

東洋古典文献研究におけるデジタルテキストの適切な記述手法について —インド学仏教学のための学術知識基盤の構築に向けて—

永崎 研宣

下田 正弘

一般財団法人人文情報学研究所

東京大学大学院人文社会系研究科

本研究では、東洋古典文献の事例として、文献資料の状況が極めて複雑となっているインド学仏教学に関連する資料を採り上げ、これをデジタル化する際に留意すべき事項について検討する。そして、これにあたって利用可能と思われる、近年普及しつつある様々なシステムや枠組み、とりわけ、遠隔コラボレーションをめぐる動向について検討し、それらの流れを反映しつつ、データの持続可能性についても配慮して実装したインド学仏教学のための総合的な学術知識基盤(Research Base for Indian and Buddhist Studies)に関して報告する。このシステムは、既存の Web 研究環境に統合されたものであり、すでにコラボレーションによるデータの蓄積が行われている。今後は、主にインターフェイスに関して利用者の意見をフィードバックさせた改良を続けることによって利用者をいっそう増やし、さらに新規利用者の意見をフィードバックしていくというサイクルを展開していくことで、より洗練されたインターフェイスを構築し、システムの完成度を高めていきたい。

Towards an Appropriate Method of Description of Digitized Texts in the Field of Eastern Classics.

Kiyonori Nagasaki

International Institute for Digital Humanities

Masahiro Shimoda

Graduate School of Humanities and Sociology
University of Tokyo

In this presentation of our research project, we discuss materials pertaining to the field of Indian Buddhist studies that include complex texts, through the example of the digitization of Eastern classics. We review some systems and frameworks that have recently become widespread, especially, regarding to Web collaboration. We will then report on plans and progress concerning the "Research Base for Indian and Buddhist Studies" which was implemented on the basis of the previously collaboratively-developed data with a view toward re-examining it from the standpoint of its sustainability. The system was integrated with an extant research environment and data input has already begun. For the future, we will develop a more sophisticated interface by way of establishing a work cycle wherein user's feedback can be implemented, further attracting new users, from whom we hope to elicit further feedback.

1. はじめに

本稿は、これまで筆者らが取り組んできた、インド学仏教学のための総合的な学術知識基盤(以下、RBIB (Research Base for Indian and Buddhist Studies, インド学仏教学リサーチベース))の構築に向けた実践的研究についての報告である。RBIB は記述と表示の自由度や Web の双方向性など、近年のデジタル技術におけるいくつかの特長を生かした学術知識共有のための基盤であり、研究支援のための環境をインターフェイスとして備えるものである。従来は紙媒体で行われてきたインド学仏教学における学術情報の共有をより効率的な形で行うことを目指して研究開発が進められている。そして、中でも特に資料の取り扱いにおいて複雑な様相を呈しているインド学仏教学に関わる問題を解決していくことによってインド学仏教学全般での利用可能性を大きく高めていくことが期待されることから、本稿では、

インド学仏教学における現時点での取り組みを主に取り扱う。

2. 本研究の目指すところ

2001年にインド学仏教学のためのサンスクリット文献の Web を用いた校訂テキスト共有システムが開発され、これ以降、様々な形でインド学仏教学に関わる学術知識基盤の構築に向けた研究が行われてきた[1][2]。異なる文脈からの取り組みも様々な存在するが、本稿で報告するのは、あくまでもインド学仏教学の方法論的要請に立脚した上での取り組みである。インド学仏教学における方法論にも様々なものがあるが、現在のところは主に文献学に立脚している。

インド学仏教学においては典拠となる校訂テキストが十分に整備されていない。著者自身の手になる書物が残っている例はほとんどなく、写本として伝えられるものがほとんどである。聖書学においても一次資料の状況は似通っている面もあるのだが、イン

ド仏教学の場合、古典中国語訳、チベット語訳においてテキストの伝承のかなり古い段階での内容が伝えられているにも関わらず、サンスクリット語をはじめとするインドの言語で書かれた写本がみつからないものや、あるいはみつかったも公開されていないものが未だに少なくない。それゆえ、現在のところ、校訂テキストを作ることができない重要なテキストがいくつも残っており、サンスクリット語写本等が発見・公開されるたびに、そういった資料にアクセスしつつ、すでにある程度研究が進められている古典中国語訳やチベット語訳のテキストを参照しながら、同時に、その資料を部分的に引用・参照している資料の断片を使って内容を仔細に確認しながら研究を進めていくというスタイルが現在も必須となっており、世界中でそのようにして研究が展開されている[3]。最近でも、密教の重要経典である『理趣経』のサンスクリット語写本が発見され、校訂テキストがようやく出版されたところだが、このテキスト校訂にあたっては、関連する古典中国語訳やチベット語訳、他の文献におけるサンスクリット語による引用など、多くの資料が参照されている[4]。

