

## 歴史情報の視覚的表現を指向した歴史学習素材の作成支援プラットフォーム

馬場裕子 永森光晴 杉本重雄  
筑波大学大学院図書館情報メディア研究科

今日、Web 上にも多くの学習用アプリケーションがあるが、それぞれのアプリケーション間で扱われているデータは独自のフォーマットで表現されているため、学習用データの再利用や共有が難しい。そこで本研究では、高等学校で行われる日本史の科目を対象に、Web 上でのデータの流通が可能な形式で、アプリケーション間で共通に使用できる学習内容データの記述のためのフォーマットの設計を行った。また、フォーマットに従って作成したデータを用いて、地図をインターフェイスに歴史を学べる歴史情報閲覧システムの構築を行った。加えてシステムの学習用データの拡充を目的に、既存のリソースを用いた半自動生成の手法の提案と、エディタの作成を行った。

### A platform based on RDF/OWL to support description of learning resources with visual representation of historical information

Yuko Baba Mitsuharu Nagamori Shigeo Sugimoto  
Graduate School of Library, Information and Media Studies, University of Tsukuba

Despite the number of software agents available for use in educational settings, sharing content is still a problem. Too many systems use their own unique schemes. To tackle this problem in a history context, we developed a metadata scheme in RDF/OWL that could be used by everyone. We also developed a historical Information browser called V-HAWKS and a metadata editor. In this paper, we propose a way to create metadata automatically for history resources found on Wikipedia.

#### 1. はじめに

今日、情報技術が発展して Web が爆発的に普及し、ビジネスや生活など様々な場面においてその恩恵が得られるようになった。教育分野においてもそれは例外ではない。その例の一つに、e ラーニングが挙げられるが、時間や地理的な制約から逃れることができ、学習者が無理のない適切な進度で自己学習が行えるなどの利点がある。また、教師が授業で用いた教材をウェブ上にアップロードし、その他の教師が授業で教材を再利用できるようにしているサイト (NICER<sup>[1]</sup>など) もある。このように Web の技術を活用した教育のためのサイトが多く存在するようになったが、これらを相互利用するのは難しい。それは、個々のサイトで用いる情報の形式が、それぞれのアプリケーションごとに異なり、アプリケーション間での情報の共有や相互利用が困難なためである。この問題を解決するには、Web 上でそれぞれのアプリケーションが利用できる共通の学習データ記述フォーマットが必要である。

これまで中等教育で行われている各科目の学習方法は、学校で行われる授業による学習や、教科書や参考資料集を用いた自己学習などが主流であった。そこで我々は、高等学校で行われている日本史の科目を対象に、Semantic Web<sup>[2]</sup>技術を用いて、歴史の学習内容の記述フレームワークの設計を行った。その形式に従い作成したデータをベースに、歴史を地図上で学習でき

る歴史情報閲覧システムの構築を進めてきた。本稿では、1) 設計した記述フレームワークの概要、2) 史実メタデータの生成手法、3) 歴史情報閲覧システム V-HAWKS<sup>[18]</sup>のアーキテクチャについて述べる。

#### 2. Semantic Web を利用した歴史学習

今や Web は爆発的に普及しその必要性は広く認識されているが、現在の Web は多くの問題を抱えている。例えば、Web 上のリソースに広く用いられている HTML 形式のデータは、情報の分割の粒度が粗く、また型付けがされていないという問題があり、これが情報の機械解釈を困難にさせている。また、HTML ではノード間を結ぶリンクに対して情報が付与されていないため、ノード間のつながりの意味は人間がリンクをたどって確認しなければわからない。このような問題を解決するために、Semantic Web 技術が注目されている。代表的なものには Resource Description Framework(RDF)<sup>[3]</sup>が挙げられる。RDF とは Web 上で利用するメタデータのためのデータモデルとその表現形式を定めるフレームワークであり、Semantic Web に関わる技術として広く利用されている。また、概念やリソースの関心の精密な記述や、推論可能な論理の記述を目的とした Semantic Web の活動によって Web Ontology Language(OWL)<sup>[4]</sup>が開発されている。そこで本研究では、RDF と OWL を用いて歴史の学習支援に必要な情報を記述するためのフレームワークを定義する。

一方、史実の流れや関連性を学習者にわかりやすく解説するために、地図上に史実を視覚的に表現することが有用である。これは、歴史情報と地理情報は密接な関連があるためであり、従来から歴史地図帳が多く存在することからもわかる。これまで、安価に地図を表示することは容易ではなかったが、現在では Google Maps<sup>[5]</sup> のような Web 上でどこからでも利用できるサービスを用いて、地図情報と歴史情報を重ね合わせて表示するといったことも簡単に実現できるようになった。

以上の観点に基づき、本研究では、中等教育の歴史の学習支援を目的として、歴史の学習内容に関する情報（史実のメタデータ）とそれを地図上に視覚的に表現するために必要な情報（視覚情報のメタデータ）を記述するためのフレームワークを設計し、それに基づいて歴史情報閲覧システムの構築を進めた。

