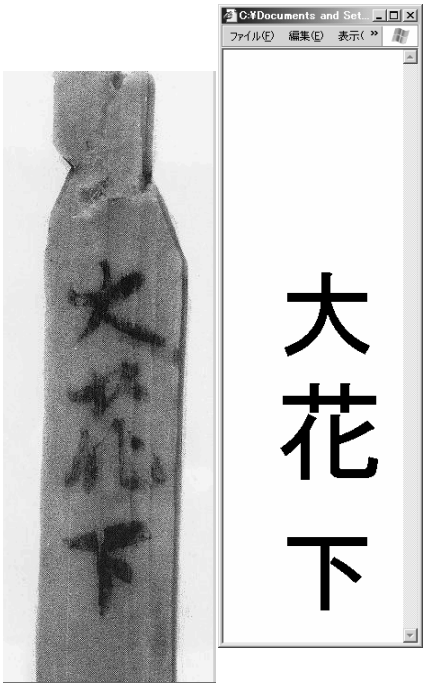

図 7.1 文献画像と注釈から生成した HTML 文書

7.2 注釈と関連付けた歴史文献画像の WWW 環境での表示

本節では，文献データを用いた歴史文献の画像に併せて関係づけられた注釈を WWW 環境で表示する方法について述べる．

歴史文献の注釈を利用者に示す方法には，以下の方法が挙げられる．

(a) 注釈の一覧を示し，注釈から関連づけられている歴史文献の画像領域を示す方法

(b) 歴史文献の画像領域から関連づけられている注釈を示す方法

歴史文献の内容理解の支援として注釈を利用する場合，後者の方法による注釈の提示の方が自然であると考えられる．本節では，文献画像から注釈を閲覧する方法について述べる．

WWW 環境の上で画像から関連づけられた注釈を表示する方法には，以下の方法が挙げられる．

(1) クリッカブルマップによる注釈へのハイパーリンクを利用する方法

(2) JavaScript などのスクリプト言語を利用する方法

一般に注釈が付けられる領域は重なるので，二つ以上の領域が重なる部分は関係する複数の注釈を一度に表示する必要がある．しかし，クリッカブルマップでは，一つの領域に複数のハイパーリンクを付けることはできないため前者の方法は利用できない．従って，JavaScript や VBScript などのスクリプト言語を用いて実現する．例として以下に示す注釈を付けた木簡の文献データを用いる．例で示す XML 文書は，本節と関連の有る部分を抜き出したものである．タグ名や構成は，付録 A に示す Relax スキーマに従う．im タグは文献画像，an タグは注釈，ct タグは対応表を表している．

例 7.5 文献データの記述例

```
1:    <im>http://foo.ac.jp/mokan20030826-1a.JPG</im>
2:    <an>
6:      <row><id>1</id><body>大</body></row>
4:      <row><id>2</id><body>花</body></row>
5:      <row><id>3</id><body>下</body></row>
6:      <row><id>4</id><body>大化５年２月...</body></row>
7:    </an>
8:    <ct>
9:      <ci><imp>61,226,186,334</imp><tp>1</tp></ci>
10:      <ci><imp>57,347,186,489</imp><tp>2</tp></ci>
11:      <ci><imp>64,519,169,613</imp><tp>3</tp></ci>
12:      <ci><imp>55,234,185,627</imp><tp>4</tp></ci>
13:    </ct>
```

(1) 文献データから HTML 文書への変換
文献データから (a) 注釈要素と，(b) 対応表要素のそれぞれの値をスクリプト言語の配列に変換する．また，HTML 文書で示す文献画像は，与えられる文献データの文献画像要素に示される URI から得ることができる．
(2) 注釈の表示処理
HTML 文書に変換した文献データを用いた注釈の表示の手順は次の通りである．

(a) 注釈閲覧のために利用者は歴史文献画像上の位置をマウスで指示する．

(b) 与えられたマウスの位置情報から対応表の領域情報と比較する．

(c) マウスの位置が対応表の領域情報に含まれる場合，その領域に関係づけられている注釈文を結果として保存する．

(d) すべての領域情報について (c) の操作を繰り返し，マウスの位置に関連あるすべての注釈文を取り出す．

(e) 操作 (c)〜(d) から得られた結果を別ウィンドウで表示する．

注釈は，XML 文書であるため HTML 文書への変換は XSLT などの変換モデルが利用できると考えられる．上記の処理に従った JavaScript の記述例の主要な部分を以下に示す．

