What Technology Skills Do Developers Need? A Text Analysis of Job Listings in Library and Information Science (LIS) from Jobs.code4lib.org.

Monica Maceli

ABSTRACT

Technology plays an indisputably vital role in library and information science (LIS) work; this rapidly moving landscape can create challenges for practitioners and educators seeking to keep pace with such change. In pursuit of building our understanding of currently sought technology competencies in developer-oriented positions within LIS, this paper reports the results of a text analysis of a large collection of job listings culled from the Code4lib jobs website. Beginning more than a decade ago as a popular mailing list covering the intersection of technology and library work, the Code4lib organization's current offerings include a website that collects and organizes LIS-related technology job listings. The results of the text analysis of this dataset suggest the currently vital technology skills and concepts that existing and aspiring practitioners may target in their continuing education as developers.

INTRODUCTION

For those seeking employment in a technology-intensive position within library and information science (LIS), the number and variation of technology skills required can be daunting. The need to understand common technology job requirements is relevant to current students positioning themselves to begin a career within LIS, those currently in the field that wish to enhance their technology skills, and LIS educators. The aim of this short paper is to highlight the skills and combinations of skills currently sought by LIS employers in North America through textual analysis of job listings. Previous research in this area explored job listings through various perspectives, from categorizing titles to interviewing employers;[1,2] the approach taken in this study contributes a new perspective to this ongoing and highly necessary work. This research report seeks a further understanding of the following research questions:

- What are the most common job titles and skills sought in technology-focused LIS positions?
- What technology skills are sought in combination?
- What implications do these findings have for aspiring and current LIS practitioners interested in developer positions?

As detailed in the following research method section, this study addresses these questions

Monica Maceli (mmaceli@pratt.edu) is Assistant Professor, School of Information and Library Science, Pratt Institute, New York.

through textual analysis of relevant job listings from a novel dataset—the job listings from the Code4lib jobs website (http://jobs.code4lib.org/). Code4lib began more than a decade ago as an electronic discussion list for topics around the intersection of libraries and technology.[3] Over time, the Code4lib organization expanded to an annual conference in the United States, the Code4Lib Journal, and most relevant to this work, an associated jobs website that highlights jobs culled from both the discussion list and other job-related sources. Figure 1 illustrates the home page of the Code4lib jobs website; the page presents job listings and associated tags, with the tags facilitating navigation and viewing of other related positions. Users may also view positions geographically or by employer.


Figure 1. Homepage of the code4lib Jobs Website, Displaying Most-Recently Posted Jobs and the Associated Tags.[4]

In addition to the visible user interface for job exploration, the website consists of software to gather the job listings from a variety of sources. The website incorporates jobs posted to the Code4lib discussion list, American Library Association, Canadian Library Association, Australian Library and Information Association, HigherEd Jobs, Digital Koans, Idealist, and ArchivesGig. This broad incoming set of jobs provides a wide look into new technology-related postings.

New job listings are automatically added to a queue to be assessed and tagged by human curators before posting. This allows manual intervention where a curator assesses whether the job is relevant to technology in the library domain and to validate the job listing information and metadata (see figure 2). Curating is done on a volunteer basis, and curators are asked to assess whether the position is relevant to the Code4lib community, if it is unique, and to ensure that it has an associated employer, set of tags, and descriptive text. Combining both software processes

and human intervention in the job assessment results in the ability to gather a large number of jobs of high relevance to the Code4lib community. As mentioned earlier, Code4lib's origins are in the area of software development and design as applied in LIS contexts. These foci mean that most jobs identified as relevant for inclusion in the Code4lib jobs dataset are oriented toward developer activities. The Code4lib jobs website therefore provides a useful and novel dataset within which to understand current employment opportunities relating to the intersection between technology—particularly developer work—and the LIS field.
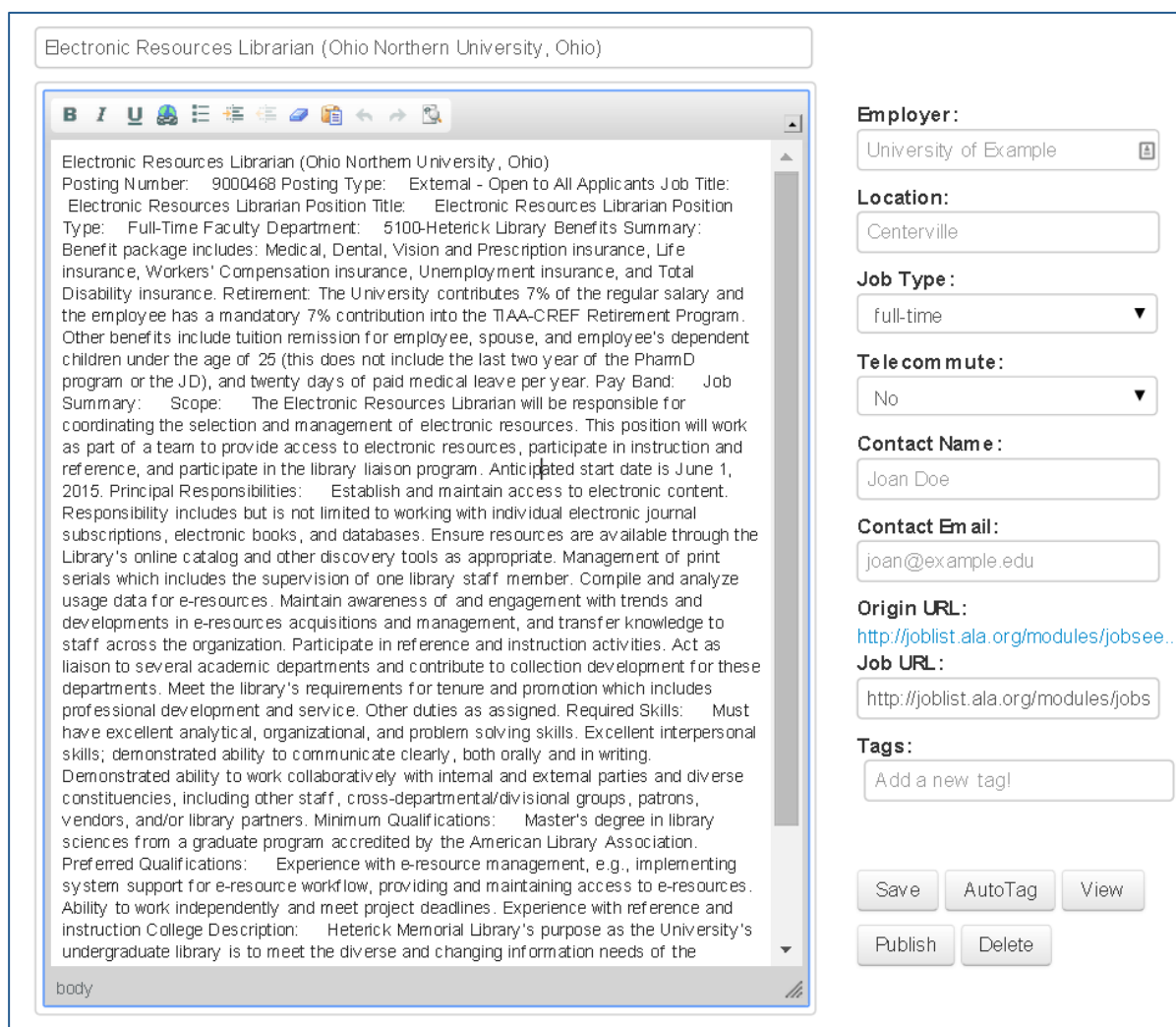
Figure 2. Code4lib Job Curators Interface Where Job Data is Validated and Tags Assigned.[5]

