

要素間の関連情報を基盤とする仏教文献 デジタル・アーカイブの可能性

永崎 研宣[†]

[†] 山口県立大学情報化推進室

インド仏教研究においては、各文献中の要素、とりわけ文字列断片同士がどのように関連づけられているかという情報は、研究にあたっての重要な基盤として機能している。したがって、この「要素間の関連情報」を記録しておくことの重要性は言うまでもない。本稿では、現状のデータ記述手法における様々な問題点を検討し、この「要素間の関連情報」の蓄積を可能とするためのデジタル・アーカイブ構築の手法を提案するとともに、これを実現するために試作した Web アプリケーションを紹介する。

Digital Archives of Indian Buddhist Philosophy Based on the Relationship between Content Objects

Kiyonori NAGASAKI[†]

[†] Yamaguchi Prefectural University

The information about the relationships between each content object within their philosophical materials is the basis for the discipline of Indian Buddhist philosophy. Therefore, it is very important to store such information. In this article, I propose to describe the structure of the digital archives to store such information on the Web and to report the implementation of it in digital archives.

1 はじめに

要素間の関連情報、とりわけ文字列断片¹同士の関連情報は、インド仏教研究²においては不可欠なものである。その理由としては、まず、この分野が研究対象とする思想的伝統そのものの持つ特徴が挙げられる。この伝統において自らの思想を表明するためには論書と呼ばれる哲学的文献を執筆する中で、自らの学派的な基本的な文献を逐語的に採り上げ、それに対する解釈を

通じて行うというスタイルが多く見られ³、その議論の過程で、対論者の言を引用し、それを論破していくというスタイルを取ることも多い。あるいは、同じ仏教徒同士の論争であれば、自らの主張の典拠として仏教経典の文言を引用することもよく見られる。これらのことが示唆しているのは、思想史の展開において文献同士の相互関係が大きな重要性を持っていたということであり、さらに、個々の文献の内容のみならず、そ

¹ 本稿では、接頭辞、単語、文、句など、様々な異なるレベルの区切りを含む一連の文字列を総称する言葉として「文字列断片」という言葉を用いる。

² ここでは主に、インド仏教の中でも、「論書」と呼ばれる文献を通じて思想を表明した一群の仏教徒による宗教哲学思想に関する研究を取り扱う。

³ たとえば、2,3世紀頃に活躍した Nāgārjuna の『中論』に対する多くの注釈書は「中観派」を形成し、その思想の広がりにはチベット・中国・日本にまで及んだ。また、7世紀頃に活躍した Dharmakīrti の『Pramāṇavārttika』に対して書かれた多くの注釈書は、いわゆる仏教論理学の伝統を形成し、他教団や他学派との論争を通じてインド思想史における論理学に大きな影響を与えたとされている。

これらの文献同士の関連を、文字列断片のレベルで仔細に検討することが重要であり、それを通じて、思想史の解明が促進されるということである。

この分野において文字列断片同士の関連情報が重要であるもうひとつの大きな理由は、現存する研究資料の様態である。この研究分野においては、典拠となるサンスクリット文献の写本が物理的に満足に伝承されていないことが少なくない。そもそも、写本そのものが見つかっていない場合もある。そのような場合にも、チベット語訳、あるいは漢訳された文献が現存していることが多い。したがって、サンスクリット写本が見つかっていない文献に関しても、それらの翻訳文献が、その内容を推測させてくれる場合が少なくない。それらの翻訳文献それ自体は、それぞれの言語圏における仏教思想の流布のために翻訳され流通したものであり、その後、それぞれに独自の思想的展開を遂げている。したがって、それぞれの言語圏での仏教思想の展開に即して理解されることがまず重要だが、一方で、上述の文字列断片間の引用・参照関係とこれらの翻訳文献を組み合わせることで、サンスクリット文献が断片的にしか残されていなかったとしても、その語句の形を推測し、インドにおいてサンスクリットを通じて展開されていた思想的展開を再構成することにもまた、多少なりとも資することが可能である。

