

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 25 日現在

機関番号：82683

研究種目：基盤研究(C) (一般)

研究期間：2012～2014

課題番号：24520060

研究課題名(和文) リソース間連携を考慮したサンスクリット写本データベースの構築に関する基礎研究

研究課題名(英文) Basic research toward the creation of Sanskrit manuscript database with resource-interoperability

研究代表者

苔米地 等流 (Tomabechi, Toru)

一般財団法人人文情報学研究所・仏典写本研究部門・主席研究員

研究者番号：60601680

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：本研究は、東京大学所蔵サンスクリット写本データベースの改良を通し、写本情報の効率的な電子的記述手法の確立を目指すものである。主たる成果としては、同データベース内容をTEI P5に準拠したXMLによって構造化し、新たなデータベースの構築に向けた技術的基礎を確認したことが挙げられる。これによって将来的、研究上の利便性がより高いサンスクリット写本データベースが実現されることが期待される。また、本研究で得られた知見が日本国内の他の写本コレクションにも適用され、資料のオンライン公開が促進されることも期待できる。

研究成果の概要(英文)：This project aims at establishing an effective method of electronic description of Sanskrit manuscripts though an attempt to improve the University of Tokyo Database of Sanskrit manuscripts. One of the major achievements of the project is the confirmation of the usefulness of XML (compliant with TEI P5 Guidelines) for structural description of Sanskrit manuscripts. It is expected that the result of this research will contribute to the realization of a new Sanskrit manuscript database which is more effective and convenient. It is further hoped that this research stimulates the creation of online databases of other Sanskrit manuscript collections in Japan.

研究分野：印度哲学・仏教学

キーワード：サンスクリット写本 データベース 仏教学 印度哲学 人文情報学

1. 研究開始当初の背景

インド学仏教学の分野では近年、写本に基づく原典研究の重要性への認識が高まっており、新出のサンスクリット写本に基づく文献研究も活況を呈している。その一方、既知のサンスクリット写本コレクションについては、ごく一部の例外を除き閲覧・研究利用における不便が解消されていない面があることは否めない。一般にサンスクリット写本は、所蔵機関において貴重書の扱いを受けることが多いため、現物を直接参照することは容易ではなく、複写取得の手続きも煩瑣である。また、写本画像データがウェブ上で公開されている場合であっても、多くの場合、データベース仕様の問題などからデジタルリソースとしてのメリットが必ずしも十分に活かされていないのが現状である。

現在運用されているサンスクリット写本画像データベースの代表例として、2006年より東京大学東洋文化研究所が公開する同大学総合図書館所蔵写本のデータベース(「東京大学総合図書館所蔵南アジア・サンスクリット語写本データベース」)があり、これが学界に裨益するところは少なくない。しかしながら、人文学研究における近年のデジタル化技術の進展状況から見ると、当該データベースはいくつかの点で問題と限界を抱えているのも事実であり、改良・拡張が望まれていた。

2. 研究の目的

本研究課題は、上記の東大所蔵サンスクリット写本データベース(以下、「東大データベース」)の機能を拡張し、研究利用上の利便性を向上させることを中・長期的目標とする。そのために、先ず以下の3点を、補助事業期間中に検討すべき当面の課題として設定した。

(1) サンスクリット写本情報を電子的に記述する効率的手法の模索

(2) 効率的かつきめの細かい検索を可能とする写本データベースの構想

(3) 他のインド学仏教学関連デジタルリソースとの連携を実現するインターフェースの実現

これらの課題を東大データベースをテストケースとして検討し、より利便性が高く、情報の更新による持続的運用が可能なサンスクリット写本データベースを構築するための基礎的手法を開発することが本研究課題の主たる目的である。

3. 研究の方法

上述のとおり本研究課題においては、現行の東大データベースを作業の基盤とする。これは、東京大学所蔵サンスクリット写本がすでに画像化データ化されているため、画像データベース構築に際して最もコスト・時間を

要するフェーズがすでに完了しており、純粋に写本情報の電子記述の問題に集中できるという利点からの選択である。

現行東大データベースは、1965年に刊行された写本目録S.

Matsunami, *A Catalogue of the Sanskrit Manuscripts in the Tokyo University Library* (以下、『松濤目録』)に記載される情報を収録したメタデータベース部分および写本の画像データを格納するストレージ部分からなるが、本研究はこのうちメタデータベースの内容を構造化し、機能拡張・情報の持続的更新を可能とすべく、以下のような手順で遂行された。

まず、データ構造化の予備調査として『松濤目録』の内容を精査し、写本情報記述に必要なデータフィールドの洗い出しを行なった。現行の東大データベースは、『松濤目録』の内容をフラットなテキストデータとして格納しているため、ここの写本にまつわる特定の情報単位(文献作者名・写本の物理的情報・年代など)をピンポイントで指定して検索するのが困難である。そのため、本研究では、『松濤目録』の記載を、個別の情報単位に分解し、それらを独立的に記述する手法を選択した。