### 3. 地図上での歴史情報の視覚化

戦国時代のある武将の領土の変遷や人の移動などのように史実のもつ性質には様々な特徴があり、それぞれに適した地図上での視覚的な表現手法は異なっている。そこで、地図上での史実の視覚的表現の手法として、どのような種類があるのか検討する必要があると考えた。そのために、歴史地図帳や Flash などのアニメーションで地図上に史実を視覚化した Maps of War<sup>[6]</sup> で用いられている表現手法を最小単位に切り分けて、本研究で用いる視覚的表現の手法を考えた。その結果、本研究では地図上での史実の視覚的表現の手法として、以下に記す5つの視覚化手段を実現する。図1に5つの視覚化手段のイメージ図を示す。各視覚化手段の詳細は以下の通りである。

#### 1) 同時発生

同時期に起きた史実を視覚化する。同時期に起きたそれぞれの史実の発生した場所（地理的な位置）を地図上にマーキングし、アイコンをクリックすると史実に関する情報をウィンドウで表示することができる。例) 2001年9月に起きた同時多発テロ事件のような、同時期に起きた史実を視覚化する。

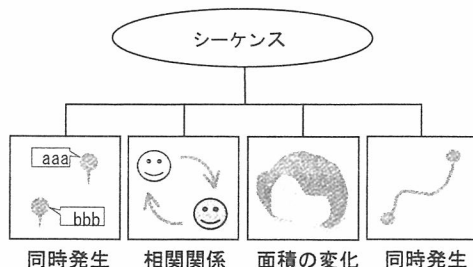


図1 本研究で提案する5つの視覚化手段

#### 2) 相関関係

歴史情報間の相関関係を地図上に視覚化する。歴史情報のもつ地理的な位置を地図上にマーキングし、関連のある歴史情報同士を、関連ごとに色分けした線で結んでグラフ化する。例) 歴史人物の人間関係のような、歴史情報間の相関関係を視覚化する。

#### 3) 面積の変化

史実の地理的な領域の変化を地図上に視覚化する。地図上に面積オブジェクトを用いて史実の地理的な領域を表現し、面積オブジェクトをアニメーションを用いて変化させていく。例) 東京湾の海岸線の変化のような領域の変化を視覚化する。

#### 4) 移動

ある歴史情報の移動を地図上に視覚化する。歴史情報の移動した経路を地図上に線で描き、その上にマークがアニメーションを用いて移動する。途中で移動を一時停止してウィンドウを出し、その地点で起こった史実に関する情報を表示することができる。例) 著作『奥の細道』において、松尾芭蕉がたどった足跡のような、歴史情報の移動を視覚化する。

#### 5) シーケンス

1) から 4) に示したそれぞれの視覚化手段は単純であるため、これらで表現できることは限られている。そこで、これらを組み合わせて歴史の流れをより豊かな表現方法で視覚化することができるシーケンスを定義する。

### 4. 先行研究

史実と視覚情報のメタデータスキーマの設計にあたり、史実を記述するためのフレームワークの設計に関する先行研究を調査した。以下に先行研究の研究概要と、スキーマの分析を示す。

#### Semantic Web for History (SWHi)<sup>[7]</sup>

SWHi プロジェクトは、RDF や OWL などの Semantic Web 技術を用いた、アメリカ史初期に起こったある史実に関する印刷物資料の効率的な発見に関する先行研究である。そのために、アメリカ史初期に刊行された印刷物資料を収集した『Early American Imprints, Series 1 : Evans』の書誌データを用意し、史実と印刷物資料のつながりを OWL を用いて記述している。SWHi で設計されているスキーマでは、1) 出版された印刷物資料とそれに関連のある史実との結びつき、2) 著者・出版年など印刷物資料の書誌事項、3) 史実のメタデータとして、時間、場所、関係者などの基礎事項が表現できる。SWHi のスキーマで定義されている語彙の数は多いが、それらは特定の時代、地域の史実（アメリカ史初期の史実）の記述に特化したものが多く、日本史のメタデータ記述には向いていない。

## Historical Events Markup and Linking Project (HEML)<sup>[8]</sup>

HEML プロジェクトでは、史実と地図上での史実の視覚化を表現するためのメタデータスキーマを設計し、XML 形式で記述したメタデータを基に XSLT や Java を用いて、史実のリスト表示や、年表・地図を用いた史実の視覚化を行っている。HEML で定義されているスキーマでは、史実とそれを地図上に視覚的に表現するのに必要な情報が記述できる。具体的には、歴史上のイベント（歴史的な何らかの事象）のメタデータとして、関係者や時間、場所などの基礎情報を記述することができる。場所に関する情報とは、ある特定の地点を示すための情報で、緯度・経度の組み合わせを用いて表現している。史実の地図上での視覚化は、この緯度・経度という情報を基に、史実が起きた位置を地図上にマーキングするという表現方法を用いて行っている。これは本研究で実現する視覚化手段の中の、同時発生に対応している。HEML のスキーマを用いて表現できる史実のメタデータは限られており、また地図上での表現手法も本研究の方が多く。