さらにまた、すでに校訂テキストが公開されているテキストであっても、新たなサンスクリット語写本の発見によって校訂テキストの内容が書きかえられる場合もある。たとえば、大乘仏教の思想を理論化した重要な論書として高等学校の教科書にも登場する龍樹(Nāgārjuna)の『中論』は、2,3世紀頃に著されたものであるとされているが、実際には、元のテキストがいかなるものであったかについてまったく決着がついていない。これは、主に、後代に著された『中論』の注釈文献において引用されたテキストから抽出して本文が再構成されているが、注釈文献はすでに様々な伝承を経た上で『中論』を引用しているため、そこに引用されている『中論』は決して一様なものではなく、注釈文献が新たに発見されるたびに、『中論』本文自体もまた問題とされることになる。すでにサンスクリット語写本が見つかり校訂テキストが出版されている場合であっても、たとえば、中論の代表的な注釈書であり、すでに校訂テキストが出版されている『Prasannapadā』[5]の場合には、校訂テキスト出版時以降に見つかった同書のサンスクリット語写本を参照することによって『Prasannapadā』の本文が書き換えられる事態になっており、さらに、主そこから抽出されて本文が再構成されているサンスクリット語『中論』もまた書き換えられる可能性があるという状況になっている。このような状況であることから、『中論』のテキストの再構成についての議論は未だまったく決着を見ていない。

あるいはまた、『中論』のような、いわゆる論書と呼ばれる、具体的な著者の存在を想定することができる文献とは異なり、大乘仏教の経典については、サンスクリット語の写本が多く見つかったにもかかわらず伝承された地域や時代等によって大きく形態が異なる場合もある。そのような場合には、オリジナルを想定すること自体が困難なこともある。

本稿で取り扱うのは、そのような様々な確定しきれていないテキストをめぐる文献資料を取り扱う研究の手法そのものを、双方向デジタルメディアとしての Web 上で展開していくためのテキストの記述手法である。さらにそれは、Web がもたらすメディアとしての新たな可能性を如何にして研究上の方法論にフィードバックし得るか、ということにもつながり得るものである。

3. 本研究での取り組み

3.1. 情報の持続可能性

これまでの一連の研究の中で特に重視してきたのは、情報の持続可能性である。デジタルメディアとして作成される情報は、処理速度や開発上の利便性、あるいは商業上の要請等から特定の処理系に依存する形で作成されてしまうという状況を長らく続けてきたが、XML の登場と普及によって状況は著しく改善され、今や MS-Office でさえも XML 形式を標準フォーマットとするようになってきている。人文学分野としても、TEI(Text Encoding Initiative)コンソーシアムではすでに XML 登場以前からのこの種の問題への取り組みを始めている[6][7]。TEI では、人文学資料のデジタル化に際して、プロプライエタリなソフトウェアへの依存をやめて標準的な規格に依拠したデータを作成することでデータの持続可能性を保持しようとする意図が当初から明確であり、最初は記述言語として SGML を用いていたが、XML の普及とともに XML へと移行し、現在に至っている。TEI は、XML による記述手法をとりまとめたガイドラインの公開を通じて人文学資料のデジタル化に際しての XML による記述方法を提示しており現在のバージョンは P5 の 1.8.0 のリリース直前という状況だが、現段階においてもすべての資料を完全に記述できるだけの規格となっているわけではなく、むしろ、より良い記述方法の確立を目指した議論の場として機能しており、その議論の過渡的な成果としてガイドラインが提供されていると見るべきだろう。TEI members' meeting 2010 (於ザダル大学)において Marjorie Burghart 及び Malte Rehbein によって報告された写本デジタル化に関わる利用者アンケート¹によれば、9 割以上の利用者が TEI ガイドラインに満足しておらず、プロジェクトごとの様々な必要性に応じたスキーマのカスタマイズも広く行われているということだが、TEI では、Roma という Web アプリケーションを通じてスキーマをカスタマイズする方法も提供しており、その点についてもある程度の配慮がなされている。もちろん、これを処理するソフトウェアについても様々な開発・提供されており、世界中でより良いソフトウェアを目指した開発が続けられている。

