

Improving Digital Record Annotation Capabilities with Open-sourced Ontologies and Crowd-sourced Workers

ARPI MARDIROSSIAN AND LINDSAY TURLEY

Abstract: The Museum of the City of New York's collections contain a tremendous amount of cultural heritage information. The Museum has thus far implemented traditional methods for describing and managing this heritage data, but with current advances in technology and a desire to make the Museum's cultural heritage resources available to the general public, there is a need and an opportunity for a more scalable and novel solution.

Currently, the Museum catalogs items in its collections with metadata derived from a range of different vocabularies, such as AAT, LoC, and in-house standards. We employ full-time staff to carry out the cataloging. While this method works well on a small scale and for specialist researchers, it presents accessibility problems for the larger audience of the general public, is labor intensive and expensive, and is unable to scale to meet the Museum's needs.

To address these problems, we are working on the following solution, funded by a Digital Humanities Start-up grant from the National Endowment for the Humanities: dividing the work of cataloging into micro-tasks that can be completed by huge labor pools available through crowd-sourced marketplaces, as well as using both algorithmic and human computing methods to generate cross-walks between elements from different vocabularies that are semantically equivalent. This proposed solution results in order-of-magnitude increases in scalability and cost efficiency for both cataloging and ontology alignment tasks.

Thus far, this new solution is providing us granular, consistent metadata with a dramatic increase in the rate of production. Working with Tagasauris, a data curation platform, we currently employ an ever-growing labor pool of approximately 7,000 workers who process an average of 3,000 images per day. In this platform presentation, you'll learn more about the Museum's project and Tagasauris.

About the authors:

Arpi Mardirossian is a Co-founder and the Director of Product Development for Tagasauris. She oversees and implements new product development, enhancements, and design. She is an accomplished engineer and innovator working at the intersection of technology and creativity. She holds a doctorate in Industrial and Systems Engineering and a master's degree in Operations Research from the Viterbi School of Engineering at the University of Southern California. Her academic work, in the area of music information retrieval, involved assessing and visualizing music similarity. Her industry work has been in user interface design, data visualization, web development, and software engineering. She is also a faculty member at the University of Southern California where she teaches courses in Operations Research.

Lindsay Turley, Manuscripts and Reference Archivist at the Museum of the City of New York, holds a Masters Degree in Information and Library Science with certificates in Archival Studies and Museum Librarianship from Pratt Institute. She has been with the Museum since 2010. Her responsibilities include developing and implementing a strategy for processing and cataloging the Manuscripts and Ephemera Collections, and overseeing the reference function of the Collections Access department, including formulating and applying the Museum's internal and external research policies and procedures, and assisting users with accessing collections through the Museum's online Collections Portal. Previously, she held the position of Assistant Archivist at the American Civil Liberties Union in New York City, where she managed the migration of several legacy databases into a single archival management system, and developed a local cataloging manual and lexicon for the organization's holdings.