Reimagining Search and Discovery for Digital Collections with Machine Learning

Speaker: Prof. Benjamin Lee
Information School
University of Washington
Friday, September 8, 2023
1-2PM, NVC 214

Abstract

Widespread efforts by libraries and archives have drastically improved digital access to cultural heritage collections. Yet, scholars and the public alike face a persistent challenge: how to explore and analyze these collections, which frequently contain millions of items and often suffer from imperfect metadata. My interdisciplinary research addresses this question by bringing together approaches from machine learning, library & information science, and the digital humanities in order to build and examine large-scale search systems for digital collections. In this talk, I will present my project, Newspaper Navigator, which reimagines search and discovery for the visual culture embedded within 16 million digitized historic newspaper pages. Begun while I was an Innovator in Residence at the Library of Congress and a Ph.D. student at the University of Washington, the project has grown to consist of three components. First, I will discuss constructing and publicly releasing the Newspaper Navigator dataset and search application, with a particular focus on open faceted search – an extension of faceted search that leverages human-AI interaction affordances to empower users to define their own facets in an open domain fashion. Second, I will present ongoing collaborations to use Newspaper Navigator in order to advance digital humanities research surrounding ethnic newspapers and community-based archives. Third, I will discuss my efforts to examine the project from a critical sociotechnical perspective, resulting in the Newspaper Navigator data archaeology.

Biography

Benjamin Charles Germain Lee is an incoming Assistant Professor in the Information School at the University of Washington, as well as a Kluge Fellow in Digital Studies at the Library of Congress. He recently received his Ph.D. in Computer Science & Engineering from the University of Washington, which was supported by a National Science Foundation
Graduate Research Fellowship in machine learning. After graduating from Harvard College with an A.B. in Astrophysics and Mathematics, he served as the inaugural Digital Humanities Associate Fellow at the United States Holocaust Memorial Museum, as well as a Visiting Fellow in Harvard’s History Department. Ben also served as a 2020 Innovator in Residence at the Library of Congress and the 2020-2021 Richard and Ina Willner Memorial Fellow in the Stroum Center for Jewish Studies at the University of Washington. His writing has appeared in WIRED, Gawker, Current Affairs, Real Life, and GoldFlakePaint.