

正倉院文書トピックマップへの知識情報の充実

後藤 真
花園大学

内藤 求
(株) ナレッジ・シナジー

筆者らは、正倉院文書トピックマップの開発について、以前に報告を行った。その後も開発と拡張・情報の充実を進めている。知識情報を経典に関するものを中心に充実させた。正倉院文書の中心は写経所の文書であり、経典の情報の充実は、正倉院文書の中心部分の研究に役立つことになる。知識情報の充実と、その抽出・可視化の手法によって、奈良時代の経典の所在や、所属の状況を明らかにできるようシステムを作った。また、正倉院文書データベース（SOMODA）の新たなバージョンとしての位置づけも加えられるよう、文書に関する情報も充実させ、より高度な情報発見を可能にするものとした。

Improvement of knowledge information to Shoso-in Monjo Topicmaps

Makoto GOTO
Hanazono University

Motomu NAITO
Knowledge Synergy Inc.

The Authors had reported the Shoso-in Monjo Topicmaps. We have been developing Shoso-in Monjo Topicmaps continuously. We expanded knowledge information mainly about the "Kyoten"(sacred book). The main part of the Shoso-in Monjo is a document of the "Shakyojo"(copying of a sutra department). Therefore, the improvement of information of the "Kyoten" helps the study for the main part of the Shoso-in Monjo. We made the system which could clarify the situation of the location and the position of the sacred book of the Nara era by method of the extraction, visualization. In addition, we placed Shoso-in Monjo Topicmaps as a new version of the Shoso-in Monjo database (SOMODA) by expanding the knowledge information about the Shoso-in Monjo. By this works, we made it possible to more advanced information discovery.

1. はじめに

筆者らは、正倉院文書トピックマップの開発について、以前報告を行った[1]。その後も開発と拡張・情報の充実を進めている（なお、この参考文献[1]を、以後は「前稿」として示す）。本発表では、正倉院文書トピックマップの前稿以後の開発状況について報告を行う。具体的には、トピックマップによる、奈良時代の経典書写の解析状況をもとにした歴史研究への新たな知見の可能性について、また、史料情報の提供の多様化による発見手法の多様化について、報告するものである。

2. 正倉院文書トピックマップの概要

正倉院文書トピックマップは、正倉院文書に関する史料情報（本文・目録情報なども含む）をオントロジベースでトピックマップ化したものである[1]。

トピックマップは、主題(概念)と主題(概念)間の関係

を用いて、知識ネットワークを構築するための情報技術である。主題を"トピック"、主題間の関係を"関連"また主題に関する情報リソースへのリンクを"出現"と表現している。情報リソースがトピックマップ記述ファイルの内部にある場合を"内部出現"、外部にある場合を"外部出現"と呼ぶ。

トピックマップは、"トピック"、"関連"、"出現"に加えて、"トピックマップ"、"トピック名"、"異形"および"関連役割"の合計 7 つの情報項目をもる。そのうち、"トピック"、"トピック名"、"関連"、"関連役割"および"出現"は型をもち、型自身もトピックで表現する。型および型の間の関係の構造をトピックマップのオントロジという。各型は、インスタンスをもちトピックマップは、型とそれらのインスタンスから構成されている。

