

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 4 月 4 日現在

機関番号：82651

研究種目：若手研究（B）

研究期間：2009～2011

課題番号：21700274

研究課題名（和文） 史料学研究支援のためのアノテーション管理基盤に関する研究

研究課題名（英文） Annotation Management Methods for Support of Study of Historical Materials

研究代表者

山田 太造（YAMADA TAIZO）

大学共同利用機関法人 人間文化研究機構・本部・特任研究員

研究者番号：70413937

研究成果の概要（和文）：

本研究では、史料学などの研究における研究過程を支援するために、研究過程で生成される意見や所見などのアノテーションである研究的“知見”を管理し利活用しえる情報基盤に関する研究を行った。アノテーションとともに関連付けられた情報リソースを管理、検索、出力できる機能を有するシステムを開発した。

研究成果の概要（英文）：

In order to support research processes in study of historical materials, this research aims to information infrastructure in which annotations such as opinions and findings generated in the research processes can be managed and used. The system which can manage, search and output annotations and information resources related with the annotations also is developed.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	800,000	240,000	1,040,000
2010年度	1,400,000	420,000	1,820,000
2011年度	600,000	180,000	780,000
年度			
年度			
総計	2,800,000	840,000	3,640,000

研究分野：総合領域

科研費の分科・細目：情報学、図書館情報学・人文社会情報学

キーワード：情報検索、情報統合、歴史情報、アノテーション、翻刻、ISAG（G）、TEI

1. 研究開始当初の背景

日本史研究で取り扱う情報リソースが変化しつつある。従来は、研究者個人が収集した歴史史料であった。現在ではそれ以外に、東京大学史料編纂所（以下、史料編纂所）歴史情報データベース（SHIPDB）やPORTA（国立国会図書館デジタルアーカイブポータル）などの Web 上での検索サービスを介してデジタル史料を扱うことが容易となった。

Web2.0 の登場により、電子的な情報を共有

できる仕組みとして、blog や wiki を代表する CMS（コンテンツ管理システム; Content Management System）のような Web システムが生み出されてきた。参加しているサイトにおいてコンテンツ作成、および共有が容易に行えるようになり、「集合知（collective intelligence）」が発展し、「知識」として頻繁に活用されつつある。しかしながら、Web2.0 などでは、研究的論拠を確実にしながら「知識」を積み上げていくような研究過程を支援

する仕組みではなく、また、従来の CSCW やグループウェアの範囲では実現しえない。集合知のように研究成果をまとめ上げた「知識」ではなく、研究過程で生成される研究者の知見や所見となりうるアノテーションをその根拠となりうる情報リソースに関連づけることが史学・史料学の研究を進める上で重要であると考えられる。そのためにはアノテーションの管理、および、その利活用を可能にする情報基盤を確立し、その実現方式について究明する必要がある。

2. 研究の目的

本研究では、Web 上のシステムやサービスにおける情報リソースから得られる研究的“知見”を管理することにより、歴史学や史料学の研究者の研究過程を支援するための情報基盤を明らかにすることを目指した。ここでの研究的“知見”は、意見や意図を示すアノテーションとその論拠となる情報リソースとのリンクなどによって構成し表現することになる。そこで本研究では、以下の問題を解決することとした。

(1) アノテーションの表現方法とその管理方法

研究過程で付与されるアノテーションは研究者がその研究目的により付与することで生成される。そのため研究者ごとに管理する必要があると考えた。また、アノテーションを付与することにより、情報リソース間での新たな関連づけが行われる可能性がある。そのような関連付けを可能とするための表現方法、および、その管理方法が必要である。

(2) アノテーションを付与するためのユーザインターフェース

史料学の研究者が、ある研究的“知見”を得たとき、そのベースとなった情報リソースにアノテーションを関連づけるために、ユーザ-マシン間のインタラクションが可能なユーザインターフェースが必要である。

(3) アノテーションを組み入れた情報リソースの検索、出力

ユーザによって付与されたアノテーションは情報リソースに関連付けられる。アノテーションとともにその情報リソースを検索可能とする仕組みが必要である。また、研究成果を出力するため、テキストおよび目録等の情報リソースとともに関連付けられたアノテーションを出力する方法も考慮する必要がある。

(4) テキスト入力への支援

アノテーションの付与は翻刻時に入力されるテキストに対して行うことを想定してい

る。そのため翻刻への支援は重要であると考えた。そこでテキスト入力への支援が必要であると考えた。

3. 研究の方法

2. で述べた、史学・史料学における研究的“知見”(情報リソース、意見・所見などのアノテーション、および、情報リソースとアノテーションとの関連)を管理・利活用できる情報基盤を明らかにするために、以下の問題を解決した。これらは、史学・史料学の研究者個別での利用事例を蓄積し、その事例に応じたプロトタイプの開発とその試用、試用におけるフィードバックの分析、の3つのプロセスを繰り返しながら本研究を進めた。

