

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年 6月 25日現在

機関番号：64302

研究種目：基盤研究（A）

研究期間：2007～2010

課題番号：19200019

研究課題名（和文）近代日本の歴史的時空間データマイニングのための基盤整備

研究課題名（英文）Infrastructure Development for Historical Time-spatial Data Mining to Modern Japan

研究代表者

山田 奨治（YAMADA SHOJI）

国際日本文化研究センター・研究部・教授

研究者番号：20248751

研究成果の概要（和文）：前近代～近代日本の地理情報として『古事類苑』「地部」のメタデータ付き全文情報、輯製図・帝国図・京都地域の仮製図・正式図の高度接合画像、各種地図のシェイプファイル、明治元年時点の国郡境界データを作成公開した。これにより日本近代の歴史事象を時空間分析するための基礎的情報の整備が進展した。また『古事類苑』「地部」の成り立ちと、京都盆地の庭園分布の水文条件についての分析を行った。

研究成果の概要（英文）：We developed the following digital resources as basic geographical information of Japan in pre-modern and modern era: full text data of *Kojiruien Chibu* with metadata; high-quality jointed digital image and shape files of several maps of Japan and Kyoto district in modern era. Through the project, some basic digital resources were prepared to conduct space-time analysis of historical events of modern Japan. At the same time, we analyzed the editing process of *Kojiruien Chibu* and the hydrological condition of the gardens in Kyoto basin.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	9,100,000	2,730,000	11,830,000
2008年度	8,500,000	2,550,000	11,050,000
2009年度	8,000,000	2,400,000	10,400,000
2010年度	7,500,000	2,250,000	9,750,000
年度			
総計	33,100,000	9,930,000	43,030,000

研究分野：総合領域

科研費の分科・細目：情報学・情報図書館学・人文社会情報学

キーワード：時空間情報、データマイニング、地理情報、全文データベース、GIS、古事類苑、旧版地図、正式図

1. 研究開始当初の背景

研究代表者はこれまでに古文書を対象とした文字認識の基礎的な研究などで科学研究費の研究代表者を努めてきた。平成11年度から進めた古文書文字認識にかんする一

連の研究では、総計20万文字を超える古文書文字データベースを作成して文字認識実験を行い、その成果を基に、電子くずし字辞典の実用化に筋道を付けることができた。平成16年度から18年度までの予定で行っ

ているデータマイニング資源の開発研究では、明治～大正時代に編纂・出版された本邦初の大百科事典である『古事類苑』の一部、約4,200頁分の全文入力を行い、それをマイニング資源とするべくメタデータの付与実験、HTML版のネットでの公開実験、Wiki版の実装実験、前近代日本の諸概念のシソーラス辞書の作成・公開、『古事類苑』から抽出した古典言語解析のための64,000語の辞書作成・公開などの成果をあげた。これらの一連の研究を通して、電子的な歴史資料とそれに対するデータマイニングのノウハウと実績を積み重ねることができた。ただし、これまでの成果は文字情報の世界に閉じたものであり、歴史研究に不可欠な時空間データマイニングには踏み込めていなかった。本研究では、これまでの研究蓄積を生かしつつ、その適用範囲を時空間に拡大する。

2. 研究の目的

この研究で取り組むことは、(1)近代以前の日本列島の地理にかんする全文データの整備とメタデータの付与、(2)古地図のデータ化、(3)地域を限定した時空間データマイニングによる知識発見の試行、の3点である。(1)に関しては、『古事類苑』地部の全文データを整備して、近代以前のあらゆる書物・記録にあらわれる地理情報にメタデータを付加していく。(2)に関しては、基本となる歴史的な地図情報を電子化し、地理情報システム上での取り扱いを可能にする。(3)に関しては、京都の近世～近代に限定し、京都の「近代化」が地理空間上、いつどのような形で進行したのかを検討する。

3. 研究の方法

研究は(1)テキストデータ作成、(2)地図データ作成、(3)分析の3部構成とする。(1)に関しては、テキストデータの作成とメタデータ付与、地図データとのリンクをはかる。(2)に関しては、国土地理院の旧版地図や京都の歴史的な市街図データを作成し、GISにインプリメントする。(3)に関しては、テキストデータ班と地図データ班が作成したデータをもとに、京都の古市街図を中心に分析を行い、京都の近代化の一側面について地理的な観点から新たな知見や研究の方向性を見出す試みを行う。