例 7.6 文献画像と注釈の関係を表示するため JavaScript の記述例

```
<head><script language="JavaScript">
1:X1[0]=61; Y1[0]=226; X2[0]=186; Y2[0]=334;
2:X1[1]=57; Y1[1]=347; X2[1]=186; Y2[1]=489;
3:X1[2]=64; Y1[2]=519; X2[2]=169; Y2[2]=613;
4:X1[3]=55; Y1[3]=234; X2[3]=185; Y2[3]=627;
5:Comment[0]="大";Comment[1]="花";Comment[2]="下";
6:Comment[3]="大化５年２月-天智３年２月の冠位";
7:function mClick(){
8:  px = mousPositionX(); py = mousPositionY();
```

```
9:  for ( cnt = 0 ; cnt < 4 ; cnt++ ) {
10:  if(X1[cnt]<px&&X2[cnt]>px&&Y1[cnt]<py&&Y2[cnt]>py)
11:  buf = buf + Comment[cnt];
12:  }
13:  alert(buf);}</script></head>
14:<body>
15:<img src="http://hoo.ac.jp/img/mokan20030826-1a.JPG"
16:  id="b0" onClick="mClick()"/>
17:</body>
```

上記の記述例の 1 行目から 6 行目は，注釈と対応表から取り出した値をこのスクリプトで利用するために配列である．X1 と Y1 が歴史文献に関連付けている注釈の長方形の左上の頂点の座標値であり，X2，Y2 は，その直方体の右下の頂点の座標値である．Comment は，注釈文の本体である．7 行目から 13 行目までが，文献画像上でマウスをクリックした時に，関係のある注釈の表示処理を行う関数である．9 行目でマウスの位置を取得し，10 行目から 12 行目までがマウスで示した位置と注釈の関連づけられている長方形を比較し関連する注釈文を取り出す．13 行で取り出した注釈文を利用者に示す．図 7.2 に実行例を示す．図 7.2 は，文献画像上の “花” の上で注釈閲覧の要求を出した例であり，結果として，注釈が関連づけられているすべての注釈文を示している．

7.3 考察

文献データモデル等に従ったデータは，XML 文書であるため XSLT などのデータ変換モデルを利用することで HTML 文書などの文書に変換することができる．例えば，文献データを HTML 文書に変換した場合は，文献データを一般的な WWW ブラウザで閲覧することが可能であり，他の HTML 文書と統合し利用することが可能であると考えられる．

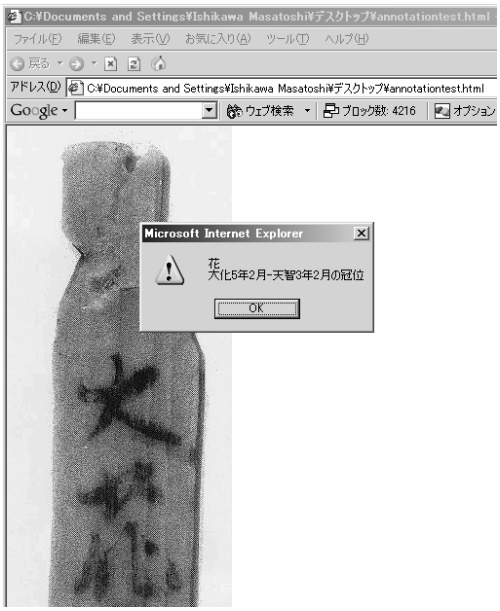

図 7.2 注釈を関連づけた歴史文献の WWW ブラウザの表示例

第8章　結論

本論文では，インターネット上で公開されている東アジア圏の歴史文献の画像と関連する書誌情報や注釈などを利用者間で集約し共有するためのデータのモデルを提案した．提案モデルは，文献データのモデルと電子スクラップブックデータのモデルとからなる．前者は，文献画像と注釈などのテキストを関連付けて管理する．後者は，互いに関連する文献データの分類を管理する．さらに文献データの注釈は，文字属性や単語の意味などの様々な種類の情報を扱う．そこで，様々な情報が記述される注釈を効率的に再利用するためのデータ構造を定義した．本論文では，提案モデルに従ったデータの記述に XML を利用するための Realx スキーマの設計方法について提案した．XML を利用することでシステムに依存しない形式でデータが記述されるため利用者間のデータ共有に役立つと考えられる．また，関係データベースを利用した検索を実現するための関係表の設計について述べた．提案モデルの操作として，注釈の編集，切り抜き，分類，共有，検索について述べた．編集操作は，歴史的文献に関連する情報の収集を支援するための操作であり，文献画像に対する注釈の追加，削除がある．切り抜き操作は，文献データの一部を利用して新たな文献データを作成する操作であり，分類操作は，利用者が収集した文献データを分類する操作である．共有操作は，サーバへのデータ登録とサーバからのデータの取り出し操作からなる．検索は，文献データや電子スクラップブックデータを複数の利用者間で共有するための機能である．文献データの検索は関係データベースを利用するので，本論文では検索条件から SQL の生成につて提案した．さらに歴史文献から得られる時間情報や地理情報に着目した文献データの検索方法を提案した．歴史的な時間情報や地理情報は，期間や境界があいまいであり値の欠落があるなどの特徴がある．そこで，本論文ではあいまいな地理情報など

