

京都大学人文科学研究所共同研究最終報告書（終了）

1. 研究課題

人文学研究資料にとっての Web の可能性を再探する

RE-EXPLORATION OF POTENTIALS OF WEB FOR RESOURCES OF THE HUMANITIES

2. 研究代表者氏名

永崎研宣

Kiyonori Nagasaki

3. 研究期間

2013 年 04 月 - 2016 年 03 月（2 年度目）

4. 研究目的

Web 技術の発展にともない、人文学資料向けの Web サービス（以下、人文系 Web サービス）においてもサービス同士の相互連携をはじめとする様々な面での新しい可能性が大きく拓けてきているが、古い設計に基づくシステムやデータの改良は容易ではなく、結果として、最先端の Web 技術が投入されたものとそうでないものが入り乱れた状態になっており、利用者にとっての利便性という観点からは改善の余地がますます大きくなってきている。本研究の目的は、人文科学研究所における各種 Web サービスを中心としつつ、本研究の共同班班員が関わっている様々な Web サービスの事例も含め、現在の Web サービスとして求め得る水準と実際のその距離を再検討することで、それを縮めるための方策を明らかにすることにある。この再検討にあたっては、各種人文系 Web サービスの研究における意義だけでなく、当初計画や依拠する規格、予算の性格、低コストな改良可能性など、学会研究会で報告されにくい部分にも焦点をあてていくことで、問題の具体的な解決策に少しでも近づけることを目指す。

As development of Web technology, Web service for resources of the humanities has been increased its possibilities regarding interoperability of various services and so on. However, it is still difficult to solve several problems such as improvement of legacy systems and data. As a result, it has become difficult for humanities scholars to use them easily among the mixed environment of legacy and advanced services. Under such circumstances, we aim to re-explore Web for resources of the humanities based on surveying ideal implementations and current various Web services which have been managed mainly by ourselves in order to clarify ways of realizing better Web services for

resources of the humanities.

6. 研究成果の概要

人文学にとって Web はどのようなものになりつつあるのか。そしてそれはどのように接していくべきなのか。もはやインフラと言っても良いほどに普及した Web に大量に蓄積された文化資料を前に、かつて紙媒体に依拠してその方法論を鍛えてきた人文学は、どのような態度を取るべきなのか。本研究班では、これを明らかにすることを目的とした 3 年間の活動を終了したところである。

研究班は、文化資料のデジタル化と研究利用に関わっている研究者を中心に構成し、各班員が関わる個々の取組みについて報告と議論をするところからはじまった。

主な報告を挙げてみると、東洋学文献類目検索の報告ではレガシーデータからの移行の困難さが提起され、CiNii の報告では大規模学術デジタル事業の運営におけるマネジメントの難しさが報告された。

そして、SAT 大蔵経データベースに関しては、20 年近くになるデータを現在にも十分に活用できるようにするための工夫が報告された。

これに加えて、2013 年度は 3 回の公開シンポジウムを開催した。

「東洋学におけるテキスト資料の構造化と Web の可能性」(2013 年 12 月 9 日開催) では、日本近代文学の草稿の構造化とデジタル化、仏典の構造的なマークアップ等、東洋学におけるテキストの扱いについての幅広い議論が行われた。

「近デジ大蔵経公開停止・再開問題を通じて人文系学術研究における情報共有の将来を考える」(2013 年 1 月 24 日開催) では、東京大学・国際日本文化研究センターからの研究者の参加に加えて、著作権問題に取り組んでいる弁護士の参加もあった。国立国会図書館におけるデジタル化資料の公開をめぐる問題にはじまり、編集著作権の扱いや人文学における校訂の論理と著作権の考え方の齟齬についての問題など、様々な議論が交わされ、大きな反響を得た。

「翻デジ 2014: クラウドソーシングによる近デジ資料のデジタル翻刻」(2014 年 2 月 19 日開催) では、国立国会図書館が提供する NDL lab においてはじまったクラウドソーシングデジタル翻刻プロジェクト、翻デジ 2014 のシステムが初公開され、同時にチュートリアルも行われ、参加者が実際に翻刻作業に参加した。

