

Digital Humanitiesにおける Web コラボレーション

橋本 雄太¹

概要: ティム・オライリーが「Web 2.0」という造語で Web の変化を表現してから 10 年以上が経過し、Web が双方向的なコミュニケーションの場であることは当たり前の事実となった。Web を介して遠隔地の専門家間の協働を実現する Web コラボレーションは、Digital Humanities 分野の主要研究テーマの一つであり、2000 年代から多数の研究が実施されている。最近では、専門家同士の協働だけでなく、多数のボランティアな市民を人文学研究に参画させるクラウドソーシング型の研究も活発である。本発表では、Web コラボレーションに関する最近の Digital Humanities 研究を複数取り上げ、その動向を報告する。

キーワード: デジタル・ヒューマニティーズ, Web コラボレーション

Web-based Collaborative Researches in Digital Humanities

YUTA HASHIMOTO¹

Abstract: More than a decade has passed since Tim O'Reilly coined a term "Web 2.0" for describing the newly evolving practices and standards of the World Wide Web in the mid-2000s; By now no one will doubt the collaborative nature of the Web. In the field of digital humanities collaborations of humanities scholars via the Web have become one of major research topics since 2000s. In this short article, I will review several recent Digital humanities projects that make uses of web-based collaborative technologies, and sum up the trends in this research field.

Keywords: digital humanities, web collaboration

1. はじめに

自然科学や工学分野と比較して、人文学分野では複数の研究者が共同で成果を発表する事例自体は多くないものの、研究者間の連携自体は活発に実施されている。たとえば歴史研究においては、史料解読に関わる負荷を分担するため、あるいは史料解釈の妥当性を担保するために、複数人で一つの史料を共有する「読み合わせ」が伝統的に実施されている。一方で、こうした研究形態は、(1) 場所的制約、(2) 時間的制約、(3) 人力的制約の 3 つに制限されてしまうため、環境によっては実施自体が困難である。

Digital Humanities 分野では、Web 技術を駆使してこれらの共同研究にかかる諸制約を解消し、研究者間の国際的あるいは異分野横断的連携を実現しようとする試みが活発に実施されてきた。以下では、こうした人文学者間の Web コラボレーションに関する研究を複数取り上げ、近年の研究動向を報告したい。

究動向を報告したい。

2. DH 分野における Web コラボレーションに関する研究

以下では、近年の Digital Humanities 分野における主要な Web コラボレーションについての研究を、いくつかのサブテーマに分けて紹介する。

2.1 データ標準化・相互運用性に関する研究

ネットワークを介してデータをやり取りするためには、データの形式と交換手続きに関する標準化が必要である。人文学で扱われる種々のデータをどのように標準化し、運用するかについては、さまざまな立場からの議論と研究が存在する。

人文情報学分野では広く知られているように、文学作品や歴史史料など人文学テキストの電子符号化においては、Text Encoding Initiative が策定した XML ベースのガイドラインが事実上の標準となっている [1]。TEI 自体、Web

¹ 京都大学大学院文学研究科 博士後期課程
Graduate School of Letters, Kyoto University

を介したデータ交換を強く志向する性質の規格であるが、後に紹介するように、最近ではソースコード共有サービスの GitHub (<https://github.com>) に TEI の XML データを保管するレポジトリを開設し、多人数でデータの共有と編集を実現する事例が増えている。

また、テキストを始めとする人文学データに対するアノテーション情報を標準化する試みとして、W3C のワーキンググループが「Open Annotation データモデル」を策定している (<http://www.openannotation.org/>) [2]。Open Annotation データモデルは RDF を始めとするセマンティック Web 技術をベースにして開発されており、テキストや画像、動画など各種のメディアに対するアノテーション情報を相互利用可能な形式で記述することを目的としている。また、人文学テキストに対するアノテーションを容易に共有・相互運用することを目的として、Open Annotation データモデルを拡張した Open Annotation Collaboration (OAC) なども提案されている [3]。