を処理するための時区間関連と領域関連を定義し，関連の判定の処理について述べた．さらに，問合せを効率的に処理するために既存のデータから値を補完する方法を提案した．

次に提案モデルに基づいた電子スクラップブックシステムのプロトタイプを実装し，文献データの閲覧等の実行例と検索処理の実行例から提案モデルの有用性を示した．提案モデルを用いれば，分散して公開されている歴史文献に関する情報を利用者が独自に集約することが可能となる．さらに注釈を利用者間で共有することで，歴史文献の相互関連の発見などが容易になると考えられる．

注釈は，歴史文献に関連する知識を集約するために利用される．そこで，文献データに従って集約した情報を WWW 環境での閲覧支援として，文献のレイアウトに則したテキストデータの WWW 環境での表示について述べた．また，文献画像と注釈の関係を WWW 環境で表現する手法について提案した．

本論文の提案によって歴史文献に関連する情報の集約が容易になるので，歴史学や考古学での研究支援だけではなく，教育や電子図書館の個人化などの幅広い分野への応用利用が可能になると考えられる．

今後の課題は，本論文の提案した電子スクラップブックシステムを利用者に広く公開することで提案したデータモデルの有効性を評価する．また，文献データモデルに従って集約した知識の GIS や教育などでの応用について考察することとする．

謝辞

本研究を行うにあたり，適切なご指導ならびにきめ細かなご配慮を頂きました植村俊亮教授に心から深く感謝の意を表し，御礼申し上げます．植村俊亮先生には，ご多用，ご多忙中にもかかわらず本研究の指導教官になって頂き，本論文の草稿に対しまして，多数のコメントを頂きました．

本研究を行うにあたり，適切なご指導を頂きました伊藤実教授に心から深く感謝の意を表し，御礼申し上げます．伊藤実先生には，ご多用，ご多忙中にもかかわらず文献データの検索処理における関係データベースの必要性についてのコメントを頂きました．

本研究を行うにあたり，適切なご指導を頂きました大阪府立大学の宝珍輝尚教授に心から深く感謝の意を表し，御礼申し上げます．宝珍輝尚先生には，ご多用，ご多忙中にもかかわらず歴史学に関する研究活動と電子スクラップブックシステムにおけるデータ共有についてのコメントを頂きました．

本研究を行うにあたり，適切なご指導を頂きました名古屋大学の吉川正俊教授に心から深く感謝の意を表し，御礼申し上げます．吉川正俊先生には，ご多用，ご多忙中にもかかわらず文献データのモデルと電子スクラップブックデータのモデルに関してのコメントを頂きました．

本研究に対する有益なご助言およびご協力をいただいた奈良先端科学技術大学院大学情報科学研究情報生命学専攻バイオ情報学領域データベース学分野の宮崎純助教授，天笠俊之助手，波多野賢治助手ならび学生の皆様に深く感謝いたします．

島根県立大学の野中保雄教授，貴志俊彦助教授，井上治助教授，江口真理子助教授，末廣泰雄助手，村尾義和非常勤講師ならびメディアセンター職員の皆様には，本研究を行うにあたり様々なご配慮を頂きました．ここに深く感謝し

ます.

最後に，島根県立大学の勝村哲也教授には，本研究に関して様々なご助言とご支援を頂いていましたが，去る平成 15 年 9 月 10 日に急逝されました．深く感謝するとともに謹んで哀悼の意を表します.

業績リスト

論文

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. 石川正敏，波多野 賢治， 天笠 俊之，植村 俊亮: “XML を用いた歴史文献データモデルの応用”, 情報考古学, Vol. 9, No. 2, pp. 11 – 23, 2004 年 3 月.

国際会議 (査読有り)

1. Masatoshi Ishikawa, Kenji Hatano, Toshiyuki Amagasa, Shunsuke Uemura, and Tetsuya Katsumura: “A Data Model for Reconstructable Kanji Documents using XML”, IASTED International Conference on Information Systems and Databases (ISDB 2002), pp. 258 – 263, Tokyo, Japan, September 25 – 27, 2002.

