

# 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

永崎 研宣<sup>†</sup>, 鈴木 隆泰<sup>††</sup>, 下田正弘<sup>‡</sup>

<sup>†</sup> 山口県立大学情報化推進室 <sup>††</sup> 山口県立大学国際文化学部 <sup>‡</sup> 東京大学大学院人文社会系研究科

大正新脩大藏經テキストデータベースは、すでにすべてのテキストの入力を終了し、その正確性を高めるという次の段階に入っている。これを効率的に行っていくためには、より効率的かつユーザフレンドリーなコラボレーションシステムの開発が有効である。コラボレーションシステムは、それまでの作業のルーティンを損なうことなく、その問題点を解消できるものでなければならない。そのため、Web ベースのシステムを基本としつつ、コア部分では、作業者のスキルの格差を考慮し、ローカル側での作業に比重を置くインターフェイスと、Web 上ですべての作業を行うインターフェイスの二種類を併用する形とした。

## The Development of a Collaboration System for the Text Database of the Taisho Tripitaka

Kiyonori NAGASAKI<sup>†</sup> Takayasu SUZUKI<sup>††</sup> Masahiro SHIMODA<sup>‡</sup>

<sup>†</sup> Yamaguchi Prefectural University

<sup>††</sup> Yamaguchi Prefectural University

<sup>‡</sup> The University of Tokyo

The text-database of the Taisho Tripitaka has entered a correction phase since inputting of all texts has been finished. It is useful to develop a more efficient and user-friendly collaboration system to solve previously experienced problems without disturbing the routine of the work. Therefore, we developed a Web-based collaboration system which has two interfaces accomodating the differences of workers' skill levels. One is based on the work done on the local computers and the other is carried out entirely on the Web.

### 1 はじめに

大正新脩大藏經は、1924-1934 年にかけて刊行され、現在もスタンダードとして幅広く用いられている、全 100 巻に渡る漢文の仏教文献の叢書である。インドで書かれたものを古典中国語に翻訳したもの、中国で書かれたもの、日本で書かれたものが含まれている。この大正新脩大藏經のテキスト部分 85 巻を電子化するプロジェクトはすでに 1986 年より続けられており、すでにほぼすべてのテキストの入力が終了し、修正の段階に入っている。この修正作業において作業を効率化することを目的として、2005 年を期に、この電子化作業

に Web ベースのコラボレーションシステムを開発・導入した。本稿では、このコラボレーションシステムに関して報告するとともに、人文科学分野における Web ベースの共同作業のあり方について検討する。

### 2 それまでの構築手法

大藏經テキストデータベース研究会 (SAT) は、1986 年より大正新脩大藏經のテキストデータベースの構築を進めてきた。構築開始当初より、学術的なニーズへの対応をひとつの目標とし、仏教学研究自身の手によるデータベース構築を重視した。このため、仏教学研究者に広く受け入れ可能な手法を用いた構築方法が採

られた。すなわち、全国に散らばる仏教学研究자들이、主にスタンドアロンな Windows や Macintosh といったパソコン OS 上で、テキストエディタ等を用いて ShiftJIS の文字コードでテキストを入力・修正していくという方法を主としたのである。

この手法は、コンピュータ・スキルとしてはもっとも基本的なものであり、データベース構築作業に対してあまり高いスキルを求める必要がないため、作業は幅広く行われ、最終的に全てのテキスト公開に到達することができた。<sup>1</sup>

しかしながら、公開されたデータは、この種のデータベースの常として、データの正確性という観点からは必ずしも十分とは言えないものがあった。したがって、より正確性を高めたバージョンのリリースを目指して次のフェーズへ移行する必要があった。データベース及びその構築にあたっての主な問題点としては、以下のようなものがあった。

- a. テキストデータそのものの正確性
- b. タグの統一及び正規化
- c. 行番号の正規化
- d. 外字の取り扱い方
- e. 作業進捗状況の把握の難しさ

これらの問題点について、以下に概観してみよう。

- a. テキストデータそのものの正確性

これは、言うまでもなく、入力されたテキストが、元となっている大正新修大蔵経と同じであるかどうか、という意味での正確性である。これに関しては、手入力をせざるを得ないため、ミスが含まれることはやむを得ないことであり、地道な修正作業が必要となる。

- b. タグの統一及び正規化

タグとは、本データベースにおいて利用されている XML ライクなタグに関することである。大きなテキストデータベースであり、入力された時期もかなり長期間に渡っていることから、入力された時期によって利用されるタグの種類やパラメータが異なっているとい