また、Open Annotation の策定ワーキンググループの活動が元となり、2016 年 7 月に、W3C がアノテーション情報を記述するための新しいデータモデル「Web Annotation データモデル」を勧告候補 (Candidate Recommendation) として公開した (<https://www.w3.org/TR/annotation-model/>)。今後この規格が勧告され、W3C 標準として流通することになるかもしれない。

2.2 クラウドソーシング

近年では、少数の専門家による協働では到底処理しきれない、大量の人文学テキストの翻刻や歴史史料画像の分類などの作業を、Web を介して参加する市民の手に委ねる形態の研究プロジェクトも活発に実施されている。こうした試みも広義の Web コラボレーション研究に含まれると考えて良いだろう。このようなプロジェクトは Digital Humanities の文脈ではクラウドソーシング (crowdsourcing) と呼ばれる [4]*1。

Digital Humanities 分野でもっとも成功したクラウドソーシングプロジェクトは、ユニヴァーシティ・カレッジ・ロンドンが主催した “Transcribe Bentham” だろう (<http://blogs.ucl.ac.uk/transcribe-bentham/>) [7]。同プロジェクトでは、哲学者 J. Bentham の大量の遺稿の翻刻をボランティアに依頼しており、2016 年までに全体の 94% を占める 15,634 ページの翻刻が完了または着手済みの状態にあるという。

クラウドソーシングやシチズンサイエンスの手法は近年さまざまな分野で活用が進んでおり、シチズンサ

イエンスのポータルサイトである Zooniverse (<https://www.zooniverse.org>) には、人文学関連の 4 つのプロジェクトが公開されている。このうちの一つ、“Operation War Diary” は、第一次世界大戦の従軍兵士の日記に含まれる場所、時間、人名といった諸情報をボランティアの手によってタギングするプロジェクトである。

最近では、2016 年 9 月に、欧州の統合デジタルアーカイブポータルである Europeana が、第一次世界大戦に関する資料を公開している “Europeana 1914-1918” の手書き文書の翻刻を、コンペティション形式で実施すると発表した (<http://pro.europeana.eu/blogpost/transcribing-europeana-1914-1918>)。

3. DH 2016 における研究発表

続いて、今年 7 月にポーランド・クラコフで開催された Digital Humanities 年次カンファレンス (DH2016) における発表をいくつか紹介したい。筆者の確認した限りでは、今年度の発表アブストラクトのうち、何らかの形で Web コラボレーションに関係する研究の数は 9 件であった。

イタリアの A. Bolioli ら [8] は、文学作品を対象とした Web ベースの共同アノテーションツール “CBook” (<https://cbook.it/>) を紹介した。Bolioli らは同ツールを高校生向けの教育に活用しており、CBook を利用することによって教員と生徒は作品テキストの一部にアノテーションを共有可能な形で付加することができる。また作品に関連する画像など他のコンテンツを共有することも可能である。作品へのアノテーションは JavaScript 製ツールの Annotator (<http://annotatorjs.org/>) を組み込むことで実現されており、アノテーション本体は TEI 形式に変換して保存される。

Web コラボレーションに関する研究発表では、分散型ソースコード版管理システムの git と、git を利用した GitHub に関係する発表が 9 件中 3 件を占めていたことが印象的であった。たとえば L. Spiro は、人文学データの保存・共有・公開プラットフォームとしての GitHub を評価する発表をしている [9]。Spiro は、機能性、操作の簡易さ、コミュニティ志向性、公開性などの面から、既存の学術レポジトリに対する GitHub の優位性を分析している。

T. Grallert による [10] は、19 世紀末から 20 世紀初頭にアラビア語で刊行された雑誌等の定期刊行物を共同翻刻するプロジェクトである。イスラム国による文化財破壊が近年問題になっている中東では、貴重な文化財をデジタル化する試みが続けられている。刊行物もこれに含まれるが、アラビア語文章への OCR 適用の難しさから、テキストの文字起こしは人力での翻刻に頼らざるを得ないという。Grallert は、共同で翻刻を実現するためのプラットフォームとして GitHub を活用し、雑誌のデジタル画像と翻刻テキストを共有可能な形式で管理している (<https://github.com/>)