国内発表

1. 植村俊亮, 小川政行, 石川正敏, 石川佳治, “映像メディアのモデルと乱呼出し”, 「高度データベース」諏訪ワークショップ講演論文集, pp. 52 – 57, 1996 年 7 月.

2. 石川正敏, 高倉弘喜, 植村俊亮, “動画像中のオブジェクトに注目したデータモデルと問合せ処理”, 情報処理学会全国大会, pp. 3-81 – 3-82, 1996 年 9 月.

3. 石川正敏, 高倉弘喜, 植村俊亮, “3 次元仮想空間を用いたコミュニケーション支援 ”, 電子情報通信学会技術研究報告, DE96-85, pp. 299 – 304, 1997 年 1 月.

4. 石川 正敏, 高倉 弘喜, 植村 俊亮: “領域定義による共有仮想空間での情報交換支援”, 夏のデータベースワークショップ’97 北海道, 情報処理学会研究報告 97-DBS-113, pp. 299 – 304, 札幌, 1997 年 7 月.

5. 石川正敏, 高倉弘喜, 植村俊亮: “領域定義による仮想空間での情報選択”, 第 55 回情報処理学会全国大会, pp. 4-101 – 4-102, 福岡, 平成 9 年 9 月 .

6. 石川正敏, 高倉弘喜, 植村俊亮: “利用者による複数共有仮想空間の空間選択手法”, 第 9 回データ工学ワークショップ (DEWS’98), 1998 年 3 月.

7. 石川 正敏, 高倉 弘喜, 植村 俊亮: “複数仮想環境の利用者による統合手法”, 第 3 回サイバースペースのためのデータベースワークショップ, 1998 年 5 月.

8. 石川正敏, 高倉弘喜, 植村俊亮: “利用者ごとの仮想空間の構築と共有”, 夏の DB ワークショップ’98 in 福井, 情報処理学会研究報告, 98-DBS-116-53, および平成 10 年度 科学研究費特定領域研究「高度データベース」福井ワークショップ講演論文集, pp. 107 – 114, 1998 年 7 月.

9. 石川正敏，天笠俊之，波多野賢治，植村俊亮：“仮想空間管理のための記述形式独立なデータモデル”，電子情報通信学会技術研究報告 MVE99-83, pp. 45 – 50, 2000 年 3 月.

10. 石川 正敏, 波多野 賢治, 天笠 俊之, 吉川 正俊, 植村俊亮, 勝村 哲也: “意味付き文字画像を用いた文献の電子化” 情報処理学会研究報告, Vol.2001, No.44, 2001-DBS-124-15/2001-FI-62-15, pp. 113 – 120, 2001 年 5 月.

11. 石川正敏, 波多野賢治, 天笠俊之, 吉川正俊, 植村俊亮, 勝村哲也: “XML を用いた再構築可能な漢字文献データモデル”, 電子情報通信学会 第 13 回データ工学ワークショップ (DEWS2002), 2002 年 3 月 4 – 6 日.

12. 石川 正敏，波多野 賢治，天笠 俊之，植村 俊亮，勝村 哲也: “XML を用いた木簡画像共有支援システム”, 日本情報考古学会第 16 回大会 発表要旨, pp. 91 – 96, 2003 年 9 月 20 – 21 日.

13. 石川正敏, 波多野賢治, 天笠俊之, 吉川正俊, 植村俊亮, 勝村哲也: “歴史文献のための電子スクラップブックシステムの設計”， 人文科学とコンピュータシンポジウムじんもんこん2003:-)．IPSJ Symposium Series Vol. 2003, No.21, pp. 227 – 234, 2003年12月.

14. 石川正敏, 波多野賢治, 天笠俊之, 植村俊亮, 勝村哲也: “歴史文献の時空間情報に関する問合せ処理”, 電子情報通信学会 第15回データ工学ワークショップ (DEWS2004), 2004年3月4 – 6日.

関連発表

1. 石川正敏: “Electronic Culture Atlas Initiative 参加報告”，島根県立大学メディアセンター年報，Vol. 2，pp. 92 – 99，2002年.

2. 石川正敏: “北東アジア歴史文献の共有利用を目指して”, 島根県立大学メディアセンター年報，Vol. 3，pp. 108 – 112，2003年.

3. 石川正敏，貴志俊彦，井上治: “北東アジア地域の社会科学研究のための資料・書誌情報データベースの紹介”，島根県立大学メディアセンター季刊報「界隈」Vol.13，pp. 6 – 7,2003年12月.

