

デジタル法寶義林における研究データの共同構築

渡邊要一郎¹ 永崎研宣² 大向一輝¹ 井野雅文¹
村瀬友洋³ 朴賢珍¹ 下田正弘¹

概要: デジタル法寶義林は、仏教研究におけるフランスと日本の共同研究の成果としての『法寶義林』をデジタル化し、地図・年表上で情報を閲覧できるサービスとして提供するだけでなく、このために構築した TEI P5 ガイドライン準拠の研究データをオープンライセンスで公開したものである。本稿では、これを構築するにあたって取り組んだデータの共同構築のプロセスを中心に、このシステムの概要と課題について報告する。

キーワード: 仏教研究, 大蔵経, TEI, 歴史地図

Collaborative Development of Research Data in Digital Hôbôgirin

Yoichiro Watanabe¹ Kiyonori Nagasaki² Ikki Ohmukai¹ Masafumi Ino¹
Tomohiro Murase³ Park Hyunjin¹ Masahiro Shimoda¹

Abstract: Digital Hôbôgirin is a Web service for Buddhist studies to conveniently show research information on a historical map with a TEI compliant XML file under open license as a result of bilateral collaborative research project between France and Japan. This presentation reports the process of development of the system and the data which were collaboratively gathered.

Keywords: Digital Hobogirin, Buddhist studies, TEI, historical map

1. はじめに

本稿は、2021年11月27日に公開されたWebサイト『Digital 法寶義林』及びそこで公開されている研究データの共同構築に関する研究報告である。本稿では、まず、Digital 法寶義林の元となった『法寶義林』及びWebシステムとしての『Digital 法寶義林』について紹介した上で、そこで用いるべく共同構築された研究データの構築手法について報告する。これにより、研究データの共同構築の事例を共有・蓄積することが本稿の目的である。

2. 『法寶義林』とは

『法寶義林』は、仏教に関する一種の百科事典であり、1920年代、仏教研究の一大拠点であったフランスにおいて計画されたヨーロッパで最初の大規模な仏教辞典である。フランス学士院、日本学士院、フランス極東学院、日仏会館、フランス国立科学研究センターといった諸機関の後援を受けて進められた共同プロジェクトとして、1929年にAからBombaiまでの項目を収めた第一分冊が刊行されて以来、2003年には第八分冊が刊行されており、その間二冊の別冊も発刊されている。一つは1978年に刊行された大正新脩大蔵経に収められた経論の目録、その著者・訳者の解説付き目録であり、もう一つは1984年刊行の第一分冊から第四分冊までの漢字索引である。

この『法寶義林』は、「第一には、インド学者には中国・朝鮮・日本における仏教教義の発展について資料蒐集を可能にし、一方、東洋学者には仏教によって伝播された多くの概念の印度起源を見出すことを可能にすることであり、第二には、西洋人に日本の同時代的な仏教学の業績を紹介すること」をその目的としており、その成果は、既存の日本の仏教辞典や研究を踏まえつつ、新規情報も織り込んだものとなっている。そしてその「広い文化史的視野」によってもたらされた学術的成果の印度学仏教学への貢献は、日本においても広く認められている[1]。

3. 『Digital 法寶義林』とは

『Digital 法寶義林』[2]とは、『法寶義林』のうち、上述の、1978年に刊行された別冊に収録されているデータをSAT大蔵経データベース研究会の協力の下でオンライン化したものである。単にデータを検索できるようにするのみならず、収録された経論(3,018件)と著者・訳者(967件)の関係や、著者・訳者の様々な事績を通じた相互関係など、この資料に含まれる各種の情報を横断的に閲覧できるようにすることで、デジタル・ネイティブ世代の初学者が仏教学の成果に触れやすくなることを目指すのみならず、資料を読むだけでは気づきにくかったことを視覚的に確認できるようにすることで、学習や研究を進めるに際して新たなきっかけを見出せるようにすることも目指している。ま

1 東京大学 2 一般財団法人人文情報学研究
3 SAT大蔵経データベース研究会

た、日仏共同事業のため、インターフェイスやマニュアルにはフランス語版・日本語版・英語版の三カ国語版を用意している。

4. Digital 法寶義林のシステムの概要

Digital 法寶義林のデータは TEI (Text Encodnig Initiative) P5 ガイドライン[3] version 4.2.0 に準拠した形式で作成して Web サーバに置いてあり、そのままダウンロードすることも可能だが、Web ブラウザの閲覧システムで表示させる際には、Web ブラウザから適宜 URL によるクエリが Web サーバ側に送出され、Web サーバ側では問い合わせの内容にあわせてこの TEI 準拠のデータを PHP のスクリプトで適宜変換して JSON データとして返戻し、それを JavaScript で Web ブラウザ上に表示処理をするという形となっている。

