

## 人文学のためのコラボレーション

—ITLR コラボレーションシステムの開発を中心的事例として—

永崎研宣, 苫米地等流 (一般財団法人人文情報学研究所)

Dorji Wangchuk, Orna Almogi (ハンブルク大学)

下田正弘 (東京大学大学院人文社会系研究科)

デジタル媒体としてのインターネットの普及と高度化により、人文学のためのコラボレーションはますます重要なものとなりつつある。本稿では、これまでのインド学仏教学分野におけるインターネットを通じたコラボレーションの事例を検討した上で、新たに開始された ITLR プロジェクトにおいて開発したコラボレーションシステムについて報告するとともに、そこにおける課題と解決策について検討する。

## Collaboration in the Humanities

— Through the Case of Development of the ITLR Project —

Kiyonori Nagasaki, Toru Tomabechi (International Institute for Digital Humanities)

Dorji Wangchuk, Orna Almogi (Hamburg University)

Masahiro Shimoda (University of Tokyo)

According to the spread and highly-development of the Internet as digital media, the significance of collaboration in the humanities has been steadily increasing. In this paper, we will treat several cases of collaboration through the Internet in the field of Indian and Buddhist studies, then report on the ITLR (Indo-Tibetan Lexical Resources) project which has been recently started and aims to realize international active collaboration on the Web and discuss its issues and solutions.

### 1. はじめに

筆者らは、2010年5月より、ハンブルク大学・東京大学・人文情報学研究所の共同研究プロジェクトとして、インド・チベットを中心とする思想研究のためのデジタル用語集 **Indo-Tibetan Lexical Resource** プロジェクト [1] (以下、ITLR) を進めてきており、これにともなって、ITLR 実現のための Web コラボレーションシステムの開発を進めてきた。本稿では、人文学におけるコラボレーションについての課題を提起するとともに、このシステムの開発における実践例を報告し、解決

策について検討する。

### 2. 人文学におけるコラボレーション

人文学研究者は個人研究が基本であり共同研究は得意ではない—という声を聞くことがある。確かに一面の真実を衝いていると思わされることもあるが、実のところ、人文学におけるコラボレーションは、敢えてデジタル媒体を挙げずとも、紙媒体の時代から様々に行われてきている。辞典編纂や索引の作成はその典型的な例の一つであり、他にも、様々な種類の調査から成果発表に至るまで、個人では成し得ない成果を世に問うことを目指すコラボレーションは様々に展開されてきた。近年増加しつつあるデジタ

ル媒体を通じてのコラボレーションについても、そのような視点から見てみるなら、利便性の向上は見られるものの、本質的にどれほどの変化があるのかという点については検討の余地があるだろう。さらに、ここ数年は、インターネットの高度化と普及に伴い、インターネットを通じたコラボレーションも技術的には十分に現実的なものとなり、盛んに行われるようになりつつある。それもまた、上述のような意味での検討の余地があると思われるが、本稿では、そうした議論に踏み込むための足場作りとして、インド学仏教学分野において展開されつつあるインターネットを通じたコラボレーションについて検討してみたい。

### 3. 仏教学におけるこれまでのコラボレーションへの取り組み

2005年より2年程の間、Webコラボレーションシステム上で大蔵経テキストデータベースの構築に関わるテキスト及び外字のデータベース共同作業が行われ[2]、2008年頃より、日本印度学仏教学会の論文書誌データベースINBUDSの構築作業がWeb上での共同作業へと移行した[3]。また、2009年頃より、大蔵経テキストデータベースWebサイト上での、複数文献の部分的な関連づけに関するデータベースの構築作業も開始された[4]。これらの3つの事例は、いずれも、一つのサーバに対してすべての共同作業者がアクセスし、統一された一つのフォーマット、一つの手続きに基づいて各作業者が自らに割り当てられた担当範囲を処理していくという形を基本としており、その点では典型的なインターネットを通じたコラボレーションである。とはいえ、それぞれに、その時点での技術的なトレンドを反映する形で異なる様相を見せている。まずはこれらのWeb上でのコラボレーションについて時系列で改めて検討してみよう。