参考文献

[1] 大藏經テキストデータベース研究会: “大正新脩大藏經テキストデータベース”, http://www.l.u-tokyo.ac.jp/ sat/japan/, 1998.

[2] Academia Shinica Computing Centre: “漢籍電子文献”, http://www.sinica.edu.tw/ftms-bin/ftmsw3, 1997.

[3] 高科 智子, 北川 生馬, 永富 亘, 澁川 雅俊: “奈良絵本デジタルアーカイブ事例報告”, じんもんこん*:-)2002*, 人文科学とコンピュータシンポジウム, IPSJ Symposium Series Vol. 2002, No. 13, pp. 153 – 160, 平成 14 年 9 月. 20 – 22 日.

[4] Tetsuo Shoji, Yoshihiro Okada, Kogi Kudara, Takashi Irisawa and Yoshihisa Oda: “Development of 3-D Contents Management System for Digital Archives”, じんもんこん*:-)2002*, 人文科学とコンピュータシンポジウム, IPSJ Symposium Series Vol. 2002, No. 13, pp. 137 – 144, 平成 14 年 9 月. 20 – 22 日.

[5] 独立行政法人国立公文書館アジア歴史資料センター: “アジア歴史資料センター”, http://www.jacar.go.jp/, 2001.

[6] International Dunhuang Project:“International Dunhuang Project”, http://idp.bl.uk/, 1994.

[7] 奈良文化財研究所: “木簡データベース”, http://acd.nabunken.jp/Open/mokkan/mokkan1.html, 1992.

[8] 柴山 守, 吉井 良邦, ベンガッシュ・ラガワン et al. : “ 近世史料アーカイブズのためのバーチャル図書館”, じんもんこん*:-)2001*, 人文科学とコンピュータシンポジウム, IPSJ Symposium Series Vol. 2001, No. 18, pp. 109 – 116, 平成 13 年 12 月. 14 – 15 日.

[9] 桶谷 猪久夫, 才藤 千津子, Delmer Brown: “簡易型タグを利用した歴史史料の英日全文連携検索システムの設計と開発 - 古事記 日本書紀における事例 -”, じんもんこん*:-)2001*, 人文科学とコンピュータシンポジウム, IPSJ Symposium Series Vol. 2001, No. 18, pp. 65 – 72, 平成 13 年 12 月. 14 – 15 日.

[10] 原正一郎:“人文科学研究支援コラボレーション機能に関する実証的研究～国文学研究資料館の資料・史料を主な対象として～”，平成 12 年度～平成 14 年度科学研究費補助金基盤研究 (B)(2) 展開研究成果報告書，平成 15 年 3 月.

[11] Ian Johnson, Andrew Wilson:“The TimeMap Project: Developing Time-Based GIS Display for Cultural Data”, Journal of GIS in Archaeology Vol 1. ESRI Inc., Redlands, 2002.

[12] Masatoshi Ishikawa, Kenji Hatano, Toshiyuki Amagasa et al. : “A Data Model for Reconstructable Kanji Documents Using XML”, IASTED International Conference on Information Systems and Databases (ISDB 2002), pp. 258 – 263, Tokyo, Japan, Sep. 25 – 27, 2002.

[13] 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.

[14] Tim Bray, Jean Paoli, C. M. Sperberg-McQueen et al.: “Extensible Markup Language (XML) 1.0 (Second Edition)”, http://www.w3.org/TR/2000/REC-xml-20001006, 6 October 2000.

[15] Adobe Systems Incorporated: “PDF Reference third edition Adobe Portable Document Format Version 1.4”, http://partners.adobe.com/asn/developer/acrosdk/docs/filefmtspecs/PDFReference.pdf, 1999.

[16] The Open eBook Forum: “The OeBF Publication Structure 1.0.1 Recommended Specification”, http://www.openebook.org/oebps/oebps1.0.1/download/, 2001.

[17] Morgan N. Price, Bill N. Schilit, and Gene Golovchinsky: ”XLibris: The Active Reading Machine ”, CHI 98, ACM Press, pp. 22 – 23, Los Angeles, CA, April 18 – 23, 1998.

[18] Shinichi Ueshima, Kazuhiro Ohtsuki, Jun-ya Morishita et al. : “Incremental Data Organization for Ancient Document Database”, *Proceeding of the Fourth International Conference on Database Systems for Advanced Applications (DASFAA'95)*, pp. 457 – 466, Singapore, April, 10 – 13, 1995.