このように、構造化データに関わる情報の持続可能性については、すでに様々な取り組みが行われて

¹ このアンケート結果は近日中に公表予定とのことである。

おり、これらの取り組みを拠り所としてさらに発展させていくことが必要だろう。

3.2. 既存の研究との連続性

構造化データと処理系というレベルではそのようにして改善されつつあるが、実際のデータ記述手法について考えるなら、個別の分野に即したより厳密なルールが必要となる。なかでも重要なのは、これまでの研究情報との連続性という意味での持続可能性という観点である。すなわち、これまで個々の研究分野において利用され共有されてきた情報の記述方法は、人文学分野全体にとって共通するものではなく、また、必ずしも必然的なものでもない場合もあるだろう。しかしながら、記録し共有するための媒体がデジタルメディアへと移行していく過程では、単にデジタルメディアにおいて情報を生成していくだけでなく、それまで流通してきた情報と可能な限り透過的に扱えるようにすることが重要となる。

この観点から、本システムにおいては、現在のところ、個々の研究情報におけるリソースの位置情報を指示する記法について、個々のリソースごとに過去の紙媒体における研究成果等における記法を検討しつつ、それらをデジタルメディア上の位置情報とマッピングさせるという形で記述するという手法を採っている。たとえば、広く用いられている古典中国語訳伝典テキストである『大正新脩大藏経』を引用した際の位置情報の記述は、第 1 巻に収録されている経典のテキスト番号 1 番のテキストの 2 ページ目の 3 段目の 4 行目、という内容が含まれる。実際にこれを記述する仕方には色々な記法があるが、デジタル化されたテキストの場合でもこの記法に基づいて位置情報を参照できなければ、著しく利便性を欠いてしまう。また一方で、デジタルテキストを直接参照した場合にも、上記の内容をうまく取り出すことができれば、紙媒体上での研究との連携がうまくいかなくなってしまう。したがって、本システムでは、上記の位置情報の値を適切に取り出せる形でタグを埋め込んだ上、デジタルデータ上の位置情報とマッピングさせるラッパーを用意する形としている。

また、位置情報はタグで埋め込んでいるため、AJAX 等による取り出しも容易であり、フォーマットの自由度も高い。そして、本システムでは、これを一步先に進めて、「その行の何文字目か」という位置情報をも記録対象とすることにしている。これについては、各文字に位置情報を付与するのではなく、指定された行の中で必要に応じてカウントするようにしている。たとえば、現時点では、Web ページ上で作業する際に、任意の文字列をドラッグすると、該当する文字列の開始位置と終了位置を JavaScript がカウントするようになっており、また、特定の文字列を呼び出す際には、呼び出しプログラムが、与えられた文字位置情報に基づいて呼び出された文字列を切り出して表示するようになっている。

なお、この手法の問題点として、依拠する文字列に修正が入った場合に、文字列の位置情報がずれてしまうということがあるが、それについては、文字列修正の際に同時に文字位置情報を修正することで対応することが可能である。

あるいは、これに変わる実装方法として、すべての文字に位置情報を事前に与えておくという方法がある。この方法の場合、依拠した文字列が修正された場合でも、文字列がなくなってしまう限り位置情報が無効化されることはないので、上記の手法に比べると実装が容易でありデータの安全性も高いように思われる。しかしながら、この手法はすでに試行済みであり[8]、小規模なデータであれば十分に対応できるが、スケーラビリティに問題があり、試行した時点ではある一定以上の規模を持ったデータへの対応が困難であると思われたため、本システムでは結果的に採用していない。あるいは、位置情報参照専用のレイヤーを動的に設けてそこで各文字に位置情報を与えるという手法もあり得るが、レイヤーを追加することによる処理系の複雑化も一方の問題となるため、現時点では採用していない。