このような形で現存する資料から思想史を再構成しようとする営みは、世界各地で行われている。こうした営みに資する研究としては、文献¹⁾において仏教論理学派の文献を対象として提示された方法論が注目している。この方法論では、文字列断片間の関係のあり方、すなわち、引用・参照の仕方がどのようであるかを12種類に分類し、その信頼性に応じて5段階に階層化することで、関連文献を含めた思想的文脈の中で対象とする哲学的文献を再構成する手法を提示している。この関係のあり方は、文献²⁾⁴⁾においては8種類に分類され、その信頼性に応じて3段階に階層化されている。本稿において検討するのは、蓄積対象となる文献を利用する(可能性のある)人々の間で利用可能な、このような文字列断片をはじめとする「要素間の関係情報」を蓄積し、共有するための枠組みであり、同時に、現段階の検討結果に基づいて試作したデジタル・アーカイブについて報告する。

⁴ pp. lii-liv.

2 文字列断片の関連情報記述のための問題点

ここではまず、要素のうち、もっとも重要であり、かつ、すでに検討が進んでいる³⁾文字列断片に限定して検討を進めていく。

文字列断片間の関連情報を記述・蓄積するにあたって必要なのは、関連づけられる2つ以上の文字列断片の内容、位置情報、そして、それらの関係についての情報である。これらのうちでまず問題となるのは、以下の2点である。

1. 位置情報の流通性
2. 位置情報の永続性

この2点について以下にそれぞれ検討する。

2.1 位置情報の流通性

ここでまず目指しているのは、「文字列断片間の関連情報」の蓄積である。これを蓄積するためには、少なくとも、相互に関連を持つ2つ以上の文字列断片と、それに対する関連性の種類についての情報を記述しなければならない。後者については、前者がきちんと記述できさえすれば、それに対するメタデータとして何らかの方法で記述できればよいのでそれほど問題にはならないだろう。ここで問題となるのは、文字列断片の記述とその流通性である。

「文字列断片間の関連情報」は、既存の研究手法においては、研究論文や校訂テキストにおける脚注等の中で、出版された文献における頁・行、あるいは目録情報が付加された写本における行などの形で提示される。したがって、依拠するのはそういった紙媒体上での位置情報ということになる。言い換えるなら、既存の研究手法においてもそのまま流通させる情報とするためには、紙媒体上での位置情報が蓄積できることが必須となる。

しかしながら、たとえ対象となる文字列断片が行の途中で始まり、あるいは途中で終わっているとしても、既存の研究手法における紙媒体上での位置情報は、多くの場合、行単位での記述までにとどまっている。これはあくまでも、最終的には人間が目を確認することを前提とした記述方法であり、コンピュータを用いて位置情報を記述しようとする場合、より詳細に記述することが望ましい。それができなかったとしたら、蓄積された「文字列断片間の関連情報」は、結局のところ、人間の目でひとつずつ再確認しなければ意味を持ったデータとならず、何の自動処理もできないということ

になってしまうだろう。したがって、位置情報として記述する必要があるのは、紙媒体上での位置情報だけでなく、対象となる文字列断片が実際にはどこからどこまでのどういう文字列なのか、という情報も必要ということになる。

実際にどういう文字列であるか、ということを実体的に提示する方法については、様々な方法が考えられる。関連する文字列断片をそれぞれ取り出して、紙媒体上の位置情報とともに記述し、関連情報を付加するという既存の紙媒体に近い方法で蓄積していくこともひとつの方法として考えられるが、その方法では、ひとつの行に同じ文字列が重なってしまうときなど、必ずしも正確に情報を記述できるとは限らず、また、詳しくは次節に譲るが、取り出した文字列断片に変更が加えられた場合にはまったく対応できなくなってしまう。こういった問題を解決し、文字列の正確な位置情報を確認できるようにするためには、ベースとなる電子テキストを用意し、そこにおける位置情報として記述していくことが考えられる。なお、位置情報の記述方法としては Xpointer などの XML ベースの技術もひとつの選択肢として考えられる。

2.2 位置情報の永続性

前節でも多少触れたが、文字列断片の位置情報は、ベースとなる電子テキストに変更が加えられた場合、その時点で無効となってしまう可能性がある。位置情報の永続性については、HTML の「リンク切れ」に典型的にみられるように、独立した電子テキストにおける位置情報を参照しようとする場合には解決しなければならない問題である。⁵

