

# セマンティック・ウェブ技術を用いた歴史研究支援システムの開発

47-126749 中村 寛  
指導教員 大和 裕幸 教授

In this study, the system to support historical study by using the technology of Semantic Web is developed. This system enables historical researchers to describe bibliographic information and research findings as metadata, and to analyze historical materials with these metadata according to each researcher's objective. Also this paper proposes a new process of historical study which enables researchers to collaborate on accumulating metadata and analyzing materials with others. The effectiveness of the proposed system is evaluated by an experiment with several researchers to analyze the process of deciding addresses of diplomatic documents.

Key words: Historical Study, Semantic Web, Metadata, Digital Materials

## 1 緒言

近年の情報技術の急速な発展を背景として、企業・教育・医療といった様々な分野で資料の電子化が進められ、その活用を目的とした研究が数多く行われている<sup>1)2)</sup>。歴史研究の分野でも電子化された史料の活用を目的とし、オントロジーを用いた関連史料の可視化システムの開発<sup>3)</sup>等がある。しかし各歴史研究者（以下、研究者）の研究目的に応じて史料や調査結果を蓄積し、それらを再利用することで分析を行う環境は整えられていない。よって本研究ではセマンティック・ウェブ技術を用いて書誌情報や研究者による調査結果をメタデータとして記述し、メタデータの利用に基づいた史料分析を可能とするシステムの開発を目的とする。なお中村<sup>4)</sup>は同目的に基づいた手法を提案しているが、本研究では中村の手法を複数研究者で共有可能とする点で拡張した。また計算機を用いたメタデータの自動付与及び分析を可能とするシステムの開発、及び研究者との実験による評価を行った点において差異がある。

## 2 既存の歴史研究プロセスと課題点の整理

三浦<sup>5)</sup>は歴史研究を「歴史学では現存する不完全なデータ（史料）を歴史学者の主観を基に分類し、その分類されたデータを時間的な流れといった関係から整理し、独自の仮説を構築し、さらにそれを裏付けるデータを探して検証を行っていく」ものと定義し、史料の収集・整理・分析の三つのプロセスが繰り返されると述べている。

史料の収集プロセスでは研究者がデジタルアーカイブで公開されている史料のダウンロードや紙媒体史料の撮影を行い、PCのフォルダ等の各研究者が利用している保存環境に史料を蓄積する。史料の整理プロセスでは研究者の目的に基づいて、画像を中心とする史料を閲覧することで必要な情報を抽出する。本作業については史料研究者へのヒアリングから、抽出した情報はMicrosoft Excel等の史料とは異なる環境に保存され、各研究者が独自に付与したID等で紐づけられていることが確認できた。史料の分析プロセスでは史料の整理プロセスにおいて得た気づきに基づいて仮説を立案し、ファイルに保存した情報との整合性を確認することで仮説の検証を行う。これらのプロセスは基本的に研究者個人によって行われていることが、歴史研究者へのヒアリングから確認された。

三浦はこれらの各プロセスの課題についても言及しており、収集プロセスでは史料が分散しているために各研究者がそれらを総覧し収集することは困難である点を課題として挙げている。整理プロセスでは研究に必要な情報を

収集した膨大な史料から抽出することは難しい点を指摘している。また分析プロセスでは整理した史料の分類や比較に基づいて仮説の立案・検証を行う必要があるが、史料を俯瞰的に分析するにはそれらが一定の基準で整理されている必要があると述べている。

## 3 提案手法

### 3.1 提案する歴史研究環境

2の課題に対して、本研究ではFig. 1に示す研究環境を提案する。本提案環境では史料や調査結果をクラウド上で管理し、研究者やアーキビスト、情報技術者等が史料の収集・整理・分析を共同で行うことを可能とする。これにより各研究者の収集や整理における労力軽減や、異なる分野の研究者による協調的な研究が実現できる。

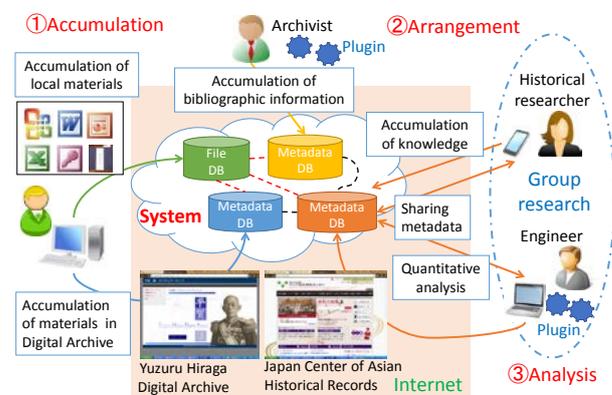


Fig. 1 Proposed method

収集プロセスでは、研究者がPC等で管理している史料やデジタルアーカイブで公開されている史料をクラウド上に構築されたDBに登録する形で行う。整理プロセスでは、収集した史料から得た調査結果を研究者が蓄積する他、画像処理技術等を用いて機械的に情報を抽出する。分析プロセスでは、DBに登録された書誌情報や調査結果を利用することで、研究目的に応じた史料の分類や比較を行う。

