

平成 30 年 6 月 27 日現在

機関番号：14301

研究種目：基盤研究(B) (一般)

研究期間：2015～2017

課題番号：15H02948

研究課題名(和文)古文書のウェブを目指して

研究課題名(英文)Towards Web of Komonjyo

研究代表者

林 晋(Hayashi, Susumu)

京都大学・文学研究科・教授

研究者番号：40156443

交付決定額(研究期間全体)：(直接経費) 7,900,000円

研究成果の概要(和文)：古文書のWEBを実現するにおいて、最も重要なことは古文書の翻刻、つまり、古文書のテキストを文字列にすることである。それにより、古文書のテキストの検索、リンクづけなど、WEBと呼ぶにふさわしい文書の集合体を作成することができる。

その実現法の一つとして、市民のボランティア参加による翻刻方法、クラウド翻刻(Crowd transcription)が知られており、英国などでの成功例が知られている。しかし、日本の古文書に対しては、成功例がなかったが、地震関係の古文書を対象にして、ボランティアが崩し字の読みを学習できるようにした、「みんなで翻刻」システムを開発し、これを初めて成功させた。

研究成果の概要(英文)：A crowdsourced manuscript transcription service "Minnade Honkoku" of Japanese handwritten documents from the Edo period or before is built. It became the first successful service of the kind, and so proved the possibility of building a Web of "Komonjo", Japanese handwritten documents from the Edo period or before.

研究分野：歴史学、人文情報学、歴史社会学

キーワード：市民参加型翻刻 人文情報学 クラウド翻刻

1. 研究開始当初の背景

この数年、歴史資料のWEBアーカイブが次々と発表されている。一枚の古文書のデジタル・イメージを一面のWEBページと考えれば、これは古文書のWWWとも考えられる。

もちろん、それが単に画像として提示されているだけならば、マークアップ、リンク、サーチなどが可能なWWWとは大きく違うものだと言わざるを得ない。

しかし、代表者林や分担者橋本、久木田などが開発したSMART-GSシステムでは、ネット上ではないものの、「画像ページ」群に、WWWの様にマークアップ、リンク、そして、サーチが可能である。

また、現在、IIIFというデジタルアーカイブの規格が広まりつつあり、SMART-GSに似たことが、WEB上で可能となりつつある。これをもとに、WEB上のデジタルアーカイブの画像をページとする、画像史料のウェブと考えられるものが生まれつつあるといえる。

言いかえれば、本研究の目的である「古文書のWEB」への道が着々と築かれつつあるのである。

しかし、現在のWWWと、この「古文書のWEB」を比較してみると、後者には、前者が持つ「検索可能性」と「多数のボランティア情報提供者の存在」が欠けている。この二つの特徴こそは、ブログやGoogle searchのような検索エンジンが象徴する、現在のWWWの大きな特徴である。この二つの特徴を欠いた古文書WEBが、現在、隆盛を極めているWWWに匹敵する様な意味を人文学にもたらすとは考えにくい。

他方で、もし将来的に古文書、石碑なども容易に検索できるようになれば、例えば、未知の過去の津波の存在や、その津波の到達点を検索で知り、費用はかかるが、より精度の高い津波堆積物調査に繋げる契機にできる可能性さえある。この様に、この二つの特徴を持つ古文書WEBの意義は学術的だけでなく社会的にも大きいと思われた。

このことから、次の「研究の目的」で示すような意味での歴史資料のWEBの構築が望まれていた。その方式としては、ロンドン大学のTranscribe Benthamの様な、ボランティアによるCrowd transcriptionが最も現実的であるが、日本国内の江戸期以前の古文書では、成功例が皆無であった。

2. 研究の目的

古文書、古地図、絵図、石碑、写真などの歴史的資料のWEB化が達成できるならば、歴史学、古地震学、歴史地理学などに革命が起きることは関係する研究者の誰もが認識するところである。

そして、そのためには、1の「背景」でも若干説明したように、検索が何より重要である。たとえば、「津波」と「安政」で、江戸時代の古文書を検索して、現在は海岸線の変

化や伝承の風化などにより、津波被害を想定し辛い場所での、安政東海地震、安政南海地震に際しての津波被害が記述されている古文書を見つけることができれば、それは大きな社会的意味を持つのである。