表示に際しては、とりわけ、事績に関連する寺院や地名 (1,009 件)、そしてその年代を地図・年表上にもプロットできるようにするために、既存のオープンソースの Javascript ライブラリ等を適宜利用した。基本的なライブラリとしては jQuery を使用し、デザインは Bootstrap を利用した。地図については Leaflet[4] を介して OpenStreetMap[5] を利用し、マーカーをプロットするようにしている。特定の人物を検索してその事績の関連する場所を地図上に表示でき、時系列で前後関係がある場合には矢印でつながるように表示している (図 1)。この地図上での人物の表示には対応するパーマリンクが用意されており、地図の左上にある「リンクを取得」ボタンをクリックするとそのパーマリンクがクリップボードにコピーされる。たとえば、図 1 のパーマリンクは以下ようになる。

<https://tripitaka.l.u-tokyo.ac.jp/hbgrn/HMap-TP203-TP216-TP679-TP554>



図 1 地図上での著者・訳者の事績のプロット

また、この地図は、デフォルト表示では、1,009 件の寺院・地名のうち座標情報を得られたデータ 952 件がマーカーとしてプロットされるようになっており、マーカーをクリック

されるとポップアップ表示される寺院名・地名をクリックするとその寺院名・地名に関連する人物を検索できるようになっている。このマーカー表示においては、近い箇所にマーカーが多く重なる場合には、Leaflet.markercluster[6] を利用して自動的にクラスタ表示されるように設定した。

年表については、Web サイトへの実装が簡単な vis.js[7] を利用した。時間情報は人物情報における事績等のいわゆるイベント情報のエレメント<event>に属性として付与されており、人物情報をサーバから JSON データとして取得した際にその一部に含まれる。これらデータを vis.js の timeline 表示用ライブラリに読み込ませることで年表上での情報の表示を実現した (図 2)。この年表上のアイテムが座標情報を持つ寺院名・地名である場合、それをクリックすると、地図上でその位置がクローズアップされるようになっている。

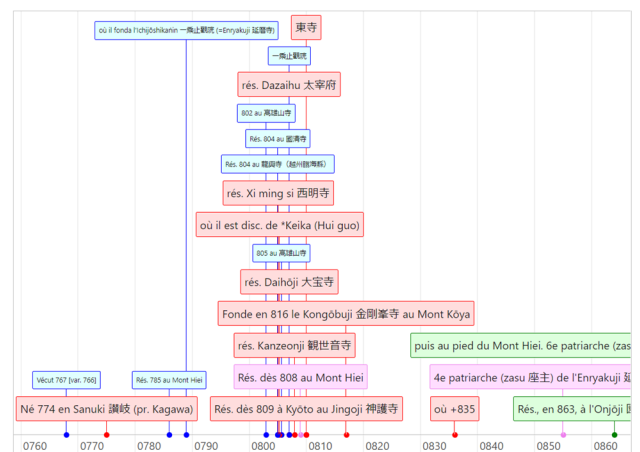


図 2 タイムライン上での表示

人物情報には著書・訳書の情報も含まれており、この書誌情報には大正新脩大蔵経のテキスト番号も含まれているため、人物情報から著書・訳書情報を介して SAT 大蔵経データベース[8]の該当テキスト情報にリンクされている。SAT 大蔵経データベースでは各テキストをベースとして様々な Web 上の仏教学関連リソースと連結されていることから、Digital 法寶義林はこれを通じてそれらとリンクした形となっている。

TEI 準拠のファイルから必要なデータを抽出して JSON データを返戻する機能については、サーバ側のプログラムとして実装し、今回採用したプログラミング言語 PHP に提供されている SimpleXML 拡張モジュールを利用した。これにより、XML ファイルをパースして PHP の配列として扱えるようになるため、配列とした上で検索を行い、出力時には配列を JSON データに変換する関数を用いて JSON データとして出力する形とした。また、仏教学においては異体字同時検索が必要となるが、これについては、文字情

報サービス環境 CHISE[9]が提供する文字関係情報のデータを PHP の検索プログラムに組み込むことで実装した。

5. データの構築

Digital 法寶義林のデータ構築は、以下の3つのフェーズで行われた。

- (1) テキストデータの作成
- (2) 人名（著者・訳者等）を起点とした書誌やイベントの TEI/XML に準拠した関連付け記述
- (3) イベントに関する地理座標と著者翻訳者の人名典拠 ID の付与

このうち(2)までについてはすでに発表しており、本発表では主に(3)のフェーズについて扱う。

5.1 データ構築の手順の方針

Digital 法寶義林では、データ単体としての幅広い応用可能性を目指し、TEI/XML に準拠したデータの構築を行ってきた。フェーズ(3) では地理座標情報や人名典拠 ID 等の、『法寶義林』の外部にある情報を付与することになる。これにあたり、作業速度をはやめるために複数人により共同でデータ付与を行うこととした。しかしながら、一つの TEI/XML のファイルを XML Editor で直接編集しようとした場合、データの同期が必ずしも容易ではなく、作業効率を高めることが難しい。そこで、Web 上での共同編集において利便性が高く、かつ一定のデータ形式の制約を編集者に課すことができる Google Spreadsheet を用いて座標情報を入力し、それを Python のプログラムにより自動的に TEI/XML ファイル全体に取り込んで全体として一つの TEI/XML ファイルを作成することとした。