う状況が生じてしまっていた。さらに、終了タグがない、タグの名前を間違えている、パラメータの記述方法を間違えている等、タグの記述ミスも多く見受けられた。

- c. 行番号の正規化

本データベースにおいては、テキスト各行にユニークな ID が付されており、これを本稿では行番号と呼んでいる。本データベースにおいて、この行番号は固定長で以下のように規定されている。

[テキスト No.][補助 No.],[巻],[頁][段落][行]

(例： 0001\_01,0001a09, 1809\_40,0511b21 等)

このフォーマットを見ればわかるように、行番号は、一定の順序に従ってならんでいなければならないが、逆順になっていたり、飛んでいたりと、あるいは適切につけられていないなど、正規でない箇所が多く存在した<sup>2</sup>。

- d. 外字の取り扱い方

本データベースは ShiftJIS で構築されており、ShiftJIS に含まれない漢字に関しては、基本的には、いわゆる諸橋大漢和番号をベースにした外字番号を、実体参照形式に基づいて記述するという形を採っている。<sup>3</sup> これに関しても、テキストエディタによる入力を中心としていたため、フォーマットに従って記述されていない例が散見された。また、外字番号に対応するフォントの利用ライセンスのために、媒体を問わず、パブリッシング全般にあまり自由に利用できないという問題もあった。

- e. 作業進捗状況の把握

本データベースは、その規模の大きさにも関わらず、仏教研究者によるボランティア的な作業を中心としていた。このため、作業進捗状況を把握する立場の人に情報が集中し過ぎて処理しきれないという事態が発生しがちであり、きちんと全体の状況を把握しきれない状態になりがちであった。

以上の問題点は、いずれも、より正確性を高めた次期バージョンをリリースするために解決されなければ

<sup>1</sup> データベースは、<http://www.lu-tokyo.ac.jp/~sat/> において公開されている。

<sup>2</sup> 管見した限りでは、このミスが 3 万箇所以上存在した。テキストデータそのものは 600 万行以上あるため、0.5%程度に過ぎないとも言えるが、このことは、テキストエディタによる作業の限界を表しているとも言えるだろう。

<sup>3</sup> 実体参照形式は、&M000001;、&M012345; 等として記述されている。なお、本データベースに登場する外字の分析や処理方法に関しては下田 [1999]<sup>1)</sup> においてすでに論じられている。

ならない点である。次に、これらの問題点を解決するための方策について検討する。

### 3 問題解決に向けて

上記の問題点のうち、a. に関しては、基本的には従来どおりの手作業で行わざるを得ないところである。ただし、部分的には別のプロジェクト<sup>4</sup>によって入力・公開されているテキストが存在することから、そういったテキストについては、双方を比較しつつ相違点に関して原本にあたることによって、より正確性の高いテキストが可能であり、すでに実施されている。

また、d. とも関係してくるが、外字に関しては「外字番号」を付しているだけであり、チェックや修正にあたっては、この外字番号を視認可能なものにする工夫が必要となる。ワープロソフトのマクロ機能などを利用することで視認可能にすることはできるが、作業としてはやや煩雑であり、より簡易な視認方法が必要であろう。

b. c. および d. の一部の問題点は、フォーマットエラーという問題としてとらえることができる。フォーマットエラーに関しては、統一的なフォーマットを確立できれば、あとはソフトウェアを用いることで自動的にチェックが可能である。ただし、c. の行番号の問題に関しては、行番号のエラーがテキストの入力ミスを反映している場合もあるため、a. の問題になってしまう場合もある。

d. のうち、フォントの利用ライセンスに関しては、利用ライセンスがより適切な多漢字フォントへの移行や、フォントに頼らない方法が考えられる。

e. に関しては、作業済みのファイルが取りまとめ担当のところに送付されるという体制ではチェックに大変手間がかかり、また、その作業情報を共有するにも大変手間がかかってしまう。この場合には、何らかのコラボレーションシステムを採用し、報告や確認の作業を作業担当者にも適切に分担することである程度の解決が可能となるはずである。