しかし、代表者たちが構築したSMART-GSにおける画像内の文字の類似度による検索、また、色々と試みられている、くずし字用のOCRなどの性能が、まだまだ低く、特に江戸期以前の古文書に対しては、十分実用になった例が知られていない現状では、これらの方式による検索は、絵に描いた餅に過ぎない。

そこで現実的な方法として、本研究では、Google Booksと同様に、テキストの画像と、その翻刻(画像としてのテキストを、文字列に起こしたもの)を併置して、翻刻を使って検索するという方法で、古文書のWEBを検索することを目的として。

この方式を採用する場合、検索自体は、すでに、Google Booksなどで実用化されているもので、その実施には、ほとんど何の困難もないといえる。

つまり、この方式における古文書のWEBの検索とは、実際には、古文書を如何に翻刻するかという問題に尽きると言ってもよいのである。

そして、その翻刻に、現状では、手書き文字OCRのようなコンピュータによる翻刻が使えない以上、考え得る方法は、ただ一つ、英国ユニバーシティ・カレッジ・ロンドンの哲学者J.ベンサムの遺稿を翻刻するTranscribe Benthamや、米国スミソニアン博物館所蔵の史料を翻刻するSmithsonian Transcription Centerなどで行われている様な、Crowd Transcription(以下、クラウド翻刻である。

このクラウド翻刻では、史料の画像が、WEBページ上に表示され、それをボランティアが翻刻して、同じページのテキストボックスに打ち込む。この様な、ネット、WEB上で、人々(Crowd)によりなされるプロジェクトを、クラウドソーシング、クラウド・コンピューティングなどと呼ぶため、これらはクラウド翻刻と呼ばれるのである。

このような海外、主に英語圏のクラウド翻刻の成功事例の一方で、我が国では、手書き史料のクラウド翻刻の成功例が全く存在しなかった。そればかりか、日本デジタル・ヒューマニティ学会による、国会図書館所蔵の活字史料のクラウド翻刻さえ、公開したものの数枚の翻刻しかなされないという状況であった。

一方で、全国に古文書の翻刻・解読に興味を持つ能力と意欲のある個人・グループは多く存在し、例えばそれを通して古地震学に貢献したいと思う人たちもいるが、実際の翻刻・解読の活動は、近隣大学の専門家の存在など、例外的に条件が整った場合に限定されている。

これは、WEB で情報発信をしたい人たちが多く存在しながら、それを容易に実現できる一般ユーザが使える技術がなかった「ブログ登場以前の WWW の状況」に似ている。本研究の目的は「歴史資料の WEB」実現のために、ブログや WEB2.0 にあたる社会的・技術的方式を生み出すことであった。

特に、「研究開始当初の背景」で述べた、江戸期以前の古文書に適用可能な方法を生み出すこと、それが本研究の最大の目的であった。

3. 研究の方法

当初は、古文書を翻刻するボランティアは、主に高齢者であるという想定で計画を立てたため、研究の軸は、古文書読解教室で使われている翻刻用紙と筆記具に近い感覚で使える、タブレット上のツールの開発であった。

つまり、古文書の WEB への第一歩は、WEB 上でのクラウド翻刻ではなく、全国に多くある、くずし字読解教室や、その様な古文書読解を行っているグループなどにおいて、開発したタブレット用アプリなどを使い翻刻をしてもらい、それを回収する。

そして、高齢世代が入れ替わり、WEB の使用も当然という世代になれば、Transcribe Bentham のような、WEB 上のクラウド翻刻に移行する、本研究は、当初、その様な方向性を目的としていた。

しかし、研究を開始してすぐに、代表者林は、自らが想定した、この研究の方向は間違っているのではないかと思うようになった。

その理由は、開発したツールの試用ユーザ・グループの中心に想定していた京都大学古地震研究会の活動を見学したことなどもあるが、最大の理由は、分担者橋本（当時は林研究室の院生）が、くずし字判読の学習の場を提供することを通して、ボランティアへの意欲を引き出すという、学習ベースの Crowd transcription という、新しいモデルを考案して、提案したことにある。

これは、橋本が、阪大文の他の科研費研究のために制作したくずし字判読学習ツールの開発と公開の経験などから、当時、流行していたゲーム「刀剣乱舞」を通して、古文書に興味を持った若い女性たちのような人たちに、くずし字判読の学習の場を提供すれば、高齢者に限らない、広く多様な層にボランティア参加を促せるのではないかというアイデアであった。

