

複数の人文系研究者による史料注釈を可能とする Web システムの試作—『東大寺要録』を用いて—

佐藤 貴文¹ 後藤 真² 木村 文則³ 前田 亮¹
¹立命館大学 情報理工学部 ²花園大学 文学部 ³立命館大学 衣笠総合研究機構

本稿では、遠隔地にいる複数の人文系研究者によって、文献資料に関する注釈情報を共同で入力するための Web システムの試作について述べる。本研究で対象とする文献は『東大寺要録』という歴史資料であり、これに対して注釈作業を行っている人文系研究者による利用を想定している。本システムの特徴として、複数の研究者が同一の史料に対して同時に注釈の付与が可能である点、複数の研究者による注釈が部分的に重なる場合や、離れた文字にまとめて注釈を付与するなど、複雑に入り組んだ注釈への対応を考慮している点が挙げられる。本発表では、実際に利用する人文系研究者の意見を得ながら進めている本システムの構築について、現状の報告および今後の展望について述べる。

Developing a Web-based System for Annotating Historical Materials by Multiple Humanities Researchers — The Case of "Todaiji Yoroku" —

Takafumi Sato¹ Makoto Goto² Fuminori Kimura³ Akira Maeda¹
¹ College of Information Science and Engineering, Ritsumeikan University
² Faculty of Letters, Hanazono University
³ Kinugasa Research Organization, Ritsumeikan University

This paper describes a prototype Web-based system for collaboratively making annotations on historical materials by multiple humanities researchers who are distant from each other. The target material of this study is “Todaiji Yoroku”, which is believed to be written in the 12th century, and the system is supposed to be used by humanities researchers who are actually making annotations to this material. The unique features of the proposed system are, 1) multiple users can make annotations to the same material simultaneously, 2) capable of making complex structured annotations such as partially-overlapped annotations, and annotations which consist of non-consecutive characters. In this paper, we describe the current status and future prospects of the system, which is under development by the authors, and is improved by getting feedback from the humanities researchers who are actually using the system.

1. はじめに

本発表は、遠隔地にいる複数の人文系研究者によって、文献資料に関する注釈情報を入力するためのシステムの試作について述べる。文献資料の注釈システムは、SMART-GSのような、個人で行い、研究を進めるシステムとして高機能なものは存在する[1]。また、Web上で議論可能な形で、歴史的図像資料への情報を付加するシステムの構築は行われている[2]。しかし、これらのシステムは、いずれも「複数の人物で」かつ「複雑な構造を持つ歴史的文献資料」を扱うことを可能としたものではなかった。