*1 ただし、元々 crowdsourcing という用語は企業のアウトソーシングの一形態を表すために使われた用語で、学術プロジェクトの分類名称として利用するには問題も多い [5, 6]。自然科学領域における市民参加型プロジェクトは一般に「シチズンサイエンス」と呼ばれる。

tillgrallert/digital-muqtabas). また翻刻テキストのマークアップには TEI が利用されている。

J. Reeve による [11] も同様に GitHub を利用した研究である。このプロジェクトは、大英図書館がデジタル化した5万点を超えるパブリックドメイン書籍のテキストデータを、GitHub 上で管理することを目標としている。Reeve は、プロジェクト・ゲーテンベルクに代表される従来の電子テキストレポジトリには、(1) 利用者による修正・編集が困難である、(2) 編集履歴が残されないことが多い、(3) テキストデータを利用するためのインターフェイスが統一されていない、などの複数の欠点があることを指摘する。これらの課題を解決するため、Git-Lit では、既存のテキストデータから自動的に GitHub レポジトリを生成するソフトウェアを開発中であるという。

4. おわりに

本稿では、近年の Digital Humanities 分野における Web コラボレーションについての研究を簡潔に紹介した。研究会発表では、本稿で紹介した研究プロジェクトやツールのデモンストレーションも交えて、より具体的な動向のレビューを行う予定である。

参考文献

- [1] Consortium, T., Burnard, L., Bauman, S. et al.: *TEI P5: Guidelines for electronic text encoding and interchange*, TEI Consortium (2008).
- [2] Sanderson, R., Ciccarese, P., Van de Sompel, H., Bradshaw, S., Brickley, D., a Castro, L. J. G., Clark, T., Cole, T., Desenne, P., Gerber, A. et al.: *Open annotation data model, W3C community draft* (2013).
- [3] Hunter, J., Cole, T., Sanderson, R. and Van de Sompel, H.: *The open annotation collaboration: A data model to support sharing and interoperability of scholarly annotations*, *Digital humanities 2010*, Office for Humanities Communication; Centre for Computing in the Humanities, pp. 175–178 (2010).
- [4] Carletti, L., Giannachi, G., Price, D., McAuley, D. and Benford, S.: *Digital humanities and crowdsourcing: An exploration*, *Museums and the Web* (2013).
- [5] Howe, J.: *The Rise of Crowdsourcing* (2006).
- [6] Dunn, S. and Hedges, M.: *Crowd-sourcing Scoping Study. Engaging the Crowd with Humanities Research* (2012).
- [7] Causer, T. and Wallace, V.: *Building a volunteer community: results and findings from Transcribe Bentham*, *Digital Humanities Quarterly*, Vol. 6 (2012).
- [8] Andrea Bolioli, R. T.: *Collaborative Annotation and Exploration of Literary Works in Learning Contexts*, *Digital Humanities 2016: Conference Abstracts*, Krakow, Jagiellonian University & Pedagogical University, pp. 746–747 (2016).
- [9] Spiro, L.: *Evaluating GitHub as a Platform of Knowledge for the Humanities*, *Digital Humanities 2016: Conference Abstracts*, Krakow, Jagiellonian University & Pedagogical University, pp. 688–690 (2016).
- [10] Grallert, T.: *The journal al-Muqtabas between*

Shamela.ws, HathiTrust, and GitHub: producing open, collaborative, and fully-referencable digital editions of early Arabic periodicals—with almost no funds, *Digital Humanities 2016: Conference Abstracts*, Krakow, Jagiellonian University & Pedagogical University, pp. 529–531 (2016).

- [11] Reeve, J.: *Git-Lit: an Application of Distributed Version Control Technology toward the Creation of 50,000 Digital Scholarly Editions*, *Digital Humanities 2016: Conference Abstracts*, Krakow, Jagiellonian University & Pedagogical University, pp. 657–658 (2016).