### 5. RDF/OWL を用いた史実記述のためのフレームワーク

史実と視覚情報のメタデータを RDF と OWL で記述するために、RDF スキーマと OWL を用いて、メタデータの記述に使用する語彙を定義した。RDF スキーマと OWL では同様の性質をもつクラスと、リソース間の関連性を明示的に表すプロパティなどの語彙を定義することができる。語彙の定義にあたり、まず、既存のメタデータスキーマを調査した。既存のメタデータスキーマとしては、広い分野で用いられている Dublin Core<sup>[9]</sup>や、人に関するメタデータの記述に用いられている FOAF<sup>[10]</sup>などがある。これらで定義されている語彙を分析し、利用できるものを用いて定義する。

史実と視覚情報のメタデータスキーマの設計にあたり、史実を表現するのに必要な語彙を検討した。まず歴史情報を属性ごとに分類して、分類ごとの対応表を作成し、対応表に属性間の関連性を列挙した。歴史情報の分類は、高等学校で用いられている日本史の教科書から、歴史的用語を抽出し、用語を性質ごとに分けて行った。またこのとき、日本十進分類法や、先行研究のスキーマの分類も参考にしている。作成した対応表の一部を表 1 に示す。対応表では歴史情報の各分類をそれぞれ縦と横に並べ、表の各項目には縦の分類を主体として捉えた際の横の分類との関連性を列挙した。例えば、縦の分類が国で、横の分類が人である場合、それらの関連性にはその国の著名人や国の首相（王）などを挙げている。この対応表を基に、史実のもつ

表 1 スキーマ分析のための各分類の対応表（一部）

	国	組織	人
国	国と国の条約 敵対関係 友好関係 植民地	国を設立した組織 国の組織機関	その国の著名人 国の首相(王) 国の首相(王)の時間的变化
組織	組織の加盟国	組織と組織の関係 組織間の関係の時系列的な変化	人の所属する組織
人	人の国籍 人の移動(国)	組織を設立した人 組織の代表者 組織に所属する人	人と人の関係 人間同士の関係の時系列的な変化

属性や史実間の関連性を分析し、史実と視覚情報の表現に必要な語彙を定義した。

次に本研究で定義したクラスについて説明する。表 1 で用いた分類を、さらに粒度を高くして性質ごとに分類した。それに基づき、歴史クラス、グラフィックスクラス、地図クラス、時間クラスを本研究で定義するメタデータスキーマの主要なクラスとして構成した。これらの主な 4 つのクラスをまとめて、全ての語彙が属するメタデータスキーマの中核となる Core クラスを定義する。以下にメタデータスキーマを構成する主な 4 つのクラスを説明する。

#### 1) 歴史クラス

歴史クラスでは、中等教育の日本史の学習内容を表現できるメタデータスキーマを定義している。本研究では、中等教育の日本史の学習支援を目的とするため、定義する語彙は、歴史の教科書に出てくる用語と、その用語間の関連性を表現できるものである必要がある。表 1 の分類を参考にクラスを定義し、またクラス間の階層構造を決めた。メタデータの記述が教科書の内容に対応するように、歴史クラスの階層化は教科書の区分に即するように行っている。歴史クラスのサブクラスには、人や団体を表現するクラスや、戦争、事件などのイベントを表現するクラス、文学や芸術などの文化を表現するクラスを主に定義した。また、史実間の様々な関連性を表現できるように、親、兄弟、敵など、関連性に関する用語を定義できる語彙を定義した。

#### 2) グラフィックスクラス

グラフィックスクラスでは、歴史クラスで記述する史実のメタデータを、地図上で視覚的に表現するためのメタデータスキーマを定義している。グラフィックスクラスのサブクラスには、1) 史実を地図上に視覚的に表現するために必要な、緯度・経度などの基本的な情報を表現できる視覚化クラス、2) 地図のズームレベルや線の色など、アニメーションの効果をカスタマイズすることができるアニメーション効果クラスを定義している。1) の視覚化クラスは 3 章で述べた 5

表2 プロパティの一覧表 (一部)

プロパティ	内容
dc:title	情報資源に与えられた名称
dc:description	情報資源の内容の説明・記述
dc:coverage	時空間情報についての記述
foaf:name	人物の名前についての記述
vhwk:visualize	視覚情報への参照
vhwk:locatedAt	地図情報についての参照
vhwk:causedBy	行動の主体についての記述
vhwk:startTime	時区間の始点についての記述
vhwk:endTime	時区間の終点についての記述
vhwk:participant	参加者についての記述
geo:lat	緯度についての記述
vhwk:transcription	漢字の読みについての記述