2005年に開始されたWeb上での大蔵経テキストデータベース構築作業の頃には、すでにWeb2.0という言葉が登場しておりWeb上でのコラボレーションは注目されつつあったが、Web上で現在のような便利なインターフェイスを提供することはまだ容易ではなく、したがって、Web上で細かな作業をするということはあまり一般的ではなかった。その一方で、各作業者のコンピュータスキルは多様であり、実際のところ、作業側からは、使いたくエディタ等のソフトウェアで作業をしたいという要望が強かった。特に、テキストそのものに関する入力や校正等の作業に関しては、一応、Web上で完結できるシステムを用意したものの、上記の

ようなユーザの要望に応えるために、完全にWeb上で作業するのではなくローカルのコンピュータに一度ダウンロードした上で好きなソフトウェアを用いて作業を行い、作業終了後にそれを再びWebサーバ上にアップロードするというワークフローとそのための機能をも用意することとなった。このことは、当時としては作業効率の向上に寄与したが、一方で、データの流れが個々の作業者のローカルコンピュータ上にまで及び、かつ、ローカルのデータとサーバ上のデータとの整合性の確保が作業者のみに委ねられるため、そこから生ずる問題の回避に際してやや苦勞する面があった。また、このときに初めて構築を開始したWeb上の外字データベースに関しては、テキストデータと異なりデータの構造が明確であり、作業内容がWeb上のみで作業するのに比較的適していたため、あらかじめ、すべてのデータがサーバ上で処理されるように設計された。この場合には、ローカルにデータをダウンロードすることまではできるが、それをサーバ上に戻すという機能を設けなかったため、整合性の確保という点ではテキストのデータベースほどの問題は生じなかった<sup>1</sup>。なお、ここでのいずれのデータベースも、コラボレーションによる構築のみを目的としており、そこから直接、一般にデータを公開することは念頭に置いておらず、別途、一般公開のためのシステムの開発が行われたということにも留意しておきたい。

2008年頃より開始された日本印度学仏教学会論文書誌データベースINBUDSのWeb上でのコラボレーションに際しては、コラボレーションシステム上でデータを一般に公開することまでできるようなワークフローが採られた。このワークフローでは、データは完全にサーバ上で完結したものとなり、作業者が一度入力したデータは数回の確認と必要に応じた修正を受けた上で同じシステム上で一般にも公開されるようになっていく。これは、先述の外字データベースと同様、データの構造が明確であったことから、作業の内容や手順を限定しやすく、それ故にWeb上で完結できたという面がある。また、一方で、この頃にはAjaxが利用しやすくなってきていたこともあり、Webメールやブログ等、あるいは、大学生活や研究生活において必要なサービスがWebブラウザを利用環境とするようになってきており、いわゆるWebアプリケーションに関する利用経験や基本的な理解を有するユーザが大多数であったことも、サーバ上でデ

<sup>1</sup>ただし、グリフデータに関してはその限りではなく、ローカルで画像を作成・修正する形となり、GlyphWiki[5]のようにグリフデータをもサーバ上で編集するには至っていなかった。

ータの流れを完結させるにあたって有益であった。そして、ここでは、Web ブラウザ上での入力支援的な機能の提供も行った。たとえば、インド学仏教学において必須となるサンスクリットやチベット語等にしばしば用いられる、いわゆるダイアクリティカルマーク付きのローマンアルファベットの入力に関して、通常は入力支援プログラム等を各自のコンピュータ上に用意しなければならないところを、Web ブラウザさえあれば入力できるようにしたのである。特に、この種の通常のコンピュータ環境ではあまり用いられない文字の入力に関しては、ローカルコンピュータ上での作業の場合には文字エンコーディングを始めとして様々なデータの齟齬が生じてしまう可能性があるところを、Web ブラウザ上での作業に一元化することで、データを UTF-8 とし、やや特殊な文字の入力に関しても一元性を確保できるようにしたのであり、このことは、インターネットを通じたコラボレーションの意義を一段階高めたと言っていいたいだろう。これらをまとめるなら、INBUDS におけるコラボレーションは、Web サーバ上でほとんどの作業が行われることでデータの様々な側面が一元化されるとともに、一度入力されたデータは Web サーバ上で完結する形になっている、と言えるだろう。

ただ、ユーザ同士のやりとりといったことや個々のデータに関する議論などに関するフォローの機能は用意されておらず、その種の事柄については電子メール等の別の手段か、あるいはオフライン等で行われざるを得ないという面もある。それに関しては、むしろ、このワークフローにおける作業者同士の連携や業務分担が適切にできているということが重要であり、それが整備されているがゆえに、現在もこのシステムが稼働し、データの追加・修正が持続的に実施されているのだと言っていいたいだろう。