そこで、本発表では、人文系研究者の実際のフィードバックを得つつ、複数の人文系研究者による注釈作成システムを試作の状況について報告する。

本発表が対象とするのは『東大寺要録』という歴史資料であり、この注釈作業を行っている人文系研究者の研究会によって最初に用いることを想定したものである。

2. 『東大寺要録』の概要とデータベース作成の経緯

『東大寺要録』は、序文によると嘉承元(1106)年に東大寺の僧が衰微した寺の再興

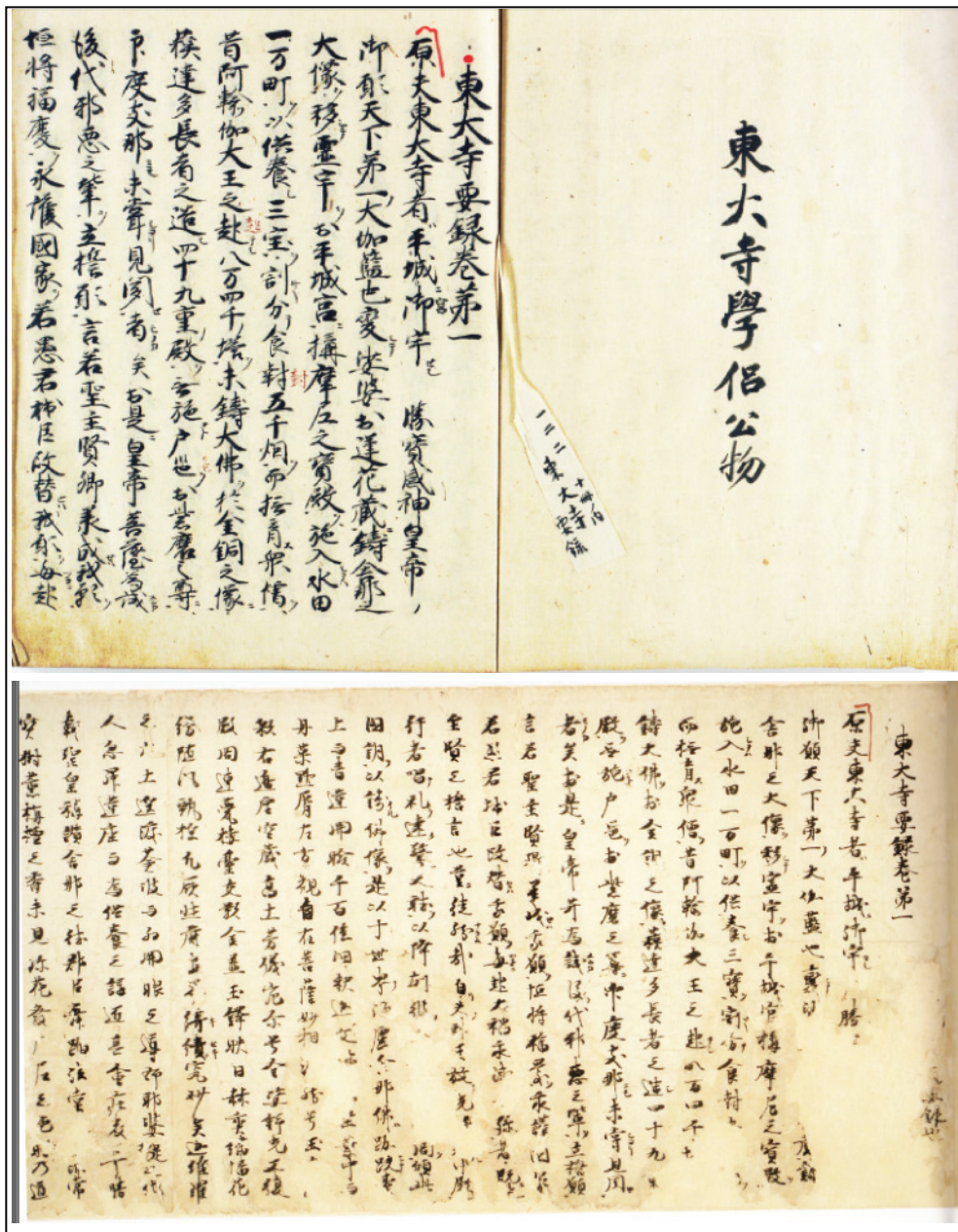


図1 東大寺本『東大寺要録』(上)と醍醐寺本(下) それぞれに書き込みがあると同時にテキストの文字関係も若干異なっている

を願って編纂したとある。しかし、実際の編者名などは知られておらず、詳細は不明である。長承3(1134)年になって観嚴が増補・再編して、現在知られる形になっている。原本・増補本ともに散逸し、現在は、醍醐寺に第1・第2巻の、東大寺に全10巻の写本が残されている。醍醐寺本がより古いものであるとされている。

東大寺本は、特に東大寺における年中行事において重要な役割を果たしており、観嚴の増補の後も、増補が続けられ、その宗教的・歴史的価値も非常に高いものである。現在、

刊本としては、筒井英俊氏による校訂がなされた『東大寺要録』[3]、および続々群書類従に収録されているもの[4]の2種類がある。

現時点においては、より良好な校訂とされる筒井氏本は入手が難しくなっている。また、この両者はいずれも原漢文をそのまま文字起こししたものであり、書き下し文や詳細な注釈がついたものとはなっていない。そのため、日本史学・仏教学・建築学・美術史学などの研究者によって研究グループ(東大寺要録研究会)が組織され、注釈作業を行っている。しかし、時間的・空間的な制約により、

通常の作業方法では、一定期間内での注釈の完成は困難であると考えられる（3章にて述べる）。そこで、筆者らは、この研究グループとの協業によって、注釈・校訂のためのシステムを作成し、研究作業の軽減を図るとともに、その成果とシステムを公開することとなった。