5.2 作業対象となるデータの構造

今回の作業対象となったのは地理座標情報と人名典拠 ID である。いずれのデータも、本件の元となる『法寶義林』に含まれる情報ではなく、作業者が外部の情報に基づいて付与するものであることから、誰がどういう情報源に基づいて付与したか、ということが重要である。また、扱われるのは歴史的な寺院や地名の座標情報であり、情報の信頼度にかかなり幅があるため、情報の信頼度についても情報を記述しておける必要がある。そのようなことから、Google Spreadsheet では、各寺院・地名などの情報、及び人名に対して独自の ID を付与し、その ID を 1 行として、以下の項目を記述することとした。

- (1) 注記 1
- (2) 注記 2
- (3) 入力者
- (4) 根拠となった情報の URL

- (5) 座標情報 (latitude, longitude)
- (6) 確認者名
- (7) 信頼度

現段階では、地理座標情報についてはこれらのデータのうち一部を利用し、元の『法寶義林』に加えて以下のようなデータの構造とした。すなわち、(3)は<geo>エレメントの@resp 属性、(4)は@source 属性、とし、(5)は<geo>のテキストコンテンツとした。また、(1)、(2)、(6) の記載については今後の課題とした。図 3 としてマークアップの例を挙げたので参照されたい。

```
<place xml:id="vh0026">
  <placeName xml:lang="ja-Latn"> Chimyōin </placeName>
  <placeName xml:lang="zh"> 智明院 </placeName>
  <location>
    <geo cert="low" resp="渡邊要" source="https://www.wikidata.org/wiki/Q535065">
      34.2142548,135.584624 </geo>
    </location>
  </place>
```

図 3 寺院情報のマークアップ例

人名と典拠 ID に関しては、地名よりも簡素にしており、VIAF と Wikidata において記載のあるものを Google Spreadsheet 上に記述し、それを取り込む形になっている。取り込んだ状態の TEI/XML ファイルの該当箇所は図 4 のようになっている。

```
<person xml:id="TP13">
  <idno type="VIAF"> http://viaf.org/viaf/358146997180518890414 </idno>
  <idno type="URI"> https://www.wikidata.org/wiki/Q1267947 </idno>
  <name type="main">
    <persPronouns xml:lang="ja-Latn"> AN SEIKŌ </persPronouns>
    <persPronouns xml:lang="zh-Latn"> An Shi gao </persPronouns>
    <persName xml:lang="zh"> 安世高 </persName>
  </name>
```

図 4 人名情報のマークアップ例

このように、TEI/XML のデータ構築においても、定型的内容については Google Spreadsheet を用いた共同作業も可能であり、共同作業システムの構築にかかるコストを考慮するならば、むしろ、このようにして作成されたデータの断片をスクリプトで TEI/XML ファイルに取り込むという手法は十分に有効である。

6. Digital 法寶義林の課題と今後の展開

本稿では、Digital 法寶義林の概要とデータの共同構築に関する報告を行った。Digital 法寶義林は、ここに至ってようやく、1978 年に刊行された資料に基づいたデータを提供したところである。ここに掲載されている様々な情報は、それ以降の研究の発展を受けて多くの追記が可能であり、部分的には修正したり、異論を並記したりする必要もある。そのような情報の記述に際しては、TEIP5 ガイドラインが

提供するエレメント・属性を利用することで適切に対応可能だが、そのような記述をインターフェイス上でどのように表示するかというのはまた別の、次の課題である。今後は、直感的に理解しやすいインターフェイスの設計にも取り組んでいきたい。

謝辞

本研究は、JSPS 科研費 JP19H00516, JP19H00526 の助成を受けたものである。また、コレージュ・ド・フランスの Jean-Noël Robert 先生、彌永信美先生のご理解とご助力により推進されたものであり、深く感謝したい。

参考文献

- [1] 松田訓典, 彌永信美, 永崎研宣, 下田正弘. 2010. 「フランス語仏教辞典『法寶義林』目録のデジタル化とその課題——TEI ガイドラインの適用を通して——」. 研究報告人文科学とコンピュータ (CH) 2010 (7): 1-5.
- [2] Digital 法寶義林 (Hôbôgirin). <https://tripitaka.l.u-tokyo.ac.jp/hbgrn/> (2022-01-26 閲覧)
- [3] P5: Guidelines for Electronic Text Encoding and Interchange. <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html> (2022-01-26 閲覧)
- [4] Leaflet - a JavaScript library for interactive maps. <https://leafletjs.com/> (2022-01-26 閲覧)
- [5] OpenStreetMap. <https://www.openstreetmap.org/> (2022-01-26 閲覧)
- [6] Leaflet.markercluster. <https://github.com/Leaflet/Leaflet.markercluster> (2022-01-26 閲覧)
- [7] Vis.js. <https://visjs.org/> (2022-01-26 閲覧)
- [8] SAT 大正新脩大藏經テキストデータベース. <https://21dzk.l.u-tokyo.ac.jp/SAT/> (2022-01-26 閲覧)
- [9] 文字情報サービス環境 CHISE. <https://www.chise.org/> (2022-01-26 閲覧)