RESEARCH METHOD

To analyze the job listing data in greater depth, a textual analysis was conducted using the R statistical package, exploring job titles and descriptions.[6] First, the job listing data from the most recent complete year (2014) were dumped from the database backend of the Code4lib jobs website; this dataset contained 1,135 positions in total. The dataset included the job titles, descriptions, location and employer information, as well as tags associated with the various
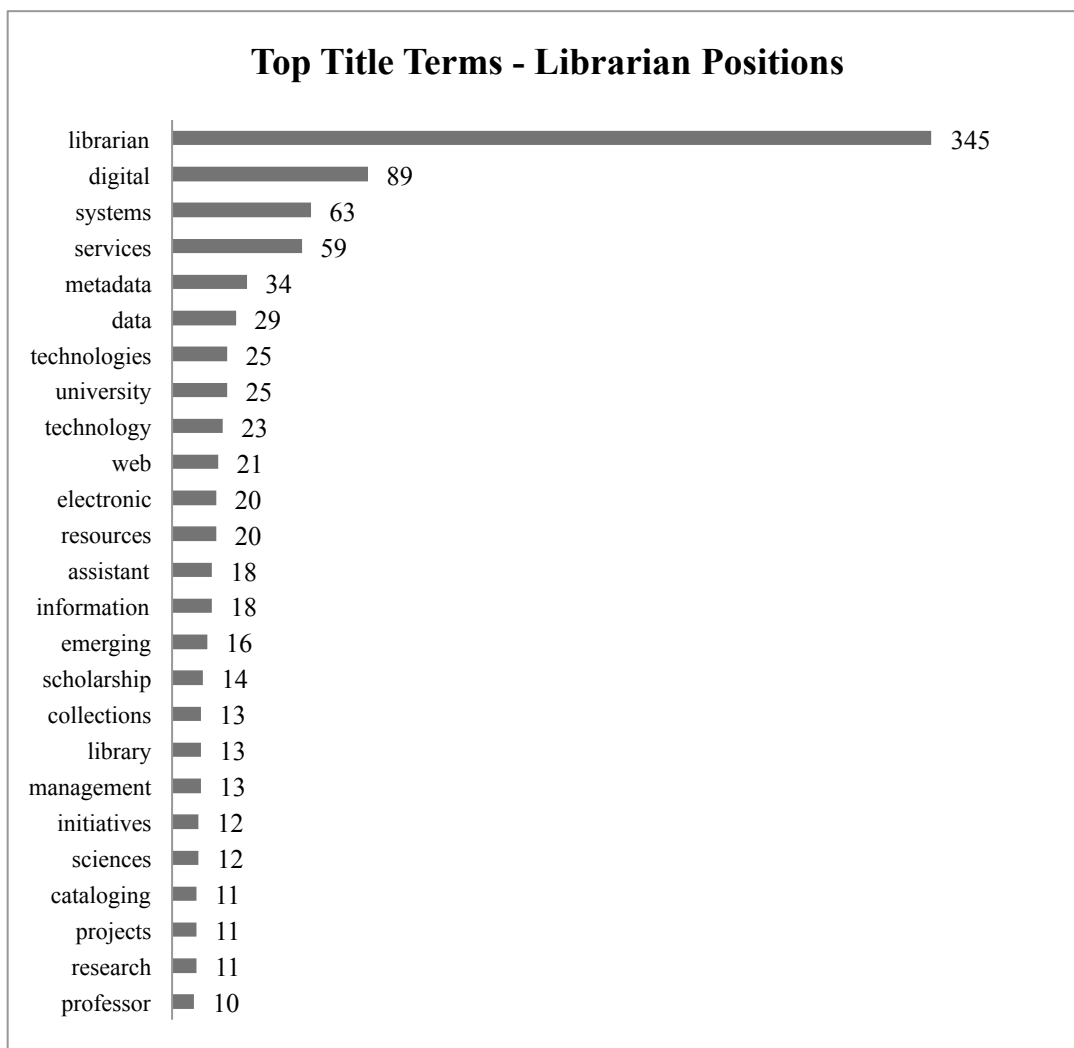
positions. The text was then cleaned to remove any markup tags or special characters that remained from the scraping of listings. Finally, the tm (text mining) package in R was used to calculate frequency, correlation of terms, generate plots, and cluster terms across both job titles and descriptions.[7]

RESULTS

Job Title Analysis

Of the full set of 1,135 positions, 30 percent were titled as a librarian position; popular specialties included systems librarian and various digital collections and curation-oriented librarian titles. Figures 3 and 4 detail the most common terms used in position titles across librarian and nonlibrarian positions.