[19] 師 茂樹:“SVG を用いた『華嚴一乘法界圖』の表現実験”, http://www.l.u-tokyo.ac.jp/ sat/japan/tech/experimental/svg.html, Nov. 2000.

[20] J. F. Allen: “Maintaining Knowledge about Temporal Intervals”, Comm. of the ACM, pp. 832 – 843, Nov. 1983.

[21] Yoshifumi Masunaga: “A Temporal Expansion to the Multimedia Object Model in OMEGA”, Proceedings of 4th International Conference on Database Systems for Advanced Applications(DASFAA'95), pp. 430 – 440, 1995.

[22] Yoshifumi Masunaga: “Temporal Multimedia Data Modeling in OMEGA”, Proceedings of International Symposium on Advanced Database Technologies and Their Integration (ADIT'94), pp. 190 – 199, Oct. 1994.

[23] M. J. Egenhofer, “Point-Set topological relations”, International Journal of Geographical Information Systems, Vol. 5, No. 2, pp. 161 – 174, Taylor&Francis, 1991.

[24] Erlend Tossebro, Mads Nygard: “Uncertainty in Spatiotemporal Databases”, ADVIS 2002, LNCS2457, pp. 43 – 53, 2002.

[25] 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, No.SIG05,pp. 118 – 133, 2002.

[26] 中村幹敏，奥井康弘: “改訂版標準 XML 完全解説 上”, 技術評論社, 平成 13 年.

[27] 師 茂樹：“電子テキスト概論 ”，電脳中国学，漢字文献情報処理研究会，好文出版，pp. 196 – 204 , 1998 年.

[28] 木簡学会: “日本古代木簡選”, 岩波書店，1990.

[29] 勝村哲也，丹羽正之: “eKanji”, http://nohara.u-shimane.ac.jp/ekanji/, 2000.

[30] 文字鏡研究会: “今昔文字鏡”, http://www.mojikyo.org/html/index.html, 1998.

[31] 岡安 光彦: “データベースを基盤とする考古学研究支援システムの研究”, 修士論文, 奈良先端科学技術大学院大学, 1996 年.

[32] 三宮 健: “出土状況モデルに基づいた考古学データベースの構築”, 修士論文, 奈良先端科学技術大学院大学，2000 年.

[33] Dublin Core Metadata Initiative: “Dublin Core Metadata Element Set, Version 1.1: Reference Description”, http://dublincore.org/documents/1999/07/02/dces/, 7/2, 1999.

[34] Marc Nanard, Jocelyne Nanard: “Cumulating and sharing end users knowledge to improve video indexing in a video digital library”, *Proceedings of the first ACM/IEEE-CS joint conference on Digital libraries*, pp. 282 – 289, Virginia USA, 2001.

[35] 川俣 晶: “XML による画像参照交換方式”, 日本工業規格協会, JIS-TRX0045, 2001.

[36] David C. Fallside :“XML Schema Part 0: Primer ”, http://www.w3.org/TR/xmlschema-0/, May 2001.

[37] Makoto Murata: “RELAX (Regular Language description for XML)”, http://www.xml.gr.jp/relax/, May, 2001.

[38] Anders Berglund, Scott Boag, Don Chamberlin et al. : “XML Path Language (XPath) 2.0”, http://www.w3.org/TR/xpath20/, December 2001.

[39] Scott Boag, Don Chamberlin, Mary F. Fernandez, Daniela Florescu, Jonathan Robie, Jerome Simeon: “XQuery 1.0: An XML Query Language”, http://www.w3.org/TR/xquery/, 2003.

[40] 玉川大学図書館: “百人一首”, 漢籍和装丁本コレクション, http://www.tamagawa.ac.jp/sisetu/tosyo/w_index.htm, 2000.