3.3. インターフェイスの重要性

また、本システムにおいて重視していることの一つにインターフェイスの使いやすさという点がある。この観点から人文学におけるコンピュータ利用の事例において近年の成功例を見るなら GlyphWiki が挙げられるだろう[9]。GlyphWiki の登場により、従来は極めて困難だった外字のグリフの共有が極めて容易になったことはすでによく知られるところだが、ここで注目したいのはインターフェイスの秀逸さである。GlyphWiki のようなグリフの動的共有の構想において鍵となる事柄の一つに、利用者が如何にして共通のフォーマットに基づくベクター画像を作成し提供するか、という問題がある。この場合、GlyphWiki 登場以前であれば、いずれかのベクター画像作成ソフトウェアをグリフ画像作成者がそれぞれ購入し、そのソフトウェアの使い方に習熟した上で、特定のルールを規定して、それに従って皆が作業を行うという形になる。このようなやり方の場合、グリフ画像作成に限らず、いずれの人文学のデジタル化プロジェクトであっても、通常はその点で足並みをそろえるだけで大きな困難があり、よほどマネージメントを適切に行わなければプロジェクト自体が挫折してしまいがちなのだが、GlyphWiki では、Web ブラウザ上でグリフそのものを作成可能とすることで、この問題をほぼ解決してしまったのである。その発想のみならず、それを実用レベルのサービスとして実現したという点でも、GlyphWiki が人文学のデジタル化において注目に値する画期的なシステムであることは言うまでもないだろう。そして、このような仕組みの背景にある、Web アプリケーションという枠組みについても後で検討する。

¹ここでは PHP で開発した独自のプログラムを用いてい

る

あるいはまた、TEI が規格としてはそれなりに整備されてきているにも関わらず未だに十分広く利用されていない背景には、それを利用するにあたって実際に研究者が眼にするインターフェイスのわかりにくさという点があることをしばしば耳にする。上述の TEI における利用者アンケートでも、ユーザフレンドリーなツールが存在しないことが利用開始当初の障害であったという回答が 40 だったことから、人文学分野でのデジタル化における良いインターフェイスの重要性がうかがえる。現状では、汎用 XML エディタである oXygen をはじめ、いくつかの XML エディタソフトが利用可能であり、また、Emacs ももちろん利用可能だが、それらを用いて記述する際の、TEI の規格としての幅広さと編集後の結果のわかりにくさが利用者を遠ざけてしまう一因となっていることは筆者もしばしば眼にしてきた。もちろん、TEI としても、ガイドライン P5 では、利用目的に応じてエレメントをモジュール化することによってスキーマをスリム化できるようにし、さらに、そのような手順を容易に実行できるようにした Web サービス Roma を公開するなど、様々な工夫を行っているが、一般的な利用者のレベルでは未だ十分にわかりやすい汎用可能なソリューションが用意されているとは言い難い面がある。一方で、TEI ガイドラインに準拠したシステムとして専用のインターフェイスを用意しているものも散見される。本システムにおいても、より目的に特化された専用の Web インターフェイスを開発することで状況を改善するという方向を目指し研究開発を進めている。もちろん、ここでのインターフェイス開発によって得た知見やそれを運用したことによる知見は、少なくとも、東洋古典文献学全般にとっても有益であると思われるため、関連する他のプロジェクトとも積極的に共有していきたいと考えている。

3.4. Web アプリケーションという枠組み

Web アプリケーションは、本システムにおいても、テキストデータの修正や外字情報の蓄積・共有など、様々な面ですでに利用している枠組みである。Web アプリケーション登場以前は、人文学資料のデジタル化の作業においては、各作業者のパソコンに作業用ソフトウェアをインストールして個別に作業を行い、一定の段階でそれを集約するというワークフローが一般的であった。しかし、この場合には、作業の進捗状況の把握が難しく、また、よほど良いソフトウェアを用意しない限り、各自が作成するデータのフォーマットをそろえることもなかなか困難であり、しかも、各自のパソコン環境は異なっていることが多いために、実際の作業結果の統合がうまくいかないことすらあり、共同のデジタル化作業はしばしば困難を伴った。ところが、Web アプリケーションの場合には、ある程度標準化されたインターフェイスを用いることができ、また、一つのデータ入出力システムを皆で共有することが可能である。これは IT 環境の高度化なくしてはなしえなかったことだが、これによって、共同作業の標準化が極めて