この事態を回避するのに必要なのは、コンピュータ上の処理手続きだけではない。ここで蓄積しようとしている関連情報の利用者、及びそれを利用する可能性がある者の間で何らかの手続きを共有する必要があることがすべての前提となることには留意されたい。

独立した電子テキストを前提とする場合、これは比較的広く用いられている電子テキストの共有方法だが、電子テキストにバージョン情報を明記し、位置情報の方でもそのバージョン情報を明示しておくことで、当該電子テキストにおける位置情報の永続性のある面では確保することができる。ただし、ここで記述された

位置情報は、特定のバージョンにしか対応できないことになる。すなわち、この方式では、位置情報に変更が加わる形での派生バージョンが登場するたびに、それに合わせてすべての「関連情報」の位置情報を変更しなければならないことになってしまう。そして、それを諦めた時点で、互換性のない様々な「関連情報」があちこちに散乱することになってしまいかねない。以下では、この問題を回避し、位置情報の永続性を、より確実に確保できる手法を検討する。

3 より有用性の高い位置情報の検討

独立した電子テキストを各自が変更して公開していくという枠組みとは別の方向性もあり得る。それは、ベースとなる電子テキストを Web サーバ上に構築した Web アプリケーションで管理し、それに基づいて位置情報を記述し、関連情報も含めて蓄積していくという手法である。この方法では、ベースとなる電子テキストに変更があった場合にも、位置情報の記述・参照方法も含めて Web サーバサイドで管理することによって、見た目は常に位置情報が変わらないようにすることができる。

3.1 Distributed Multivalent Encoding

これに近い考え方としては、DME(Distributed Multivalent Encoding)¹⁰⁾ がある。DME は、記述的マークアップにおける「overlap」問題回避のための有力な手法の一つである。TEI/XML に基づく記述的マークアップは、文書の構成要素が「OHCO(Ordered Hierarchy of Content Objects)」を成していることを前提としている。⁶⁾ しかしながら、実際には、そもそも「OHCO」を構成する要素の定義でさえ分野によって異なる場合があり、さらに、その必ずしも明確でない定義に基づいて実際の文書から構成要素を読み取って記述的にマークアップするという仕事は、マークアップをする人自身の解釈を反映したものであるという側面が否定できない。したがって、複数の利用者によってマークアップされ、利用されることを前提とする場合には、「OHCO」を構成できないことになってしまうこともあり得る。⁶⁾ こういった問題は「overlap」問題として議論されており、マークアップが overlap した場合でも記述や XML によるベースを可能にするためのいく

⁵ W3C における Xpointer の議論では、「リンクの永続性」として採り上げられたことがある⁴⁾。また、HTML の「リンク切れ」の一般的な問題については文献⁵⁾において言及した。

⁶ この問題については文献⁷⁾において、電子テキストデータベース構築に際しての問題として詳細に論じられている。

つかの手法が提案されている。⁷ DMEは、このような議論の文脈において出てきたものであり、Webアプリケーションを用いることで、マークアップのoverlapを回避しつつ解釈の複数性を共存させ、それを共有可能にしようとする枠組みである。DMEにおいては、マークアップの情報は参照のベースとなる電子テキストとは別に保管される。そして、マークアップ情報は、Webインターフェイスを通じて入力され、参照のベースとなる電子テキストの位置情報に対して保存され、表示に際しても、Webインターフェイスを通じ、必要に応じて必要なマークアップ情報が、場合によっては電子テキストと共に表示されることになる。この表示の仕方については、インターフェイスを必要に応じてプログラミングすることによって、TEIをはじめ、様々な記述方法に対応できることになる。

3.2 参照のベースとなる電子テキスト

スタンドアロンな環境で利用者が個々に何らかのアプリケーションを利用するという形態が前提となる場合⁸、OSの相違が原因で同じアプリケーションを使えず、結果として、文書の共有がうまくいかなくなってしまふことがあり、あるいはまた、商用アプリケーションであれば、費用の問題で共有できない場合がある。さらに、データフォーマットの規格が公開されていない場合には、アプリケーション開発者/企業の多様性が失われ、アプリケーションの機能が不十分なまま放置されたり、不当な価格高騰を招いたりすることにもなりかねない。TEIをはじめとする記述的マークアップが発展してきた背景にそういった事態を打開しようとする意図が存在したことは確かである⁶。