Top Title Terms - Librarian Positions

Term	Count
librarian	345
digital	89
systems	63
services	59
metadata	34
data	29
technologies	25
university	25
technology	23
web	21
electronic	20
resources	20
assistant	18
information	18
emerging	16
scholarship	14
collections	13
library	13
management	13
initiatives	12
sciences	12
cataloging	11
projects	11
research	11
professor	10

Figure 3. Most Common Terms Used in Librarian Position Titles.

Top Title Terms - Non-Librarian Positions

Term	Value
digital	182
developer	141
library	116
manager	90
specialist	86
software	68
web	65
archivist	59
services	59
technology	59
engineer	55
director	52
data	49
systems	49
analyst	40
coordinator	40
information	40
senior	40
metadata	38
administrator	35
lead	34
project	34
head	33
programmer	32
research	24
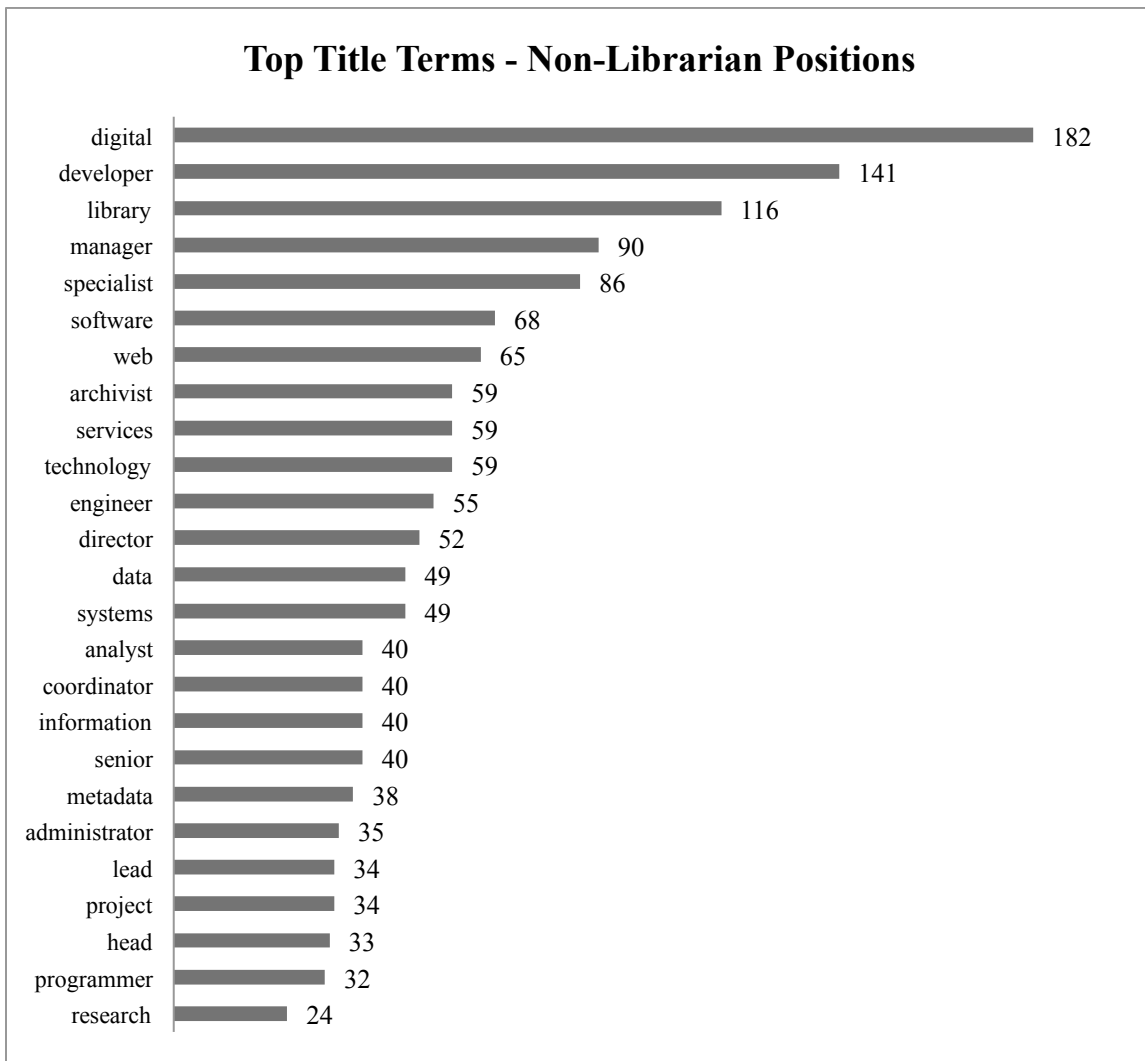
Figure 4. Most Common Terms Used in Nonlibrarian Position Titles.