つの視覚化手段に対応する、同時発生クラス、相関関係クラス、面積の変化クラス、移動クラス、シーケンスクラスをサブクラスに定義した。例えば移動クラスのメタデータスキーマは、内容の基本事項(タイトル、概要、日時など)と、移動する主体となる歴史情報と、主体が移動する経路の情報について記述することができる。移動する経路の情報は、緯度・経度で示される地点を、移動する経路順に繰返し記述する、という形式でメタデータスキーマを決めている。

### 3) 地図クラス

地図クラスでは、地点やある領域など、地図上で地理的特徴のモデリングに関する情報を表現できるメタデータスキーマを定義している。地図クラスのサブクラスには、ある位置を表現する地点や、領域、ルートを表現するクラスを定義した。これらの各クラスでは、地理的特徴の情報を緯度・経度を用いてメタデータを記述する。地図クラスでは、Google MapsやGoogle Earth<sup>[11]</sup>に表示するポイント、線などの地理的特徴をモデリングして保存するKML<sup>[12]</sup>のスキーマを参考にしている。

### 4) 時間クラス

時間クラスでは、時間に関する情報を表現できるメタデータスキーマを定義した。時間情報の表現には、主に時代や世紀、年月日を用いて表現する方法がある。そこで時間クラスのサブクラスには、時代クラス、世紀クラス、期間クラス、曖昧な期間クラスを定義した。期間クラスと曖昧な期間クラスでは年月日の表記を用いて、ある一時点や期間、また開始時点などの時間の範囲について記述することができる。

次に、本研究で定義したプロパティについての説明する。表1の対応表に列挙してある史実の関係性から、性質が類似する関係性同士をまとめて、各クラスで汎用的に使用できるように、

一般性の高いプロパティを定義した。またそれぞれのプロパティに対してプロパティが関連づけるリソースがどのクラスに属するのかわかる情報も示し、各プロパティのもつ役割を明確にした。本研究で定義したプロパティには、Dublin CoreやFOAFの語彙も利用している。表2に定義したプロパティの一覧表を示す。接頭辞がdcとなっているものがDublin Coreで定義されているプロパティ、foafはFOAFで、vhwkは本研究で独自に定義したプロパティである。

以上に示したクラス、プロパティをOWLを用いて記述した。

## 5.1 メタデータの記述例

以下に本研究で設計したメタデータスキーマに基づいて、著作『奥の細道』における松尾芭蕉の足跡を表現する史実と視覚情報のメタデータを、RDF/XML形式で記述した例を示す。

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:foaf="http://xmlns.com/foaf/1.0/"
  xmlns:vhwk="http://pe.tsukuba.ac.jp/~s0821671#">
  <vhwk:Movement rdf:ID="route_of_hosomichi">
    <dc:title>『奥の細道』での松尾芭蕉の足跡</dc:title>
    <dc:description>元禄15年に松尾芭蕉が著した紀行文。日本の古典における紀行作品の代表的存在であり、松尾芭蕉の著作のなかで最も有名な作品である。</dc:description>
    <vhwk:causedBy rdf:resource="#松尾芭蕉"/>
    <vhwk:visualize>
      <rdf:Seq>
        <rdf:li>
          <rdf:Description>
            <vhwk:locatedAt rdf:resource="#隅田川"/>
            <vhwk:timePoint>1689-03</vhwk:timePoint>
            <dc:description>草の戸も 住み変わる代ぞ 雛の家</dc:description>
          </rdf:Description>
        </rdf:li>
        <rdf:li>
          <rdf:Description>
            <geo:lat>35.73926718561</geo:lat>
            <geo:long>139.8243713</geo:long>
          </rdf:Description>
        </rdf:li>
        (...中略...)
      </rdf:Seq>
    </vhwk:visualize>
  </vhwk:Movement>

  <vhwk:Location rdf:ID="隅田川">
    <dc:title>隅田川</dc:title>
```

```
<geo:lat>35.73926718560775</geo:lat>
<geo:long>139.526413</geo:long>
</vhwk:Location>
```