図 1 に、正倉院文書トピックマップのオントロジ図を示す。四角がトピック型、線が関連型を表していて、

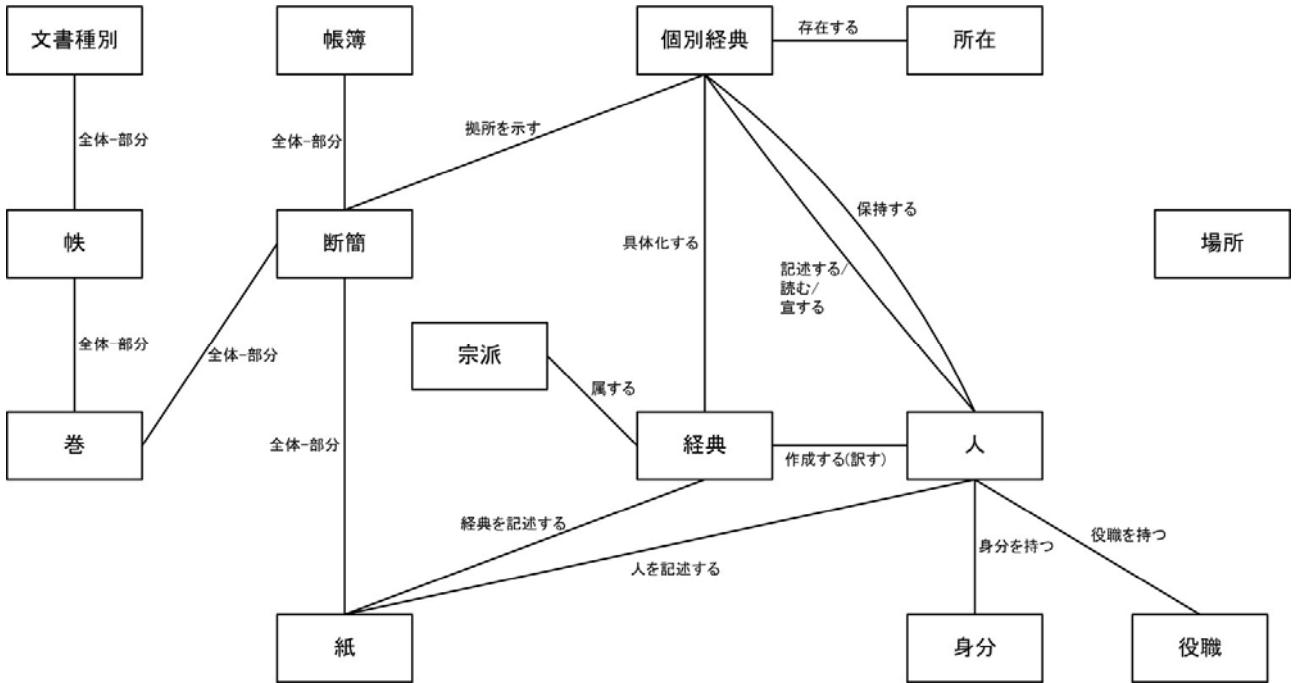


図 1 正倉院文書トピックマップのオントロジ図 2013 年版

それぞれ名前が付けられている。本論文執筆時点での各型の数とサンプルを、以下に示す。数は括弧内に示す。

- トピック型(14): 宗派, 経典, 個別經典, 人, 所在など
- 関連型(12): 作成する(訳す), 属する, 具体化するなど
- 関連役割型(24): 作成者, 作成対象, 所属宗派など
- 出現型(4): Text, 画像, SATID, よみ

図 1 には、関連役割型、出現型は、描かれていない。上記、オントロジに基づき、正倉院文書トピックマップを作成した。さらに、それに基づき、正倉院文書トピックマップ Web アプリケーションを開発している。

開発には、トピックマップアプリケーション開発・運用プラットフォーム Ontopia を利用している[2]。

Ontopia は、オープンソースであり、多くの特筆すべき機能を持っている。その一つ「問合せ言語 tolog」は、本研究とも特に関わりが深い[3]。tolog は、宣言型論理プログラミング言語 Datalog をベースに、SQL の特性を加えた問合せ言語である。Datalog は、構文的にはプログラミング言語 Prolog のサブセットである。問合せの一般的な形式は以下のとおりである（問合せ式の最後には、"?"が必須である）。

**関連型(関連役割プレーヤ 1 : 関連役割型 1,
関連役割プレーヤ 2 : 関連役割型 2)?**

例えば、トピック型"人"のインスタンス「元暁」とトピック型"經典"のインスタンス「法華疏」の間に、二項関連「作成する(訳す)」があり、この"関連"において、元暁は、"作成者"という"関連役割"をもち、法華疏は、"作成対象"という"関連役割"をもつとする。その場合、元暁と法華疏は、それぞれの"関連役割"のプレーヤになる。関連型、関連役割型、関連役割プレーヤは、以下のようになる。

- 関連の型: 作成する(訳す)(make)
- 関連役割プレーヤ 1: 元暁
- 関連役割型 1: 作成者(maker)
- 関連役割プレーヤ 2: 法華疏
- 関連役割型 2: 作成対象(made)

そして、その場合の問合せ式は、以下のようになる。

make(元暁 : maker, 法華疏 : made)?

作成する(訳す)関連で結ばれている不特定多数の人と經典の集合を問合せする場合は、関連役割プレーヤの部分を変数に置き換える。tolog では、変数は、先頭に\$をつけて表現する。問合せ式は、以下のようになる。

make(\$PERSON : maker, \$SCRIPTURE : made)?