2014 年度は、前年度に引き続き、各班員の個々の取組みについての報告と議論を中心として進められた。題材として採り上げられた主なものとしては、東京大学史料編纂所データベース群、唐代拓本文字データベース等があった。

また、8 月 5 日には漢デジ 2014: デジタル翻刻の未来と題する公開シンポジウムを日本学術振興会科学研究費補助金基盤研究(C)「大規模漢字集合フォント対応のデータベース構築による平安時代漢字字書研究」(代表: 池田証壽) との共催で開催した。主として漢字の符号化に関わる事例が様々に提供され、活発な議論が展開された。

2015年度は、成果物を刊行するための議論と原稿作成を中心とした研究会開催となったが、主なものとしては、国立国語研究所の通事コーパス、NewsMLを中心とした共同通信社の記事作成システム等の報告と議論が行われた。成果物については現在まとめを行っており、近日中の刊行を目指して鋭意作業中である。

本研究班の3年間の活動を通じて明らかになったことは様々だが、中でも、デジタル技術やそれに関わる制度が日進月歩であり可塑的であるという状況において、長期利用を前提とする人文学向けの文化資料に関わるデータやプログラムをどのようにして持続可能なものとしていくかということが大きな課題であり続けるという点は、研究班開始当初より継続した通奏低音であったように思われる。

紙媒体においては読者に委ねられていた自由な解釈の余地は、デジタル化に際して効率的効果的な機械可読性を求められる中で、分類や符号化といった形で捨象されていくことになる。しかし、それに際してのルールが十分に記述されていなければ、効率や効果どころか、かえってデータの解釈を誤らせ、さらにはデータそのものが失われることさえある。それを避けるには、ルールを適切に記述しておく必要がある。ルールの記述に際しては、内容が適切かどうかというだけでなく、ルールを記述する形式が持続性の高いものとなっていることも重要である。成果物においては、こういった点を基礎として、人文学にとってのWebの可能性とそれにあたって人文学側が成すべきことについて、研究会で集約された様々な論点をたどりつつ、提示していく予定である。

なお、末筆ながら、拙い研究会運営にも関わらず、多忙な中で少しでも時間を工面して研究会に参加して下さった班員の方々、いつも迅速に研究班開催に関わる手続きを進めて下さった安岡孝一副班長に、感謝の意を表して、本報告を締めくくりたい。

8. 共同研究会に関連した公表実績

・ JADH2015 (International Conference of Japanese Association for Digital Humanities) 1-9, September, 2015, Kyoto University.

・永崎 研宣, 苫米地 等流, A.Charles Muller, 下田 正弘「持続可能なデジタルアーカイブに向けて—SAT 大蔵経データベースにおける取り組みを通じて」『じんもんこん 2014 論文集』(2015年12月), pp. 219-224.

・永崎研宣「SAT 大蔵経テキストデータベース 人文学におけるオープンデータの活用に向けて」『情報管理』Vol. 58 (2015) No. 6 pp. 422-437. doi:10.1241/johokanri.58.422.

・永崎研宣「「デジタル・アーカイブ」の利活用可能性を高めるために—仏典画像統合検索 API の構築を通じて」『情報処理学会研究報告.人文科学とコンピュータ研究会報告』, 2015-CH-107(3), pp. 1-4.

・永崎研宣「「翻デジ」とNDL」『情報処理学会研究報告.人文科学とコンピュータ研究会報告』 2015-CH-106(12), pp. 1-4.

・永崎研宣「仏教文献のための構造的なデジタルテキストの記述と活用」『印度学仏教学

研究』第 63 巻第 2 号(2015 年 3 月), pp. 1088-1094.

・永崎研宣, Paul Hackett, 苜米地 等流, A.チャールズ・ミュラー, 下田 正弘「人文学にとっての「リンク」の意義 SAT 大蔵経データベースを手がかりとして」『じんもんこん 2014 論文集』(2014 年 12 月), pp. 17-22.

・Kiyonori Nagasaki, A. Charles Muller, Toru Tomabechi, and Masahiro Shimoda, “Exploring Possibilities of Digital Environments for Buddhist Studies”, 17th Congress of the International Association of Buddhist Studies (IABS), Vienna (Austria), (2014/8/23), <http://iabs2014.univie.ac.at/academic-program/list-of-sections/18-information-technology/>

・Kiyonori Nagasaki, A. Charles Muller, Toru Tomabechi, and Masahiro Shimoda, “Bridging the Local and the Global in DH: A Case Study in Japan” Digital Humanities 2014, Lausanne (Switzerland), (2014/7), pp. 279-280.