### 3.2 セマンティック・ウェブ技術を用いた史料管理手法

セマンティック・ウェブとはWebページに記述された内容について、それが何を意味するかを表す情報(メタデータ)を一定の規則に従って付与し、情報を構造化する構想である。本研究ではこの技術を応用し、史料内容を管理する。Fig. 2にセマンティック・ウェブ技術を用いた史料管理手法を示す。各史料に書誌情報や調査結果等をメタデ

ータとして付与し、さらに研究対象分野における概念をオントロジーによって記述し、それらをメタデータとして用いることで史料間の関係性を体系化する。

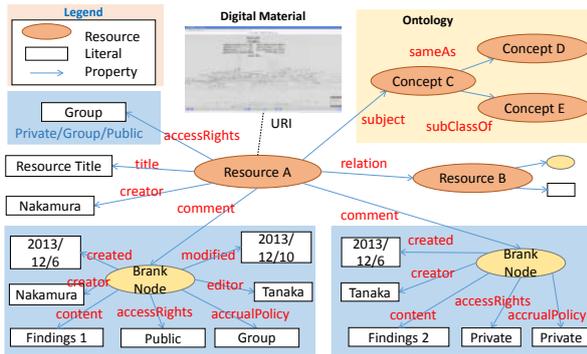


Fig. 2 Management of historical materials by use of the technology of Semantic Web

また研究者へのヒアリングから、史料や調査結果等を複数研究者で共有する場合、各研究者が得た知見は研究の独自性・新規性につながるという観点からその公開範囲を制限したい、また他の研究者の調査結果についてはその信憑性等の観点から情報の出所を明示化したい等の必要要件が確認された。よってこれらを満たすため、本研究ではブランクノードを導入し、史料に付与された登録日や登録者といった情報をさらにそのメタデータという形で管理する手法を提案する。例えばメタデータの公開範囲を個人、グループ、全体から選択することで、史料に付与された各メタデータについて他の研究者による閲覧範囲を制限できる。またメタデータの更新日や更新者情報を付与することで、それがいつ誰によって更新されたのか等の情報を明示化する。

## 4 開発したシステム

### 4.1 開発したシステムの概要

開発したシステムの全体像を Fig. 3 に示す。

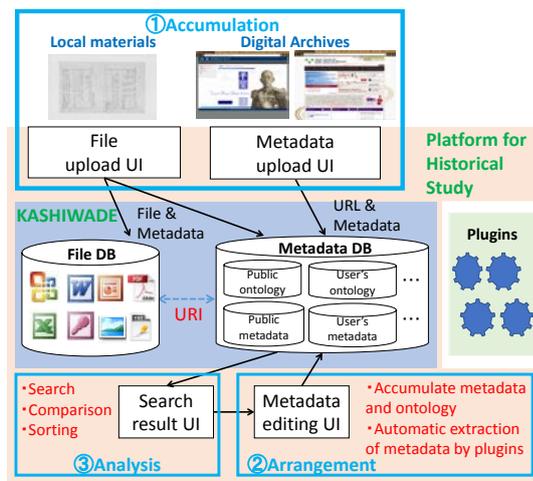


Fig. 3 Developed system

システムはウェブ上に構築され、各研究者はブラウザを介してシステムを利用する。システムは 2 で述べた史料の収集・整理・分析の各プロセスを支援する機能を提供す

る。各研究者は PC 等に保存された史料やデジタルアーカイブで公開されている史料を登録し、これら史料は Fig. 3 に示した形式に変換される。また研究者は Fig. 4 に示すインタフェースを用いて史料の整理を行う。



Fig. 4 Interface to edit metadata

本インタフェースでは左部に史料が持つメタデータ、右部に史料が表示されるため、研究者は史料の閲覧と並行して書誌情報や調査結果をメタデータとして蓄積可能である。3.2 で説明した各メタデータの公開範囲や編集権限の設定は、図中①に示すアイコンからそれぞれ選択することで行う。図中②には更新者や更新日に関する情報が表示される。さらに本インタフェースでは 4.2 で後述するプラグインの実行機能 (図中③) を備え、目的に応じたプラグインを選択することで、対象史料に対して機械的にメタデータを付与する。そして研究者は蓄積されたメタデータを用いて、史料の分類や並び替え、比較等の分析を行う。