2008年4月には、前出の SAT テキストデータベースの Web 公開版がリリースされた。これにより、85 巻、600 万行の主に漢文仏典のテキストが Web 上で全文検索できるようになり、さらに、後述するような様々な機能が付加されることになるが、ここではまず、2009 年頃よりこの SAT Web 公開版に追加され、開発が続けられている「関連情報データベース」に関するコラボレーションについて検討してみたい。

この「関連情報データベース」は、2006 年頃より開発が続けられているものであり[6]、テキストにおける任意の文字列を他の任意の文字列と関連づけ、それに何らかの意味を持たせるというものであり、また、画像の任意の箇所と任意のテキストとのリンク機能も実装されていた。このシステムの目指すところは様々であり、たとえば、仏典テキストにおけるチベット訳と漢

訳との関係をより細かく記述するために、文章等の単位で関連づけ、個々の文章同士の関係について記述し、それを集積していくことで、チベット訳と漢訳との関係をより可視的に確認できるようにしたり、あるいは関係について統計的に分析できるようにすることや、あるいは、現代語訳と漢訳等とを関連づけて、読者に読みやすさを提供したり現代語と漢訳との関係について検討するための基礎データを蓄積するといった用途が想定されている。これが 2009 年頃より、SAT Web 公開版に統合され、さらに開発が続けられているのである。

このコラボレーションシステムでは、作業者は、二つのテキストの任意の部分をそれぞれドラッグして選択し、その関係についてプルダウンメニューから選択したり直接入力したりした上で、送信ボタンをクリックするとサーバに蓄積されるという形になる。そして、INBUDS において開発された Web コラボレーションシステムと同じく、サーバ上に入力された後には、データはサーバからダウンロードされることなくいくつかのプロセスを経て公開に至るという形になる。そして、作業者間でのやりとりに関しても INBUDS と同様に、他の仕組みを使って行うことを前提としている。テキストの一部の箇所をドラッグしただけでそのテキストの位置情報を文字単位で取得してサーバに送信できるようにしていることは、ユーザによる入力形式の一元性を高めるものであり、それもまた INBUDS において志向された方向性を踏襲していると言える。

また、これまでのコラボレーションシステムに比較した場合のこのシステムの一つの特徴としては、既存の SAT Web 公開版に対していわゆる Web API を用いてデータを取り出しているという点である。SAT Web 公開版と関連情報データベースは、インターフェイス上は統合されているために一見すると区別するのが難しいが、実際の所、この関連情報データベースは既存のテキストデータベースからは完全に独立しており、テキストの特定の箇所を関連づけの対象として指し示す際には、その位置情報を URI によって指定して取得するという形になっている。この点をコラボレーションとの関連で考えるなら、一つのワークフローの外側で作成されたデジタルデータとの相互運用ということができ、いわば、ワークフロー外 (=他のプロジェクト) との、緩やかなコラボレーション、と言ってもいいだろう。なお、この頃には、jQuery をはじめとする Ajax ライブラリが大変充実してきていたため、このインターフェイスでは、ほとんどの Web ブラウザ上での作業はページ遷移をしないままに完遂することができるようになってきており、Ajax のメリットは、このシステムでは全面的に生かされている。

さらに、Web API 等を用いた相互運用を緩やかなコラボレーションとして捉えるなら、前出の「関連情報データベース」のみならず、仏教学分野においてはすでに様々な形でコラボレーションが展開されていることになる。これについては、世界的な規模で見ると枚挙にいとまがないのだが、筆者らが関わっているプロジェクトにおいてもその典型がある。前出の SAT Web 公開版がまさにそれであり、INBUDS、Digital Dictionary of Buddhism、CiNii など、様々な Web サービスとの連携による便利なサービスを提供している[7]。たとえば、テキスト本文をドラッグするだけで Digital Dictionary of Buddhism の項目とその英訳が調べられる機能や、テキスト本文を読みながら INBUDS の論文書誌データを検索し、さらに、CiNii Web API の利用により、ヒットした論文情報の全文にあたるものが Web 上に公開されているかどうかを容易に確認できるようになっている。いずれの情報も、それぞれの Web API 等を介して接続され、それぞれのシステムの運用者が継続的に情報を追加・修正する体制になっており、先述のような緩やかなコラボレーションが形成されているということが出来るだろう。