3. 歴史的文献に対する注釈の課題

『東大寺要録』のような日本史文献テキストへの注釈作業工程は、以下のとおりである。

1. 原漢文の校訂作業を行う
2. 原漢文を読み下し、書き下し文を作成する
3. 書き下し文の用語の中から想定される読者の理解に必要な言葉をピックアップ
4. 言葉について、他の文献から調査を行い、注釈を付す
5. 書き下し文および注釈について、研究会（この場合は、東大寺要録研究会）にて、発

表を行い、その妥当性について研究者相互で討論

6. 討論成果をフィードバックして、注釈内容の変更

7. 最終的な書き下し文・注釈の確定・公開

これらをまとめると図2のように示すことができる。1～4までの作業は、通常研究者のコンピュータ上で行われるが、5の研究会での作業は、会場を確保し、一定の時間を確保して行われる。しかし、検討には、ごくわずかな量であっても、多大な時間を使う。また、研究会を開催するためのスケジュール調整や、費用なども大量に発生する。さらに、研究会では、裏付けとなる他の史料にすぐにアクセスできない、短時間であるため、見落としが生じるなどの問題も生じてしまう。

また、6・7の作業も、最終的には取りまとめが必要になるが、Wordのデータなどでは、成果の取りまとめが難しい面があった。その

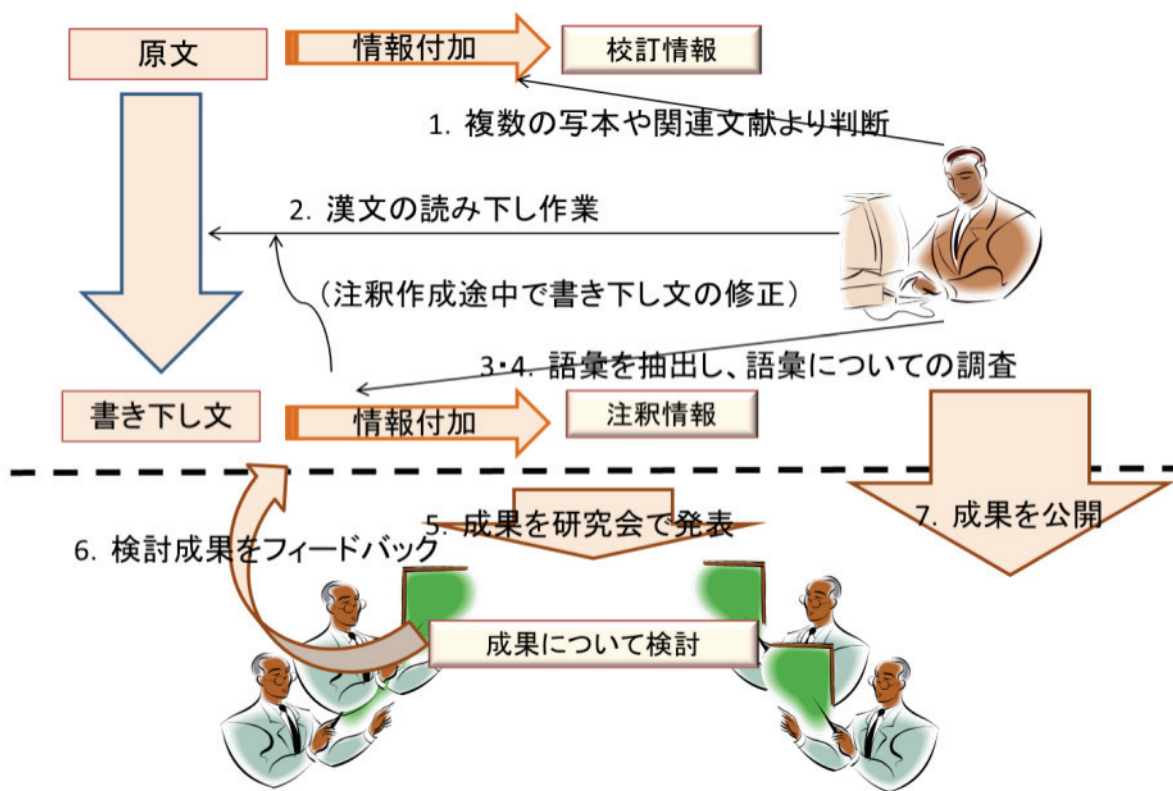


図2 東大寺要録注釈作成作業イメージ図

破線部より上が研究者個人での作業になり下が共同作業となる

ため、筆者らは、前述の研究グループと協業して、実用性の高い Web システムを作り上げることにした。

この Web システムには、以下のような要件が求められる。

1. コンピュータに習熟しない人文系研究者が、ストレスをなるべく少なく入力可能であること。
2. 少なくとも入力段階では、同じ言葉に、複数の研究者の異なる意見を入れられること。
3. あわせて、注釈作業の履歴を保存し、必要に応じてロールバックしたり確認作業ができること。
4. 「建五重塔、又置皇后宮職并施薬院」とある場合に、「施薬院」と「置」のように離れた文字に対してまとめて注釈を付したい場合や、(この場合はあり得ないが)「職并」と「皇后宮職」といった、入れ子形式ではない形で注釈を付したい場合がある。このように、原史料に対して自由に注釈を付すことができること。
