

セマンティック・ウェブ方法論を適用した 由緒帳データベースの提案

林 正治^{*1} 堀井洋^{*2} 堀井美里^{*3} 沢田史子^{*4} 宮下和幸^{*5} 中野節子^{*5}

^{*1}一橋大学情報基盤センター ^{*2}北陸先端科学技術大学院大学知識科学研究科

^{*3}金沢大学社会環境科学研究科 ^{*4}金沢星稜大学大学総合研究所 ^{*5}金沢大学人間社会研究域

我々はこれまで「先祖由緒井一類附帳」と呼ばれる歴史資料のデータベース化を試みてきた。この試みの中で、それぞれ目的の異なる三種類のデータベースを構築してきた。しかしながら、同一の史料を対象にしたデータベースであれば統一的に扱えた方がよい。本稿では、三種類のデータベースの概要と課題について述べ、セマンティック・ウェブ方法論を応用することでそれらのデータベースを統合する手法を提案する。

A proposal of Historical materials database creation method on semantic web

Masaharu Hayashi^{*1} Hiroshi Horii^{*2} Misato Horii^{*3} Ayako Sawada^{*4}
Kazuyuki Miyashita^{*5} Setsuko Nakano^{*5}

^{*1}Information and Communication Technology Center, Hitotsubashi University

^{*2}School of Knowledge Science, Japan Advanced Institute of Science and Technology

^{*3}Graduate School of Socio-Environmental Studies, Kanazawa University

^{*4}Research Institute, Kanazawa Seiryō University

^{*5}College of Human and Social Sciences, Kanazawa University

We attempted to construct a historical materials database and created three type databases from one historical material. In generally, historical database reflect a personal viewpoint. Therefore, various types of database are constructed from one historical material. However, these databases should be integrated into one database. In this paper, we proposed a historical database creation method for integration of small historical databases on semantic web.

1. はじめに

現在、我々は加賀藩の藩制組織構造の解明を目的とした、金沢市立玉川図書館近世史料館所蔵「先祖由緒井一類附帳」（以降、由緒帳と略す）のデータベース化プロジェクト¹を推進中である。

由緒帳は前田家に関する歴史を編纂するために明治初年から御家禄方および前田家編輯方が収集した「加越能文庫」と呼ばれる史料群の一つで、加賀藩士の由緒や家族関係を知る上で重要な史料である[1]。しかしながら、これまで由緒帳の全容解明は殆ど行われておらず、その特徴を明らかにする必要がある[2]。

本プロジェクトでは由緒帳のデータベースを構築・分析することでその全容解明を目指している。これまでに、由緒帳目録データベース、由緒帳研究データベース、研究ノートデータベ

ースの設計・構築を始めた。

一般的に、研究利用のためのデータベースは、歴史学研究者の背景知識・研究方法論に深く依存するため、個別に構築されることが多い[3,4]。

同一の史料を対象としたデータベースであれば、統一的に扱えた方がよい。ただし、個別的把握を特徴とする歴史学では、各データベースの構造を統一したり、統合したりすることは難しい。八村氏が指摘するように、ハイパーメディア型の歴史データベース[3]の構築が必要である。

一方、ハイパーメディア型のデータベースとして、ウェブにおけるデータ処理を目的としたセマンティック・ウェブに関する研究がある。従来のウェブでは個別分散的に作られるウェブページ内のデータやページ間の関係性を統一的に表現する手法が欠けていた。セマンティック・ウェブはウェブのメタデータ記述方法や処理方法を定義することで、これらの課題の解決を目指している。

我々はこのセマンティック・ウェブ方法論を応用すればハイパーメディア型の歴史データベースを実現することが可能ではないかと考えた。

¹平成21年度科学研究費補助金 基盤研究(C)「由緒帳データベースによる藩政組織構造の解明に関する提案」(研究代表：堀井洋)の支援による。

そこで、本稿ではセマンティック・ウェブ方法論を適用した由緒帳データベース構築手法を提案する。

まず、これまで我々が構築してきたデータベースを観察し、その課題を明らかにする。そして、それらの課題を解決するセマンティック・ウェブ方法論を適用した由緒帳データベース構築手法について述べる。