この問合せの結果として、変数を列とするテーブルが得られる（図 2・次ページ）。

PERSON	SCRIPTURE
円測	因明正理門論疏
聖德太子	法華經義疏
仏陀耶舍	四分律
元康	三論疏
宗一	華嚴經疏
文備	正理門論
志徳	金光明經疏

図2 tolog による問い合わせ結果

また、図3に、問合せ式と関連、関連役割、関連役割プレーヤの関係を示す。

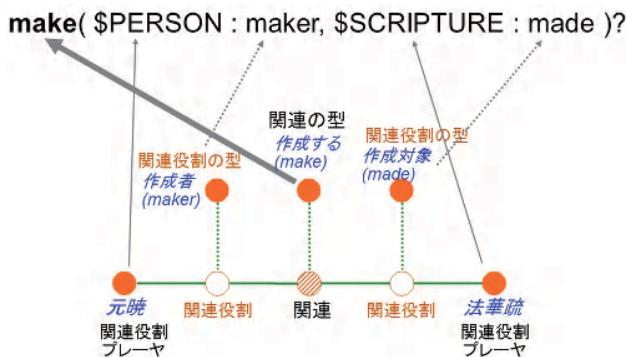


図3 tolog 問合せ式と関連、関連役割等との関係

正倉院文書トピックマップ Web アプリケーションは、tolog を用いて各 Web ページを動的に生成している。また、問合せ式入力フィールドから、利用者が任意の問合せ式を実行できるようになっている。本アプリケーションは、本稿執筆時点で、以下の機能をもつている。

- 各トピック型のインスタンスの一覧
- インスタンスの詳細表示
- 関連に基づくインスタンス間の航行
- 文字列検索
- 問合せ言語 tolog を用いた問合せ
- グラフ表示
- リモートアクセスプロトコル(TMRAP) のインターフェース

また、トピックマップにおいては、各トピックに、主題識別子を割り当てることができる。主題識別子を公開したものを公開主題識別子(PSI: Published Subject Identifier)と呼ぶ。PSI を用いることにより、ネットワーク上で、不特定多数の利用者間、システム間で主題の同定/識別、共有が可能になる。

3. 正倉院文書トピックマップの発展

- 前稿から、大きく進展した点は以下の通りである。
- 正倉院文書に記述された経典の情報の整備
 - 経典に関する情報の追加付与による知識状況の解明
 - 帳簿をもととした、復原情報の追加
 - 目録情報の PSI 作成による史料へのアクセス方法の追加

以下、順に説明したうえで、これらの進展による人文科学への貢献の可能性について述べる。

a. 正倉院文書に記述された経典の情報の整備

前稿において、正倉院文書に現れる経典類を、実体(奈良時代当時、モノとして存在していた)としての個別経典と、テキストの中身を指す経典の二種に分け、データの入力について述べた。この経典数を大幅に増やし、正倉院文書に現れる経典類を原則としてすべてトピック化した。そして、経典のトピックに対し、大正新脩大蔵經の番号を関連づけた。この関連付けには経典名の一致による単純なマッチングを行っている。この点についての問題点は後述する。これにより、大正新脩大蔵經の番号から奈良時代の経典一覧を探すことが可能となる。現在、経典数 5940 (同一経典の別名有) に対し、992 の経典が大正新脩大蔵經の経典名と一致したため番号を付している。同一経典名で作者(訳者)とテキストの中身が異なる場合には、(わかる場合は) 人間が作者・巻数等を同定した上で、人間の手で ID を振りなおした。

なお、正倉院文書内には、経典名からは、どの経典か同定できないものもある。たとえば『法華經疏』は奈良時代の経典としては、同名で異なるものが 15 ある。作者が異なるものである。厳密には経典名が異なるものもあるが、法華經の注釈書という意味で、法華經疏と表現されているのである。このような場合は、もっとも若い大正蔵の番号を仮に付与した。

この手法には以下の問題点がある。

- 奈良時代の経典名と大正新脩大蔵經の経典名が一致しない場合に表示しない
- 内容は異なるが、たまたま奈良時代の経典名と大正新脩大蔵經の経典名が一致した場合に誤った情報を提示してしまう

この両者の問題点の解決を短期的に行なうことは困難である。その理由は、多くの場合、奈良時代の経典はその内容まで残されていないことが多く、たとえ不自然な状況であっても、それが「誤りである」という証