以上をまとめるなら、ここで採り上げた事例における Web 上でのコラボレーションの在り方は、次のように言うことができる。

- A) Ajax 環境の発展によって Web 上で可能な作業の範囲が拡大してきており、結果として、ワークフローの領域がローカル側からサーバ側へと移動している。
- B) ユーザの裁量を制限することによってデータ形式の統一が容易となってきている。
- C) Web API を利用することにより、他のシステムとの緩やかなコラボレーションが行われるようになってきている。
- D) 実際の作業以外の事柄についてはシステム外で行っている。

次章では、このような観点を踏まえた上で、ITLR プロジェクトにおけるコラボレーションについて検討してみよう。

#### 4. ITLR におけるコラボレーション

ITLR は、デジタルメディア時代に即した、インド・チベット思想関連のデジタル専門用語辞典の構築を目指すプロジェクトとして開始された。世界中の専門家による国際的な共同作業によって構築され、分量の制約を受けず、多様

な解釈や登場箇所を可能な限り併記でき、研究状況の進展に応じて随時のアップデートが可能となっており、さらに、すでに公開されている様々な関連デジタルリソースとの連携による効率的な情報の拡充を実現する。そのようなものを志向して ITLR のプロジェクトは展開しているところである。

これを実現するために、ITLR では、専用の Web コラボレーションシステムを開発した。Web 上にシステムを構築する理由は、前章に述べたように、サーバ上でデータの流れを完結させ、データを着実にコントロールできるようにするとともに、世界中の専門家による共同での入力作業を容易に可能とするためである。分量の制約については、サーバ上にストレージを置くことで、比較的容易に、ユーザ側からは見えない形で分量を増やすことができ、その意味でも Web の採用は妥当な選択だろう。

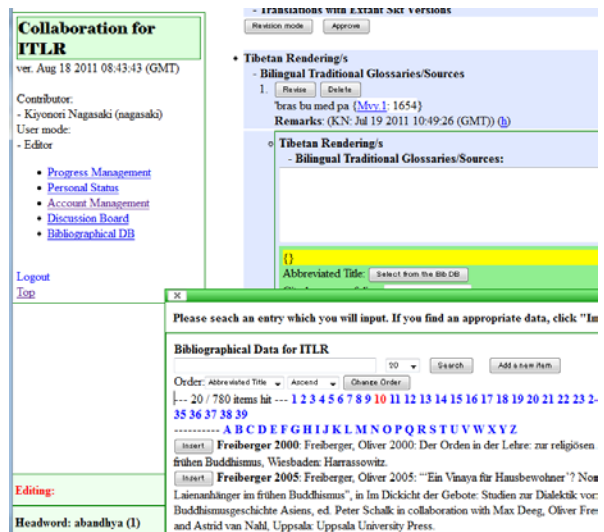
多様な解釈の記述や登場箇所の可能な限りの併記を可能とするには、一つの用語を起点としたツリー構造を基本としつつ、必要に応じて追記する情報を増やせるような仕組みを用意することになる。ただし、どのようなツリー構造を作り、どの項目では複数併記を可能とするか、といったことは、ITLR の学術上の要請に沿ったものでなければならない。また、基本的には、ITLR はインド・チベットにおける思想研究を中心とした専門用語を収録対象とするものであることから、少なくともサンスクリットとチベット語による翻訳を対象としなければならない。それだけでなく、パーリ語等のサンスクリット以外の古いインドの言語も対象とする必要がある。さらに、研究上、古典中国語訳や現代語訳なども参照できることが望ましい。もちろん、翻訳の場合には必ずしも訳語・訳文が一对一で対応するわけではなく、一つの言語への訳語だけでも複数種類の記述が必要であり、さらに、それぞれの訳語に関する出典情報が必要となる。また、出典情報についても、複数の版があればそれぞれ少しずつ異なる場合があり、場合によっては異読情報やテキストの修正情報等も必要となる。異読情報の記述については、すべての版をそれぞれ別の資料として入力して並列的に登録する方法と TEI ガイドライン[8]に従ってテキストの中に個別のアノテーションをマークアップする方法とが考えられる。ただし、テキストの修正情報等に関しては、マークアップによって対応した方が効率が高い場合があるため、ITLR においては、ツリーの末端でアノテーションが必要になった場合には、TEI に沿ったタ

