Coding for Career Planning: Using Python Programming and Data Visualization to Enhance Job-Search Capabilities for Memory Institution Professionals

Scott Richard St. Louis
2020 Joint Virtual Meeting of the
Council of State Archivists and the Society of American Archivists

DISCLAIMER: Due to scheduling disruptions caused by COVID-19, this poster outlines a potential forthcoming personal project in metadata/programming, rather than a completed one. Thank you for your feedback!

> By understanding job postings as information objects holding curation possibilities of distinct usefulness to the individual, early-career memory professionals with foundational skills in Python programming and descriptive metadata might be equipped to analyze and evaluate their own career possibilities in helpful ways computationally.

Quick Description

This poster outlines a potential forthcoming personal project that may be of interest to fellow graduate students (and others new to memory institutions) who seek to refine their career trajectories with technology. In consonance with the 2020 conference theme, "Creating Our Future," I will share how I plan to organize and manipulate a collection of more than 100 job postings gathered over nearly the last two years, specifically with an approach that could strengthen the quality and efficacy of any full-time job search for early-career memory professionals with basic proficiency in Python programming and metadata creation.

Materials

I began this personal project outline without realizing it in 2018, when I first started compiling job postings relevant to individuals with graduate training in archives, libraries, and information science using position announcement services including Archives Gig, INALJ, the ALA/ACRL scholarly communication listserv, and the digital curation Google group. I began doing so as I considered the available career opportunities that most excited me while envisioning my own possible future in graduate school.

Since then, I have compiled more than 100 job postings in a variety of allied specialties across a diverse range of employers, including:

- Archives and libraries at academic institutions with different Carnegie classifications.
- Businesses with hiring needs in digital curation.
- Nonprofit organizations with a focus on supporting the work of memory institutions.
- Nonprofit organizations with curation needs of their own.

Methods and Tools

Please find below a sequence of the methods and tools through which I am presently considering commencement of this project:

- Returning to coursework and project files from SI 666 (Organization of Information Resources) at the University of Michigan. Considering possibilities for a more capacious set of inclusion/exclusion criteria through which to filter job postings, consistent with the wide range of entry-level possibilities that a graduate degree affords in libraries, archives, and information science.
- Creating a homegrown metadata schema to ensure capture of all information relevant to my career goals, including (but not necessarily limited to) title, salary, employer, location, required/preferred qualifications, and position responsibilities.
- Developing a highly detailed guide for crosswalking my homegrown schema to an appropriate existing schema, for example https://schema.org/JobPosting
- Considering controlled vocabulary possibilities for various metadata fields, such as the LoC Name Authority File for location.
- Creating metadata records in CSV file, followed by cleaning in OpenRefine.
- Developing Python code to export CSV records to SQLite database viewable in DB Browser for SQLite.
- Finally, developing Python code to analyze job-posting metadata records for key questions. Possibilities might include the geographical distribution of jobs, average salary by employer type, Carnegie classifications of academic employers, most commonly required qualifications or technological proficiencies, et cetera. Other than Plotly, I am currently wondering what Python libraries for data analysis and/or visualization might be useful here?

Limitations

Naturally, the major limitation of embarking on such a personal project would be its usefulness only to one individual at a time. Even so, similar skills might be of broader usefulness to the profession through a study that systematically reviews and analyzes recent entry-level job postings in a specific subfield. For example, a study of entry-level position announcements in scholarly communication could make for an interesting submission to the *Journal of Librarianship and Scholarly Communication,* or a similar peer-reviewed publication. I am thankful to Colby Lewis at the University of Michigan for bringing this possibility to my attention.

Acknowledgments

I wish to thank Professor Andrea Thomer and graduate student instructor Michael Lenard at the University of Michigan for teaching me about metadata crosswalking in the Fall 2019 semester for SI 666 (Organization of Information Resources). A metadata project in this class inspired the creation of this outline.

Thanks are also due to Professors Anthony Whyte and Mark Newman, as well as their superb instructional teams, for teaching Python in SI 506-507, also at the University of Michigan (Fall 2019 and Winter 2020).

Matthew Carruthers, Metadata Projects Librarian at Michigan, taught my classmates and me the basics of OpenRefine during a guest lecture in Winter 2020 for SI 667 (Foundations of Digital Curation).

Colby Lewis, Digital & Scholarly Initiatives Librarian at the University of Michigan Law School, encouraged me to think about the eventual possibility of using Python programming to conduct a more systematic analysis of recent entry-level job openings in a specific subfield of archives and libraries (i.e. scholarly communication).