図 4 智懶の所持する經典の一部と宗派の関係を Vizigate したもの

明ができないのである。そのためデータベースとしては、同一經典名については暫定的に関連付ける以外に方法がない。最終的な研究成果としては、「正解である」証明ができない場合には、その情報は提示すべきではないが、データベースとしては、情報の可能性をより広く提案することが肝要であるため、このような措置を行った。

この作業により、正倉院文書に出てくるほぼすべての經典が、単純な文字列検索による発見ではなく、經典の作者や、大正新脩大藏經の情報および関連情報からの発見が可能となった。さらに後述する情報の付与によって、より複雑な検索を可能とした。

b. 経典に関する情報の追加付与による知識状況の解明

さらに、經典情報には、当時の史料に基づいた「宗派」の情報を付与した¹。正倉院文書の中には、經典の所蔵・所有の状態を記した史料が、複数存在する（經典目録）。その經典目録の中には、当時の宗派ごとに經典を分類したものが残されている。そのため、この宗派ごとに經典が所蔵された史料の情報を用いて、經典のトピックに対して「所属する」という関係性を持た

せた「宗派」というトピックをあらたに作成した（図 1）。これにより、ある宗派が、どのような經典群を用いて仏教を学んでいたかをデータ上で確認できることとなる。

また、正倉院文書ではないが、平安初期の經典目録（『五宗錄』）も同様に入力を進めており、この両者を比較することで、宗派に属する經典の時間的変遷を見ることが可能になる。また、現時点では、經典には、人および実体としての個別經典が関連付けられている。經典と人は、「作成する（訳す）」関連で結び付けられ、經典が「作成対象」、人が「作成者」の関連役割を果たしている。（「作成者」は、表現としては適切ではないが、漢訳・疏類の著述などをすべてまとめて表現するために、このような書き方を行っている。）經典と個別經典は、「具体化する」関連で結び付けられている。そして、個別經典には所在地や所有者の情報が関連付けられている。そのため、ある宗派の經典をどの寺院が多く持っていたか、もしくは、誰が多く持っていたかということが、視覚的に分析可能となった（図 4）。

図 4 は、本稿執筆段階で東大寺の僧である智懶が所持していた經典の中から、宗派のデータとの関連があるものを tolog Query で抽出し、Vizigate したものである。以下に、その tolog の問合せ式の例とその結果を示す（図 5・次ページ）。問い合わせ式の 1 行目で prefix

¹ なお、宗派の情報分析は、科学研究費基盤 C 「經典目録よりみた古代国家の宗教編成策に関する多面的研究」（代表・中林孝之）の研究分担の成果の一環である。

を指定している。問合せ式において、関連型、関連役割型等は、その型に割り当てられた PSI を利用して記述することができる。しかし、その結果として、問合せ式が長く煩雑になってしまう。それを解消するために prefix を用いている。

DENOMINATION	SCRIPTURE	INDIVIDUAL	PERSON
俱舍	俱舍論疏	俱舍論三十巻	智慧行
法性(経)	香王菩薩陀羅尼咒經	顯揚論疏十巻	慈訓師
法性(論疏)	一切經要集	一切經要集三十一巻	願教
花巖	華嚴孔目	花巖經方軌一部五巻	不明
花巖	起信論別記	起信論疏	玄幾
花巖	起信論別記	起信論疏一巻	法玄
花巖	起信論疏	起信論私記一巻	智慧行

図 5 宗派と人との関係についての問合せ結果

```
using psi1 for i"http://somoda.jp/psi/shoso-in/"
psi1:belong($DENOMINATION : psi1:belonging,
             $SCRIPTURE : psi1:belonged),
psi1:manifest($SCRIPTURE : psi1:manifesting,
              $INDIVIDUAL : psi1:manifested),
psi1:hold($INDIVIDUAL : psi1:held,
          $PERSON : psi1:holder)
order by $DENOMINATION, $SCRIPTURE, $INDIVIDUAL,
$PERSON?
```