グを付与するという方針を採用した。TEI のタグを付与する理由は、人文学資料における異読情報の記述などの比較的簡素なマークアップにおいては TEI のタグセットが十分に有効であること、ITLR に関わるコミュニティにおいてすでにある程度 TEI が知られていること、システム全体としてはデータ出力時のフォーマットとして TEI を一つの重要な選択肢として位置づけており、ツリーの末端部分での TEI によるマークアップはその際にも効率的であること、などである。これに際しては、やはり、入力手法の簡素化とユーザ操作の限定によるフォーマットのより容易な統一化が必要となる。ここでは、すべてを Web サーバ・Web ブラウザ上で行うか、あるいは、oXygen XML Editor 等の XML エディタを用いてローカルで記述したものを HTML Form にコピー&ペーストするか、という選択肢があり得る。前者についてはすでに HTML Form で TEI/XML の記述を可能とする Ajax モジュールを開発済みであり、システムに組み込むための詳細な検証を行っているところだが、後者についても、すでに TEI/XML で多くのデータを蓄積している作業員もいるため、それをそのままコピー&ペーストできるようにするという選択肢も残しておくことが現段階では効率的であると思われる。

また、ツリー構造については、ツリーの起点から枝分かれしていただくだけでなく、他の用語、あるいは、他の用語のアイテムとして記述された情報などにリンクすることも可能としている。とりわけ、著作物のエントリにおける著者情報、著者のエントリにおける著作物情報は、それぞれ、相互のエントリへとリンクすることを基本としている。そして、他のエントリ等へのリンクづけを効率的かつ確実にするために、Ajax を用いて同画面上で容易に選択操作ができるインターフェイスを用意している。なお、ここでは、個々のエントリが URI によって識別されるようになっており、クリック操作でエントリを選択するとその URI が記述されるようになっており、それによってデータ入力の正確性を高めている。

書誌情報の入力に際しても、情報入力時の情報の正確性・統一性を高めるため、書誌情報データベースを別に持つ形にしつつ、ID を指定することによって書誌情報を参照する仕組みを用意している(Fig.1)。なお、書誌情報データベースは、現在のところは、独自のものとなっているが、将来的には INBUDS や SARDS[9]等の他の書誌情報データベースと連携しつつ、さらに、可能なものについては DOI 等のスタンダードな識別子を参照できるようにすることで、上

述のような緩やかなコラボレーションを実現していく予定である。



(Fig. 1: 書誌 ID 入力画面の一部)

データ入力に際しては、2種類の権限を設定し、ワークフローを整理している。編集権限を持つグループ(Editor)と一般入力者(Collaborator)との二種類に分かれており、Editorには、ユーザ登録の権限から他の入力者のデータの改変・削除といった様々な権限が与えられている。一方、一般入力者であるCollaboratorの権限はかなり制限されており、自分のパスワードの変更はできるがユーザ登録に関する権限はない。さらに、特に留意しているのは、Collaboratorの場合には他の入力者の入力データを上書き(変更)・削除できないように制限しているという点である。これは、入力者が行った貢献が可能な限り削除されないようにすることで、プロジェクト参加者のインセンティブを確保することが狙いである。このことは、人文学分野での協働においては重要な要素であり、このような方法に限らず、常に様々な形で配慮が求められるところである。

ITLRのデータが随時のアップデートを必要とする理由は、ITLRが対象とする研究分野においては既存の研究成果が新発見資料等によって更新される機会が少なからずあり、継続的な更新が可能でなかったとしたら比較的是やく内容の有効性が失われていってしまう可能性があるためである。印刷された書籍においては常に発生する問題だが、とりわけ、ITLRのような、多くの資料を参照情報として提示することを目指すもの場合、内容の有効性が失われる速度はさらに高まる可能性がある。ここでは、上述の、EditorとCollaboratorをベースとするワークフローを持続的に運用し続けることによる対応を目指しており、システム上はこれを実現することを可能としている。ただし、この課題は、