5. 研究グループ内の他の研究者に向けて、注釈を依頼できること(たとえば、仏教関係の用語が出現した場合、歴史研究者から仏教研究者に注釈を依頼する)。

以上のような条件が必要となる。注釈作業は、いわば文献分析のための基礎的な「実験」である。そのためには、テキストに対し可能な限り自在に情報を入力することが肝要となる。また、実験的な情報を付加するため、その情報が妥当かどうかを、他の研究者によって検証することが必要となる。場合によっては議論を重ねることもでてくるであろう。それらを実現するために上記のような要件が求められることとなる。

また、最終的に入力したデータは公開されることになる。その際には、注釈の単純な表示だけでなく以下のような要件が求められる。

1. 原漢文と書き下しが対比的に読めるよう

になること。

2. 注釈そのものが検索できること。
3. 関連情報へのリンクがあり、自由に移動できること
4. 複雑に入り組んだ注釈を、ストレスなくみられること

複雑に入り組んだ注釈を、短時間で効果的に見るためには、書籍よりはコンピュータのほうが向いている側面がある。そのため、そのコンピュータの効果を最大限発揮できるインターフェースが重要であると考える。

4. 『東大寺要録』共同注釈 Web システム

3章で挙げた要件をふまえ、Web ブラウザ上で稼働する史料注釈の Web アプリケーションの実装を行った。本システムは、歴史資料研究者向けのシステムであるため、コンピュータに習熟しない人文系研究者でも使いやすいうように、直感的に操作できるインターフェースを構築することに重点を置いている。

図 3 は、『東大寺要録』共同注釈 Web システムの概要図を示している。『東大寺要録』共同注釈 Web システムは、「Web インタフェース」と『東大寺要録』データベースから構成されている。『東大寺要録』データベースは、「原文テーブル」「書き下し文テーブル」「注釈テーブル」「ユーザ管理テーブル」の関係表から構成されている。

『東大寺要録』共同注釈 Web システムを閲覧する、または注釈を投稿する利用者は、まず Web インタフェースにアクセスを行う。内容の検索や注釈の投稿などの利用者の操作要求は、Web インタフェースを介してシステムに伝達される。システムは利用者から受け取った操作要求に対し、その結果を提示するために必要な情報を『東大寺要録』データベースから取得し、Web インタフェースへその結果を返す。返された結果は Web インタフェース上で利用者に提示する。