智慧行は、華厳宗学僧とされているが、俱舍関係の經

典を保持していることも（仏教系研究者が史料を見ればわかる部分ではあるのだが）、明確にデータ上で表現ができている。奈良時代における「六宗兼学」の状況がデータ上でより明瞭に見受けられる事例であるといえるであろう。ただし、これが、智慧行が所持している經典の特徴を示すと言い切れるかは、今後史料上・データ上の状況などとの総合的な判断を行う必要がある。正倉院文書トピックマップは、それらの情報へのリンクも用意した（図 6）。図 6 は、図 4 の情報に加えて、經典が書かれた史料へのノードの連関を追加表示したものである。図 6 の例では、「起信論私記」および、俱舍論のいずれもが、続々修 12 帚 9巻第 4 断簡表に書かれていることを示している。そして、この断簡が「応写疏本勘定目録」という帳簿に含まれることがわかる。「応写疏本勘定目録」がどのような帳簿なのかは、ここから当該のトピックのページに移動することで閲覧が可能になる。これにより、これらの經典知識情報が、どのような史料に書かれていたかという前提をより容易に確認することができるようになる。

帳簿の情報及び断簡情報の追加と充実の詳細内容については、c.および d.で述べる。

また、図 7（次ページ）は、宗派と所在の関係について Vizigate したものである。Tolog の問い合わせ式の例とその結果を以下に示す（図 8・次ページ）。