本稿の構成は次の通りである。まず、第二章では我々がこれまで設計・構築してきたデータベースを概観し、その課題を明らかにする。第三章では提案手法であるセマンティック・ウェブ方法論を適用した由緒帳データベース構築手法について述べる。第四章ではまとめと今後について述べる。

2. 三種類の由緒帳データベース

2. 1. 由緒帳目録データベース

由緒帳の目録は『加越能文庫解説目録 上巻』[5]に含まれている。目録では、由緒帳を当主の氏名の五十音順で並べており、各由緒帳に対して九項目の解説が付けられている。歴史学研究者はこれらの項目を手掛かりにして、目的の由緒帳を探し出す。表1に各項目の詳細を示す。このような目録は多くの歴史学研究者の膨大な努力による作られており、史料へのアクセシビリティを格段に向上させる。しかしながら、書籍ベースの目録には幾つかの欠点がある。

まず、出版に関わる問題から、内容の修正が困難である。例えば、誤植や史料の読み違いを部分的に修正することは可能だが、索引を修正するなどの大幅な修正は難しい。たとえ、出版元に修正箇所が報告されたとしても、直ちに反映されることはない。

つぎに、項目の内容が歴史学研究者の利用方法を反映していない問題がある。『加越能文庫解説目録 上巻』内の目録の場合、その内容は由緒帳成立時の当主を基準に作られている。しかしながら、由緒帳成立時の当主ではなく、その祖先を調べることの方が多。その場合、別名や通称が大きな手掛かりとなるのだが、目録にはそのための索引は存在しない。そのため、幾つか候補を列挙し、史料自体を解読しながら目的の由緒帳を探し出すことが必要になる。

目録データベースは、これらのことを背景に設計・構築された。

本システムは、由緒帳を目録の各項目で検索することができる。また、各項目の索引を作成し、キーワード補完機能として利用することで、索引の不足を補っている(図1参照)。

本システムでは目録データベースと史料画像

データベースを関係付けている。それにより、場所や時間の制約を受けずに史料にアクセスすることが可能となる(図2参照)。さらに、目録の内容を電子データとして記録することで、内容の修正が書籍の場合に比べて容易になる。

表1 由緒帳目録の項目

項目名	説明
帙番号	保管した箱(帙)の番号
当主の氏名	由緒帳成立時の当主氏名
別名・通称	当主の別名や通称
父の名前	当主の父氏名
秩禄	明治以前の知行
成立年代	由緒帳の成立年代
年齢	由緒帳成立時の年齢
檀那寺	檀那寺の名称
丁数	由緒帳の丁数
冊数	由緒帳の冊数

別名・通称：

年齢：

成立年代(Mは明治)：

丁数：

図1 キーワード補完機能

以上が、目録をデータベース化することの利点である。書籍であった目録をデータベース化することで、その利便性は飛躍的に向上する。

本システムを歴史学研究者に実験的に公開したところ、検索機能の利便さは実感してもらうことができた。

しかしながら、目録にある情報だけでは、使い物にならないという厳しい指摘も受けた。目録にある情報だけでは、その史料が使えるかどうかを判断することは難しく、結局のところ、史料原本を調べる必要が出てくるということであった。たとえ、史料画像が関連付けられていても、その内容を検索することはできないため、現状の書籍ベースの目録と利便性は大きく変わらないということであった。