容易になる道筋がひらかれたのである。とはいえ、Wikipedia のように、比較的汎用な Web アプリケーションの場合には、書式の標準化という面でまだやや困難さを残している。そこで登場してきたのが Ruby on Rails 等の Web アプリケーションフレームワークである。これにより、個別の特化された作業にあわせた任意のフォーマットを Web アプリケーションで比較的容易に複数人で作成することが可能となったのである。

Web アプリケーションのメリットの一つは、実際に格納するデータの形式と入出力するデータの形式が異なっても構わないという点である。これにより、内部のデータ形式がいかなるものであれ、Web アプリケーション側で出力用プログラムを用意するだけで、JSON や XML 等、様々なフォーマットのデータを出力できるようになる。このことは、データベース同士の相互連携という面で極めて有利に働くことになる。本システムでは、この枠組みを様々な形で利用することで、ユーザサービスのみならず Web API も提供することが可能となっている。

また、Web アプリケーションにおいては、サーバ側の処理とクライアント側の処理をいかにして切り分けるかということも重要である。かつて、クライアント側のパソコンの性能がそれほど高くなかった時期には、サーバ側で多くの処理を行い、最終的に整形された HTML ファイルのみをクライアント側に送信するというスタイルが基本であり、文字に関しても、Unicode が一般的でなかった頃は文字画像までもサーバ側で生成してクライアント側に送り込んで表示させる場合さえあった。しかしながら、パソコンの高性能化や Unicode の普及、さらには、AJAX の登場によって、クライアント側でできる処理は圧倒的に増えている。現在では、サーバ側で XML や JSON 形式のデータを作成・送信し、クライアント側はそれを受け取った後に整形して表示させる、といったことさえ一般的に行われている。このような状況では、クライアント側にいかにしてうまく処理を割り当てることで利便性を高めるか、という観点が重要になってくる。本システムでは、この側面について、常に AJAX 等の動向に注意しながら随時改良を続けている。

3.5. 関連情報の記述と表示の考え方

本システムが現時点で目指しているのは、「ある資料のある箇所についてどういう研究が行われているか」という情報の蓄積と共有である。これについてはすでに何度か発表してきたが、それをさらに深めつつより良い研究環境として提示することを目指して改良を続けているのが現在の段階である。これを以下に具体的に述べてみよう。

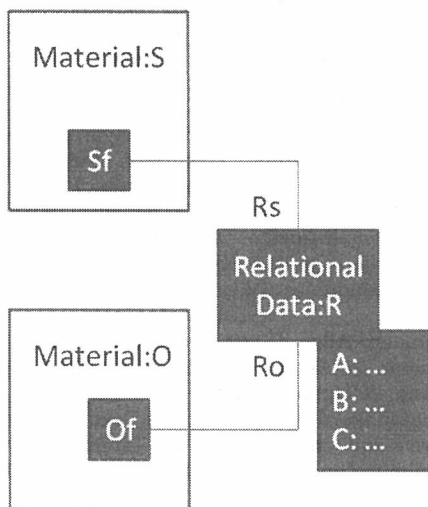
まず、ある資料のある箇所に関連する研究情報（以下、関連情報）を蓄積可能とする。これは、一次資料、二次資料のいずれをも対象とする。Web 上に参照可能な情報があれば、それを参照できるようにするが、紙媒体しか流通していない場合には紙媒体への適切な参照情報を蓄積できるようにする。こ

の蓄積に際してのインターフェイスは、可能な限り、研究者が常用する Web インターフェイス等とシームレスに利用可能なものとする。

次に、上記の蓄積された関連情報に対して、様々な View を提供することで、多様な研究成果を容易に参照しやすくしつつ、研究成果から資料を参照すること、あるいは資料同士の関連について確認することも容易にする。

以上のような仕組みを、実用性を確保した形で実装したものについては、すでに、古典中国語訳仏典テキストとそれについての現代日本語訳との間の任意の単位での対応付けについての関連情報を記述可能としつつ、その機能を SAT 大蔵経テキストデータベース（以下、SATDB）のインターフェイスに統合することで利便性にも配慮したシステムを構築しており、昨年のじんもんこんシンポジウムにおいてもその概略を発表した [2]。そして、現在では、この仕組みの適用範囲をさらに拡大する形で、チベット語訳仏典テキストとの間で関連情報の付与を実現している。以下に、このシステムについて見ていこう。

このシステムでは、関連情報 R には、ある資料 S の断片 Sf の位置情報 Rs、その資料と関連づけられる資料 O の断片 Of の位置情報 Ro に加えて、以下のような情報が与えられるようになっている。