現状の『東大寺要録』共同注釈 Web システムでは、Web インタフェースを介して以下

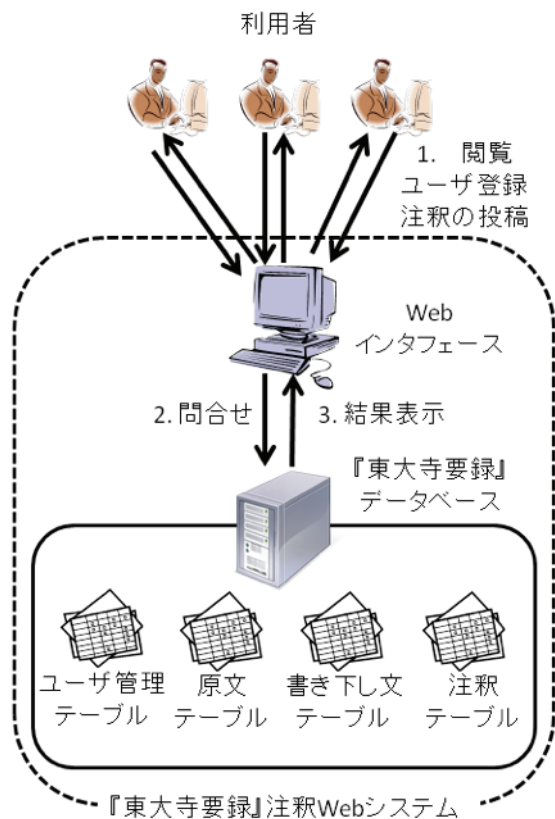


図 3 システム概要図

の 4 種類の操作を行うことが可能である。

- ユーザ登録
- 原文・書き下し文の閲覧
- 注釈の閲覧
- 注釈の投稿

4.1. ユーザ登録

本システムでは、『東大寺要録』に付与された注釈に対して、議論を行うことを目的として構築している。そのため、複数の利用者が本システムを利用することを前提としている。しかも、それぞれの利用者がただ閲覧するだけでなく、注釈の投稿を行うことも想定している。そのため、それぞれの利用者を識別する必要がある。本システムでは事前にユーザ登録を行い、固有のユーザ名を発行することで各ユーザの識別を行えるようにしている。

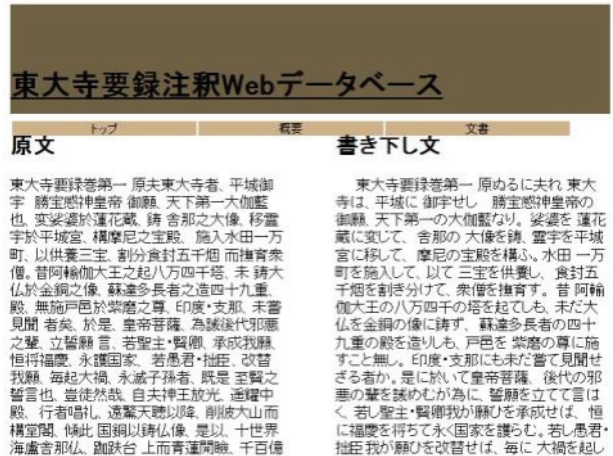


図 4 原文と書き下し文の並列表示画面

4.2. 原文・書き下し文の閲覧

歴史的文献注釈システムに求められる要件の一つである、「原漢文と書き下し文が対比的に読めるようになること」を実現するために、原文・書き下し文の閲覧画面では、直感的に分かりやすいように、原文と書き下し文が左右に並列に提示されるようになっている (図 4)。

原文および書き下し文は行単位で『東大寺要録』データベースの各テーブルに格納されており、利用者が指定した行から、画面に表示できる行数の原文と書き下し文の両方がデータベースから取得され、閲覧画面に提示される。

4.3. 注釈の閲覧

図 5 は、原文に対して付与された注釈を閲覧する画面である。原文の下にひかれた下線は、その原文の文字列に対して注釈が付与されていることを示している。その下線にマウスポインタを合わせると、注釈がポップアップで表示される。また、注釈が同じ箇所に複数付けられることや一部重なって付けられる可能性があるが、その際にはその文字列に対して複数の異なる色の下線によって注釈であることが示される。

本システムでは、複数の注釈が部分的に重複する、入れ子形式ではない重複が起こるような場合についても対応している。HTML や XML では、注釈を付与する箇所の重複は、

東大寺要録 宝感神皇帝、原破、依図本補
 宇 勝宝感神皇帝 御願、天下第一大伽藍
 也、変娑婆於蓮花蔵、铸 舍那之大像、移靈
 宇於平城宮、構摩尼之宝殿、施入水田一万
町、以供養三宝、割分食封五千烟而撫育衆
 僧。昔阿輸伽大王之起八万四千塔、未 铸大
 仏於金銅之像、蘇達多長者之造四十九重、
 殿、無施戸邑於紫磨之尊、印度・支那、未嘗

図 5 注釈表示画面

入れ子形式になる場合でなければ対応できない。しかし、歴史的文献注釈では入れ子形式ではない重複が起こるような注釈の付与を行うことがありうる。このような場合についても注釈を付与し、それを閲覧可能であることは、本システムの特徴の一つである。