次いで、写本情報の各単位を電子情報として格納する手法について検討し、XML(拡張可能なマークアップ言語)を採用することに決定した。この際、既存のXMLスキーマを応用するか、新規に専用スキーマを設計するかを選択する必要があったが、検討の結果データの汎用性を重視し、人文学分野でのデータ構造化に広く利用されているTEI P5をスキーマとして採用することを決定した。TEI P5では、写本の記述を目的としたモジュールが用意されており、検討の結果『松濤目録』の記載情報もこのモジュールを用いて概ね記述可能と判断した。

上述の情報内容分析およびTEIを用いた電子化手法に基づき、『松濤目録』記載の写本518点に関する情報のXML化を開始した。この作業は、サンスクリット語およびサンスクリット写本に関して十分な知識を持つ入力作業員2名に依頼し、最初の2ヶ年度中に全写本についてのXML化をほぼ達成したが、これと並行してデータの構造化手法の再検討をつづけ、入力作業員からのフィードバックに基づき随時手法の最適化につとめた。

『松濤目録』は刊行からほぼ50年を経過しており情報内容が古く、加えて目録自体の誤りもある。そのため、単純に当該目録を再現するだけでは、学術研究用データベースとしては不十分である。本研究では、データ構造化手法の検討と並行し、写本情報をアップデートするための調査も行ない、目録刊行以後に発表された、写本に関する研究成果についての情報を収集した。

本来は、最終年度に新データベースの実験的実装を行なう予定であったが、新たなメタ

データベースと連携して稼働すべき東大データベースがセキュリティ上の問題から長期間停止を余儀なくされたため、計画を変更、すでに入力された XML データの見直しに注力し、入力作業者にデータの再校正を依頼することとなった。

4. 研究成果

本研究で得られた主要な知見としては、XML、特に TEI P5 による構造化が写本情報記述の手法として、完璧ではないまでも十分に有効であることが確認された点が挙げられる。『松濤目録』自体の記載内容は十分に構造化されておらず、必ずしも構造的に連関しない情報が同一の項目下に列挙されているが、目録の内容を意味的・論理的単位に分解して構成し直した XML データとすることによって、より柔軟なデータベース実装を実現できるものと考えられる。また、このように構成された XML データは、情報の追加によるアップデートが容易であり、データベースの持続的な運用に寄与するであろうと思われる。

また、TEI を用いた構造化によって、現行東大データベースにおいて問題となっていた、『松濤目録』とデータベース間の文献番号に関する齟齬を解消することができた。この問題は、一つの写本に複数の文献が収録されている場合、『松濤目録』は[写本番号]+[写本内での文献番号]という形で階層化しているのに対し、現行東大データベースは全文に個々の写本を跨ぐ通し番号を与えているために生じたものであるが、XML を用いた構造化によって『松濤目録』の階層構造を保存することが可能となった。

上述した一連の構造化手法が他のサンスクリット写本コレクションにも適用可能であることを、京都大学文学部所蔵サンスクリット写本目録の精査によって確認しており、本研究で得られた知見が他のサンスクリット写本コレクションのオンライン公開の機運を高めることが期待される。

なお、上記の構造化手法については、H26 年度の日本印度学仏教学会学術大会において研究発表を行ない、その内容は同学会誌に論文として掲載された。

プロジェクト全体としての達成度については、前述の理由によって計画を一部変更したため、当初の目標であった新データベースの試験的実装を補助事業期間中に実現できなかったことが悔やまれる。また、他の電子リソースとの連携についての検討も未だ不十分であるし、『松濤目録』に対するアップデート情報を XML データに追加する作業も今後継続して進めていかねばならない。これらについては、現行東大データベースを管理運用する東京大学東洋文化研究所との連携を通じて、早期に実現したいと考えている。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 1 件)

(1) 苜米地 等流、「梵文写本データベースの XML による構造化の試み—東京大学所蔵写本データベースの改良作業を通して—」、『印度学仏教学研究』63-1、2014 年、pp. 544-538、査読有
<http://ci.nii.ac.jp/naid/110009899968>

〔学会発表〕(計 1 件)

(1) 苜米地 等流、「梵文写本データベースの XML による構造化の試み—東京大学所蔵写本データベースの改良作業を通して—」、日本印度学仏教学会第 65 回学術大会、2014 年 8 月 31 日、武蔵野大学(東京都)

〔図書〕(計 件)

〔産業財産権〕
出願状況(計 件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況(計 件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

6. 研究組織

(1) 研究代表者

苜米地 等流 (TOMABECHI, Toru)
一般財団法人人文情報学研究所・仏典写本研究部門・主席研究員
研究者番号：60601680

(2) 研究分担者

加納 和雄 (KANO, Kazuo)
高野山大学・文学部・准教授
研究者番号：00509523

宮崎 泉 (MIYAZAKI, Izumi)

京都大学・文学研究科・准教授
研究者番号：40314166

永崎 研宣 (NAGASAKI, Kiyonori)
一般財団法人人文情報学研究所・人文情報学
研究部門・主席研究員
研究者番号：30343429

(3)連携研究者
永ノ尾 信悟 (EINOO, Shingo)
東京大学・東洋文化研究所・教授
研究者番号：40140959
(H24年度まで)

馬場 紀寿 (BABA, Norihisa)
東京大学・東洋文化研究所・准教授
研究者番号：40431829