<Figure 1: An example of relational data>

A 軸：（オーソライズ情報）

- A-1 研究者による入力（データ著者）
 - 著者名、時間、内容、備考
- A-2 オーソライズ組織によるチェック
 - チェック者名、時間、内容、備考
- A-3 オーソライズ組織による公開
 - 公開者名、時間、内容、備考

B 軸：要素間に関連の仕方についての情報

- B-1 翻訳（訳の対応・精度—逐語訳、意識等）
 - B-1-1 各要素の言語情報(基本は各文献の通りだが上書きの場合に記述)
- B-2 引用（引用の対応—ほぼ同一、少し異なる、要約等）
- B-3 異読（異読の対応—少し異なる、大きく異なる、不足、余分）
- B-4 注釈（注釈の対応・性質—入力者の任意）
- B-5 直接対応（片方が辞書や知識ベースのようなものであり一意に関連づけ可能な場合）

C 軸：データの重み付け情報

- C-1 閲覧者による重み付け操作
- C-2 データの利用状況による重み付け

上述のように、現在のシステムでは、古典中国語訳での複数のテキスト、現代日本語訳[10]、及びチベット訳のテキストとの関連付けを行っている。システム上はいずれの資料をも O として利用可能となっているが、ここでは、個々の古典的テキストを Sf、Of として相互に直接に関連づけるのではなく、それぞれの古典的テキストを S とし、現代日本語訳を O とした上で、Of を介して Sf 同士が関連づけられる View を提供している。この場合の O をここでは仮に「抽象テキスト」と位置づける。一つのテキストに関連づけられ得る様々なテキストは、伝承の過程で変化を遂げている場合があり、必ずしも一様の内容を持つとは限らず、増広や省略・欠損など、テキストの各部分に着目すると、様々な形があり得るのみならず、関連付けが不可能になってしまう場合もある。しかしながら、ここでは、現代日本語訳を「抽象テキスト」と位置づけ、それを介することで、増広・欠損部分を比較的容易に扱えるようにしている。この場合、この事例ではあくまでも仮定だが、抽象テキストは、いずれかの時点で本来そのようなあったと仮定されるテキストという位置づけとなるため、それに対する関連情報は、そのようなテキストに対する増広・欠損といった位置づけが可能となる。この仕組みの場合には、抽象テキスト自体も一つではなく、いくつもの抽象テキストを仮定し、それに対する何らかの関連付けを行うという形にすることも可能である。これは、資料をデジタル化しただけのものからは少なくとも一歩踏み出すものであり、一次資料となるテキストとは異なるレイヤーに位置づけられるべきものだろう。そして、データの蓄積システム上ではなく、運用上の記述・表示手法として実装されるべきだろう。本システムでは、現在ではこれらを並列的に扱っているが、そのような形で、各資料の階層化を行うことも予定している。

また、参照元となる資料 S において大幅な欠損がある場合、関連情報 R を記述しようとしても、位置情報 Rs がまったく記述できず、関連情報 R の位置づけ自体が難しくなってしまうという問題もある。代表的な例としては、『大正新脩大蔵経』[11]（以下、大正蔵）における『大乘中観釈論』の後半部欠損問題がある [12]。大正蔵を基盤とする SATDB を

S とするシステムでは、単純に S の欠損領域に対して関連づけるか、あるいは、S の欠損部分に何らかの架空の領域を新たに付加することが必要となる。この場合にも、先に挙げた「抽象テキスト」としての O は有効だろう。すなわち、『大乘中観釈論』の「抽象テキスト」O を仮定した上で、それに対して欠損部分に本来入るべきテキストを関連づけていくのである。これには、たとえば、高麗大藏經等がある。この「抽象テキスト」は、紙媒体の校訂テキストにおける本文に近い概念とも言えるが、ここでは、一定のルールに基づく可変性を含み得るものとする。ここで、デジタルメディアの特性を生かせることが重要となる。それについては、別稿を期したい。

3.6. 実装

実装については、サーバ側は Linux を中心としたフリーソフトウェアに、PHP による独自開発のプログラムを組み合わせている。また、クライアント側のインターフェイスとしては、主に、いわゆる AJAX ライブラリの一つである Jquery を用いることで、XML データ等を容易にクライアント側で扱えるようにするとともに、利便性の高いインターフェイスを提供している。