以上のことから、こうした問題点の多くは、適切なツールを用意することである程度解決し得るものであ

る。ツールを用意するにあたって現在よく用いられる手法としては、スタンドアロンな環境で利用可能な特定の OS 向けのものか、あるいは、Web 上で利用可能な、ある程度マルチプラットフォーム的なもの、ということになるだろう。この選択には、ソフトウェアやネットワークといった技術的な制約と、実際に作業している作業者のコンピュータ・スキルによる制約という二つの側面から考えていく必要があるだろう。また、同時に、作業者のコンピュータ・スキルを向上させるということも考慮しなければならない。<sup>5</sup>

### 4 問題点の具体的な解決方法

すでに筆者は、サンスクリットの仏教哲学文献をテキスト校訂するためのコラボレーションシステムを開発・運用したことがある<sup>3</sup>。このシステムは Web ベースのシステムであり、データベースとしてフリーの RDBMS である PostgreSQL を利用し、ミドルウェアとして PHP を使ったシステムとなっていた。このシステムはあくまでもテキスト校訂を前提としているが、底本となるテキストに対する修正もできるものであった。修正作業の際には各作業者ごとにユーザ認証を行い、校訂や修正などの作業に関してはすべて履歴を残せる形になっていた。このシステムでは、作業はすべて Web 上で行うようになっており、スタンドアロンな環境で行う作業は前提とされていなかった。これには、テキストそのものの扱いやすさや、作業者が極めて限定されていたため作業者スキルの想定が容易だったことなど、前提となるいくつかの環境的要因があった。Web ベースでシステムを構築したことと利点としては、作業できる範囲や内容をシステム側で限定するために作業手順の統一や周知徹底が行いやすかったこと、データを一箇所に集約したため、バックアップが容易であり、また、最新データへのアクセスが容易だったこと、さらに、一般に向けての公開用のデータの作成に関しても、公開用の Web ページを用意して公開用データを RDBMS から抽出するだけで済むため、

<sup>4</sup> Chinese Buddhist Electronic Text Association (CBETA) が印度撰述部、中国撰述部の電子テキストに関しては公開を行っている。

<sup>5</sup> この点について、人文科学のデータベース構築のために提唱された BBDB は、研究者が日常的に行っている資料やデータの整理作業の延長として行えるようにすることで DB 構築を容易なものにするという手法を提示している。<sup>2)</sup>

公開のために毎回作業を行うという手間を省くことができた、といった点があった。

ところが、今回の大正新修大蔵経のテキストデータベース構築の場合、規模が極めて大きいため、作業者の人数も多く、また、様々な事情により作業者の入れ換えも少なくない。このため、作業者のスキルをどのレベルに設定するかというのは容易な問題ではない。また、すでにスタンドアロンな環境を中心にして長期間に渡って作業を続けてきたという経緯があり、Webベースの作業に切り替える場合には、作業手順の新たな習得や作業手順の変更によるモチベーションの低下と言った観点を考慮に入れる必要があった。

#### 4.1 コラボレーションシステムの基本構成

本データベースのためのコラボレーションシステムは、Xeon 2.8GHzを2基搭載したメモリ4GBのサーバ上で動作している。RDBMSとしてPostgreSQL8.1を採用し、これに600万行に渡るテキストデータと行番号を入力し、作業者はPHPとApacheを通じてWebブラウザからRDBMSを限定的に操作するという仕組みにした。RDBMSを採用したのは、今後、別のデータと連携させることになった場合やデータベース内での連携を行うことになった場合にシステムを拡張しやすいからである。PostgreSQLを採用したこともまた、フリーソフトという前提の中で、可用性が高く、将来的な拡張に備えることを考慮したためである。PHPとApacheは、やはり、フリーソフトという前提の中で、WebからPostgreSQLを扱うにあたって、必要な関数とセキュリティを備え、十分な速度で動作するもの一つだからであり、採用するに至った強い理由はない。<sup>6</sup>

テキストデータベース内では、ShiftJISのテキストデータはコード番号に変換し、10進数として扱ってい

<sup>6</sup> 今回の場合、時間的な問題があり、このデータベースのほとんどの部分の開発は構想から2ヶ月程度で行われた。こうした短期間での開発の場合、開発者の確保以外に初期費用がかからないフリーソフトを採用することは、とりわけ予算の少ない人文系の学問にとっては重要である。

<sup>7</sup> テキストデータベーステーブルの概略：  
行番号 | 本文 | ユーザ名 | 変更時間 | 優先順位 | ファイル名

<sup>8</sup> 外字データベーステーブルの概略：  
外字番号 | 合成表記 | 部首番号 | 部首 | 画数 | 備考 | 同定番号 | ユーザ名 | 変更時間