しかしながら、前節のように、Webアプリケーションを介して情報の蓄積・共有を行う場合、少なくともWebアプリケーション内では、スタンドアロンな環境において懸念されることのいくつかはあまり問題とならない。そして、W3Cの継続的な努力を通じてWebの規格がかなり共通化され、FirefoxというOS環境をあまり選ばないフリーのWebブラウザソフトが公開さ

れ、さらに、Webページ上でのインタラクティブな機能を補完する枠組みとしてAJAXが登場してきたことなど、様々な環境の変化により、ほとんどすべての利用者が同時に同じアプリケーションを使えるという状況がWebにおいて現出しつつある。Webアプリケーションに蓄積された電子テキストやそれに付随する様々な情報（紙媒体上の位置情報、様々な記述的マークアップ、当該文献やそれに関連する文献の書誌情報等々）は、利用者の必要に応じて必要な形で提供され、さらに、それに対して、利用者が必要に応じて情報を追加していくことも技術的には可能となっているのである。

それを前提とした上で、参照のベースとなる電子テキストはどうあるべきか。まず、少なくとも、流通性を確保するためには、頁・行単位での紙媒体上の位置情報がきちんと再現できることが必要である。さらに、紙媒体上での文字の位置まで確定できるようになっていけば、位置情報の流通性は、より確実なものとなるだろう。参照のベースとなる電子テキストをXML文書とすればXpointerを比較的容易に利用できるようになるが、ここでは、DMEの考え方を徹底し、ベースとなる電子テキストは、紙媒体上の位置情報を反映したプレーンテキストとする。その上で、文字単位で指定可能なポイント情報を付加する。そして、ベースとなる電子テキストのポイント情報を位置情報として、

- (i) 「2つ以上の文字列断片の位置情報とそれらの関連情報」
- (ii) 「文字列断片の位置情報と記述的マークアップ情報」

を記述できるようにする。この仕組みによって、「overlap」を気にすることなく、DMEの考え方に基づく記述的マークアップを行いつつ、さらに要素間の関係情報も蓄積していくことができるのである。そして、ユーザが閲覧する際には、必要に応じて、電子テキストとともに、蓄積されたマークアップや関連情報が閲覧できることになる。なお、このようなデータ形式に対する検索機能は一見すると困難なことになりそうだが、プログラミングによって様々な形で追加できるため、本質的な問題ではない。⁹

⁷ DeRoseは「TEI-style Milestones」として1対のemptyタグを開始・終了タグとして用いる解決策を紹介している⁸。この場合、XMLとしてパースが可能、すなわち、XML向けアプリケーションの多くを利用できるという処理上のメリットは残るが（しかもこのメリットは有意義なものだが）、「OHCO」モデルという意味では有効性を失ってしまっている。

⁸ インターネットが一般に普及する以前の人文科学研究者の研究環境においては、このようなコンピュータ環境を前提とせざるを得なかった。

⁹ ただし、データが大きくなると検索に時間がかかるという実用上の問題が発生することがある。たとえば、大蔵経コラボレーションシステム⁹においては、約600万行の電子テキストを対象とした検索を行う機能が提供されているが、システムを運用しているサーバのハードウェアスペックがあまり高くないこともあり、データ全体に対する検索の場合、一度の検索に数分を要している。

3.3 いくつかの問題点

ベースとなる電子テキストは、まったく不変というわけにはいかない。一応、紙媒体上の位置情報との対応を前提として作成されているため、活字による版本に関しては、変更が加えられることはそれほど多くはないと思われるが、写本、手稿等を直接電子化する場合には、電子化作業者の解釈の余地があり得ることから、変更が加えられることも想定しておかねばならない¹⁰。この場合には、ベースとなる電子テキストに対するポイント情報の付加の仕方を工夫することである程度は対応することが可能だろう。ただ、あまりに大きく変化してしまう場合には、「文字列断片間の関連情報」における位置情報の変更が必要となるかもしれない。

あるいはまた、複数の写本等の資料がありながらも断片のあるいは不完全なものであり、ベースとして利用し得る文献が存在しない場合もあり得る。この場合には、それらの資料に基づいて参照のベースとなる電子テキストを新たに構成する手法を検討する必要があるかもしれない。ただし、この仕組みを提供することは、結果として、資料とは異なる文脈での恣意的な解釈を許容することにもつながりかねない。あくまでも、その文献に即した方法論に厳密に基づいた上で、適切にその手法を定義できる場合に限定されなければならないことには注意されたい。¹¹