ユーザインターフェイスとしては、テキストを上下に二つ表示させ、それぞれの任意の箇所をドラッグして選択すると、それぞれに位置情報が一定のフォーマットで表示される。あとは、それらの二つのテキストの関係をプルダウンメニューで選択し、サーバに送信する、というだけである。これは、二つのテキスト断片間の関連付けという作業を直観的に行うことを第一に優先して設計したものであり、これまでの利用者からはわかりやすいとの評価を得ている。

なお、前述のように、ドラッグされた文字列の開始・終了位置を Web ブラウザ上で計数しているが、これに関わる Javascript 機能は、Web ブラウザごとに微妙に異なっている。このため、現時点では、比較の様々なクライアントで利用可能な Firefox だけに対応するという形になっており、このままでは互換性に問題を残したままになってしまう。しかしながら、XML の属性値としてテキストの位置情報を扱えるようになれば AJAX のフレームワークの中で処理できるため高い互換性を確保できる見通しができる。そこで、すでに述べたように、近い将来には、テキストをサーバから送信する際に、一つ一つの文字に位置情報の属性値を与えることで、XML データとして位置情報まで操作できるようにする予定である。現時点では 1 ページ分を表示しただけでブラウザが重くなって使い勝手が悪くなってしまうので、今しばらくのクライアントの高性能化を待つ必要があるが、折を見て対応したい。

4. 終わりに

本研究は、統合的な学術知識基盤の構築を目指す試みの中の一部である。学術知識基盤には、適切な

情報の記述・蓄積とともに、利用者にとって利便性の高いインターフェイスが必要である。現在の AJAX に依拠した Web アプリケーションは、そのようなインターフェイスを実現するにあたり、非常に強力な枠組みとなっている。良いインターフェイスの提供は、人文学のデジタル化への参加者をおおいに広げることにつながっていき、参加者が広がることは、検証の機会の増加につながり、そして、より洗練されたシステムへと結びついていくことだろう。

謝辞

本研究の一部は科学研究費補助金（課題番号：22700255）の助成を受けて遂行されたものであり、また、一部は科学研究費補助金（課題番号：22242002）の助成を受けて遂行されたものである。

参考文献

- [1] NAGASAKI, Kiyonori[2008] "A Collaboration System for the Philology of the Buddhist Study", *Digital Humanities 2008*, pp. 262-263.
- [2] 永崎研宣, 白須裕之, 下田正弘[2009], 大藏經における多言語対訳コーパスの構築, 人文科学とコンピュータシンポジウム論文集, (社) 情報処理学会, pp. 129-134.
- [3] Steinkellner, Ernst[1988], Methodological Remarks On The Constitution Of Sanskrit Texts From The Buddhist Pramāṇa-Tradition, *Wiener Zeitschrift für die Kunde Südasiens*, Band XXXII, 1988, pp. 103-129.
- [4] Tomabechi, Toru ed.[2009], *Adhyardhaṣṭikā Prajñāpāramitā*, China Tibetology Publishing House and Austrian Academy of Sciences Press.
- [5] La Valée Poussin, Louis de ed. [1903], *Mūlamadhyamakakārikās de Nāgārjuna avec la Prasannapadā Commentaire de Candrakīrti*, *Bibliotheca Buddhica IV*, St. Petersburg, 1903-1913.
- [6] Renear, Allen H.[2004], *Text Encoding, A Companion to Digital Humanities*, Blackwell Publishing, 2004, pp. 218-239.
- [7] Cummings, James [2007], *The Text Encoding Initiative and the Study of Literature, A Companion to Digital Literary Studies*, Blackwell, pp. 451-476.
- [8] 永崎研宣[2007], 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, 情報処理学会研究報告, 2007-CH-75, pp. 31-38.
- [9] 上地宏一[2007], 漢字グリフ管理 Wiki システム (GlyphWiki) の構築, 人文科学とコンピュータシンポジウム論文集, pp.237-244.
- [10] 下田正弘 [1993], 藏文和訳『大乘涅槃經』1, 山喜房佛書林.
- [11] 高楠順次郎編[1924], 大正新脩大藏經, 大正新脩大藏經刊行会, 1924-1934.
- [12] 三枝充應[1984], 中論—縁起・空・中の思想(上), 第三文明社.