<sup>9</sup> 認証情報データベースのテーブルの概略：  
ユーザ名 | 本名 | パスワード | メールアドレス | 変更時間 | モード | 所属

る。これは、テキストの中にPostgreSQLのShiftJISでは対応できない文字が入っていた場合にエラーが発生してしまうため、それを避けるための措置である。また、テキストの変更が行われた場合に履歴をたどれるようにするため、変更日付、テキストの優先順位、ユーザ名(変更者)が記録されるようになっている。<sup>7</sup>

データが大きくなってしまったために動作が若干遅くなる場合があるが、これは、PostgreSQLに対するメモリ割り当てを大きくしつつ、利用方法に応じてIndexを作成することで対応している。

テキストデータベース以外に、外字作業のための外字データベースも構築した。これは、外字のデータを容易に共有するとともに、文字情報の検索やWeb上での視認も可能とするためのものである。このデータベースには、任意に割り当てられた外字番号以外に、画数、部首、合成表記、音などの情報が登録されている。<sup>8</sup>

また、認証情報データベースも用意されている。データベースへのWeb経由でのアクセス時にはBasic認証が要求されるようになっており、その際に、このテーブルを使って認証を行うようになっている。これは、単に認証だけでなく、作業者の作業記録を残す際に作業者自身による工数を増やさないようにするという意味もある。<sup>9</sup>

以上が、コラボレーションシステムの基本構成である。

#### 4.2 コラボレーションシステムのインターフェイス

コラボレーションシステムは、WebのフォームとPHPのスク립トによって操作されることになる。この場合、作業者に対して、どこまで、どのように操作させるのかというのを限定することによって、作業のルー

ティンを比較的確実に構築することが可能である。以下では、どのような形でそれを構築したかについて言及していく。

まず、認証情報に関しては、Webブラウザからのユーザ登録やユーザ情報の変更が可能である。また、管理者モード・作業モード・閲覧者モードというパラメータを用意し、管理者モードに設定されているユーザのみが他のユーザの登録や情報変更をできるようになっている。作業者がアクセスした場合、自分の情報の一部のみが変更できるようになっている。このようにすることで、Webブラウザの基本的な操作ができる人であれば誰でもユーザ登録等の作業に携われるようになっている。

ユーザ名	名前	所属	E-mail	ID	最終変更	登録	パスワード
nagasaki	永島研彦	山形県立大学	nagasaki@ypu.jp	管理モード	2005-11-04 04:21:22.937701	変更	変更
				作業モード	2005-12-09 20:01:02.758181	変更	変更

[認証情報のインターフェイス]

外字データベースに関しては、文字に関する様々な属性を一覧できるようにし、さらに、それらの属性を検索することもできるようになっている。また、文字画像表示システム「文字焼き」を利用することで、外字の文字画像を表示させられるようにしている。「文字焼き」は、Web上でうまく表示できない文字を画像で表示させるためのシステムであり、このコラボレーションシステム全体で利用できるようにカスタマイズを行った。<sup>10</sup>

SAT 外字 DB [トップ画面](#)

Search   検索終了 [新規外字番号追加](#)

Search Results for "にくづき" (10 件ヒット)

[国](#) [国](#) [国](#) [国](#) [国](#) [国](#) [国](#) [国](#) [国](#) [国](#)

画像	外字番号	合成表記						情報	作業者	出現箇所
	同定番号	部首 No.	部首	内画	総画	音	変更日時		変更	
𪛗	&MT00007;	月+音+文						𪛗(&M029375)の 諸構同意	nagasaki	参照
	&MT00007;	130	にくづき	12	16	サン	2005-12-31 20:15:38		<a href="#">変更</a>	
𪛘	&MT00009;	月+(備-用)×巾)						𪛘(&M029827)の 異体字?	nagasaki	参照
	&MT00009;	130	にくづき	9	13	ヨウ	2005-12-31 20:16:03		<a href="#">変更</a>	

[外字データベースのインターフェイス]

<sup>10</sup> 「文字焼き」は、インド・アラビア系の語学教育教材を開発するにあたってそれらの言語で用いられている文字をサポートしていないシステム上で文字を画像として表示させるために開発されたシステムだが<sup>4)</sup>、現在でも、携帯電話での多国語表示や、Windows でサポートされていない文字を Web で表示するのに有効な手段となっている<sup>5)</sup>。