3.4 文字列以外の媒体

仏教文献デジタル・アーカイブにおいては、文字列として位置情報を特定できない資料(画像、音声、動画など)も対象としなければならない可能性があり得る。あるいは、文字列であっても、不鮮明な、あるいは不明瞭な文字(列)を扱う場合には、典拠としての写本等の実際の画像が提示されることによって根拠が確認しやすくなり、結果として議論が行いやすくなる。このような、文字列以外の媒体に関する関連情報、すなわち「要素間の関連情報」を蓄積する場合にも、重要となるのは位置情報である。既存の媒体上に還元可能な形で位置情報を記述可能にすることで、「要素間の関連情報」の流通性と永続性は、文字列断片の場合と

¹⁰ この変更の権限が誰に与えられるべきものなのか、という点については、DMEにおいても問題として提起されているが、それについては別稿を期したい。

¹¹ 仏教研究におけるこの種の問題としては、文献¹³⁾において指摘されている「歴史的ブグダの抽出」の問題を参照されたい。

ほぼ同様のレベルで確保できることになるだろう。¹²

これに必要なのは、既存の媒体上に還元可能な形で位置情報の記述の仕方と、それを実際に記述し、さらに必要に応じて表示できるようにするためのインターフェイスの作成である。前者に関しては、それぞれの媒体を扱う研究分野において運用されているルールに従うことができれば最善である。ただし、このルールは必ずしも明示的に運用されているとは限らず、その場合には、採用するルールを、当該研究分野の文脈における根拠とともに明示した上で、それに従って提示していくのが適切であろう。後者に関しては、プログラミング作業が若干大変かもしれないものの、技術的には十分解決可能である。

また、この種の、文字列以外の資料の場合、種々の権利上の問題から内容を公開できないものが多いと思われるが、位置情報と要素間の関係についての情報が適切に提示できれば、それだけで十分に有意義である。

4 試験的実装

今回は、電子テキストと画像データを対象として、以上のような考え方に基づいて「要素間の関連情報」を蓄積できる仕組みを試験的に実装した Web アプリケーションを構築した。

4.1 ソフトウェアの構成

この Web アプリケーションのためのベースとなる主なソフトウェアは、Linux(CentOS4)、Apache2、PostgreSQL8.1、PHP5.2、である。これらはいずれもオープンソースによって開発されているフリーソフトウェアである。オープンソースソフトウェアを採用した理由は、主に費用面でのスケラビリティが高いという点である。

個々のソフトウェアの採用の理由に関して簡単に触れておくと、ここ数年は動作が安定しているという理由で PHP4 系統で開発していたが、今回の Web アプリケーションでは、以下の二つの主な理由により、比較的動作が安定してきている PHP5.2 系統を採用した。ひとつは、今回の Web アプリケーションでは画像上の領

¹² ただし、文献を様々な媒体でよりよく提示したからと言って、それで文献の状態のすべてを完全に再現できているわけではないことには留意されたい¹⁴⁾。たとえば、資料の特徴の一部として文献の素材を記述する手法も開発されつつある¹⁵⁾。現在の観点では研究対象とならない一次資料上の情報が、新しい観測の登場によって、記述され蓄積されるべき重要な情報となり、それが新しい研究成果をもたらすこともあるということをお忘れではない。

域指定をもターゲットとしていることから画像に関する関数が比較的豊富に用意されている PHP5 系統の方が開発しやすいという点であり、もうひとつは、画像ファイルを扱うことから、CGI の FORM 経由での複数画像ファイルの同時アップロードに対応できるようにするためには、PHP5.2 で新たに実装された ZIP 関数の有用性が高いという理由である。また、PostgreSQL は、データ型として Polygon 型を持っており、多角形の頂点座標のデータを使って多角形を対象とした様々な比較演算が可能であることから、画像データ上の領域情報の扱いの際には非常に便利である。ユーザ側のインターフェイスとなる Web ブラウザとしては、Firefox2 を前提としているが、Internet Explorer6 でも動作確認している。