これらのことを総合すると歴史学研究者の要求に沿ってカスタマイズ可能なデータベース、すなわち、データベースの個性化[6]を実現する必要がある。

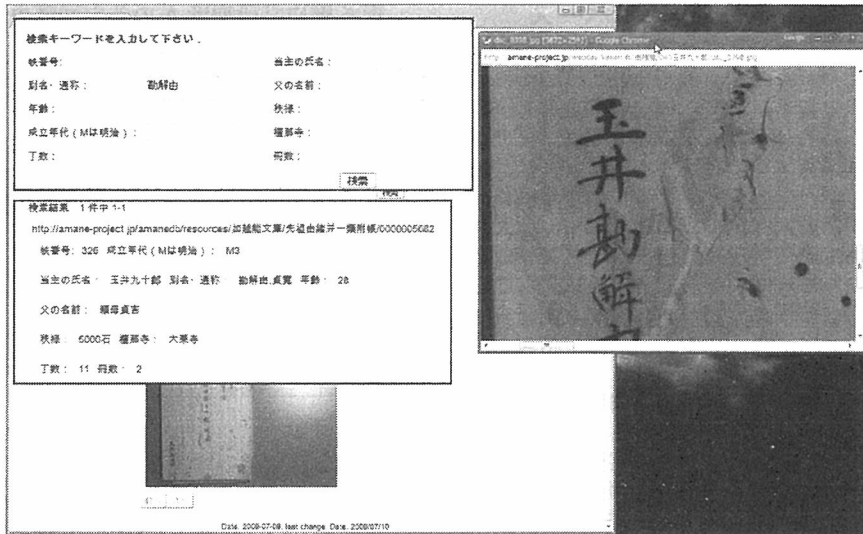


図2 目録データベースの画面写真

2. 2. 由緒帳研究データベース

由緒帳研究データベースは由緒帳による加賀藩制構造の解明を目的に、歴史学研究者により設計されたデータベースである。前述の目録データベースを使うことで、秩禄や檀那寺などを利用して、基本的な統計処理を行うことはできる。しかしながら、加賀藩士の家族関係や姻戚関係からくる、役職の変遷など、歴史学研究者の研究テーマに沿った分析は不可能である。

そこで、由緒帳研究データベースでは、教人の歴史学研究者が集まり、テーマについて議論した後、データベースの設計を始めた。

テーマは人持組以上の家における役職の変遷と姻戚関係の解明である。このテーマに沿って「個人基本データテーブル」（表2参照）と「詳細データテーブル」（表3参照）を設計した。

表2 個人基本データテーブルの例

項目名	例
続柄	本人
出身・肩書き	玉井頼母嫡子
本国	濃州
生国	加州
菩提寺	禅宗大乘寺
姓	玉井
いみな	貞寛
読み仮名	サダヒロ
別名	勘解由, 九十郎
没年月日	明治三年
没年齢	二十八
備考	由緒帳作成時年齢

表3 詳細データテーブルの例

項目名	例
開始年	元治元年
月	三月
役職・賞罰等	小松御城番兼帯
終了年	明治三年
月	三月
コメント	

データ入力作業を効率化するために、個人基本データテーブルと詳細データテーブルのテンプレートを作成し、データ入力作業者に配布した。また、解読対象の由緒帳はサーバ上で共有されており、参照可能な状態にしている。

また、由緒帳の解読作業およびデータ入力作業は日本史を専攻する金沢大学の学生6名が行い、6名の歴史学研究者がデータチェックをする体制で実施する。現在は、解読作業の割り当てを行い、解読作業を始めた段階である。

由緒帳研究データベースの構築は始まったばかりであり、今後の状況次第ではテーブルの変更もありうる。歴史学研究は試行錯誤的に行われることもあり、柔軟な対応が求められる。

もう一つの問題として、歴史学研究者にとり、データベースの構築自体は研究の本質ではないことが挙げられる。そのため、データベースの構築だけに時間を取られるわけにもいかず、必然的にデータベースの規模も制限されてくる。

データベースを構築できるのは、史料を解読できる歴史学研究者だけであり、構築に関わる作業支援が求められる。

2. 3. 研究ノートデータベース

前述してきたデータベースは、複数の歴史学研究者の協力が必要なものであり、その規模も膨大なものとなる。また、概念の同意も必要になってくるため、個別的把握を特質とする歴史学分野では構築すること自体が難しい。