4.4. 注釈の投稿

図 6 は、原文の文字列に対し注釈を付与し、それを投稿する画面である。

注釈を投稿する利用者は、まず原文から注釈を付与する文字列をマウスで選択する。次に右クリックを行うと、注釈投稿ウインドウが表示される。注釈投稿ウインドウの上の欄には、選択した原文の文字列が自動的に表示される。注釈投稿ウインドウの下欄には、付与したい注釈の内容を記述する。注釈内容の記述が完了し、投稿ボタン (submit ボタン) をクリックすると、注釈の投稿が行われ、

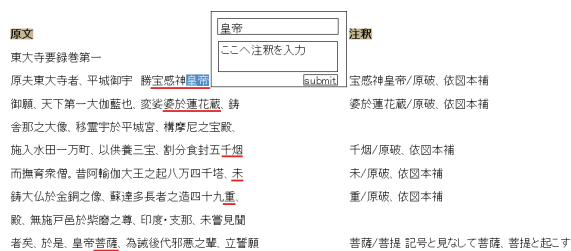


図 6 注釈の投稿画面

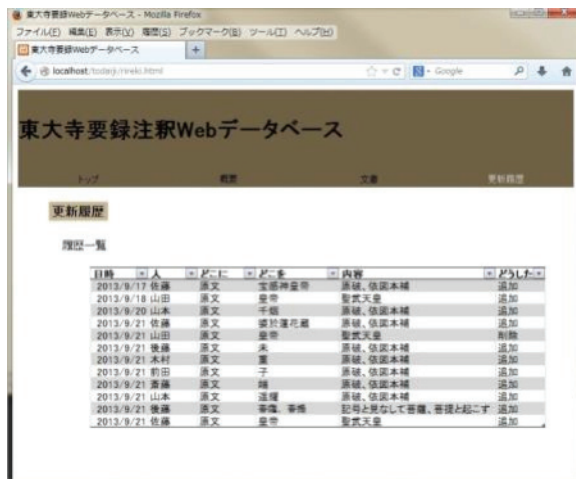


図 7 注釈の更新履歴確認画面

その注釈は『東大寺要録』データベースの注釈テーブルへ格納される。

投稿された注釈は、注釈対象の文字列、注釈内容だけでなく、注釈を付与したユーザ、日時、操作内容 (追加, 修正, 削除) などの情報も併せて格納される。注釈を付与したユーザ情報には、ユーザ登録の際に発行したユーザ名を用いる。これらの情報により、注釈の更新履歴の管理も行う。注釈の更新履歴は、図 7 に示した注釈の更新履歴確認画面により確認できる。

5. 期待される成果と現状の課題

現在、東大寺要録研究会では、『東大寺要録』の校訂作業を、直接会合することで行っている。そのため、会合の時間に対して議論すべき項目が多く、すべてについて詳細まで議論を深めることが困難な状況である。また、議論すべき項目は会合の場で初めて明らかになるため、議論内容の整理も容易ではない。本発表において提案したシステムが実現することにより、Web を介して議論が可能となるため、直接会合することなく議論が可能となり、時間的制約も大幅に緩和できることが見込まれる。直接会合したうえでの議論が必要となった場合においても、Web を介して議題の整理や優先順位の決定が事前に可能となるため、会合時間を有用に使うことが可能となる。

図 2 で示す、破線部の上部の作業段階で、複数の研究者の目に事前に触れさせることができる。そのため、破線部下の作業がより容易になるのである。

これは、一つのまとまった史料を用いて、かつ両者の研究者が多数関与することで、なかなか困難であった技術開発による汎用性と人文科学の実用性のコラボレーションを、実質的なものとして実現できている点が、本研究の最大の特徴である。

しかし現状としてシステムが作成中であり、以下のことが実現すべき課題として挙げられる。

1. どのようなインタフェースが人文系の研究者にも使いやすいかを考慮する。
2. 原文と書き下し文を並列に表示させる際に、書き下し文のほうが長くなってしまいう関係により完全に対応して表示させることが難しくなっている。
3. ブラウザ上で複数の研究者が編集することができるようにした際の、注釈のバージョン管理をどのようにするか。