4. 研究成果

平成19年度は、(1)『古事類苑』「地部」全文データの校正・公開と地名シェイプフ

イルの作成、(2)『古事類苑』「地部」に登場する「山」についての空間分析、(3)大正期帝国図海岸線シェイプファイルの作成、

(4)明治期京都地図の結合・正規化とメッシュ化、(5)京都にある庭園のリストアップと経緯度情報の付与等を行った。(1)に関しては、すでに校了になっている「地部1」の全文データを公開し、「同2・3」の校正作業を進めた。また、「地部」に登場する人工物・自然地名に対して経緯度情報を与えてシェイプファイル化し、当該シェイプから『古事類苑』の全文データベースへリンクする機能を試験した。(2)に関しては、「地部」に登場する「山」の経緯度を空間分析し、『古事類苑』の「山」の選定基準に潜んでいる文化的な要素について考察した。(3)に関しては、『古事類苑』を時空間分析し、あるいは地図から検索するためのベースマップに適切な地図を精度面から比較検討し、すべての輯製図・帝国図の画像をデジタル化するとともに、帝国図の海岸線をシェイプファイル化した。(4)に関しては、京都の近代化過程をデータマイニングするための基礎資料とするために、2万分の1仮製図と正式図をデジタル化し、複数の図葉についてゆがみを精密に補正しながら結合し、3次メッシュに分割する作業を実施した。(5)に関しては、時空間データマイニングの基礎資料とするために、京都市内に現存する日本庭園をリストアップし、その経緯度と様式のデータを作成した。

平成20年度は、テキストデータの作成に関しては、『古事類苑』「地部1」の全文データを公開し、懸案だった「同2・3」についても校了となり、公開まであとは外字のユニコード変換テーブルの作成と変換を残すのみとなった。地理データの作成に関しては、20万分の1帝国図ベースマップにした海岸線データと旧国郡境界データを作成し、『古事類苑』が編纂された明治末～大正期の国土のシェイプファイルを整備した。また、『古事類苑』地部に記された人工物・自然地名の位置比定と経緯度情報の作成作業、シェイプファイル化作業をあらかた終えた。また、京都地域の近代化と都市化の考察のための基礎資料とするべく、京都地域の2万分の1仮製図と正式図の高精度接合作業を行なった。データ分析にかんしては、京都盆地北部に現存する164の日本庭園のリストを作成し、GISを用いて地図化した。庭園築造数を歴史的にみると、江戸時代と大正時代にそのピークが認められた。土木技術が発達した江戸時代には大規模な庭園が数多く築造されたが、池などの地表水を欠く庭園の多くが扇状地扇央に分布しているのに対し、地表水のある庭園は湧水のみられる東山丘陵の山麓に集中していることが確認できた。明治

～大正時代に建設された庭園のほとんどは1890年に完成した琵琶湖疏水沿いに位置しており、池・流れ・滝などの水景が重視されている。これらのデータは、京都盆地に分布する庭園が自然的にも人為的にも水文条件の影響を強く受けていることを示していることがわかった。

平成21年度は、テキストデータ作成については、『古事類苑』「地部2・3」約2800ページの全文情報を「古事類苑全文データベース」の一部として計画通り公開した。また、同書の伊勢國の記述を対象に、そこから手作業で抽出した地名と『大日本地名辞書』の地名を文字列マッチングで比較対照することにより、『古事類苑』の本文から地名を自動抽出する可能性について検討を行った。その結果、異体字を集約するプロセスを経れば、地名自動抽出に有効なツールを作ることができる可能性が示された。地図データ整備の面では、明治後半に作られた20万分の1帝国図から作成した旧国郡境界、自然地名、人工物地名シェイプファイルの検証を行い、明治元年の国郡境界を復元する必要性を認識し、そのための基礎情報の収集を行った。また、京都地域の2万分の1正式図(10葉)にある、道路、鉄道、駅、行政界、水厩線、土地利用、地名のすべての情報をシェイプファイル化した。これにより、GIS分析に供することが出来る、明治なかば頃の京都のデジタル地図がほぼ完成した。データ分析の面では、『古事類苑』の編纂事歴を精査し、「地部」に採録された事項の情報量に地域的な偏りがあることをあきらかにした。