一方で、個々の歴史学研究者は、通常の研究業務として、膨大な数の史料を解読し、研究ノートや私的な歴史学研究データベースを作成している。これらは歴史学研究者ごとに異なる視点で毎に異なる視点で記録されており、決まった形式はない（以降、このような歴史情報を非定型歴史情報と呼称する）。

非定型歴史情報の多くは、共同研究のような特殊な場合を除き、共有されることはない。しかし、同じ史料を対象とした非定型歴史情報の中には、共通部分が含まれる可能性が高くなる。また、これら非定型歴史情報を統合することで、ある史料を対象とした歴史学研究データベースの完成度を高めることが可能となる。例えば、前述した由緒帳目録データベースと由緒帳研究データベースの個人基本データテーブルは相互補完可能な関係にある。

本研究では、実際に二人の歴史学研究者から由緒帳に関する非定型歴史情報を提供してもらい、その可能性を探っている（以降、前者を歴史学研究者 A、後者を歴史学研究者 B と記す）。

歴史学研究者 A による非定型歴史情報は由緒帳成立時の当主とその父親について、直臣・陪臣の別と役職の変遷について網羅的に調べたものである。これらの非定型歴史情報は『加越能文庫解説目録 上巻』のコピーに直接メモされている（図 3 参照）。

表 4 『加越能文庫解説目録 上巻』の例

項目名	内容
帙番号	52
当主の氏名	上杉 半平
別名・通称	易彰, 半之進
父の名前	梶之助易順
秩禄	9 人扶持
成立年代	M3(明治 3 年)
年齢	37 才
檀那寺	本長寺
丁数	5 丁
冊数	2 冊

上杉半平(易彰, 半之進) 梶之助易
 家 9人扶持 M3 37才 本長寺 5丁

前田三吉(三左衛門, 土佐守) 家来 (同家来)

図 3 研究メモの例

表 5 歴史学研究者 A の非定型歴史情報

項目名	内容
本人	前田三吉 (三左衛門, 土佐守) 家来
父	同家来

表 6 歴史学研究者 B の非定型歴史情報

項目名	内容
由緒帳コピー有	○
由緒帳有無	○
帙・名前	52・上杉半平
読み仮名	うえずぎ
氏名	上杉半之進
年齢	37
身分	給人
禄高	9 人扶持
履歴	嘉永 5 年 (7 人扶持)、安政 5 年加増 (2 人扶持)
居住地	下屋敷
備考	早田弥作次男
先祖	2 世
由来・由緒	直方弟の妾腹
土佐守家出仕時点	直方, 2 世, 小将

もう一方の歴史学研究者 B による非定型歴史情報は、人持組頭前田土佐守家の家臣を調べたものである。これらの非定型歴史情報はマイクロソフト社のエクセルデータとして記録されている。

例えば、『加越能文庫解説目録 上巻』に記された「上杉半平」（表 4 参照）について、両者の非定型歴史情報を見てみると次のような違いがある。

表 5 は歴史学研究者 A による加賀藩士「上杉半平」に関しての非定型歴史情報である。歴史学研究者 A は、『加越能文庫解説目録 上巻』のコピーに直接メモしているため、その内容は非常に簡素であるが、「上杉半平」は前田家の陪臣であり、「前田三吉（三左衛門，土佐守）」の家来，すなわち前田土佐守家の家臣であることが解る。また、「上杉半平」の父もまた前田土佐守家の家来であることが解る。

表6は歴史学研究者Bによる「上杉半平」に関する非定型歴史情報である。歴史学研究者Aと比較して、かなり詳細な記録であることが解る。「上杉半平」は「上杉半之進」として、前田土佐守家に給人として使っていたことが解る。ただし、由緒帳コピー有や由緒帳有無など個人的な情報も含まれており、ノイズも多い。

両者の非定型歴史情報の共通部分は誰の家臣であるかを調べている点のみである。しかしながら、これだけの情報を調べるためにも、史料自体を解読しなくてはならず、両者のどちらかが共有されていれば、その手間を省けたはずである。