たとえば、東大寺要録研究会でシステムのデモを行った際には、以下のような要求が出されており、これにあわせて修正作業が行われている。

1. ポップアップが本文を隠さないようにしてほしい。注釈を確認する際に、本文は重要な情報であり、同時に閲覧することが必要となる。
2. 注釈の全体像がみえるウインドウが必要である。注釈がどの程度ついていいるのかを確認し、作業の状況が確実に見えるようにしたい。
3. 自分の作業状況を確認するために、注釈だけを見る場合があるので、注釈だけの一覧が欲しい。
4. 作業時も同じ語句に注釈をつけようとしたときには、サジェストが欲しい。具体的には「聖武天皇」に注釈をつけようとした場合に、「すでに**巻××行に注釈がある。この注釈内容でよいか」などの情報が出ると、

同じ語句に何度も同じ作業をする必要がなくなる。

他にも多くの要望があるが、ここでは、サンプルにとどめる。1~3 はインタフェースの問題であり、比較的容易な対応が可能な部類である。一方、4 は紙の作業ではありえない、コンピュータ独特の作業への期待であると言えよう。上記の例の場合、たとえば「聖武皇帝」や「天皇」のみなどのような表記に注釈を加える場合に、どのような処理が必要かなど、テキストの類似度を考える必要も生まれてくると思われる。

このような、複雑な対応まで含めて、いわば「人文系のリアル」に対応しつつシステムの構築を進めている点が、本システムの最大の特徴であると言えよう。

6. 展望

本システムは、東大寺要録研究会の議論をやっと得る段階になりつつあるものである。そのため、今後のフィードバックによる改良こそが、本質であり、中途の段階であるのは事実である。しかし、歴史・宗教・建築・美術などの研究グループとの実質的なコラボレーションが開始され、Web システムの形で結実する仕組みができたことが、重要である。今後、この結実したデータベースをどのようにブラッシュアップさせていくかが課題となる。また、要録のテキストデータを TEI などに合わせていくことも、今後の重要な課題となるであろう。

TEI に対応させることで、『東大寺要録』のデータを、注釈を含めて、テキストデータの標準化が可能となる。当初は、内部システムとして作成するが、資料情報の流通の現状を考えれば、いずれは外部システムからも検索をおこなったり、ポイントを示したりすることが可能なシステムにすべきであろう。

そのためには、より汎用的なデータ形式を持つことが望ましいと考えられる。現時点で、TEI はもっともテキストに関しては汎用性の高い形式であると考えられる。また、一方

で、複数の版を持つ手書きテキストである東大寺要録は、日本の記録文書の TEI 化の好例ではないかと考える。TEI にこのような日本における典型的なテキストの例をフィードバックすることは、将来的に TEI の仕様に日本のテキストに向けた形のものが入ることが期待できる。

史研究所所長の榮原永遠男氏のご協力をいただいている。記して感謝申し上げる。

7. おわりに

テキストへの注釈作業それ自体は、決して珍しいものではない。また、その成果のデジタル化及び公開の事例も多く見受けられる。しかし、多くの人々が共同で、そして作業の面に着目するという点では、多くの研究がなされてきたとは言い難いのではなかろうか。

今後、このような作業の観点に着目したシステム[5]が多くあらわれることで、人文科学研究と情報学のコラボレーションがより進むことを期待するものである。

参考文献

- 1) SMART-GS <http://sourceforge.jp/projects/smart-gs/> (参照 2013-09-10)
- 2) 倉持基, 研谷紀夫, 津田光弘, 馬場章「デジタルアーカイブを利用した歴史写真の情報学的研究」『情報処理学会シンポジウムじんもんこん 2005 論文集』 pp.113-120,(2005.12).
- 3) 筒井英俊『東大寺要録』国書刊行会, 初版 1943 年, 2003 年再版
- 4) 続群書類従完成会編『東大寺要録』1969 年
- 5) 山田太造, 横山伊徳, 綱川歩美, 高橋典幸, 林譲「採訪史料管理システム」『情報処理学会シンポジウムじんもんこん 2010 論文集』 pp.113-120,(2010.12).

謝辞

本システムの開発に際しては、東大寺要録研究会のみなさまより忌憚なきご意見をいただくことで進められている。また、同研究会代表である大阪市立大学名誉教授・東大寺