その結果、テキストデータベースのテキスト中に外字番号が入っていた場合、諸橋大漢和番号に関しては「GT フォント」<sup>11</sup>を利用して文字画像を表示し、それに含まれない「外字」に関しては、外字データベース

に登録されている外字画像を使って表示できるようになっている。また、この文字画像は、発見されたものから順次追加されるようになっている。

```
1548_28,0578c02.<#16/>珂貝璧玉珊瑚璣性寶貝珠沙石草木枝葉 [Change]
1548_28,0578c03.莖節。及餘外非受堅。是名外地界。如是内 [Change]
1548_28,0578c04.外水界。是名地外字画像。二水界。 [Change]
1548_28,0578c05.内水界。云何内水界。若此身内受。水 [Change]
1548_28,0578c06.涎瘡。<#18/>膽汗肪髓腦脂。<#19/>嗽唾膿血小便及餘 [Change]
1548_28,0578c07.此身内受水潤等。是名内水界。云何外水界。 [Change]
1548_28,0578c08.若外水界非受。<#20/>蘇油蜜石。<#21/>蜜黑石蜜乳酪 [Change]
```

[テキスト中の文字画像]

テキストデータベースは、ファイル名一覧とテキスト本体との二つのテーブルから成っている。ファイル名一覧は 2800 の文献=テキストファイルの一覧であり、そこから各ファイルに対する様々な作業が行えるようになっている。また、作業の進捗状況等の注記もその一覧において閲覧・変更できるようになっている。作業上の注記には、チェック項目を用意したものもあり、それらも含め、注記に関しては、Web 上での変更だけでなく、エクセルで作成した CSV ファイルのアップロードによる書き換え機能も用意することで、ローカルと Web 上で、書き換え作業が二度手間にならないような工夫も行っている。

テキスト本体の修正作業に関しては、二つのアプローチを行った。この理由は、人文科学系全体にも言えることだと思われるが、研究者の間に存在するコンピュータ・スキルの大きな格差である。本データベースは研究者の手で構築されることを重視しており、データベースの大きさから、かなり多くの研究者に協力を求めることになる。したがって、作業者のスキルにはかなりのばらつきが生じることになる。このばらつきを少しでも有効に生かすための方策として、以下の二つのアプローチを行ったのである。

一つ目のアプローチとしては、これまでの作業のやり方をなるべく維持したままにするため、テキストの

修正・確認作業そのものは従来通りテキストエディタ等のスタンドアロンな環境で行う形とした。そして、作業の進捗状況を全体として把握・共有しやすくするために、コラボレーションシステム上に作業済みテキストファイルを Web ブラウザ経由で丸ごとアップロードすることによって、差分行のみをデータベースに取り込み、同時にそれらの行の作業者名等を記録するようにした。この手法では、作業者に対する新たな手順の周知としては、作業後のアップロードと最新版入手のためのダウンロードだけで事足りる。このアプローチでは、上に挙げた問題点のいくつかを解消することができるが、必ずしも十分ではない。このやり方の場合、\*.mst という拡張子のファイルをテキストファイルとして扱わなければならないため、拡張子に関する知識が必要となる。新たに参加した作業者の中には、拡張子についての知識がなく、\*.mst.txt というファイル名でアップロードするなど、作業内容を十分に周知するのに手間がかかってしまうケースもあった。また、行番号の部分の誤って非正規なものに変更してしまった場合も、事前の予測の範囲を超えていたためにうまく対応できないことがあった。すなわち、スキルのあまり高くない作業者にとっては有効なシステムとは言えない面があった。

<sup>11</sup> GT フォントは日本学術振興会と東京大学多言語処理研究会が作成・公開している多漢字フォントであり、非営利目的に限って利用が許されているため、この場合には利用可能であった<sup>6)</sup>。

もう一つのアプローチとして、Web上でテキストの修正作業を行うという方法も用意した。これは、単純に、変更したい箇所をWebブラウザ上で表示させて、それを修正するというものである。この方法では、作業者が修正できる部分は本文部分だけとなり、極めて限定されることになる。したがって、ファイル名やファイルの正規化に関して作業者が考慮する必要がなくなるため、スタンドアロン環境で発生しがちないわゆるケアレスミスをほとんど排除することができる。外字に関しても、上述のように、「文字焼き」を介することで、Webブラウザ上で特別な操作をせずに外字を画像で視認しながら作業することができるため、その意味でも作業の正確性を高めることに役立つ。