The most popular job title terms were then clustered using Ward's agglomerative hierarchical method (dendogram in figure 5). Agglomerative hierarchical clustering, of which Ward's method is widely used, begins first with single-item clusters, then identifies and joins similar clusters until the final stage in which one larger cluster is formed. Commonly used in text analysis, this allows the investigator to explore datasets in which the number of clusters is not known before the analysis. The dendograms generated (e.g., figure 5) allow for visual identification and interpretation of closely related terms representing various common positions, e.g., digital librarian, software engineer, collections management, etc. Given that job titles in listings may include extraneous or infrequent words, such as the organization name, the cluster analysis can provide an additional view into common job titles across the full dataset in a more generalized fashion.
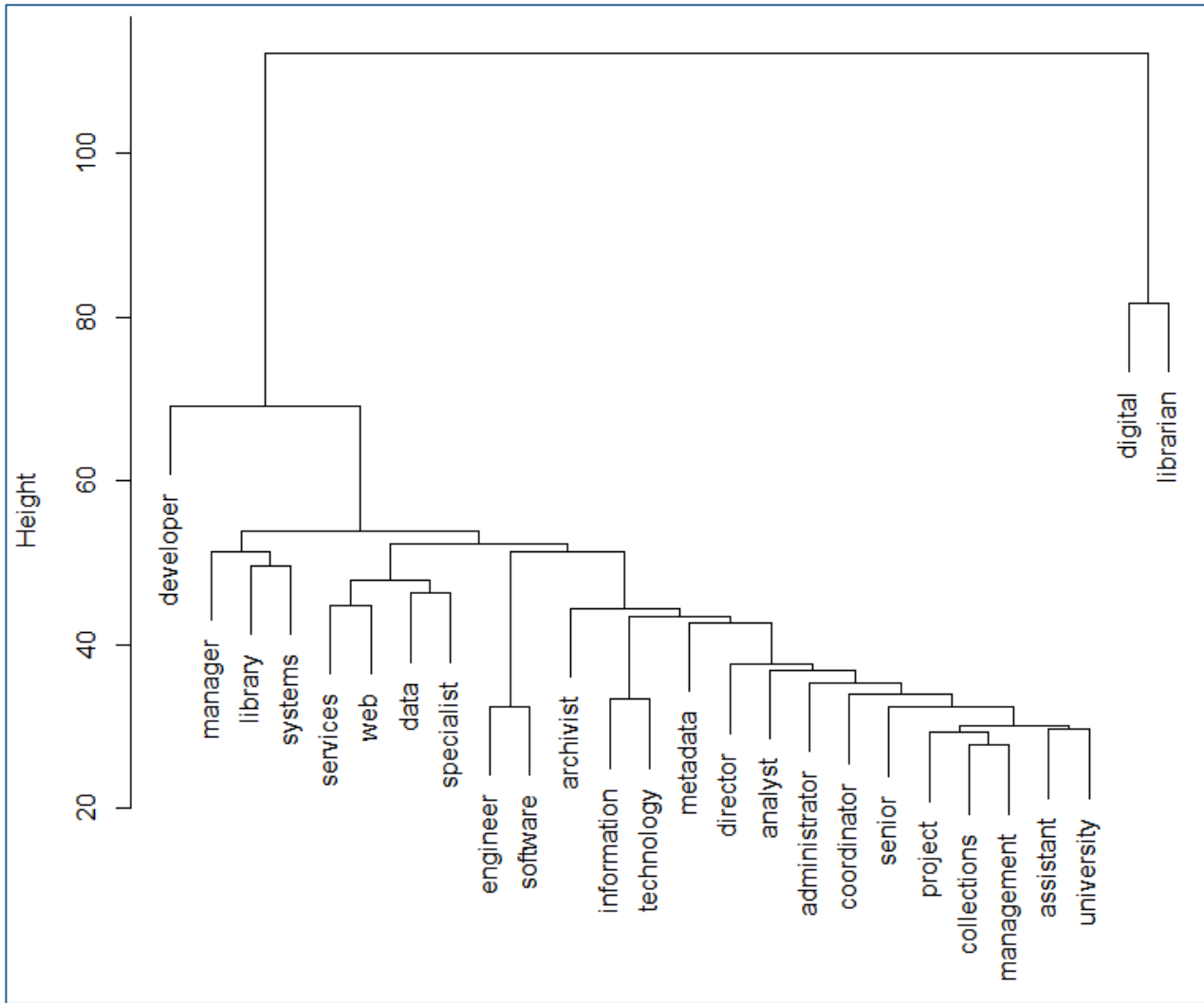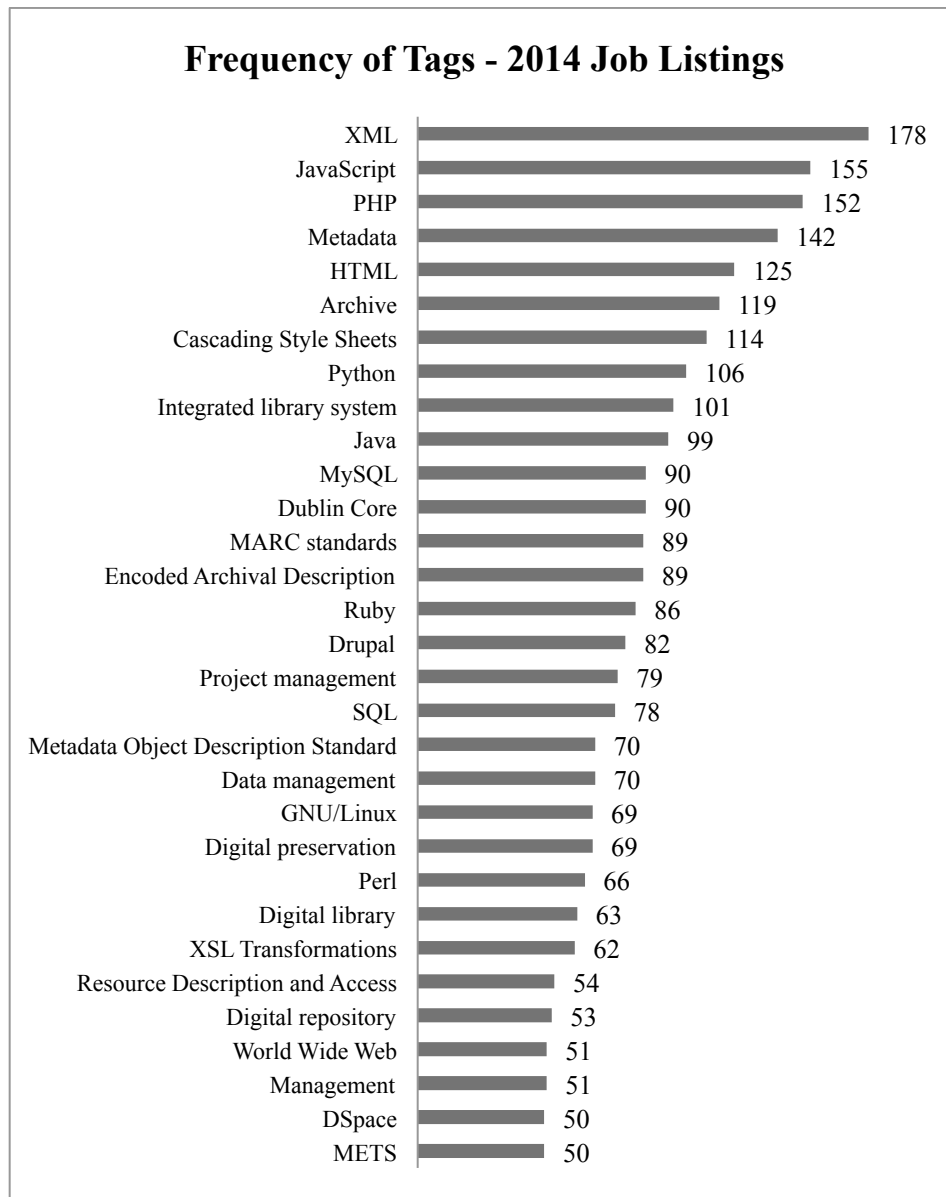
Figure 5. Cluster Dendrogram of Terms Used in Job Titles Generated Using Ward's Agglomerative Hierarchical Method.