(1) アノテーションの表現方法とその管理方法

(2) アノテーションを付与するためのユーザインターフェース

(3) アノテーションの検索、および、アノテーションを組み入れた情報リソースの検索

(4) テキスト入力への支援と構造化

(1)については、情報リソースとして、史料編纂所の個々の研究者が史料調査により収集した史料、および、SHIPSDBにおける「所蔵史料目録DB」(HI-CAT)、および、それと関連する史料画像を扱った。また、「アーカイブハブ」および「日本古文書ユニオンカタログ」に対する適用も検討した。

アノテーションを表現する上で、史料編纂所の史学研究者を対象に、史学・史料学研究を推進する上での研究者による情報整理および分析方法などのデータを収集し分析した。これにより、研究過程で生成されるアノテーションのタイプ、それと関連づけられる情報リソース、および、その関連などのデータ構造を確立した。また、その管理に関するデザイン方法を検討した。

(2)については、(1)においてデザインされるアノテーションを、対象とする史料へのアノテーション付与の方法とそのユーザインターフェースについて検討し、アノテーション付与機能を開発した。

(3)については、(1)および(2)にもとづいて蓄積されるアノテーションを用いた検索方法について検討した。この検索では、付与したアノテーションに対して、誰がどのようなアノテーションを付与したかを検索しえるよう検索インターフェースを開発した(アノテーション検索機能)。また、情報リソースとアノテーションの関連を考慮した検索を行うことを想定しているため、単にアノテーションを検索結果として返すのではなく、関連づけられた情報リソースとともに検索結

果として返す方法について検討した。また、研究成果として情報リソースとともにアノテーションを出力する方法についても検討し、開発した（出力機能）。

(4)については、SHIPSDBにおけるフルテキストデータベースにおけるテキストを学習データとして利用することで、現在入力したい文字を検索することが可能な、文字 n-gram ベースな候補文字検索機能として提案し、これを実装した。

4. 研究成果

研究成果としては以下の3つが挙げられる。

(1) システム開発

2009年度は、情報リソースとして、史料編纂所の個々の研究者が史料調査により収集した史料を対象に、翻刻を支援する翻刻エディタをプロトタイピングした。翻刻エディタは、テキストの自動構造化、出力機能、および、候補文字検索機能を備えるスタンドアロンタイプのアプリケーションである。

翻刻エディタをベースに、科学研究費補助金基盤研究(A)ネットワーク環境における前近代日本史史料の翻刻・編纂フレームワークの確立（研究課題番号：21240022、代表者：加藤友康）では翻刻支援システムを開発した。そこで、2010年度に、このシステムに対し、翻刻エディタにおけるテキスト出力機能、および、候補文字検索機能を付与し、さらに、2009年度に検討したアノテーション付与機能、および、アノテーション検索機能を搭載した。

2011年度は2010年度まで搭載した各機能に対して検証し、その結果をフィードバックすべく、システムに反映させた。

(2) アノテーションと情報リソースの出力

2009年度にプロトタイピングした出力機能では史料の目録・画像・テキスト・アノテーションを関連付けて、XML形式で出力する。これに対し、翻刻支援システム外でも利用できる出力形式とすべくTEI (Text Encoding Initiative)による出力方法を検討した。TEIによる出力は研究成果の出力だけでなく、外部システムとの連携や内部の目録・画像・テキスト・アノテーションの解析を進める上で重要であると考えている。

(3) 候補文字検索機能

初期は文字 n-gram モデルをシンプルに用いた方法(具体的には学習データとして3-gramを用いた方法)を実装していた。これに対し、2009年度は翻刻対象の史料年代に応じた方法を提案した。また2010年度はスムージングを用いた方法を提案した。提案方法を検証

すべく行った実験において、スムージングによる方法が最もよい精度で検索できることがわかった。そこでスムージングによる候補文字検索機能をプログラミングし、翻刻支援システムから呼び出せるように実装した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計11件)

山田太造、近藤成一、野村朋弘、日本古文書ユニオンカタログ - 古文書情報を網羅するための“古文書リンケージ”プラットフォーム -、研究報告人文科学とコンピュータ(CH)、査読無、Vol.2012-CH-93、No.1、2012、pp.1-8

Taizo Yamada、Satoshi Inoue、Tamaki Endo、Noriko Kurushima、A Support Method for Text Structuring of Japanese Historical Documents、proceedings of Osaka Symposium on Digital Humanities 2011(OSDH2011)、査読有、2011、p40