### 4.2 プラグインによるメタデータの自動付与と利用

プラグインの実行環境を Fig. 5 に示す。

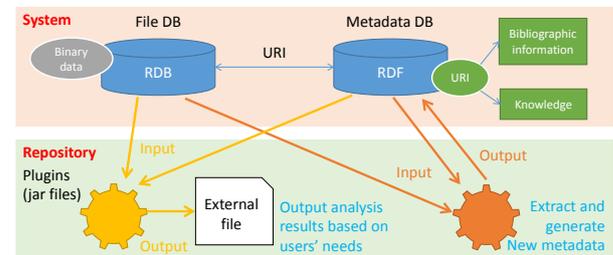


Fig. 5 Execution environment for plugins

プラグインはサーバ上のシステムとは異なるリポジトリで管理され、システムは指定したプラグインを呼び出す形で実行する。各プラグインはシステムに登録されたリソース (バイナリファイル) やメタデータ (文字列) を引数とし、新たに抽出したメタデータの付与や外部ファイルへの出力を行う。プラグインはシステムから独立しているため、システムを改変することなく、目的に応じて自由に追加や修正することが可能である。

## 5 実験: 外交文書の送付過程分析

### 5.1 実験目的

本実験では開発したシステム上で複数研究者による歴史研究を行い、提案手法の有用性の検証を行う。本実験は人文社会系研究者修士学生一名 (以下、研究者 X)、著者 (以下、技術者 Y) および史料収集補助のための学生 2 名で行った。

5.2 実験内容

本実験では第一次世界大戦後の来電に関する送付処理過程の分析を行う。来電とは外務省内で閲覧された後、総理大臣・海軍・陸軍等の他機関へと送付される外交文書の一つであり、外務省はこれらの情報の送付先の決定権を有していた。しかしこの来電の送付先を決定していた人物については明らかになっていない。これに対して研究者 X は当時の来電を分析し、埴原正直（以下、埴原）が外務次官を務めた 1919 年 9 月から 1922 年末において、来電の送付先の決定過程に変化が生じていることを発見した。当該期間以前では外務次官によって来電の送付先が指示されていることが示唆されたが、当該期間では外務次官以外の人物が送付先を指示し、また他の人物がそれを訂正している来電が確認された。しかしこの先行研究では 50 件弱の来電の分析に留まり、具体的な人物は明らかになっていない。よって本実験では来電に対する量的な分析を複数研究者で行い、これらの人物を特定することを目的とする。

5.3 対象史料と分析方法

本実験では外交史料館が保有しているパリ講和会議に関する来電を収録した簿冊 10 件、来電 471 件を対象とした。来電の例を Fig. 6 に示す。

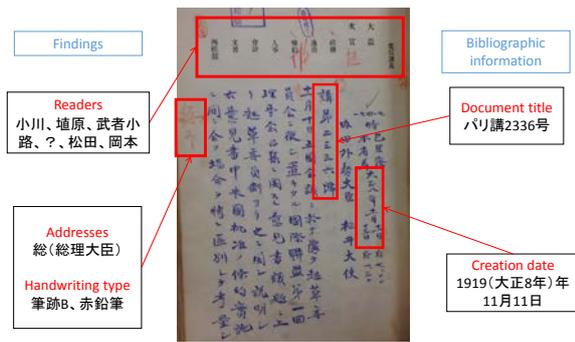


Fig. 6 Example of diplomatic document

各来電には史料名と作成日の他、史料上部にはその来電を閲覧した人物のサイン、史料左部にはその来電の送付先が記述されている。また送付先に訂正線が記述されている来電や異なる筆跡で新たな送付先が追記されている来電が存在する。よって対象史料群に対して閲覧者や筆跡に関する情報をメタデータとして整理し、それらを用いて史料を分類・比較することで史料の分析を行う。

5.4 メタデータフィールドの定義

複数研究者による史料の整理を可能とするため、まず史料に付与するフィールドを二種類の目的で定義した。一点目はウェブや文書の書誌的な情報を記述するための語彙「Dublin Core」を用い、表題や作成日等のフィールドを定義した。二点目として、本実験では送付先の筆跡に基づいた人物特定を行うため、来電の「送付先筆跡」や、筆跡に基づいた来電の特徴を分類するため、「閲覧者」「送付先」等の読み取り可能な情報をフィールドとして定義した。