Tag Analysis

As described earlier, the Code4lib jobs website allows curators to validate and tag jobs before listing. The word cloud in figure 6 displays the most common tags associated with positions, with XML being the most popular tag (178 occurrences). Figure 7 contains the raw frequency counts of common tags observed.

Figure 6. Word Cloud of Most Frequent Tags Associated with Job Listings by Curators.

Frequency of Tags - 2014 Job Listings

Tag	Frequency
XML	178
JavaScript	155
PHP	152
Metadata	142
HTML	125
Archive	119
Cascading Style Sheets	114
Python	106
Integrated library system	101
Java	99
MySQL	90
Dublin Core	90
MARC standards	89
Encoded Archival Description	89
Ruby	86
Drupal	82
Project management	79
SQL	78
Metadata Object Description Standard	70
Data management	70
GNU/Linux	69
Digital preservation	69
Perl	66
Digital library	63
XSL Transformations	62
Resource Description and Access	54
Digital repository	53
World Wide Web	51
Management	51
DSpace	50
METS	50

**Figure 7.** Frequency of Commonly Occurring Tags (frequency of fifty occurrences or more) in the 2014 Job Listings.

Job Description Analysis

The job description text was then analyzed to explore commonly co-occurring technology-related terms, focusing on frequent skills required by employers. Figures 8, 9, and 10 plot term correlations and interconnectedness. Terms with correlation coefficients of 0.3 or higher were chosen for plotting; this common threshold chosen broadly included terms with a range in positive relationship strength from moderate to strong.

Plots were created to express correlations around the top five terms identified from the tags: XML, Javascript, PHP, metadata, and HTML (frequencies in figure 7). Any number of terms and