しかし一方で、このアプローチの場合、すでにスタンドアロンな作業環境を整えているスキルの高い作業者からは、置換などの一括操作をはじめ、使い慣れたツールを使った手元での十分な作業ができなくなってしまおうという意見も出た。これに関しては、スキルの高い作業者のニーズを満たせるようなインターフェイスを用意することで対応可能であることから、当面は前者のアプローチによる手法を続けてもらうようにしつつ、今後の検討事項としたい。

## 5 終わりに

このような形で、コラボレーションによるテキストの修正システムを効率的に運用していくことは、ある程度は可能である。また、テキスト・データベースや外字情報の公開に関しても、データ公開用のプログラムを作成してデータ公開用ページとして公開しておけば、継続的に最新情報を提供することができるようになり、その意味でも効率的である。とはいえ、このテキスト・データベースを実際にどのように研究利用していくかというのはまた別の次元の話である。いわゆる「デジタル・アーカイブ」として、最新版や修正履歴を研究者や一般に向けて公開することは少なからぬ意義があると思われる。行番号等を用いて動的な参照性も確保すれば、さらに有用性は高まることだろう。しかし、その一方で、とりわけ、原典資料に忠実であら

ねばならないという研究者の立場からは、このデータベースをそのまま研究の典拠として利用するわけにはいかない。あくまでも、テキスト検索や所在情報の暫定的な確認といった用途に限られるだろう。その場合にも、データベースの正確性・継続性といった面での信頼性を高めていくことで、有用性を高めていくことはできるはずである。また、今後の展開としては、先に触れたチベット大蔵経やサンスクリット仏教文献のデータベースとも連携させることで、統合的な仏教文献データベースを構築することも視野に入れていきたい。<sup>12</sup>

謝辞本研究は、平成17年度科学研究費補助金 成果公開促進費（データベース）『大正新脩大蔵経』テキストデータベース（No. 178002）の一部として遂行されたものと、平成17年度科学研究費補助金（若手研究（B））「古典文献学における電子テキスト共有基盤の研究」（No. 17700253）の一部として遂行されたものを含んでいる。また、これまでこのプロジェクトに参加してきた多くのスタッフの努力によって本コラボレーションシステム及び本研究が成り立っていることを感謝とともに記しておく。

## 参考文献

- 1) 下田正弘 師茂樹, 大正新脩大蔵経テキストデータベース (SAT) における外字問題, 人文学と情報処理 No.25, 勉誠出版, 1999.10, pp.35-43.
- 2) 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.
- 3) 永崎研宣, デジタルアーカイブと校訂テキスト— Webを用いた Sanskrit テキストの電子校訂テキスト共有システムを通じて—, 人文科学とコンピュータシンポジウム論文集 Vol. 2003, No.21, (社) 情報処理学会, 2003.12, pp. 1-8.
- 4) 永崎研宣, インド系文字による Web 環境での情報の共有, 人文科学とコンピュータシンポジウム論文

<sup>12</sup> ただし、それには、各文献の関連付けデータ等の膨大な情報をデータベースに追加していかなければならない。eDicプロジェクトにおいても試みられているが<sup>7) 8)</sup>、統合的なデータベースを構築することになると、かなりの時間を要するものと思われる。

- 集 Vol. 2001, No. 18, (社) 情報処理学会, 2001.12, pp. 213-220.
- 5) Kiyonori NAGASAKI, Makoto MINEGISHI, Izumi HOSHI, Displaying multi-script data on the Web, Proceedings of the Glyph and Typesetting Workshop, 21st Century COE Program "East Asian Center for Informatics in Humanities - Toward an Overall Inheritance and Development of Kanji Culture - " Kyoto University, 2004.2, pp. 44-51.
- 6) 東京大学多言語処理研究会 , <http://www.l.u-tokyo.ac.jp/GT/> , 2006.4.23 閲覧.
- 7) 鈴木 隆泰, Tibetan-Sanskrit 構文対照電子辞書 eDic, 明日の東洋学, 2000.4, pp. 5-7.
- 8) 相場徹 生出恭治, 東洋古典電子テキストの利便性向上にむけて, 人文科学とコンピュータシンポジウム論文集 Vol. 2005, No.21, (社) 情報処理学会, 2005.12, pp. 201-208.
- 9) 高楠順次郎編, 大正新脩大藏經, 大正新脩大藏經刊行会, 1931.