山田太造、日本史史料における翻刻データの作成支援と共有手法、漢字字体史研究国際シンポジウム「字体規範と異体の歴史」、査読無、2011、111-120

山田太造、井上聡、遠藤珠紀、久留島典子、日本史史料読解支援のための候補文字検索、じんもんこん 2011 論文集、査読有、Vol.2011、No.8、2011、pp.43-50

大内英範、山田太造、高橋典幸、綱川歩美、林譲、保谷徹、山家浩樹、横山伊徳、Hi-Cat Plus: デジタル史料の検索・閲覧システム、じんもんこん 2011 論文集、査読有、Vol.2011、No.8、2011、pp.105-110

山田太造、横山伊徳、綱川歩美、高橋典幸、林譲、デジタル史料写真帳: 収集史料のデジタル化と検索・閲覧システム、画像電子学会年次大会予稿集、査読無、Vol.39、No.T1-2、2011、-

山田太造、井上聡、遠藤珠紀、久留島典子、日本史史料における翻刻テキストの構造化支援手法、情報処理学会研究報告(人文科学とコンピュータ研究会報告) 2011-CH-91、査読無、Vol.2011、No.5、2011、pp.1-16

山田太造、井上聡、遠藤珠紀、久留島典子、日本史史料における翻刻データの管理と編集支援、情報科学技術フォーラム講演論文集、査読無、Vol.9、No.4、2010年、pp.475-480

山田太造、横山伊徳、綱川歩美、高橋典幸、林譲、採訪史料管理システム、情報処理学会シンポジウム論文集、査読有、

Vol.2010、No.15、2010、pp.145-150
山田太造、デジタル資料情報記述モデル
によるデータ記述について、国立歴史民
俗博物館共同研究「デジタル化された博
物館資料に関する情報記述法の研究」公
開研究会資料集、査読無、-、2009、
pp.23-26
山田太造、赤石美奈、石川徹也、古文書
を対象とした翻刻支援システム、画像電
子学会年次大会予稿集、査読無、Vol.37、
No.T1-2、2009、-

〔学会発表〕(計 10 件)

山田太造、日本古文書ユニオンカタログ
- 古文書情報を網羅するための“古文書
リンケージ”プラットフォーム -、第 93
回人文科学とコンピュータ研究発表会、
2012 年 1 月 27 日、奄美市立奄美博物館
(鹿児島県)
山田太造、日本史史料における翻刻デー
タの作成支援と共有手法、漢字字体史研
究国際シンポジウム「字体規範と異体の
歴史」、2011 年 12 月 17 日、東京外国語
大学アジア・アフリカ言語文化研究所
(東京都)
山田太造、日本史史料読解支援のための
候補文字検索、人文科学とコンピュータ
シンポジウム「じんもんこん 2011」、2011
年 12 月 10 日、龍谷大学大宮キャンパス
(京都府)
Taizo Yamada, A Support Method for Text
Structuring of Japanese Historical
Documents, Osaka Symposium on Digital
Humanities 2011(OSDH2011)、2011 年 9
月 14 日、The Graduate School of
Language and Culture, University of
Osaka (大阪府)
山田太造、日本史史料における翻刻テキ
ストの構造化支援手法、第 91 回人文科
学とコンピュータ研究発表会、2011 年 7
月 30 日、大阪電気通信大学寝屋川キ
ャンパス(大阪府)
山田太造、デジタル史料写真帳：収集史
料のデジタル化と検索・閲覧システム、
画像電子学会 2011 年度年次大会、2011
年 6 月 25 日、くにびきメッセ(島根県)
山田太造、探訪史料管理システム、人文
科学とコンピュータシンポジウム「じん
もんこん 2010」、2010 年 12 月 12 日、東
京工業大学大岡山キャンパス(東京都)
山田太造、史学研究をいかに支援するか
- 歴史情報の生成・管理と利活用の支援
方法 -、立命館創始 140 年・学園創立
110 周年記念 アート・リサーチセンタ
ー連続講演会第 5 回、2010 年 12 月 10
日、立命館大学アート・リサーチセンタ
ー(京都府)

山田太造、日本史史料における翻刻デー
タの管理と編集支援、第 9 回情報科学技
術フォーラム(FIT2010)、2010 年 9 月 9
日、九州大学伊都キャンパス(福岡県)
山田太造、古文書を対象とした翻刻支援
システム、画像電子学会 2009 年度年次
大会、2009 年 6 月 25 日、旭川市勤労者
福祉総合センター(北海道)

6. 研究組織

(1) 研究代表者

山田 太造(YAMADA TAIZO)
大学共同利用機関法人 人間文化研究機
構・本部・特任研究員
研究者番号：70413937