研究ノートデータベースの利点は歴史学研究者の研究蓄積を共有する事で、史料発見に関する負担を軽減することができる点である。

例えば表4と表5、そして表6のデータが統合されていれば「早田弥作次男」に関する情報を探している歴史学研究者に対して、史料の所在やその人物が「上杉半平」であること、そして前田土佐守家の家臣であることを知らせることが可能となる。

しかしながら、非定型歴史情報の共有を実現するには幾つか課題がある。

まず、非定型歴史情報を集約させるための情報が必要な点である。本稿での例では、歴史学研究者Aと歴史学研究者Bの両者ともに『加越能文庫解説目録 上巻』の目録情報に関係付けられていたため、統合することができた。しかしながら、この様な情報がなければ、非定型歴史情報の統合は困難である。

つぎに、非定型歴史情報の信用度の問題がある。非定型歴史情報は個人的な記録であるため、誤りが存在する可能性もある。また、個人的な記録であるため、共有を望まない歴史学研究者も出てくる可能性がある。非定型歴史情報とそれを作成した歴史学研究者双方の信用度を担保するための仕組みが必要である。

現在、研究ノートデータベースについては、歴史学研究者Aによる非定型歴史情報のデータベース化作業を進めている段階である(2009年11月時点で4091件)。

3. セマンティック・ウェブ方法論による由緒帳データベースの構築

これまで述べてきたデータベースから、歴史学分野におけるデータベースには次のことが求められる。

まず、歴史学研究者が自由にカスタマイズ可能なデータベースの個性化を実現する必要がある。歴史学研究者の多様な視点の一つのデータ

ベースに反映させることは困難である。従って、歴史学研究者の要望に応じて、データを追加・編集でき、検索項目を必要に応じてカスタマイズ可能なデータベースが必要である。

そして、データベースの構築に関わる負担を分散するための仕組みが必要である。

つまり、個別にデータベースを定義することが可能でありながら、定義されたデータベースを統合できる仕組みを実現する必要がある。これを実現する仕組みがハイパーメディア型のデータベースである。個別分散的に定義されたデータベースをハイパーリンクにより接続することで、それらを統合して利用することが可能になる。

我々はセマンティック・ウェブ方法論を用いれば、このような仕組みを実現できるのではないかと考えた。

セマンティック・ウェブではリソースという単位に対して、メタデータを付与することで、情報システムによるリソース処理の高度化を試みている。リソースとはURI¹で識別されるものを意味しており、ウェブページに記述されたデータであったり、画像データであったりする。セマンティック・ウェブではメタデータの付与手法や処理手法を定義している。その手法に従えば、リソースをデータベースのように検索することや、オントロジを利用した意味処理も可能となる。

セマンティック・ウェブの肝となる部分は、メタデータの記述方法を定義している点とそのメタデータをリソースの作者が作成する点である。そして、そのリソースはURIで識別される点である。

言い換えると、自由にリソースを作成しても、そのリソースを処理するためのメタデータを用意しておけば、セマンティック・ウェブの世界ではそのリソースは処理可能であることが保証される。そして、そのリソースはURIにより識別することが可能で、セマンティック・ウェブのどこからでも参照可能である。

ただし、セマンティック・ウェブの世界においても、リソースが独立して存在するのでは意味がない。それぞれのリソースがメタデータにより接続されて初めてデータベースのように利用できるようになるのである。

そこで、本研究では由緒帳の目録データベースを起点として、セマンティック・ウェブ方法論の適用を考えた。

まずは目録データベースの1レコードに対して、一意に識別可能なURIを付与する。レコー

¹ Uniform Resource Identifier の略。

ドに定義された各属性値はメタデータとして記述する。表 4 をメタデータとして表現すると図 4 のようになる。この図では <http://amane-project.jp/amanedb/resources/加越能文庫/先祖由緒并一類附帳/0000000831> がレコードを識別する URI となる。