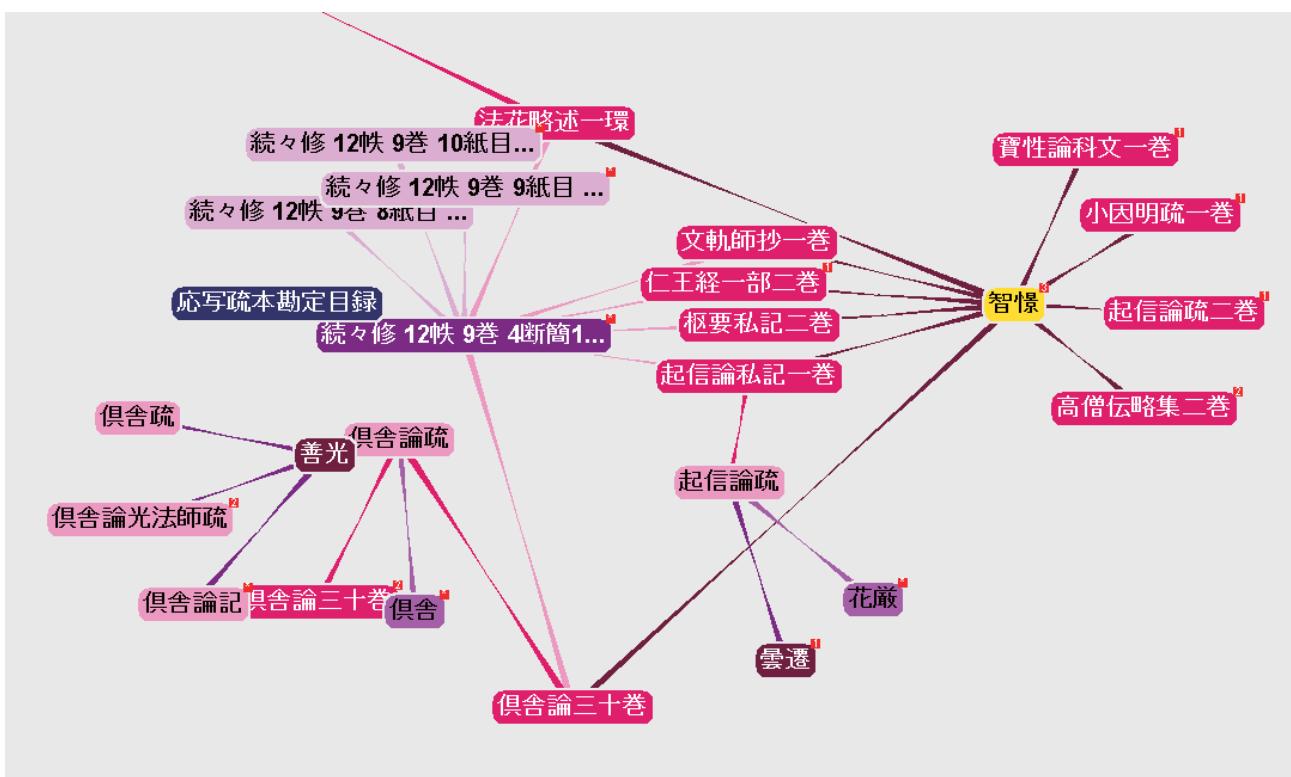


図 6 図 4 の情報に加えて対象となる經典がどの断簡に書かれており、その断簡がどのような帳簿と関係するかを示したもの

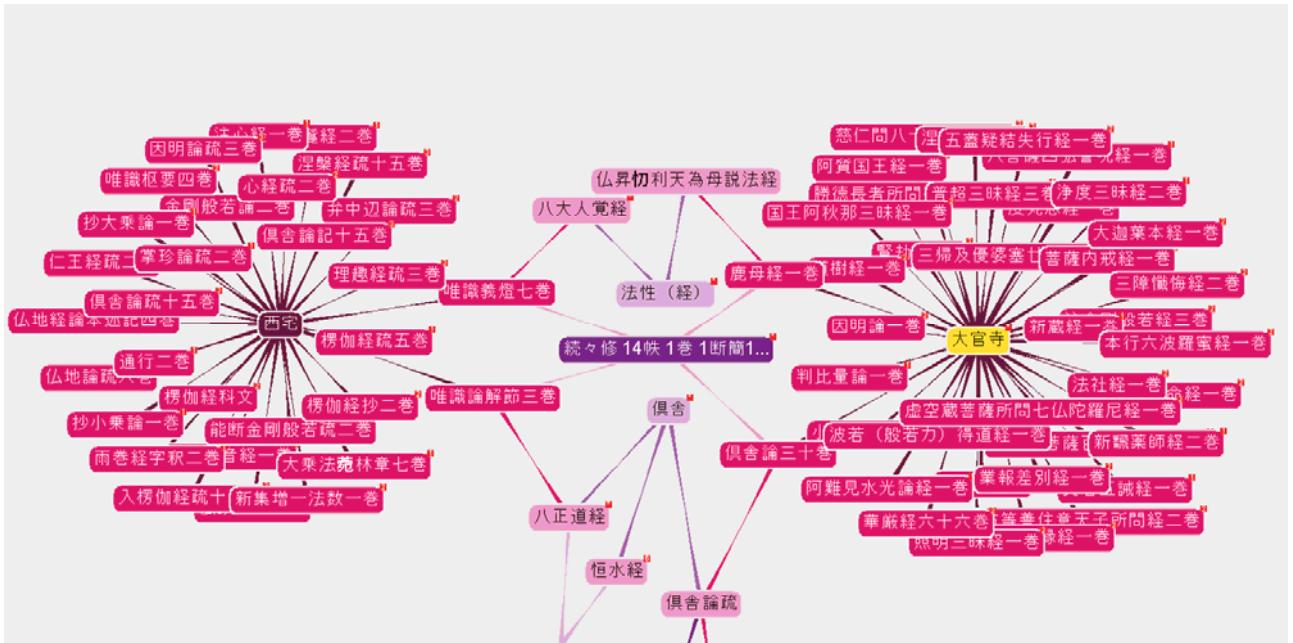


図 7 経典の所在地の情報と、宗派の情報の関係の一部を Viziage したもの

```

using psi1 for i"http://somoda.jp/psi/shoso-in/"
psi1:belong($DENOMINATION : psi1:belonging,
             $SCRIPTURE : psi1:belonged),
psi1:manifest($SCRIPTURE : psi1:manifesting,
              $INDIVIDUAL : psi1:manifested),
psi1:exist($INDIVIDUAL : psi1:existed,
           $LOCATION : psi1:existing)
order by $DENOMINATION, $SCRIPTURE, $INDIVIDUAL,
$LOCATION?

```

DENOMINATION	SCRIPTURE	INDIVIDUAL	LOCATION
俱全	五陰譬喻經	業報差別經一卷	大官寺
俱全	俱舍論疏	俱舍論三十卷	大官寺
俱全	八正道經	唯識論解第三卷	西宅

図 8 宗派と人との関係についての問合せ結果（一部）

こちらも俱舍と法性（經）の両者の宗派が、大官大寺および西宅（一説には県犬養三千代の関連施設）に關係している状況が見て取れる。所蔵場所によつても、宗派が厳密に分かれている状況が、コンピュータによって明確されている。

これらの解析情報は、今後より情報の正確性を増すと同時に、データの量を増やすことによって、新たな傾向が見えてくると思われる。また、現時点においても、ある種の傾向を示すものであるとはいえるであろう。