5.5 史料の収集

本実験では技術者 Y、および史料収集補助学生 2 名が 5.3 で挙げた史料を外交史料館にて 3 日間撮影し、それらをシステムに登録した。登録時には簿冊毎に「グループ」フィールドに簿冊名を与え、複数画像を一括で登録した。

5.6 史料の整理

研究者 X と技術者 Y は史料収集と並行して、5.4 で定義したフィールドに基づいて史料の整理を行う。本プロセスは Fig. 4 で示したインターフェースを用いて行い、Fig. 7 に示すようなメタデータを各来電に付与した。

項目	内容	公開範囲	編集権限	更新者	編集
グループ	五国会議/第3巻	[アイコン]	[アイコン]	yoshida 2015/12/17	更新
送付先指示	総陸海参軍電伊	[アイコン]	[アイコン]	yoshida 2015/12/17	更新
送付先指示筆跡	印	[アイコン]	[アイコン]	yoshida 2015/12/17	更新
作成日	1919 年 10 月 11 日	[アイコン]	[アイコン]	yoshida 2015/12/17	更新
送付先訂正	総陸海参軍	[アイコン]	[アイコン]	yoshida 2015/12/17	更新
ノート	埴原の筆跡と訂正線の筆跡が一致	[アイコン]	[アイコン]	nakamura nakamura 2014/1/19	削除
閲覧者	澤田, 埴原, 武者小路, 松田, 岡本	[アイコン]	[アイコン]	yoshida 2015/12/17	更新

Fig. 7 Example of registered metadata

各来電には表題や作成日(図中①)といった書誌情報と、「澤田、埴原、…」といった閲覧者(図中②)、及び「筆跡 B、赤鉛筆」や「印」といった送付先指示筆跡(図中③)といったメタデータを蓄積した。また本実験では研究者 X と技術者 Y を予め同じグループに割り当てたため、各メタデータの公開範囲、編集権限を「グループ」として登録することで、それぞれが登録した史料やメタデータを相互に閲覧・編集することができる。

5.7 史料の分析

5.6 における史料整理の結果、送付先指示筆跡として A、B、S、印の 4 種類が確認された。これら各筆跡の人物を特定するため、筆跡と作成日の関係、および筆跡と閲覧者の関係の 2 点の観点から分析を行った。

5.7.1 筆跡と作成日の関係に基づいた分析

まず来電を時系列に並び替えることで、筆跡の出現時期に基づいた分析を行う。この分析結果を Fig. 8 に示すが、これは技術者 Y がプラグインを開発し、471 件の来電から筆跡と作成日に関するメタデータを引数とし、月別の送付先指示筆跡の出現頻度を出力したものである。

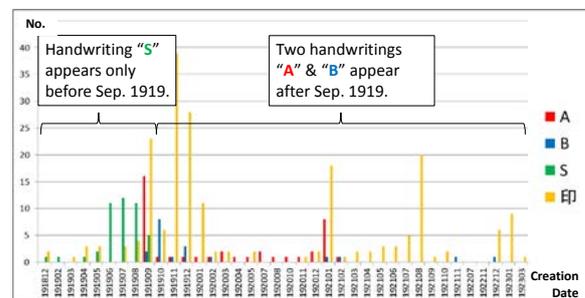


Fig. 8 Frequency of target handwritings by months

この結果から筆跡 S は 1919 年 9 月以前にのみ登場していることが確認できる。これは幣原喜重郎(以下、幣原)が外務次官を務めていた時期であり、当該期間以前では外務次官が送付先を決定しているという、研究者 X が先行研究で立案した仮説を支持する結果となった。また埴原が

外務次官に就任した9月以降では、少なくとも筆跡 A、B の二種類の筆跡によって送付先が指示されていることが視覚的に確認できるため、外務次官以外の人物が来電の送付先指示に関わっていることがわかる。また当該期間では「印」による送付先指示が増大していることが確認できるが、これは「革新同志会」らの若手が企図した文書処理改革に起因するという新たな仮説が立案された。

### 5.7.2 筆跡と閲覧者の関係に基づいた分析

次に筆跡別の来電における閲覧者の登場割合に基づいた分析を行う。これは研究者 X が史料の整理プロセスにおいて、埴原が来電閲覧時に記入するサインと筆跡 B が類似するという気づきを得たためである。この分析結果を Fig. 9 に示すが、これについても 5.7.1 と同様に技術者 Y が開発したプラグインによって出力した結果である。