frequencies can be plotted from such a dataset; to orient the findings closely around the job listing text, a focus on the top terms was chosen. These plots illustrate the broader set of skills related to these vital competencies represented in the job listings.
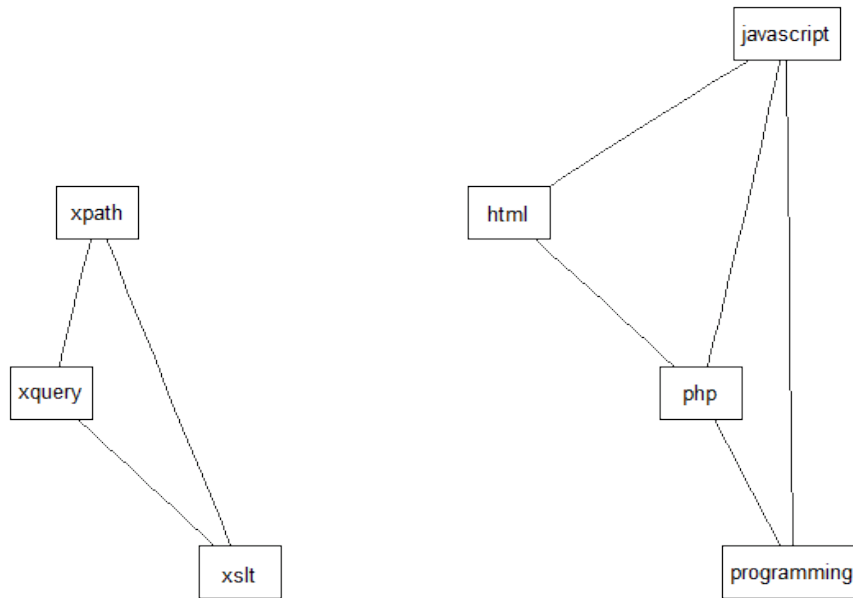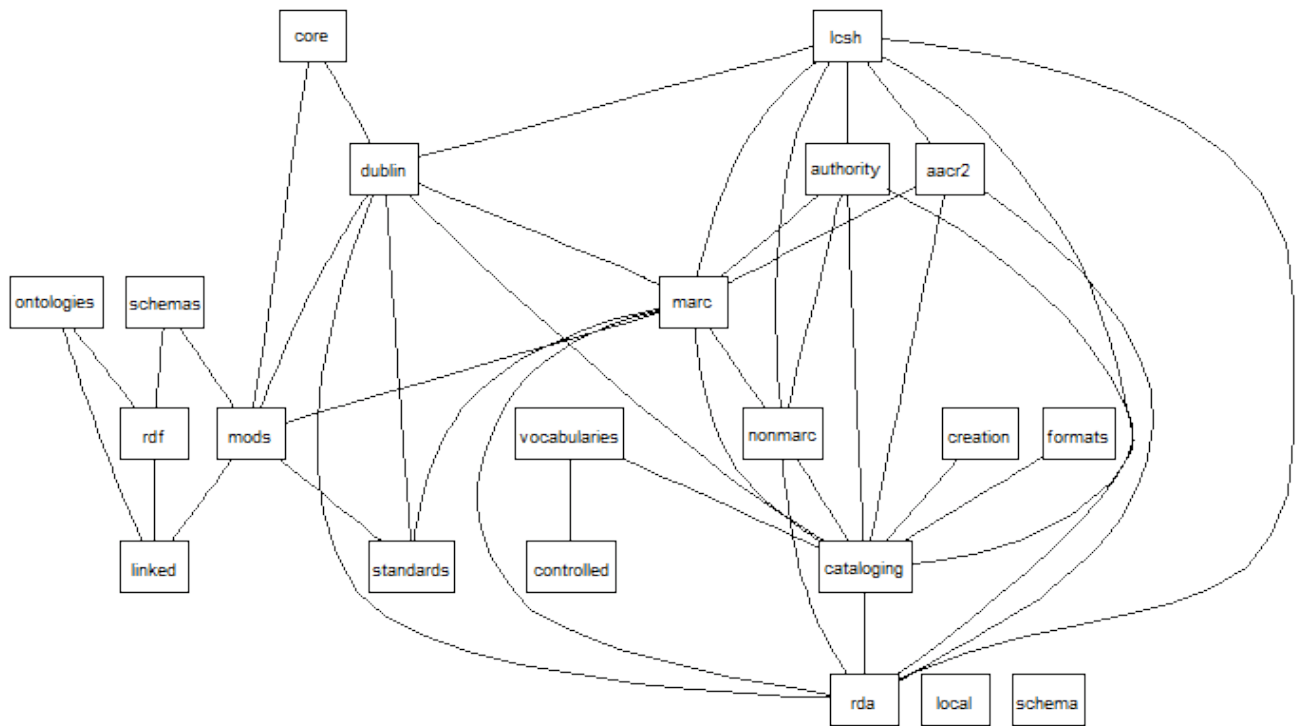