・Masahiro Hori, Osamu Imahayashi, Tomoji Tabata, Keisuke Koguchi, Miyuki Nishio, and Kiyonori Nagasaki, “The Development of The Dickens Lexicon Digital and its Practical Use for the Study of Late Modern English” , Digital Humanities 2014, Lausanne (Switzerland), (2014/7), pp. 479-480

・永崎研宣「大正新脩大蔵経とデジタル時代の学術情報流通」『DHjp』No.3, 勉誠出版 (2014 年 4 月) , pp. 11-20.

・永崎研宣「「オープン」と DH」『DHjp』No.4, 勉誠出版 (2014 年 8 月) , pp. 10-17.永崎研宣「日本語クラウドソーシング翻刻に向けて」情報の科学と技術 64(11), 475-480, 2014-11-01.

・永崎研宣「クラウドソーシングによるテキスト翻刻の実践に向けて」人文科学とコンピュータ研究会報告 2014-CH-102(6), 1-5, 2014-05-24.

・永崎研宣「デジタル技術を活用した人文学研究の現在」『日本語学』通巻 432 号 (第 33 巻 14 号) .

・山田太造「前近代日本史史料をベースとしたテキストデータベースの特徴と課題」『日本語学』通巻 432 号 (第 33 巻 14 号) .

・永崎研宣「「翻デジ」と NDL」『情報処理学会研究報告.人文科学とコンピュータ研究会報告』 2015-CH-106(12), pp. 1-4.

・田中 僚,松村 敦,宇陀 則彦, 原資料の構造を反映したデジタルアーカイブの構築, 情報処理学会研究報告. 人文科学とコンピュータ研究会報告, 2015-CH-106(9), p.1-2, 2015-05-09.

・森嶋 厚行,川島 隆徳,原田 隆史,宇陀 則彦, クラウドソーシングってどうですか?Crowd4U×NDL データの事例, 情報処理学会研究報告. 人文科学とコンピュータ研究会報告 2015-CH-106(13), p.1-4, 2015-05-09.

・永崎研宣「「デジタル・アーカイブ」の利活用可能性を高めるために—仏典画像統合検索 API の構築を通じて」『情報処理学会研究報告.人文科学とコンピュータ研究会報告』,

2015-CH-107(3), pp. 1-4.

・ Kiyonori Nagasaki, Paul Hackett, A. Charles Muller, Toru Tomabechi, Masahiro Shimoda, “Significance of Linking between past and present, east and west, and various databases” , Digital Humanities 2015, Sydney (Australia), (2015/7).

・ Christian Wittern 「Multiple Views and Modes of Engagement with a Repository of Digital Texts」 , Digital Humanities 2015, Jul 2015.

・ Taizo YAMADA, Satoshi Inoue. Detection of People Relationship Using Topic Model from Diaries in Medieval Period of Japan, Proceedings of DH2015, -, 2015. (査読有)

・ 永崎研宣「研究活動としての人文学デジタル化 —仏教学知識基盤を事例として」、大規模学術フロンティア促進事業「日本語の歴史的典籍の国際共同研究ネットワーク構築計画」第1回 日本語の歴史的典籍国際研究集会プログラム「可能性としての日本古典籍」、於国文学研究資料館 (2015年7月31日)。

・ Taizo YAMADA, Analysis of Archaeological Information Using Topic Model Technologies, International Workshop on Application of Science and Technology for Cultural Studies(IWASTCS2015), Venue: 6th Floor, Princess Maha Chakri Sirindhorn Anthropology Centre (SAC), Bangkok, Thailand, 13 Nov 2015.