```
<foaf:Person rdf:ID="松尾芭蕉">
  <foaf:name>松尾芭蕉</foaf:name>
  <dc:description>現在の三重県伊賀市出身の江戸時代前期の俳諧師。蕉風と呼ばれる芸術性の高い句風を確立し、俳聖と呼ばれる。</dc:description>
  <dc:coverage>
    <rdf:Description>
      <vhwk:startTime>1644</vhwk:startTime>
      <vhwk:endTime>1694</vhwk:endTime>
    </rdf:Description>
  </dc:coverage>
</foaf:Person>
</rdf:RDF>
```

上記の例は、移動クラスを用いて『奥の細道』で松尾芭蕉がたどった足跡の視覚情報のメタデータと、人クラスを用いて松尾芭蕉のメタデータを記述した例である。移動クラスでは、緯度・経度で表される地点を、移動する経路順に繰返し記述することによって、移動の視覚情報を表現した。上記の例を見ていくと、<vhwk:Movement rdf:ID="route\_of\_hosomichi">は移動クラスのインスタンスである。これを主語にもつ視覚情報を表すプロパティ<vhwk:visualize>の目的語に、視覚情報のメタデータを記述した。RDF コンテナのひとつである<rdf:Seq>では、リソースをグループ化することができ、またグループのメンバーに順序を与えることができる。RDF コンテナと個々のメンバーを結ぶプロパティである<rdf:li>を用いて、移動する経路順に地点の緯度・経度を記述している。一方、移動の主語となる松尾芭蕉は歴史クラスのサブクラスである人クラスのインスタンスであり、上記の例には氏名、概要、生没年のメタデータを記述している。

## 6. 史実メタデータの作成

5章で示したメタデータスキーマに基づいて、まず試験的に、史実と視覚情報のメタデータを人手で300件分作成した。300件のメタデータ作成にかかった時間の平均をとると、1件あたりの作成時間は約20分であった。歴史の学習支援を行うには、300件というデータ量は不十分であるため、データの拡充が必要である。しかし、人手でのメタデータの拡充にはかなりのコストがかかるため、少ないコストでメタデータを拡充するには、メタデータの作成支援が求められる。そこで本研究では、史実と視覚情報のメタデータの作成支援のために、1) 既存の Web 上のリソースを機械的に解釈し、メタデータの自動生成・修正を行う、2) 人手によるメタデータ作成のためのメタデータエディタを作成する、という2つの方法をとる。

推古天皇		{基礎情報 天皇
第 33 代天皇		名 =推古天皇
在位期間:593年1月15日 - 628年4月15日		代数=第 33
在位中の時代	飛鳥時代	在位=[[593年]][[1月15日]]
在位中の皇居	豊浦宮・小墾田宮	・[[628年]][[4月15日]]
別名	豊御食炊屋比売命	時代=[[飛鳥時代]]
出生	豊御食炊屋姫尊 <sup>[1]</sup>	皇居=[[豊浦宮]]・[[小墾田宮]]
死没	554年	諱 =額田部
死没場所	628年4月15日	幼称=
陵墓	植山古墳→磯長山田陵	別名=豊御食炊屋比売命 豊御食炊屋姫尊
皇子女	荒道貝蛸皇女	生年=[[554年]]
	竹田皇子	没年=[[628年]][[4月15日]]
	小墾田皇女	没地=[[小墾田宮]]
	尾張皇女	陵墓=植山古墳→磯長山田陵
	田原皇女	先代=[[崇峻天皇]]
夫	敏達天皇	次代=[[舒明天皇]]
父親	欽明天皇	子 =[[荒道貝蛸皇女]] [[竹田皇子]] [[小墾田皇女]] {(後略)}
母親	蘇我堅塩媛	}}
※ 最初の天皇号使用者 (異説あり)		
欠史十代の一。		

(左図) Wikipedia の infobox  
(右図) Wiki 書法による infobox のデータ

図2 Wikipedia の infobox

## 6.1 DBpedia

本研究では、オンライン百科事典 Wikipedia<sup>[13]</sup>を用いて史実メタデータを生成した。その理由として、1) 様々な分野の情報資源があること、2) Wikipedia にあるデータは Wiki 書法を用いて記述されているため、データに構造があること、3) 一つの記事に対してカテゴリが付与されており、そのカテゴリは階層構造を持っていること、などがある。まず、Wikipedia のデータ抽出に関する先行研究として DBpedia<sup>[14]</sup>を調査した。DBpedia とは、RDF 形式に変換した Wikipedia のデータを提供しているプロジェクトである。英語や日本語を始め、11 言語の Wikipedia をカバーしている (ただし、英語以外は全ての記事をカバーしていない)。DBpedia では、Wikipedia が提供している記事の中から、タイトルや概要、画像、infobox、カテゴリのような構造をもつデータを抽出して RDF 形式に変換している。infobox とは、属性名と属性値で構成される、記事に関する基礎的な情報の表である。(図2参照)。徳川家康の記事を RDF に変換し、n-Triple 構文で表現したデータの一部抜粋を図3に示す。n-Triple<sup>[15]</sup>の記法は単純に主語、述語、目的語の URI を『<>』で記述し、その順番に並べたものである。文末にはピリオドを置く。図3の1行目と2行目はそれぞれ異なるトリプルである。1行目も2行目も主語は同じリソースであり、これは徳川家康を表している。1行目のトリプルは、記事の名称を示すプロパティと、その目的語に『徳川家康』の文字列が ASCII コードで書かれている。2行目は、記事に関す



