

クラウドソーシングによる 歴史災害史料の大規模テキスト化と 構造化の試み

Large-scale Transcription and Structuralization
of Historical Disaster Records through Crowdsourcing

橋本雄太

HASHIMOTO Yuta

はじめに

①研究の背景

②本研究の目的とアプローチ

③『みんなで翻刻』システム

④『みんなでマークアップ』システム

おわりに

【論文要旨】

本稿では、国立歴史民俗博物館開発型共同研究「歴史災害研究のオープンサイエンス化に向けた研究」（2018～2020 年度）を通じて実施した、災害史料の構造化データ化に向けた取り組みについて述べる。本研究の目的は、地震を中心とする前近代の災害史料を対象として、クラウドソーシングにより大規模なテキスト化と構造化データ化を実現することであった。この目的のために開発した文献史料のクラウドソーシング翻刻システム『みんなで翻刻』では、2019 年 3 月までに東京大学地震研究所図書室が所蔵する 499 点の災害史料の全文テキスト化を実現した。また、翻刻テキストの構造化のために、テキスト中の重要情報をマークアップし、外部データソースと結びつけるクラウドソーシングシステム『みんなでマークアップ』を開発した。クラウドソーシングによる災害史料の大規模な構造化データ化は未達成であるが、システムの開発を通じて災害史料から時空間情報と被害情報を抽出し、各種の外部データソースと連結することで地震被害記録を構造化データへと変換するワークフローを整備した。この過程を通じて、現時点での問題点を整理し、解決の手がかりをつかむための考察をおこなった。

【キーワード】 歴史災害、クラウドソーシング、マークアップ、エンティティリンキング