

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

HASHIMOTO Yuta

This paper describes a series of our effort to convert historical disaster records into machine readable structural data that can be processed and analyzed by computers. This study was a part of a collaborative research program “Enhancing Historical Disaster Research with Open Science Methodologies”, which had been conducted by National Museum of Japanese history between FY 2018-2020. The objective of this program was to achieve large-scale transcription and structural data conversion through crowdsourcing for pre-modern disaster archives, with a focus on earthquakes. The crowdsourcing system “Minna de Honkoku,” which was developed for this purpose, achieved full transcription of 499 disaster historical materials owned by the Earthquake Research Institute of the University of Tokyo by March 2019. In addition, for structuring the transcribed texts, a crowdsourcing system “Minna de Markup” was developed to markup critical information in the text and link it to external data sources. Although we have not yet achieved large-scale conversion of disaster archives into structured data through crowdsourcing, we have developed a workflow for converting earthquake damage records into structured data by extracting spatio-temporal information and damage information from disaster records and linking them to various external data sources through the development of the system.

Key words: historical disasters, crowdsourcing, markup, entity linking