

Computer-Assisted Appraisal of Email: RATOM

CHRISTOPHER LEE

Abstract: Despite significant progress on various technologies to support both digital preservation and archival description, we have still seen relatively little progress on software support for the core activities of selection and appraisal. There are numerous data elements within born-digital materials that could (but currently are not) be used to support more effective and efficient appraisal processes. This presentation will report on work to apply machine learning to the appraisal and management of email collections. The Review, Appraisal and Triage of Mail (RATOM) project is a collaboration between the University of North Carolina and the State Archives of North Carolina. RATOM draws on expertise from past research and projects in areas with different use cases but related technologies. These include email curation guidance and case studies; tools for curation of email; the development of open-source digital forensics software platforms for LAMs; the use of NLP in LAMs to facilitate email review and research; machine learning to identify and classify important messages and entities; and appraisal and selection of electronic records. I will report on the project's rationale, objectives and progress to date.

About the Author:

Christopher (Cal) Lee is Professor at the School of Information and Library Science at the University of North Carolina, Chapel Hill. He teaches courses and workshops in archives and records management; understanding information technology for managing digital collections; and digital forensics. His primary research focus is the long-term curation of digital collections. He is particularly interested in the professionalization of this work and the diffusion of existing tools and methods into professional practice. Cal developed “A Framework for Contextual Information in Digital Collections,” and edited and provided several chapters to I, *Digital: Personal Collections in the Digital Era*. He has served as Principal Investigator of the Digital Acquisition Learning Laboratory (DALL), BitCurator, BitCurator Access, BitCurator NLP, BitCuratorEdu, and Review, Appraisal and Triage of Mail (RATOM) projects. He has been Co-PI on OSSArcFlow, as well as several projects focused on digital curation education: DigCCurr, DigCCurr 2, Closing the Digital Curation Gap (CDCG), Educating Stewards of Public Information in the 21st Century (ESOPI-21), and Educating Stewards of Public Information Infrastructure (ESOPI2). Cal was also Senior Personnel on the DataNet Federation Consortium funded by the National Science Foundation. He is a Fellow of the Society of American Archivists, and he serves as editor of *American Archivist*.