そして、橋本と加納は、このモデルのもと、自らが中心メンバーである京都大学古地震研究会の活動として「みんなで翻刻」という、東大地震研の古地震史料の Crowd transcription service を WEB 上に構築して公開した。

その結果、「みんなで翻刻」は、次の4の「研究成果」で述べるように、開発者の橋本の予想も覆すほどの成功、最も成功した

Crowd transcription のひとつとされている、Transcribe Bentham を遥かに凌ぐ成果をあげたのである。

橋本の説得力のある新提案と、それに基づく「みんなで翻刻」の成果を得て、また、最終年度には、橋本が科研費研究者の資格を得たので、橋本を分担者として、予算のほとんどを橋本の「みんなで翻刻」の開発費にあてた。つまり、本研究は当初の方法を破棄して、新しい方法により、当初の目的を達成する方向に舵を切ったのである。ちなみに、最後の年度に、橋本が、直接研究費を使えるようになって以後に達成されたのは、翻刻の全文検索モジュールの開発などである。

4. 研究成果

以上の様な経緯で、本研究は、研究開発当初には、想像すらできなかった、大きな成果を得て終了した。

本研究は、分担者橋本と加納の優れた研究により、我が国特有の問題である、くずし字を如何にクラウド翻刻するかという、学術的にだけでなく、防災などの社会応用にも大きな意味をもつ問題における、大きなブレークスルーをもたらすことで出来たと自負している。

橋本は、「みんなで翻刻」の成功の理由を、アンケート調査などで分析したが、その結果、「みんなで翻刻」は、地震史料の翻刻だから、多くの人たちの興味を引いたのではなく、それが参加者の学習意欲を引き出し、また、学習の機会を与えるものであるために、多くの人々を引き付けたことが判明した。

これは、「みんなで翻刻」が、古文書一般に適用可能で、本研究が目指した、古文書の WEB の基礎技術たりえることを強く示唆しているといえるだろう。また、橋本は、さらに、「みんなで翻刻」の方式を教育学の協調学習の概念を使って、デザイン・パターンのようなパターン化、モデル化してもある。

「みんなで翻刻」は、本研究が終了した後も、開発・運用が続けられており、従来では考えられなかった驚異的なスピードで古地震史料の翻刻が進んでいる。

具体的には、東大地震研所蔵の古地震関連史料の内、デジタル化済みの 472 点の内、2018 年 6 月 3 日の段階で、446 点が翻刻済みである。これは画像数でいうと、6966 枚の内、6440 枚である。

これを Transcribe Bentham と比較すると、みんなで翻刻のスピードの大きさがわかる。Transcribe Bentham は 2010 年にプロジェクトが開始され、2017 年 8 月時点で公表された数値では、その時点までで 18,775 枚が翻刻されているという。およそ 7 年で、1 万 8 千枚程度ということである。

これに対して、みんなで翻刻は、2017 年 1 月 10 日に公開されているので、1 年半程度で、6 千枚以上翻刻されている。単純に計算すると、Transcribe Bentham は 2500 枚/年、み

んなで翻刻は、4000 枚/年である。史料の性格が大変に異なるため、この様な単純比較には、客観的な意味は薄いのだが、みんなで翻刻以前には、日本語テキストの Crowd transcription の成功例が、著作権が切れた小説などの活字テキストを対象とする「青空文庫」しかなかったことを考慮すると、みんなが翻刻が成し遂げたことの意味は非常に大きい。

それは、古文書の WEB の実現という、本研究が掲げた遠い目標にむけての、力強い第一歩なのである。

現在、分担者橋本のもとには、京都府立京都学・歴史館が所蔵し、デジタル化して、WEB で公開もされている、国宝東寺百合文書の翻刻に、「みんなが翻刻」を使えないかなどという問い合わせが、寄せられている。

地震、津波、噴火など、近年、日本社会において、強い関心が寄せられているものに関連した古文書群にたいして、大きな成果をあげた本研究であるが、橋本の分析結果からすれば、おそらく、その様な対象古文書の性格を超えて、東寺百合文書のような、日本史の重要資料にも適用可能であろうと思われる。