平成22～23年度は、公開済みの『古事類苑』「地部」全文情報の誤謬をチェックし修正した。また「地部」の編纂事歴を調査し、その成り立ちについて新たな知見を得ることができた。帝国図シェイプファイルについては、明治元年時点の郡界を可能な限り復元してベクトルデータ化した。京都地域の2万分1正式図のシェイプについては、2葉分を追加でデータ化した。またそれと比較するための2.5万分1現代地形図の対応領域について土地利用をシェイプファイル化した。京都の「仮製図」「正式図」「現代図」とのあいだで、土地利用の変遷に着目した近代化過程の基礎的な考察を行った。9月には歴史的時空間情報の現状についての公開シンポジウムを開催した。成果公開の方法と今後の方針について、3月に研究協力者らと東京で2回会議を行う予定であったが、3月1日の大震災とそれにつづく計画停電により会議の延期を余儀なくされた。そこで2回分の旅費を平成23年度に繰り越した。最終的には、帝国図海岸線・行政界と同明治元年時点の復元郡界は、シェイプファイルで公開することとなった。京都地域の2万分1正式図と2.

5万分1現代地形図の土地利用のシェイプファイルは、希望のある研究機関に対して個別に配布することにした。

5. 主な発表論文等

〔雑誌論文〕(計7件)

- ① 原正一郎、柴山守 著、地域情報学の構築と時空間情報解析ツール、人文科学とコンピュータシンポジウム論文集、査読有、2007(15)、2007、71-78
- ② 桶谷猪久夫、人文分野における日本地名辞書の構築と地名属性の特徴分析、人文科学とコンピュータシンポジウム論文集、査読有、2007(15)、2007、79-86
- ③ 山田奨治、尾方隆幸、古事類苑・地部データへの経緯度情報付与と地図化、情報処理学会研究報告、査読無、47、2008、85-92
- ④ 李偉、尾方隆幸、山田奨治、京都盆地の庭園分布のGIS分析—庭園の属性と地形・水文条件、情報処理学会研究報告、査読無、4、2009、1-8
- ⑤ 山田奨治、中西和子、尾方隆幸、古事類苑・地部GISデータの作成、情報処理学会研究報告、査読無、2009(83)、2009、1-8
- ⑥ 山田奨治、相田満、古事類苑全文データベースの構築について、人間文化研究資源共有化研究会報告集1、査読無、1、2010、151-159
- ⑦ Ogata Takayuki et. al., A Mapping and Geographical Analysis of Japanese Gardens in the Kyoto Basin, Japan Review、査読有、22、2010、213-221

〔学会発表〕(計4件)

- ① 山田奨治、古事類苑・地部データへの経緯度情報付与と地図化、情報処理学会人文科学とコンピュータ研究会、2008年5月23日、立命館大学
- ② 李偉、京都盆地の庭園分布のGIS分析—庭園の属性と地形・水文条件、情報処理学会人文科学とコンピュータ研究会、2009年1月23日、キャンパスプラザ京都
- ③ 山田奨治、中西和子、『古事類苑・地部』の採録文献にみる特性、2009年人文地理学会大会、2009年11月8日、名古屋大学
- ④ 中西和子、編纂経緯からみる古事類苑・地部—二人の編集者、三浦千畝と加藤才次郎一、シンポジウム「日本の歴史的時空間情報の現在」、2010年9月11日、国際日本文化研究センター

〔図書〕（計1件）

- ① 山田奨治、柴山守、東京堂出版、CD-ROM版くずし字解読用例辞典、2007、CD-ROM1枚

〔その他〕

ホームページ等

<http://ys.nichibun.ac.jp/kojiruien/>

<http://www.nichibun.ac.jp/~shoji/cgi-bin/dmh/wiki.cgi>

6. 研究組織

(1) 研究代表者

山田 奨治 (YAMADA SHOJI)

国際日本文化研究センター・研究部・教授
研究者番号：20248751

(2) 研究分担者

宇野 隆夫 (UNO TAKAO)

国際日本文化研究センター・研究部・教授
研究者番号：70115799

(H20→H22 連携研究者)

柴山 守 (SHIBAYAMA MAMORU)

京都大学・東南アジア研究所・教授
研究者番号：10162645

(H20→H22 連携研究者)

原 正一郎 (HARA SHOICHIRO)

京都大学・地域研究統合情報センター・教授

研究者番号：50218616

(H20→H22 連携研究者)

桶屋 猪久夫 (OKETANI IKUO)

大阪国際大学・国際コミュニケーション学部・教授

研究者番号：90169269

(H20→H22 連携研究者)

(3) 連携研究者

尾方隆幸 (OGATA TAKAYUKI)

琉球大学・教育学部・講師

研究者番号：50548287