正倉院文書トピックマップの経典に関する知識情報の充実により、奈良時代における仏教史研究はもちろん

ん、当時、文化的・思想的な基層をなしていた仏教のあり方を実態に即して、かつ史料の情報を同時に確認しつつ分析するための基盤情報とツールの提供が可能となった。

c. 帳簿をもととした、復原情報の追加

正倉院文書トピックマップの基盤となっている正倉院文書データベース（SOMODA）[4]は、正倉院文書のバーチャルな復原を可能としたデータベースであった。その復原の機能を引き継ぐべく、復原されたものの形態である「帳簿」に断簡を関連付ける情報を充実させた。これにより、正倉院文書の復原情報を一覧形式で確認できることとなる。SOMODA とは異なり、自由な復原への利便性は多少下がるもの、どのような種類の帳簿が正倉院文書にはあるのかが一見して閲覧可能となるというメリットを得た。

現時点では、帳簿から当該の断簡を閲覧し、断簡→紙と遷移することで、テキストの閲覧が可能となっている。テキストは”外部出現”とし、1紙1ファイルのテキストデータを作成している。逆方向、すなわち、紙から断簡、断簡から帳簿へと関連を遷移することで、紙がどの帳簿に含まれていたのかを見ることができる。

また、断簡ごとに目録情報を作成しているため（d.にて説明）、断簡情報を検索し、そこから帳簿の情報へと移動することが可能となった。これらの自由な移動は、SOMODA が研究者向けで、あくまでもテキスト全文検索および大日本古文書の巻頁などという専門性の高い検索手段だけを提供していたのに比べて、より

多様な検索手段を提供することに成功している。

さらに、SOMODA と比較すると低下している利便性も、今後クエリや表示方法の工夫により、改善を行う。

d. 目録情報の PSI 作成による史料へのアクセス方法の追加

前稿では、正倉院文書のテキスト情報を紙単位で提供し、識別子を付したことについて述べた。それに加え、新たに正倉院文書の断簡に関する情報を SOMODA の基盤データの一つである目録情報から作成した。これに SOMODA の断簡 ID ごとに識別子を付した。そして、目録情報のみの簡易な検索・表示システムを作成し、正倉院文書トピックマップ中の断簡トピックへのリンクを可能にする予定である。その目録情報の詳細なテキストについては、断簡 ID ごとの XML データを閲覧・取得できる形式とした。

これにより、断簡ごとに詳細な情報（断簡に含まれる紙の長さや、文字の訂正、罫線がある場合はその情報など）を閲覧することが可能となる。

SOMODA では、目録情報からの検索を可能としていなかった。その検索を可能とすることで、正倉院文書の情報へアクセスする手段を多くすることを可能とした。また、PSI 情報に関しては、正倉院文書グローバリー[5]との連携による英語への翻訳も目指す。

正倉院文書の関連情報などの英訳化は正倉院展などの展示では進んでいるが、一方で、Web 上では正倉院の英語情報は少なく、テキストデータベースは、事前に漢字を入力しなければならない点が、海外からのアクセスの障壁になる可能性がある。

その際、トピックマップでは、英訳化された用語から、原史料へのアクセスが可能になる。（無論、最終的には漢文テキストを読まざるを得ないが）

また、本トピックマップを介して、日本国内のデータベースへとリンクする手段も確保することができ、正倉院文書研究からはじまる関連研究へのリンクの確保も可能となる。

4. トピックマップ発展の意義と課題

今回の開発によって、経典情報を中心とした、正倉院文書研究の関連知識を記述することが可能となった。正倉院文書のなかでも、写経所文書を中心とした研究へのアクセス手段の多様化と、情報の結びつけによる、新たな解析を可能とすることが、より可能になった。

引き続き日本史学のみならず関連諸学の方法論に即し

た形で、情報の充実を目指す必要がある。とりわけ、史料にあらわれる一次的な用語のトピックだけではなく、歴史学用語など二次的な用語のトピックを作成することも、今後重要になってくるであろう。

また、現在の重要な問題として、「よみ」の扱いがある。トピックマップで「よみ」を表現する方法として以下のようなものが考えられる。

方法 1. 「よみ」をトピック名として表現する。そして、有効範囲(scope)を設定して、漢字のトピック名と区別する。例えば、「よみ」、すなわち、ひらがなのトピック名に対して reading という scope を設定し、漢字のトピック名に対して kanji という scope を設定しておけば、問合せ式にて、reading scope を指定すれば、結果として、ひらがなのトピック名を取得でき、kanji scope を指定すれば、結果として漢字のトピック名を取得できる。