技術的問題というよりはむしろプロジェクト運営上の問題という側面が強く、その点については別稿を期したい。

また、他の Web サービスとの緩やかなコラボレーションに関しては、ITLR の場合には、すでに公開されている様々な関連デジタルリソースとの連携による効率的なデータの拡充とサービスの充実化、ITLR から既存の、あるいは新規の他のサービスへのデータの提供、といったものが考えられる。

既存のデジタルリソースとの連携に関しては、ITLR の場合には、まず、類似のプロジェクトとの差別化という課題がある。前出の *Digital Dictionary of Buddhism* のみならず、今日、世界的には複数の仏教学用語辞典作成プロジェクトが立ち上がり、それぞれに活動を展開している。すでにそれらのプロジェクトの多くとの対話と差別化は進められており、ITLR の位置づけはある程度確立されてきている。そこで、現在では、他の Web サービスの情報を効率的にリンクできる仕組みを検討中である。

また、他のサービスへのデータの提供に関しては、上述のように、すでに ITLR では、各エントリ・各アイテムが URI によって指定でき、それによってデータを取り出せる仕組みとなっており、この仕組みは、他の Web サービスからでもそのまま利用可能である。そこで、今後は、Web システムの一般公開にあわせてドキュメント等を整備する予定である。

ITLR では、前章で採り上げたいくつかの事例に加えて、掲示板やメッセージングシステムを導入することで、作業員間のコミュニケーションをもシステムに統合しようとしている。これまで、こういったものは既存の SNS やメールリスト等で代替してきたが、システムの中に作業員間のコミュニケーションシステムを採り入れたなら、議論の対象となるエントリやアイテムへのリンクが容易となる等のメリットが考えられる。なお、それを生かすことで作業がどの程度効率化されるか、あるいは、どのくらいより良い成果に結びつけられるか、という点についても、いずれ別稿を期したい。

## 5. 終わりに

このように、インド学仏教学分野におけるインターネットを通じたコラボレーションは、ユーザとしての研究者のニーズをシステムやデータ構造に反映しつつ、より高いユーザビリティを目指して展開してきているが、同時に、IT 技術の発展と制約に少なからぬ影響を受けている。インターネットを通じたコラボレーションが人文学にとって容易に採用し得る選択肢になるた

めには、この流れがある程度安定する必要がある。**Linked Open Data** の登場はその傾向を示唆するものではあるが、まだしばらくの時間がかかるだろう。

なお、本稿で採り上げたもの以外にも、インド学仏教学分野としても、人文学分野全体としても、インターネットを通じたコラボレーションは様々な場面で展開されており、ここでは、あくまでもその断片を切り出して検討してみたに過ぎない。しかしながら、人文学分野におけるインターネットを通じたコラボレーションについての検討や議論の場は決して多くはなく、本稿がそうした場を形成していく一助になれば幸いである。

## 6. 謝辞

本研究の一部は、科研費(30343429)の助成を受けたものであり、他の一部は、科研費(50272448)の助成を受けたものである。

## 参考文献

- [1] <http://www.kc-tbts.uni-hamburg.de/index.php/en/projects/33-what-is-the-indo-tibetan-lexical-resource-itlr> (2011/11/9 参照).
- [2] 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」『情報処理学会研究報告』CH-70(2006年5月), pp. 33-40.
- [3] 下田正弘, 永崎研宣「大藏經と人文系データベース」『情報処理学会研究報告』CH-82(2009年5月).
- [4] Kiyonori Nagasaki, Toru Tomabechi, Masahiro Shimoda "Toward a Digital Research Environment for Buddhist Studies", *Digital Humanities* 2011, (2011/6), pp. 342-343.
- [5] 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.
- [6] Kiyonori Nagasaki, A Collaboration System for the Philology of the Buddhist Study, *Digital Humanities* 2008, (2008/6), pp. 262-263.
- [7] Kiyonori Nagasaki, Charles A. Muller, Masahiro Shimoda, Aspects of the Interoperability in the Digital Humanities, *Digital Humanities* 2009, (2009/6), pp. 375-377.
- [8] Lou Burnard and Syd Bauman ed., TEI: P5 Guidelines <http://www.tei-c.org/Guidelines/P5/> (2011/11/9 参照).
- [9] South Asia Research Documentation Services 3 (SARDS3), <http://www.indologie.uni-halle.de/sards/> (2011/11/9 参照).