もちろん、この予想は、それが現実化されるまで、単なる予想に留まるので、我々は、今後とも、本研究で得られた成果の発展に努力していきたいと考えている。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 0 件)

[学会発表](計 10 件)

1. 加納靖之, 古地震研究とみんなが翻刻, 古典籍文理融合シンポジウム(第2回古典籍文理融合研究会), 国立極地研究所・国文学研究資料館・統計数理研究所, 2018年, 招待講演
2. 加納靖之, 「みんなが翻刻」これまでとこれから, じんもんこん 2017 (人文科学とコンピュータシンポジウム), 大阪市立大学, 2017年, 基調講演
3. Y. Kano, Y. Hashimoto, I. Nakanishi, J. Ohmura, T. Amano, T. Kuba, H. Sakai, K. Ito, K. Hattori, Y. Odagi, M. Nishikawa, H. Horikawa, K. Mizushima, R. Yasukuni, and M. Yamamoto, Minna de Honkoku: online transcription project of historical earthquake documents, IAG- IASPEI 2017, S23-P-09, Kobe, 2017年
4. 加納靖之, 橋本雄太, 中西一郎, 大邑潤三, 天野たま, 久葉智代, 酒井春乃, 伊藤和行, 小田木洋子, 西川真樹子, 堀川

晴央, 水島和哉, 安国良一, 山本宗尚, みんなが翻刻: 市民参加の地震史料解読プロジェクト, 日本地球惑星科学連合 2017年大会, MGI27-P10, 千葉市, 2017年

5. 橋本雄太, 変体仮名文字情報基盤を利用した学習システム開発, シンポジウム「変体仮名のこれまでとこれから」, 国立国語研究所, 2017年11月25日, 招待講演
6. 橋本雄太, 亀田 堯宙, 池田 光雪, デジタルアーカイブの情報技術, デジタルアーカイブサミット 2017, 2017年9月10日, 招待講演
7. Yuta Hashimoto, SMART-GS: Image Processing Technology Meets Humanities, Tsukuba Global Science Week 2017, Tsukuba University, 22 September 2017, 招待講演
8. Yuta Hashimoto, Minna de Honkoku: A Crowdsourced Transcription Project of Historical Earthquake Records, Digital Approaches in Japanese Studies, Frankfurt University, May 2015, 招待講演
9. 橋本雄太, くずし字の学習支援と市民参加翻刻, 第2回 CODH セミナー くずし字チャレンジ~機械の認識と人間の翻刻の未来~, 国立情報学研究所, 2017年2月10日, 招待講演
10. 橋本雄太, 学習を動機付けに利用した前近代災害史料のクラウドソーシング翻刻, じんもんこん 2016, 国立国語研究所, 2016年12月11日(査読あり)

[図書](計 1 件)

加納靖之, アプリで学ぶくずし字 くずし字学習支援アプリ KuLA の使い方, pp. 48-51, 飯倉洋一(編) 2017年, 笠間書院

[産業財産権]

出願状況(計 0 件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
国内外の別:

取得状況(計 0 件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕

林晋，失われた時のウェブを目指して，平成 29 年度京都大学図書館機構講演会講演，2017.10.17

加納靖之，みんなで翻刻 これまでとこれから，レポート笠間，63，53-56，笠間書院，2017c .

京都大学古地震研究会・加納靖之，「みんなで翻刻」市民参加のオンライン翻刻プロジェクト，地震本部ニュース，2017 年夏号，8-9，2017d .

加納靖之・橋本雄太・中西一郎，みんなで翻刻，日本地震学会ニュースレター，69，NL5，48，2017f .

京大古地震研究会
WEB サイト「みんなで翻刻」
<http://honkoku.org/>
橋本と加納が作成運用

6 . 研究組織

(1) 研究代表者

林晋 (HAYASHI, Susumu)
京都大学・文学研究科・教授
研究者番号：40156443

(2) 研究分担者

久木田水生 (KUKITA, Minao)
名古屋大学・情報科学研究科・准教授
研究者番号：10648869

加納靖之 (KANOU, Yasuyuki)
京都大学・地震予知研究センター・助教
研究者番号：30447940

橋本雄太 (HASHIMOTO, Yuta)
国立歴史民俗博物館・研究部・助教
研究者番号：10802712

(3) 連携研究者

該当せず

(4) 研究協力者

該当せず