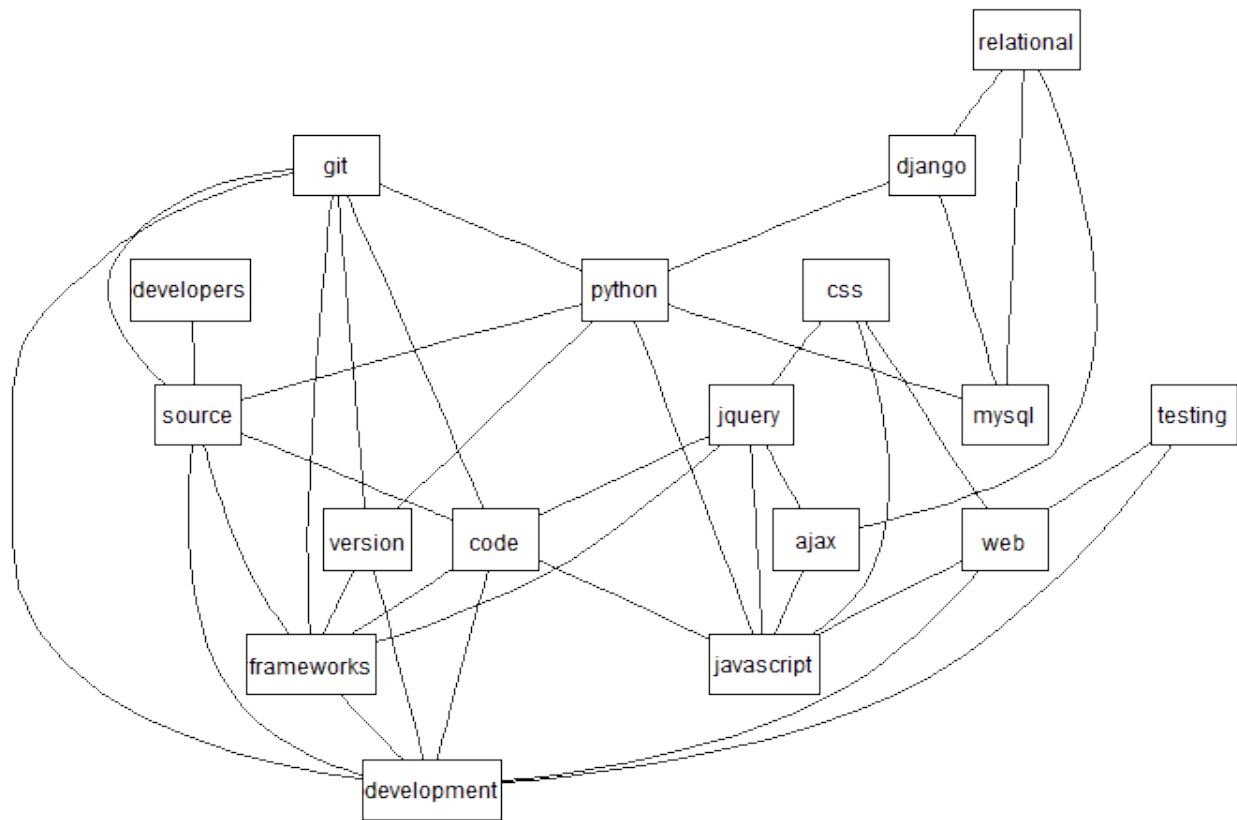**Figure 8.** Job Listing Terms Correlated with "XML" (most popular tag).
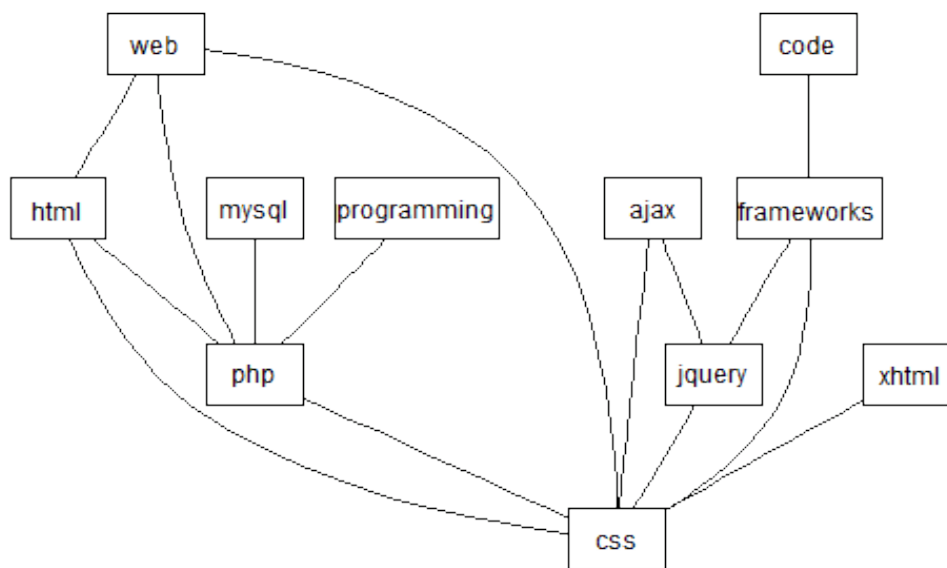
Figure 9. Job Listing Terms Correlated with "Javascript" (Second Most Popular Tag), including "PHP" and "HTML" (third and fifth most popular tags, respectively).
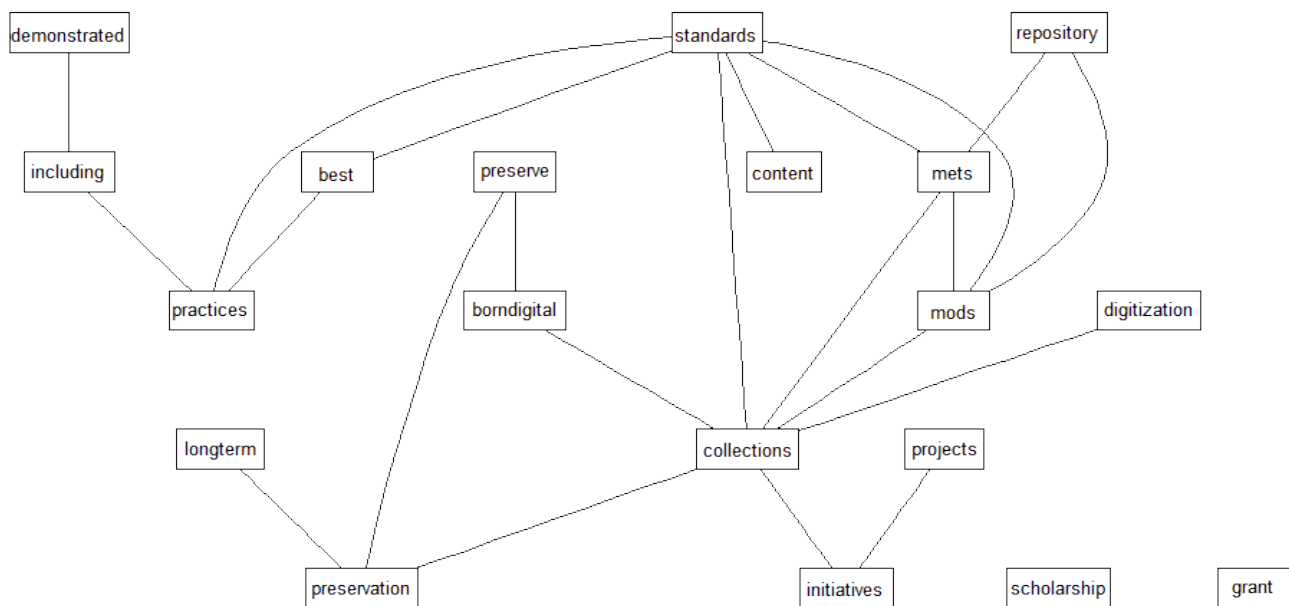
Figure 10. Job Listing Terms Correlated with "Metadata" (fourth most popular tag).