4.2 参照のベースとなる資料

参照のベースとなる資料としては、『中論』に関連するいくつかの文献の電子テキストとその頁画像を利用した。¹³

電子テキストは、それぞれのデータの最小単位ごとに、ポイント情報やその他の付加情報とともにデータベースに記録した。頁画像データについては、頁画像ごとのメタ情報をデータベースに登録すると同時に頁画像データは UNIX ディレクトリに格納した。¹⁴また、それぞれの文献に関する書誌情報を登録しておいた。¹⁵

4.3 「要素間の関連情報」の記述

まず、上記の参照のベースとなる資料群に対して、

- (i) 1 領域の位置情報とそれに対するメタデータ（及び記入者・記入日時などの若干の付加情報）
- (ii) 2 領域の位置情報とそれらの関連に関する情報（及び若干の付加情報）

を記述・蓄積できるようにすることを目指し、そのためのインターフェイスを用意した。また、(i) に関しては、蓄積されたメタデータを HTML タグに変換できる機能を暫定的に設け、電子テキストを画面表示する

際に、若干のメタデータを視覚的に反映させられるようにした。¹⁶

位置情報の記述に関しては、電子テキストと画像データの 2 つに分けて説明する。電子テキストに関しては、対象となる文献の書誌 ID、及び、頁・行番号、さらに、頁・行ごとに開始され、最小単位ごとにカウントされるポイント情報を始点と終点として記述することで対象となる範囲を記述できるようにしている。したがって、この情報だけを取り出して紙媒体と照合してもある程度の流通性を確保できるはずである。画像データに関しては、対象となる文献の書誌 ID、画像の ID¹⁷、画像データ上の対象となる領域（多角形）の頂点座標を記述することで、電子テキストと同様に、紙媒体と照合してある程度は利用できることを目指している。¹⁸

インターフェイスは、画面に表示された資料の当該箇所をクリックしていくことで位置情報、すなわち、電子テキストであればポイント情報を、画像データであれば多角形の頂点座標を記述していき、最終的に、それらの位置情報に対して、関連情報、記入者、記入日時、備考などを記入し、蓄積できるようになっている。

なお、関連情報の記述・蓄積の際に、これまでに記述され蓄積されている位置情報を確認できるようにしておくことで、位置情報の正規化を促し、関連情報をより効果的に処理できることを目指している。

4.4 「要素間の関連情報」の表示

「要素間の関連情報」の表示を行うにはいくつかの方法がある。まず、(i) 特定の資料上で、その特定箇所に関連づけられた情報を表示させる方法である。また、(ii) 「要素間の関連情報」のみを取り出せるようになる方法もある。あるいは、(iii) 権利上、表示させても問題ないものだけは内容まで表示させ、それ以外については位置情報のみを表示させる、という中間的な方法もある。さらに、(iv) 要素間の関連情報を取り出してその関係をグラフとして表示することによって視覚的に文字列断片間の関連を確認することもできる。

今回のデジタル・アーカイブにおいて重要なのは、「要素間の関連情報」を蓄積できるようにするための仕組み

¹³ あくまでも試験的に構築することが目的であるため、網羅的ではないが、チベット語文献に関しては、ACIP が公開しているデルグ版の電子テキストと筆者が入力した北京版のテキストを利用し、漢訳文献に関しては、SAT で入力された大正新修大蔵経の電子テキスト⁹を用い、サンスクリット文献に関しては、底本としてよく用いられる校訂テキスト¹⁶を筆者が入力したものを用いた。

¹⁴ ただし、アクセス制御のため、Web からの直接アクセスはできないディレクトリである。

¹⁵ 一連の記述の仕方はこれまで開発してきた仏教哲学文献データベースでの記述方法³¹¹を発展させたものであり、文献¹²において「抽象テキスト」としてモデル化されている。