```

1: <http://dbpedia.org/resource/Tokugawa_leyasu> <http://www.w3.org/2000/01/rdf-schema#label> "u5FB3u5DDD u5BB8 u5EB7n"@ja
2: <http://dbpedia.org/resource/Tokugawa_leyasu> <http://xmlns.com/foaf/0.1/depiction> <http://en.wikipedia.org/wiki/Image:Tokugawa_leyasu.jpg>
3: <http://dbpedia.org/resource/Tokugawa_leyasu> <http://www.w3.org/2000/01/rdf-schema#comment>
4: "u4EAB u5E7474 u6B73 uFF0873 u6B73 u6CA1 uFF09"@ja .
5: <http://dbpedia.org/resource/Tokugawa_leyasu> <http://dbpedia.org/property/relatedInstance>
6: <http://dbpedia.org/resource/Tokugawa_leyasu/%E5%85%88%E4%BB%A3%E6%AC%A1%E4%BB%A31> .

```

図3 徳川家康に関する DBpedia のデータ(n-Triple 形式)

る画像を示すプロパティで、目的語に画像の URI が示されている。このように DBpedia では、ある一つの記事に対して一つのリソースを生成している。Wikipedia から抽出したデータがもつ属性(タイトル、概要など)を示すために、DBpedia 独自のプロパティを定義し、目的語に抽出したデータを示すリソースや文字列を生成する、という手法で RDF を生成している。infobox のデータについては、属性名を用いてプロパティを生成し(例えば、名という属性名であれば、dbpedia:名というプロパティが生成される)それに対応した属性値を目的語として RDF を生成している。

## 6.2 DBpedia を用いた史実メタデータの自動生成

DBpediaの日本語版のデータには、大きく分けて2つの問題がある。まず1つ目は、日本語版 Wikipediaの一部の記事しかデータを抽出していない点である。DBpediaでは英語版 Wikipediaの記事については全て抽出しているが、それ以外の言語のデータは、英語の記事からリンクが張ってある場合のみ抽出している。日本語版 Wikipediaの記事は一部しか抽出されていないため、史実に関する記事も全て抽出されていない。2つ目の問題点は、日本語の記事において infoboxのデータが適切に抽出されていない場合があることである。日本語版 Wikipediaには infoboxのデータを記述するテンプレートの中に、英語版 Wikipediaにはない特殊な構造を持ったテンプレートがある。日本語版 Wikipediaのテンプレートの種類を調査したところ、図2に示すような典型的なテンプレートの他に、構造の特殊なテンプレートは2種類あることがわかった。しかし DBpediaでは、構造の特殊なテンプレートを用いて記述された infoboxのデータを、正しく変換することができていない。以上の問題点を踏まえ、日本語版 Wikipediaにある全ての記事を対象にデータを抽出し、RDF形式に変換して、日本語版 DBpediaのデータを作り直した。

DBpediaでは Wikipediaの記事を抽出して、RDF形式に変換するプログラムを提供している。しかし、そのままでは特殊な構造をもつ infoboxテンプレートのデータは抽出できないため、DBpediaの抽出プログラムを改良する必要がある。そこで、その他2種類の構造の特殊なテン

プレートも変換できるように、DBpediaで提供されているデータ抽出プログラムを改良した。

本研究では改良したデータ抽出プログラムを用いて、日本語版 Wikipediaにある全ての記事を対象に、約53万件分の記事のデータを抽出し、RDF形式のデータに変換した。

## 7. 歴史情報閲覧システム V-HAWKS の開発

6章で作成した史実と視覚情報のメタデータを基に、地図をインターフェイスに史実を学習することができる、歴史情報閲覧システム V-HAWKS(Visiting History Along the Way, Knit your Stories)を開発した。システムの開発には Ruby on Rails、Javascript、Google Maps API<sup>[16]</sup>、XSLT スタイルシート、Simile Timeline API<sup>[17]</sup>を利用している。V-HAWKS システムでは、RDF/XML 形式で記述したメタデータからユーザの情報要求に応じて、XSLT スタイルシートを用いて必要な情報を抽出する。そのデータを基に Google Maps API を用いて史実の地図上での視覚化や、Simile Timeline API を用いて年表での視覚化を行う。歴史の学習に必要な機能として、本システムでは以下の機能を備えている。

### 1) 地図上での歴史情報閲覧機能

3章で示した5つの視覚化手段の表現方法を用いて、アニメーションにより史実を地図上に視覚化して、歴史を学習できる機能である。図4は5.1章で示した松尾芭蕉の著作『奥の細道』での足跡に関する史実とその視覚情報のメタデータを V-HAWKS の歴史情報閲覧機能を用いて視覚化した実行画面である。このように松尾芭蕉のたどったルートを、地図上でアニメーションを用いて視覚的に確認することができる。



図4 V-HAWKS の歴史情報閲覧機能の実行画面

## 2) 検索機能

メタデータにある情報を検索する機能であり、ワード検索とカテゴリ検索の2通りの方法で検索が行える。カテゴリ検索はディレクトリ表示を用いて行っている。