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-
rdf-syntax-ns#"
  xmlns:j.0="http://amane-
project.jp/amanedb/schemes/
  加越能文庫解説目録/">
<rdf:Description rdf:about="http://amane-
project.jp/amanedb/
resources/加越能文庫/先祖由緒并一類附帳
/0000000831">
  <j.0:年齢>37</j.0:年齢>
  <j.0:秩禄>9人扶持</j.0:秩禄>
  <j.0:冊数>2</j.0:冊数>
  <j.0:父の名前>梶之助易順</j.0:父の名前>
  <j.0:秩番号>52</j.0:秩番号>
  <j.0:当主の氏名>上杉半平</j.0:当主の氏
名>
  <j.0:丁数>5</j.0:丁数>
  <j.0:成立年代>M3</j.0:成立年代>
  <j.0:別名>
  <rdf:Bag>
    <rdf:li>易彰</rdf:li>
    <rdf:li>半之進</rdf:li>
  </rdf:Bag>
  </j.0:別名>
  <j.0:檀那寺>本長寺</j.0:檀那寺>
</rdf:Description>
</rdf:RDF>
```

図 4 メタデータの例

この URI で識別されるリソースを起点に、歴史学研究者 A が作成したリソースと歴史学研究者 B が作成したリソースを繋ぎ合わせることで、小さなセマンティック・ウェブが形成される。

現在は由緒帳目録データベースの開発にセマンティック・ウェブ方法論の適用を終えたところである。今後は由緒帳研究データベース、研究ノートデータベースへの適用し、それらを統合する予定である。

4. おわりに

本稿では、これまで構築してきた三種類の由緒帳データベースの課題を明らかにした。その上で、セマンティック・ウェブ方法論を適用した由緒帳データベース構築手法について述べた。

歴史学分野のデータベースは歴史学研究者の研究方法論に依存しており、その構造は多様である。しかしながら、そのようなデータベースを共有されることを想定した形で構築していくことで、それらのデータベースを統合することが可能となる。

また、提案手法は非定型歴史情報の統合と共有を実現するための第一歩になると考えている。多くの歴史学研究者による非定型歴史情報は共有されずに、失われるままである。このような非定型歴史情報には歴史学研究データベースにとって、また、他の歴史学研究者にとっても、有益なデータが含まれている。非定型歴史情報が共有されることにより、歴史学研究者はより自身の研究を深めることができる。

しかしながら、提案手法を実現するには、リソースの管理手法やデータの入力手法、そして信頼性を保証する仕組みなど解決しなければならない課題は多い。今後も由緒帳データベースの構築を続けながら、歴史学データベースのためのセマンティック・ウェブの実現を目指していきたい。

謝辞

本研究にあたり、前田土佐守家資料館の竹松幸香氏、石川県立図書館の鷲澤淑子氏には、歴史学や史料のデータベース化について貴重な助言を頂いた。元金沢市史編集委員の長山直治氏には、本研究のために貴重な研究成果を提供して頂いた。本研究の史料電子化は、総務省の戦略的情報通信研究開発推進制度の地域 ICT 振興型研究開発「ユニバーサルな知識表現による地域歴史観光 ICT の研究開発」（平成 20 年～21 年度）の一部として行われており、歴史研究データベース及びそのための情報システムのための研究は科研費（基盤研究(C)21500247）の助成を受けている。関係者各位に謹んで感謝の意を表する。

参考文献

- [1] 金沢市史編さん委員会編：金沢市史 資料編 5（近世 3），金沢市，2003。
- [2] 宇佐美孝：加賀藩「先祖由緒并一類附帳」について—由緒帳の成立と変遷—，加能地域史，Vol.7，pp.1-2，1982。
- [3] 八村広三郎：人文科学とデータベース，情報処理学会誌，Vol.38，No.5，pp.377-382，1997。
- [4] 宝珍輝尚，都司達夫：考古学データベースの現状と課題，電子情報通信学会誌，Vol.85，No.3，pp.171-175，2002。
- [5] 金沢市立図書館編：加越能文庫解説目録 上巻，金沢市立図書館，1975。
- [6] 福田豊彦：パソコンは便利な文房具だ，日本歴史，No.597，pp.42-49，1998。