¹⁶ このメタデータ-HTML タグのバインディングも利用者が Web インターフェイスを通じて設定できるようになっている。

¹⁷ ここでは暫定的に、ファイル名にバインドされた頁番号を使っている

¹⁸ 画像データに関しては、特に、そのまま公開することは困難な場合が多いと考えられるため、領域情報だけで通用するようにしておくことは重要である。

みであり、この仕組みにおいて重要となるのは (i) である。この機能は、現時点では、電子テキストや画像データ等の資料を Web ブラウザ上に表示させ、表示された資料上のどこかを右クリックすると、その箇所に対して関連づけられた「要素間の関連情報」がコンテキストメニューに表示されるようになっていく。そしてさらに、そのコンテキストメニューにおける個々の関連情報の表示から、より詳細な関連情報へとリンクされている。この右クリックコンテキストメニューではいわゆる AJAX を利用しており、コンテキストメニューを表示させた時点で、その箇所に関連する情報を検索するクエリーがデータベースに対して発行され、その問い合わせ結果がコンテキストメニューの中に表示されるようになっていく。すなわち、関連情報は必要に応じて必要なものだけが随時サーバから読み出されるという仕組みになっており、これによって、資料を表示した時点では関連情報をクライアントに送信する必要がなく、関連情報が多くなっても対応しやすいようになっていく。また、関連情報の随時読み出しという点のみを挙げれば、これまでに試験的に構築してきた仏教哲学文献データベースでもすでに採用していたが³⁾、この時点では、この機能に関しては HTML によるフレーム表示を利用していたため、ただでさえ表示範囲に制約があるコンピュータのモニタ上での表示範囲をさらにせびめてしまうことになっていたが、この点に関しても、今回の右クリックコンテキストメニューの実装で解消された。

4.5 参照のベースとなる電子テキストの変更

参照のベースとなる電子テキストを変更する場合には、いくつかのケースが考えられる。今回の Web アプリケーションでは、電子テキストに対して頁・行・文字に関する追加と削除が発生するという前提で検討した。追加の場合には、頁・行・最小単位の各データに付与されたポイント情報に対して、追加された箇所の数値 ID 以降の小数点以下の数値 ID を付与していく。これによって、これまでの「要素間の関連情報」の位置情報を維持しつつ、新しい情報を付与していくことも可能となる。また、削除の場合には、単純に削除することで対応できる。

5 今後の課題

今後の課題としては、すでに触れてきたとおりだが、まず、XML をはじめとする既存の様々なデータフォー

マットでの出力に向けての、データ入力作業の流れの整備が挙げられる。上述のように、関連情報の記述・蓄積の際に、既入力的位置情報を確認できるようにしておくことで、位置情報の再利用を促し、関連情報をより効果的に処理できることを目指しているが、これはあくまでもデータ入力者の判断に多くを委ねてしまっている。これをどこまでアプリケーション側から制限していくか、どこからデータ入力者に任せるか、ということについてはまだ検討の余地が大きい。

この問題とも関連するが、様々な観点から、利用者の分類とその作業範囲をより細かく定義しなければならない。人文科学のためのデータベースの作業者の問題については、様々な議論が行われているが、管見の限りでは、データ入力からそのエンジニアリングに至るまで、人文科学研究者（≒エンドユーザー）が行えるようにすることを目指したものが多くに思われる。たとえば、BB-DB(Bare Born DataBase) は、人文系の研究者が DB のエンジニア、DB の管理者を兼ねられることを目指して開発されたものである¹⁸⁾。また、「漢讀」は、使用者が作成者にもなれることを目指し、ファイルメーカーを用いて構築されている¹⁹⁾。

利用者の定義に応じた閲覧可能な範囲も定義しなければならない。著作権をはじめとする権利上の問題のため、デジタル・アーカイブに掲載したものをすべてを誰でも閲覧できるということになると、掲載できる資料がかなり限られてしまう。したがって、掲載はするが部分的にしか表示しない、あるいは、データ入力のためだけに資料を掲載してデータ入力者のみが作業のためだけに閲覧できるようにして、閲覧ユーザに対してはメタデータのみを表示させるという方法も考えられる。¹⁹⁾いずれにせよ、この問題については別稿を期したい。

謝辞本稿執筆及び本 Web アプリケーションの試験的構築にあたっては、多くの研究者に様々なご助言・ご助力をいただいた。すべての方々をここで挙げるのは困難だが、中でも特に、参考文献リストにお名前が挙がっていない方々として、小野基先生、Karin Preisendanz 先生、Birgit Kellner 先生、Marcus Bingenheimer 先生、