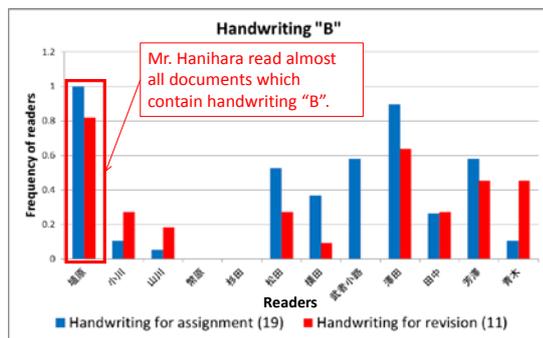


Fig. 9 Frequency of readers

この結果から筆跡 B によって送付先が指示 (19 件) または訂正 (11 件) されたすべての来電において、埴原が閲覧していることが分かる。これは研究者 X が立てた筆跡 B が埴原であるという仮説を支持する一つの量的な結果である。さらにこの発見に基づき、筆跡 A による来電に対して同様の分析を行った結果、電信課長であった澤田が高い割合で閲覧していることが分かった。また筆跡 A と澤田の閲覧時のサインを比較した結果、比較的高い精度でこれらが一致していることも確認できた。この発見から埴原次官時代には電信課長である澤田が送付先を仮決定し、埴原が訂正することで最終的な送付先を決定していたという仮説を立案することができた。またこの仮説の検証を目的として、5.7.1 で行った月別の筆跡の出現頻度を分析した結果、送付先訂正筆跡として埴原の筆跡 B のみが確認され、澤田の筆跡 A による訂正は 0 件であることが確認でき、立案した仮説を支持する結果を得ることができた。

## 6 考察

### 6.1 史料の収集

本研究ではクラウド上で史料や調査結果を管理することで、複数研究者が共同で史料を収集可能な環境を構築した。実験では、複数研究者が分担して撮影した史料をシステムに登録することで、500 件弱の史料を短時間で収集することが可能となった。

### 6.2 史料の整理

史料の整理プロセスでは複数研究者が共同で調査結果を蓄積することで、各研究者単独による整理プロセスと比較して労力を軽減することができた。また従来別ファイル

で管理されていた史料と調査結果がメタデータによって一意に紐づけられ、史料の閲覧と同時に史料の整理が可能となる等の効率化が図られた。

### 6.3 史料の分析

史料の分析では蓄積されたメタデータを利用可能な環境を構築することで、研究者の目的に応じた史料の分類や比較、計算機による分析が可能となった。これにより、蓄積した知見の可視化を通じて、研究者の新たな知見の創出へ寄与することができた。また歴史研究者が構築した仮説に対して、情報技術者が量的な検証を行うといった異なる分野の研究者による協調的な研究が可能となった。

### 6.4 来電における標準メタデータセットの提案

本実験では来電に対して表題や作成日等の書誌情報の他に、来電の送付先や閲覧者の情報をメタデータとして付与し、これらに基づいた分類や比較等の分析を行うことで、来電の送付先決定者の人物特定が可能となった。よって本研究では、来電における「送付先」「閲覧者」、およびそれらの「筆跡」を来電における標準メタデータセットとして提案する。一方、膨大な史料へのメタデータの付与には多大な労力を要するため、本研究で提案したように、計算機によるメタデータの自動抽出を行うことで史料情報を効率的に付与する環境が必要となる。

## 7 結言

本研究ではセマンティック・ウェブ技術を用いて書誌情報や調査結果をメタデータとして記述し、それらの利用による史料分析を支援するシステムの開発を行った。特に各研究者の負荷や暗黙知による属人性等の課題の解決を目的とし、複数研究者による研究を可能とする環境を構築した。また開発したシステム上で複数研究者による実験を行うことで、史料収集・整理プロセスを省力・効率化し、新たな仮説の立案や量的な検証等を可能にした。今後はより多くの研究者を対象とした実験を行うことで提案手法の評価や改善を行う。また企業内文書や電子教材といった他分野の資料管理に開発したシステムを応用することで、本提案手法の他分野への発展の可能性を検証する。

### 文献

- 1) N. Ur-Rahman, J.A. Harding : Textual data mining for industrial knowledge management and text classification: A business oriented approach, Expert Systems with Applications, 39, 4729-4739, (2012).
- 2) Tobin, W. Kenneth : Using a patient image archive to diagnose retinopathy, Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, 5441-5444, (2008).
- 3) 研谷紀夫 : 文化資源オントロジの構築とその活用(<特集>第 15 回(2007 年度)年次大会(研究報告会&総会)), 情報知識学会誌, 17, 129-134, (2007).
- 4) 中村寛:セマンティックウェブ技術を用いたデジタルアーカイブ史料研究支援システムの構築, 卒業論文, (2011).
- 5) 三浦崇 : 歴史学研究支援システムの構築, 情報処理学会研究報告,(1997).