付録

A　Relax スキーマによる文献データのモデルの記述

文献データのモデルに対応した Relax スキーマの記述例を以下に示す．行 2-8 は，文献データのルート XML 要素である．行 10-19 は，提案モデルの 5 つの構成要素のルートの定義であり，bib は文献情報，im は文献画像，te は本文，an は注釈，ct は対応表を表す．行 21-70 は，文献情報の属性であり，XML 要素 bib の子として定義する．各属性の属性名は，Dublin Core Metadata Element Set Ver. 1.1 に定義された名前を用いた．行 72-73 は，文献画像の定義であり，本論文ではこの XML 要素の値として文献画像の URI を記述する．行 75-76 は，本文の定義であり，この XML 要素の値は文献画像の内容をテキストデータとして記述する．行 78-92 は，文献データのモデルの注釈を XML 要素で定義した例である．注釈は，注釈の識別子 (行 89) と注釈の本体 (行 91) の組の集合であるので，Relax スキーマでは，識別子と本文の組を保存するため XML 要素 row (行 83) を定義した．行 94-109 は対応表の定義である．対応表は文献画像の領域 imp (行 101) と注釈の識別 tp (行 102) の組の集合であり，これらの組を保存するために XML 要素 ci を定義した．文献画像の領域の記述は，各頂点の座標値を文字列として記述する．

```
1:  <?xml version="1.0" encoding="UTF-8"?>
2:  <module
3:  moduleVersion="1.0"
4:  relaxCoreVersion="1.0"
5:  xmlns="http://www.xml.gr.jp/xmlns/relaxCore">
6:  <interface>
```

```
7:  <export label="ip"/>
8:  </interface>
9:  <!-- 文献データのモデルのルート -->
10:  <tag name="ip"/>
11:  <elementRule role="ip">
12:  <sequence>
13:  <ref label="bib"/>
14:  <ref label="im"/>
15:  <ref label="te"/>
16:  <ref label="an"/>
17:  <ref label="ct"/>
18:  </sequence>
19:  </elementRule>
20:  <!-- 文献情報 -->
21:  <tag name="bib"/>
22:  <elementRule role="bib">
23:  <sequence>
24:  <ref label="Title"/>
25:  <ref label="Creator"/>
26:  <ref label="Subject"/>
27:  <ref label="Description"/>
28:  <ref label="Publisher"/>
29:  <ref label="Contributor"/>
30:  <ref label="Date"/>
31:  <ref label="Type"/>
32:  <ref label="Format"/>
33:  <ref label="Identifier"/>
34:  <ref label="Source"/>
35:  <ref label="Language"/>
```

```
36:  <ref label="Relation"/>
37:  <ref label="Coverage"/>
38:  <ref label="Rights"/>
39:  </sequence>
40:  </elementRule>
41:  <tag name="Title"/>
42:  <elementRule role="Title" type="string"/>
43:  <tag name="Creator"/>
44:  <elementRule role="Creator" type="string"/>
45:  <tag name="Subject"/>
46:  <elementRule role="Subject" type="string"/>
47:  <tag name="Description"/>
48:  <elementRule role="Description" type="string"/>
49:  <tag name="Publisher"/>
50:  <elementRule role="Publisher" type="string"/>
51:  <tag name="Contributor"/>
52:  <elementRule role="Contributor" type="string"/>
53:  <tag name="Date"/>
54:  <elementRule role="Date" type="string"/>
55:  <tag name="Type"/>
56:  <elementRule role="Type" type="string"/>
57:  <tag name="Format"/>
58:  <elementRule role="Format" type="string"/>
59:  <tag name="Identifier"/>
60:  <elementRule role="Identifier" type="string"/>
61:  <tag name="Source"/>
62:  <elementRule role="Source" type="string"/>
63:  <tag name="Language"/>
64:  <elementRule role="Language" type="string"/>
```

```
65:  <tag name="Relation"/>
66:  <elementRule role="Relation" type="string"/>
67:  <tag name="Coverage"/>
68:  <elementRule role="Coverage" type="string"/>
69:  <tag name="Rights"/>
70:  <elementRule role="Rights" type="string"/>
71:  <!-- 文献画像 -->
72:  <tag name="im"/>
73:  <elementRule role="im" type="string"/>
74:  <!-- 本文 -->
75:  <tag name="te"/>
76:  <elementRule role="te" type="string"/>
77:  <!-- 注釈 -->
78:  <tag name="an"/>
79:  <elementRule role="an">
80:  <ref label="row" occurs="+"/>
81:  </elementRule>
82:  <tag name="row"/>
83:  <elementRule role="row">
84:  <sequence>
85:  <ref label="id"/>
86:  <ref label="body"/>
87:  </sequence>
88:  </elementRule>
89:  <tag name="id"/>
90:  <elementRule role="id" type="string"/>
91:  <tag name="body"/>
92:  <elementRule role="body" type="string"/>
93:  <!-- 対応表 -->
```

```
94:  <tag name="ct"/>
95:  <elementRule role="ct">
96:  <ref label="ci" occurs="+"/>
97:  </elementRule>
98:  <tag name="ci"/>
99:  <elementRule role="ci">
100:  <sequence>
101:  <ref label="imp"/>
102:  <ref label="tp"/>
103:  </sequence>
104:  </elementRule>
105:  <tag name="imp"/>
106:  <elementRule role="imp" type="string"/>
107:  <tag name="tp"/>
108:  <elementRule role="tp" type="string"/>
109:  </module>
```