¹⁹⁾ 文献²⁰⁾では、利用制限があるコーパスデータの公開方法として、データは Web サーバ上に置いておきつつ、利用者に対しては必要な情報のみ、それも、文字の長さや大文字小文字の区別、子音と母音の区別などの情報のみを提供するという手法を提示し、それを実現するために Java で作ったツールを紹介している。

守岡知彦先生、酒井真道氏のお名前を感謝とともに挙げさせていただきます。

参考文献

- 1) Ernst Steinkellner, Methodological Remarks On The Constituion Of Sanskrit Texts From The Buddhist Pramāṇa-Tradition, Wiener Zeitschrift für die Kunde Südasiens, Band XXXII, 1988, pp. 103-129.
- 2) Ernst Steinkellner, Helmut Krasser, Horst Lasic ed., Jinendrabuddhi's Viśālamalavatī Pramāṇasamuccayaṭīkā Chapter 1 Part I, CHINA TIBETOLOGY PUBLISHING HOUSE, AUSTRIAN ACADEMY OF SCIENCES PRESS, BEIJING- VIENNA, 2005.
- 3) 永崎研宣, シラブルを最小単位とする仏教哲学文献データベースについて, 情処研報, 2006-CH-71, pp. 33-40.
- 4) Eve Maler, Steve DeRose , XML Pointer Language (XPointer), World Wide Web Consortium Working Draft 03-March-1998, <http://www.w3.org/TR/1998/WD-xptr-19980303> .
- 5) 永崎研宣, 『文科系のための情報発信リテラシー』, 東京電機大学出版局, 2004年2月.
- 6) Allen H. Renear, Text Encoding, A Companion to Digital Humanities, Blackwell Publishing, 2004, pp. 218-239.
- 7) 豊島正之, XMLの骨抜き利用法, 『論集「情報処理」「古典学の再構築」研究成果報告集IV』, 2003年3月, pp. 178-187.
- 8) Steven DeRose, Markup Overlap: A Review and a Horse, Extreme Markup Languages 2004: Proceedings, 2004, <http://www.mulberrytech.com/Extreme/Proceedings/html/2004/DeRose01/EML2004DeRose01.html>.
- 9) 永崎研宣, 鈴木隆泰, 下田正弘, 大正新脩大藏經テキストデータベース構築のためのコラボレーションシステムの開発, 情処研報, 2006-CH-70, pp. 33-40.
- 10) Paul Caton, Distributed Multivalent Encoding, Digital Humanities 2007 Conference Abstracts, 2007, pp. 33-34.
- 11) 明星聖子, 永崎研宣, 編集文献学に基づく人文科学資料エディティング・システム構築の試み, 情処研報, 2007-CH-74, pp. 25-32.
- 12) 白須裕之, 歴史記述に対する概念分析の試み, 情報処理学会研究報告, 2007-CH-74, pp. 1-8.
- 13) 下田正弘, <物語られるブッダ>の復活—歴史学としての仏教学を再考する—, 『仏教とジャーナ教』, 2005年11月, pp. 357-379.
- 14) Peter L. Shillingsburg, From Gutenberg to Google, CAMBRIDGE UNIVERSITY PRESS, 2006.
- 15) 相井孝仁他, 顔料特性を考慮した表面下散乱モデルによる紙本のCG表現, 情処研報, 2007-CH-74, pp. 97-103.
- 16) L. de la Valée Poussin ed. , Mūlamadhyamakakārikās de Nāgārjuna avec la Prasannapadā Commentaire de Candrakīrti, Bibliotheca Buddhica IV, St. Petérsbourg, 1903-1913.
- 17) David Seyfort Ruegg, Two Prolegomena To Madhyamaka Philosophy, Studies In Indian And Tibetan Madhyamaka Thought Part 2, UNIVERSITÄT WIEN, 2002.
- 18) 及川昭文 山元啓史, Web公開のためのデータベース・エンジニアリング, 情処研報, 2000-CH-49, pp. 49-56.
- 19) ウィッテルン・クリスティアン他, 漢讀: 新しいテキスト・モデルに基づいた東洋学文献研究の支援ツール, 情処研報, 2007-CH-74, pp. 9-16.
- 20) Georg Rehm, et al., Corpus Masking: Legally Bypassing Licensing Restrictions for the Free Distribution of Text Collections, Digital Humanities 2007 Conference Abstracts, 2007, pp.166-169.