・ 永崎研宣「SAT 大蔵経データベースをめぐる漢字情報」、シンポジウム「字体と漢字情報」—HNG 公開 10 周年記念—, 於国立国語研究所, (2015年11月22日)

・ Taizo YAMADA. Extraction and Management of Spatiotemporal Term from Field Notes and Data Structuring for its Sharing in Area Studies, Proceedings of PNC2015, -, 2015. (査読有)

・ 岩崎陽一「学術データベースの開発手順について」早稲田大学オペラ／音楽劇研究所研究会, 早稲田大学, 2015/11/28.

・ 山田太造「地域研究史資料を対象とした時空間的特徴の抽出と場面の構造化」、第14回情報科学技術フォーラム講演論文集, vol.14, no.4, pp.409-410, 2015.

・ 山田太造「フィールドノートに記述された場面を特徴づける—語彙による知識処理—」, 特集 第20回情報知識学フォーラム「地域情報学における知識情報基盤の構築と活用」, 情報知識学会誌, Vol.25, No.4, pp.315-324, 2015.

・ 永崎 研宣 , 苜米地 等流 , A.Charles Muller , 下田 正弘「持続可能なデジタルアーカイブに向けて—SAT 大蔵経データベースにおける取り組みを通じて」『じんもんこん 2014 論文集』(2015年12月), pp. 219-224.

・ Kiyonori Nagasaki, “The SAT database and the future of digital humanities.” , WORKSHOP: Nirvana Sutra Workshop, UC Berkeley, (7-8, Jan. 7-8, 2016).

・ 山田太造「前近代日本史史料に関わる人名情報の収集・蓄積に関する考察」, 研究報告 人文科学とコンピュータ (CH), Vol.2016-CH-109, No.2, pp.1-4, 2016.

・ 永崎研宣「テキストマイニングをめぐるデジタル・ヒューマニティーズの課題」, 第4

回九州大学異分野融合テキストマイニング研究会シンポジウム ― テキストマイニングとデジタル・ヒューマニティーズ, 於九州大学, (2016年1月30日) .

・永崎研宣「デジタルヒューマニティーズ:大藏経テキストデータベースプロジェクト」, 東京大学大学院情報学環 DNP 学術電子コンテンツ研究寄附講座 開設記念シンポジウム「これからの学術デジタル・アーカイブ」, 於東京大学福武ホール, (2016年2月9日) .

・山田太造「テキストデータを使うとどのようにフィールドが分類できるか?」, 日本人口学会開催地域部会 2015 年度研究会, 2016年3月5日, 総合地球環境学研究所.

・山田太造「じんもんこん 2015 開催報告 -議論沸騰の3日間 in 京都」, 情報処理, Vol.57, No.4, pp.404-405, 2016.

・ウィッテルン・クリスティアン (編): センター研究年報 2015 特集 漢籍リポジトリ、京都大学人文科学研究所所属東アジア人文情報学研究センター, 2016.

・Kiyonori Nagasaki, “Re-Creation of Buddhist Studies in the Digital Era” , International Symposium: MEMORY, the (Re-)Creation of Past and Digital Humanities, Keio University, (15 Mar 2015).

・山田太造「東京大学史料編纂所の編纂とその業務にともなうデータベース」, 歴博公開シンポジウム「資料がつなぐ大学と博物館―「研究循環アクセスモデル」の構築にむけて―」予稿集, 2016.

・柳澤雅之, 高田百合奈, 山田太造「地域情報学の読み解き-発見のツールとしての時空間表示とテキスト分析」, 地域研究, Vol.16, No.2, pp.267-291, 2016. (査読有)

・Kiyonori Nagasaki, “Crisis of Humanities in Japan,” CEAL2016, Sheraton Seattle Hotel, (31 Mar 2016).

15. 研究成果公表計画および今後の展開等

本研究の成果は、一義的には、通常の学術研究と同様、学会等における研究発表や学術雑誌における論文公表等によって行われる。しかしながら、本研究の目標はそれにとどまらず、人文系研究の成果が Web で公開されるに際してのガイドラインを公表するところにある。このガイドラインは、より広く知られることを重視すべく、まず、これまでの研究会・シンポジウムにおける議論の成果を平易にまとめ、Web サイトにおいて適切な構造を以て公表する。さらに、それをよりわかりやすく解説する一般向けの啓蒙書を共同執筆し公開する予定である。この研究成果の公開後も、Web サイト上のガイドラインについては引き続き更新を続けていく予定である。