Finally, a series of general plots was created to visualize the broad set of skills necessary in fulfilling the positions of interest to the Code4lib community. As detailed in the title analysis (figures 3 and 4), apart from the generic term *librarian*, the two most common terms across all job titles were *digital* and *developer*. Correlation plots were created to detail the specific skills and requirements commonly sought in positions using such terms. Figure 11 illustrates the terms correlated with the general term of *developer*, while figure 12 displays terms correlated with *digital*. The implications of these findings will be discussed further in the following discussion section.

Figure 11. Job Listing Terms Correlated with "Developer."

Figure 12. Job Listing Terms Correlated with "Ddigital."

DISCUSSION

Taken as a whole, the job listing dataset covered a quite dramatic range of positions, from highly technical (e.g., senior-level software engineer or web developer) to managerial and leadership roles (e.g., director or department head roles centered on digital services or emerging technologies). These findings support the suggestions of earlier research,[8] which advocated for LIS graduate programs to build their offerings not just in technology skills but also in technology management and decision-making. However, the Code4lib jobs dataset is a one-dimensional view into the employment process and is focused largely on the developer perspective. Additional contextual information, including whether suitable candidates were easily identified and if the position was successfully filled, would provide a more complete view of the employment process. Prior research has indicated that many technology-related positions in LIS are in fact difficult to fill with LIS graduates.[9] While LIS graduate programs have made great strides in increasing the number of courses and topics covered that address technology, these improvements may not benefit those already in the field or wishing to shift towards a more technology-focused position.

In the common tags and terms analysis, experience with specific LIS applications was relatively infrequently required, with the Drupal content management system a notable exception. More generalizable programming languages or concepts, e.g., Python, relational databases, XML, etc., were favored As with technology positions outside of the LIS domain, employers likely seek those with the ability to flexibly apply their skills across various tools and platforms. This may also relate to the above challenges in filling such positions with LIS graduates, with the goal of opening up the position to a larger technologist applicant base.

Common web technologies popular in the open-source software often favored by LIS organizations continued to dominate, with a clear preference for candidates well versed in HTML, CSS, JavaScript, and PHP. Relating to these skills, web development and design practices were often intertwined with positions requesting both developer-oriented skillsets as well as interface design (e.g., figure 7). Technologies supporting modern web application development and workflow management were evident as well, e.g., common requirements for experience with versioning systems such as Git, popular JavaScript libraries, and development frameworks. Also striking was the richness of the terms correlated with metadata (figure 10), including mention of growing areas of expertise, such as linked data.

Interestingly, the general correlation plots expressing the common terms sought around "digital" and "developer" positions were quite varied. While the developer plot (figure 11 above) provided a richly technical view into common technologies broadly applied in web and software development, the terms correlated around digital were notably less technical (figure 12 above). While there was a clear focus on digital preservation activities and common standards in this area, mention of terms such as "grant" indicated that these positions likely have a broad role. The term digital was frequently observed in librarian job titles, so these roles may be tasked with both technical and administrative work.

Finally, there are inherent difficulties in capturing all jobs relating to technology use in the LIS domain that introduce limitations into this study. While the incoming job feeds attempt to broadly capture recent job posts, it is possible that jobs are missed or overlooked by the job curators. Given the lack of one centralized job-posting source regardless of the field, this is a common challenge to research work attempting to assess every job posting. And as mentioned above, there is also a lack of corresponding data as to whether these jobs are successfully filled and what candidate backgrounds are ultimately chosen (i.e., from within or outside of LIS).

CONCLUSION

This assessment of the in-demand technology skills provides students, educators, and information professionals with useful direction in pursuing technology education or strengthening their existing skills. There are myriad technology skills, tools, and concepts in today's information environments. Reorienting the pursuit of knowledge in this area around current employer requirements can be useful in professional development, new course creation, and course revision. The constellations of correlated skills presented above (figures 8–12) and popular job tags (figure 7) describe key areas of technology competencies in the diverse areas of expertise presently needed, from web design and development to metadata and digital collection management. In addition to the results presented in this paper, the Code4lib job website provides a continuously current view into recent jobs and related tags; this data can help those in the LIS field orient professional and curricular development toward real employer needs.

ACKNOWLEDGEMENTS

REFERENCES

1. Janie M. Mathews and Harold Pardue, "The Presence of IT Skill Sets in Librarian Position Announcements," *College & Research Libraries* 70, no. 3 (2009): 250–57, http://dx.doi.org/10.5860/crl.70.3.250.

2. Vandana Singh and Bharat Mehra, "Strengths and Weaknesses of the Information Technology Curriculum in Library and Information Science Graduate Programs," *Journal of Librarianship & Information Science* 45, no. 3 (2013): 219–31, http://dx.doi.org/10.1177/0961000612448206.

3. "About"" Code4lib, accessed January 6, 2014, http://jobs.code4lib.org/about/.

4. "code4lib jobs: all jobs," Code4lib Jobs, accessed January 12, 2015, http://jobs.code4lib.org/.

5. "code4lib jobs: Curate," Code4lib Jobs, accessed January 17, 2015, http://jobs.code4lib.org/curate/.

6. R Core Team, R: The R Project for Statistical Computing, 2014, http://www.R-project.org/.

7. Ingo Feinerer and Kurt Hornik, "tm: Text Mining Package," 2014, http://CRAN.R-project.org/package=tm.

8. Meredith G. Farkas, "Training Librarians for the Future: Integrating Technology into LIS Education," in *Information Tomorrow: Reflections on Technology and the Future of Public & Academic Libraries*, edited by Rachel Singer Gordon, 193–201 (Medford, NJ: Information Today, 2007).

9. Mathews and Pardue, "The Presence of IT Skill Sets in Librarian Position Announcements."