B Relax スキーマによる電子スクラップブックデータのモデルの記述

電子スクラップブックデータのモデルを Relax スキーマで定義した記述例を以下に示す．行 1-8 は，電子スクラップブックデータのモデルのルート XML 要素を定義である．行 10-16 は，電子スクラップブックデータの構成要素である電子スクラップブック情報 (行 13) と，グループ (行 14) の定義である．グループ は，電子スクラップブックデータ内に 1 つ以上ある．行 18-67 は，電子スクラップブック情報 の属性に対応する XML 要素を定義している．属性名は，Dublin Core Metadata Element Set Ver. 1.1 に従う．行 69-87 は，グループの属性に対応した XML 要素の定義である．グループの属性は，文献データの識別子を記述する XML 要素 ref (行 78）と複数の文献画像を一度に閲覧す

るための配置情報 x，y (行 76-77) の組の集合である．この組を保存するために， XML 要素 grow (行 71) を定義した．

```
1:  <?xml version="1.0" encoding="UTF-8"?>
2:  <module
3:  moduleVersion="1.0"
4:  relaxCoreVersion="1.0"
5:  xmlns="http://www.xml.gr.jp/xmlns/relaxCore">
6:  <interface>
7:  <export label="gp"/>
8:  </interface>
9:  <!-- 電子スクラップブックデータのモデルのルート -->
10:  <tag name="gp"/>
11:  <elementRule role="gp">
12:  <sequence>
13:  <ref label="cp"/>
14:  <ref label="sg" occurs="+"/>
15:  </sequence>
16:  </elementRule>
17:  <!-- 電子スクラップブックデータ情報 -->
18:  <tag name="cp"/>
19:  <elementRule role="cp">
20:  <sequence>
21:  <ref label="Title"/>
22:  <ref label="Creator"/>
23:  <ref label="Subject"/>
24:  <ref label="Description"/>
25:  <ref label="Publisher"/>
26:  <ref label="Contributor"/>
27:  <ref label="Date"/>
```

```
28:  <ref label="Type"/>
29:  <ref label="Format"/>
30:  <ref label="Identifier"/>
31:  <ref label="Source"/>
32:  <ref label="Language"/>
33:  <ref label="Relation"/>
34:  <ref label="Coverage"/>
35:  <ref label="Rights"/>
36:  </sequence>
37:  </elementRule>
38:  <tag name="Title"/>
39:  <elementRule role="Title" type="string"/>
40:  <tag name="Creator"/>
41:  <elementRule role="Creator" type="string"/>
42:  <tag name="Subject"/>
43:  <elementRule role="Subject" type="string"/>
44:  <tag name="Description"/>
45:  <elementRule role="Description" type="string"/>
46:  <tag name="Publisher"/>
47:  <elementRule role="Publisher" type="string"/>
48:  <tag name="Contributor"/>
49:  <elementRule role="Contributor" type="string"/>
50:  <tag name="Date"/>
51:  <elementRule role="Date" type="string"/>
52:  <tag name="Type"/>
53:  <elementRule role="Type" type="string"/>
54:  <tag name="Format"/>
55:  <elementRule role="Format" type="string"/>
56:  <tag name="Identifier"/>
```

```
57:  <elementRule role="Identifier" type="string"/>
58:  <tag name="Source"/>
59:  <elementRule role="Source" type="string"/>
60:  <tag name="Language"/>
61:  <elementRule role="Language" type="string"/>
62:  <tag name="Relation"/>
63:  <elementRule role="Relation" type="string"/>
64:  <tag name="Coverage"/>
65:  <elementRule role="Coverage" type="string"/>
66:  <tag name="Rights"/>
67:  <elementRule role="Rights" type="string"/>
68:  <!-- グループ -->
69:  <tag name="sg"/>
70:  <elementRule role="sg">
71:  <ref label="grow" occurs="+"/>
72:  </elementRule>
73:  <tag name="grow"/>
74:  <elementRule role="grow">
75:  <sequence>
76:  <ref label="x"/>
77:  <ref label="y"/>
78:  <ref label="ref"/>
79:  </sequence>
80:  </elementRule>
81:  <tag name="ref"/>
82:  <elementRule role="ref" type="string"/>
83:  <tag name="x"/>
84:  <elementRule role="x" type="string"/>
85:  <tag name="y"/>
```

```
86:  <elementRule role="y" type="string"/>
87:  </module>
```