## 3) 年表表示機能

史実のメタデータの中から時間情報をもつものを抽出し、年表上に視覚化する機能である。年表表示機能では、ユーザの情報要求に応じて年表に表示するデータをカスタマイズすることができる。

## 4) メタデータ登録機能

本研究で設計したメタデータスキーマに従い、RDF/XML形式で記述したメタデータを、V-HAWKSのデータベースに登録する機能である。メタデータ登録権限をもったユーザは、ユーザ認証をした上で作成したメタデータを登録することができる。

本研究では歴史の学習において、ウェブ上のリソースを有効活用できるように、メタデータにある用語を検索語として、Google サーチエンジンをういた検索結果のリストや Wikipedia などの関連リソースへの自動生成を行っている。

## 8. 考察

本研究では、史実と視覚情報のメタデータを表現できるメタデータスキーマの設計と、メタデータの作成支援として、DBpedia を用いた史実メタデータの自動生成を行った。また、アニメーションを用いて、地図上での史実の視覚化機能をもつ歴史情報閲覧システム V-HAWKS の実装を行った。本研究は以下の2つの観点から考察を行った。

### 1) 先行研究のメタデータスキーマとの比較

本研究では先行研究として、SWHi と HEML で設計されたメタデータスキーマを調査した。SWHi のスキーマは、史実と関連のある印刷物資料を結びつけることを目的に設計されていることがわかった。特に印刷物資料の書誌事項が詳細に記述できるように設計されている。SWHi のスキーマは、本研究のスキーマと比較して、語彙数が多く定義されている。その語彙の多くはリソースの型付けを行うクラスであるが、アメリカ史初期の社会体系に特化した語彙である。そのため、SWHi のスキーマで記述できる史実は、アメリカ史初期の史実に限定されている。また、SWHi で定義されているプロパティは、書誌事項を記述するための語彙が多く、一方で、史実を記述するための語彙は少ない。そのため、SWHi スキーマでは、史実を詳細に記述できない。それに対し、本スキーマは史実と視覚情報の記述に特化したものである。スキーマの語彙数は、SWHi のスキーマより少ないが、抽象度の高い語彙を定義したため、広い時代、地域の

史実を記述することができる。また、プロパティでは史実を記述するための語彙を多く定義しているため、詳細な史実のメタデータを記述することができる。また、本研究のスキーマは高等学校で使用している日本史の教科書を参考にしたため、クラスの階層構造は日本史の教科書の区分に対応させた。このことから本研究で設計したメタデータスキーマは、歴史の学習内容を記述するのに特化しているといえる。

次に HEML のスキーマは、史実を地図上で視覚化するために必要な情報を記述するスキーマを定義している。HEML スキーマでは、参加者、時間情報、イベントの起きた場所（緯度・経度）を用いて、ある一地点を示すための情報）に関する情報が記述できる。HEML では場所の情報をういて、地図上に史実の視覚化を行う。また、イベントに関する関連資料への参照として、一次資料の書誌事項を記述することができる。すなわち、HEML のスキーマは詳細な史実のメタデータの記述はできず、史実を地図上に視覚化することに特化しているといえる。しかし、HEML のスキーマで記述できる、視覚的表現の方法は、史実の地理的な位置を地図上にマーキングし、アイコンをクリックすると史実情報をウィンドウで表示する、というものだけである。そのため、HEML のスキーマと比較すると、本研究のスキーマの方が、史実のメタデータの記述に適しており、またより豊かな視覚的表現を用いた、視覚情報の記述を行うことができる。

### 2) DBpedia のメタデータとの比較

DBpediaで提供されている日本語版のデータには、1) 日本語の記事は全ての記事が抽出されていない、2) 特殊な構造をもつinfoboxテンプレートのデータが正確に抽出されていない、という問題点があった。そこで本研究では、まずDBpediaが提供しているデータ抽出プログラムを改良し、それをういて全ての日本語の記事を対象にデータを生成した。その結果、DBpediaが提供している日本語版のデータは、日本語版のWikipediaの記事の全体の約38%（2008年11月時点）しかなかったが、本研究では、日本語版Wikipediaの全記事のデータを作成した。DBpediaの日本語版のデータは、一部の記事しか抽出対象になっていなかったため、史実に関する記事について、抽出もれがあった。しかし、本研究では全ての記事のデータを生成したため、史実に関する記事も網羅することができた。

現在の問題は、作成したデータがDBpediaのスキーマの形式になっていることである。しかし、V-HAWKSで作成したデータを扱うには、本研究のスキーマの形式に変換しなくてはならない。そのため、1) DBpediaのスキーマと本研究のスキーマのプロパティの対応付け、2) 記事と本研究のスキーマのクラスとの対応付け、

を行う必要がある。1) に関しては、Wikipedia の infobox のテンプレートは、現在 552 個ある。しかし、それぞれのテンプレートでは同じ意味をもつ属性名にも表記のゆれがある。例えば、天皇の infobox のデータを記述するための雛形 (テンプレート) では、人の生年月日を示す属性名は『生年』である。一方、武士のテンプレートの場合は『生誕』である。そのため、infobox のテンプレート全てを対象にプロパティの対応付けを行わなければならない。プロパティの対応付けには、文字列のもつ意味を解釈できる人間が介在する必要がある。しかし、552 個の infobox のテンプレート全てを対象としたプロパティの対応付けを人手で行うのはコストがかかってしまう。2) に関しては、記事と本研究のスキーマのクラスとの対応付けには、記事のカテゴリを用いることができる。Wikipedia のカテゴリは、記事の分類やキーワードの役割を果たすため、クラスの対応付けには有用であるといえる。しかし、Wikipedia のカテゴリは、ユーザが自由に付けることができるため、記事の分類やキーワードの役割を果たす、という観点では有用性の低いカテゴリも多い。そのため、記事の分類を機械的に判断してクラスに対応付けするには、今後カテゴリの扱い方を検討していく必要がある。