方法 2. 「よみ」を”内部出現”として表現する。

方法 3. トピック名に型を指定する。例えば、「よみ」型トピック名、「漢字」型トピック名として表現する。

方法 4. 「よみ」を漢字のトピック名の”異形”として表現する。

しかし、そもそも、トピックマップのオントロジとして、漢字が表すトピックと読み仮名が表すトピックを同じトピックとして扱ってよいのかという問題を抱えている。また、奈良時代において、史料の漢字表記について、現在と同じ読みをしていたかどうかの保証はない。「ある人名の、ある時点での読み方」ということで、複数の「よみ」が存在する可能性もある。そのような状況の中、「よみ」の扱いについては、いろいろ議論の余地がある。

方法 1 ではツール上の問題もある。Ontopia に含まれているツールの一つである Vizigator(グラフィックビューワ)では、Scope の制御が困難であり、本来の意図とは異なる表示（トピック名が漢字で表示されたり、読み仮名で表示されたりします）がされてしまう。これは、研究上の情報提供として、決定的な瑕疵ではないが、表示のありかたとしては、決してきれいではない。方法 3,4 も同様の問題を含んでいる。

方法 2 では、「よみ」と「時点」のペアを”内部出現”として複数もつことができる。ただ、「よみ」をソフトキーにしたい場合、複数の中のどれを使用したらいいのか、また、現時点でツールとしてどこまで扱えるのか検証する必要がある。現時点では、方法 2 にて表現している。

さらに、トピックマップでの情報を、より汎用的に

利用できるようにするためには、RDF などへの書き出しなどの必要性も生じてくる。現在、正倉院文書トピックマップでは、外部に RDF を書きだすボタンを置いたり、RDF を提供したりはできていない。しかし、Ontopia に含まれるツールの一つである Omnipigetor や API を使用することで、RDF の書き出し自体は不可能ではない。しかし、当然ながら RDF とトピックマップではデータモデルが異なるため、トピックマップから RDF への変換がどこまで有効に行われるか、特にトピック型、関連型等、型の部分については、検証する必要がある。今後の検討課題である。

- 『人文科学とコンピュータシンポジウム 2012』 pp125-132, 2012 年
- 2) Ontopia: <http://code.google.com/p/ontopia/>
 - 3) tolog A topic map query language: <http://www.ontopia.net/topicmaps/materials/tolog.html>
 - 4) 正倉院文書データベース (SOMODA) <<http://soma.media.osaka-cu.ac.jp/>> (参照 2013-08-26).
 - 5) Guide to Shōsōin Research <<https://my.vanderbilt.edu/shosoin/>> (参照 2013-08-26).
 - 6) 江上敏哲『本棚の中のニッポン—海外の日本図書館と日本研究』(笠間書院, 2012 年)

5. おわりに

トピックマップのような、メタな情報を史料や関連知識につける動きは、試みはあるものの、決して多いわけではない。その理由は、まず大量にある史料については、デジタル化の作業を優先させるべきであり、実際に中身を読んで情報を付加するという作業は二の次であると考えられているからであろう。ただし、日本古代史、それも奈良時代史のような、比較的史料の限られたところで、このような試みを行っておくことは、将来の史料情報へのアクセス可能性を担保するという意味で、重要なものであると考える。

また、単純なデジタル化作業を行うのは、現在の Web をとりまく環境を考えれば、技術的障壁ではなく、「物量による単純作業」の費用が捻出可能かどうかだけの段階に来ていると考えてよい。

一方で、史料のデジタル化は進んでも、それが最終的な成果に直結しうるものはいまだにないのが現状である。この現状を解決するためには、解析ツールの発展も望まれるが、あわせて「質の異なる情報発見」や「今までやられていたことのモデル化」が必要な段階に来ているのではないだろうか。

日本史学の史料へのアクセスの需要自体は、国内外関わらず多いが[4]、その需要を満たすにはいまだに至っていない。アクセス手段をより多様化するため、情報を多様に結びつけた研究を支援するためのデータ・コンテンツのあり方を洗練させていくことで、人文科学研究にたいする情報学の寄与の度合いは、今後もますます高まっていくであろう。

本稿は科学研究費若手研究 B「奈良時代知識 TopicMaps 辞書構築による正倉院文書研究方法論の記述の試み」（課題番号 25730199・代表者後藤真），および基盤研究 A「正倉院文書の多元的解析支援と広領域研究資源化」（課題番号 24242025・代表者山口英男）の成果によるものである。

参考文献

- 1) 後藤真・内藤求「正倉院文書データベースへのトピックマップ応用による奈良時代知識情報構築の試み」