## 9. おわりに

本研究では中等教育での歴史の学習支援のために、それぞれの Web 上の歴史学習のためのアプリケーションが共通で利用できる形式である史実と視覚情報のメタデータスキーマの設計を行った。また、学習支援に用いる史実メタデータの作成のために、Wikipedia のデータを用いた史実メタデータの自動生成や、地図をインターフェイスとしたメタデータエディタの作成を行った。さらに、作成した史実メタデータを用いて、地図をインターフェイスに歴史を学習する歴史情報閲覧システム V-HAWKS の構築を行った。

現在、DBpediaを用いて史実メタデータの自動生成を行っているが、プロパティの対応付けや、カテゴリの扱いなどの課題が残っている。この2つの課題の根本的な原因は、様々なユーザがWikipediaの情報を自由に編集できる点にある。そのため、表記のゆれや同義語の問題など、自然言語処理の分野で取り組まれてきた研究課題であるといえる。そこで、今後は自然言語処理で使われている既存の技術を用いて、これらの問題を解決していくことを検討している。また本研究では、人手による簡易なメタデータの編集を目的とした史実と視覚情報のメタデータエディタの開発を進めている。特に視覚情報のメタデータについては、地図上に曲線や領域を

描くために、緯度・経度に関する多量のデータが必要である。そこで本研究では、地図をインターフェイスに用いた史実と視覚情報のメタデータエディタを構築している。GUIによるデータ入力操作を用いて、データ入力の簡易化を図る。

## 参考文献

- [1] 教育情報ナショナルセンター. National Information Center for Educational Resources. <<http://www.nicer.go.jp/>>. (参照:2008-11-17)
- [2] W3C. W3C Semantic Web Activity. <<http://www.w3.org/2001/sw/>>. (参照:2008-11-17)
- [3] W3C. Resource Description Framework / W3C Semantic Web Activity. <<http://www.w3.org/RDF/>>. (参照:2008-11-17)
- [4] W3C. Web Ontology Language OWL / W3C Semantic Web Activity. <<http://www.w3.org/2004/OWL/>>. (参照:2008-11-17)
- [5] Google. Google マップ - 地図検索. <<http://maps.google.co.jp/>>. (参照 2008-11-17)
- [6] MMVI. Maps of War :: Visual History of War, Religion, and Government. <<http://www.mapsofwar.com/index.html>>. (参照:2008-11-17)
- [7] Digital Library Department - University Library of Groningen. Semantic Web for History (SWHi) - UB RUG. <<http://semweb.ub.rug.nl/>>. (参照:2008-11-17)
- [8] Heml Project. Historical Event Markup and Linking Project. <<http://heml.mta.ca/heml-cocoon/>>. (参照:2008-11-17)
- [9] DCMI. Dublin Core Metadata Initiative (DCMI). <<http://dublincore.org/>>. (参照:2008-11-17)
- [10] FOAF Project. The Friend of a Friend (FOAF) project | FOAF project. <<http://www.foaf-project.org/>>. (参照:2008-11-17)
- [11] Google. Google Earth. <<http://earth.google.co.jp/>>. (参照:2008-11-17)
- [12] Google. KML. <<http://earth.google.co.jp/kml/>>. (参照:2008-11-17)
- [13] Wikipedia. メインページ - Wikipedia. <<http://ja.wikipedia.org/wiki/>>. (参照:2008-11-17)
- [14] Freie Universität Berlin. wiki.dbpedia.org : About. <<http://wiki.dbpedia.org/About>>. (参照:2008-11-17)
- [15] W3C. N-Triples. <<http://www.w3.org/2001/sw/RDFCore/ntriples/>>. (参照:2008-11-17)
- [16] Google. Google Maps API - Google Code. <<http://code.google.com/intl/ja/apis/maps/>>.
- [17] Simile Project. Simile - Timeline. <<http://simile.mit.edu/timeline/>>. (参照:2008-11-17)
- [18] 馬場裕子, 永森光晴, 杉本重雄. RDF/OWL を利用した歴史情報閲覧システムの構築. 情報処理学会研究報告 人文科学とコンピュータ. 2008-CH-79, 2008 年, pp.49-56.