The New Library User: Machine Learning

I became a librarian in 1992. Mosaic, the first popular Graphical User Interface to the World Wide Web, was introduced a year later. It sometimes feels as though my entire career has been an effort to understand and deal with the impact of the web on my profession and on the information lives of the faculty and students I serve. There have been other significant technological developments since then, of course: the dominance of search engines such as Google and the concomitant development of discovery services; the move to mobile and wireless; the migration of the journal literature from print to digital; and users’ shifting expectations of service as a result of online shopping and social media. But none of these have felt as paradigm-shifting as the transition from the print and internet library of the 1980s to the web-based, web-infused library of today. I used to have to explain to library users that a URL is like a library call number. Now I have to explain that a call number is like a URL.

As I find myself in the “final quarter” of my career, as described by Theresa Rowe in a 2018 Viewpoints column, librarians are on the cusp of a change that will be at least as significant as the move to the web. If, during this final quarter, I am able to focus on the kind of “intentional” change that Rowe advocates in her column, I hope it be this: to help my colleagues in the academic libraries and the institutions they serve—places that have given me such a rewarding and fulfilling career—prepare to understand and deal with the impact of the rise of big data, and machine learning, and artificial intelligence. Increasingly in our daily lives, with services such as Google Maps and Google Translate, we find ourselves aided by or collaborating with systems imbued with artificial intelligence and machine learning. These kinds of collaborations are occurring in librarianship as well. For instance, using statistical analytics generated from data gathered from the full range of system users, Ex Libris’s Data Analysis Recommendation Assistant (DARA) recommends specific process improvements to library customers.

The impact of big data, machine learning, and artificial intelligence on libraries falls into three buckets: assisting users (both machine and human); making collections accessible; and preserving data sets and the products of research.

Assisting Users (Both Machine and Human)
Librarians have always envisaged the human user of their services. Three of S. R. Ranganathan’s “five laws” of library science explicitly mention the reader, but he clearly meant the human reader, the only reader available to him and the Indian librarians he sought to educate and train in the 1930s. Today we need to welcome another set of users into the library. These machine learning, algorithmic, analytic users will be collaborating with human users, crunching and filtering the data and presenting the information needed by the human users. Human users will also be seeking access to the rich data that enables them to train algorithms and to conduct research using these sophisticated statistical techniques. Our librarians and staff who work directly with faculty and students in the classroom and beyond need to be prepared to help users find the data sets—that is, the training data—they are looking for. The library online platforms need to be designed so that machine users can gain unmediated access, where appropriate, to the data resources they and their human collaborators seek.

Perhaps even more significant for librarians’ direct service with users will be the impact of artificial intelligence on users’ expectations. Over the past decades, consumers’ experiences with online shopping, search engines, and e-books changed their expectations of library services. Libraries responded with faster acquisitions, speedier interlibrary loan, single search-box discovery services, and one-click access to full text. Another shift in expectations is ahead. Students may wonder why they can’t just ask Alexa or Siri to select and retrieve what they need, or they may balk at being asked to do basic evaluation and selection work that in other realms of their information lives is being handled by their intelligent digital assistants.
Making Collections Accessible
As librarians license access to content from vendors, we need to ensure that contracts do not preclude our users from conducting text and data-mining research, algorithmically based research, and machine learning. At the moment, many vendors write contracts that assume large-scale automated crawling and other techniques are an a priori misuse of their services. Concerned with reaping value from data science and machine learning techniques, they seek to control these rights. On behalf of our users, librarians need to press vendors for access to platforms and data within controlled environments; ideally, however, the content and platforms we license should be computationally accessible.

As we continue to build our own digital libraries, we also need to envision the machine user alongside the human user and to consider our own digital libraries to be “Collections as Data.” Like the commercial providers of information resources, academic librarians and other information professionals should think carefully about how openly accessible they want to make, or are able to make, these resources. This is important both ethically and legally. They also need to determine how comfortable they are with providing the raw materials from which others will reap value.

Another challenge will be the impact of machine learning on cataloging and description. Librarians have realized huge efficiencies by moving the description process, and collaborating, online. Examples of this are the shared cataloging within OCLC’s WorldCat and shelf-ready acquisitions from companies such as GOBI Library Solutions. However, behind all this online collaboration are human catalogers. Especially in archival description, we are beginning to see interesting efforts to automate portions of this descriptive work. As these efforts develop, the role of catalogers and processing archivists will continue to change, in terms of both the expertise we require of technical services staff and the work they do.

Preserving Data Sets and the Products of Research
All this access to collections as data and informed assistance from librarians able to work equally with machine and human users will result in additional scholarly production. Collections as data quickly come full circle to data as collections. We now have decades of experience in determining what from our digital lives and work should be preserved and made accessible for the long term. The same issues will need to be addressed for the products of computational use and analysis of these large data sets. Should we sustain both technical and intellectual access over the long term, and if so, how? How do we address the ethical and legal rights not only of the users of these data sets and the creators of scholarship but also of those represented in the data sets? How do we support the replication of studies and the auditing of data sets, especially those used as training data in machine learning systems, for bias?

Are You Ready?
In the migration of libraries to the web, much of the early work was done by research universities. Today much of the engagement with data science, machine learning, and artificial intelligence is also happening at those institutions. However, libraries at smaller institutions and those more focused on teaching cannot avoid this looming structural change both in the profession and in the information experiences of library users. Vendors are already incorporating artificial intelligence and machine learning into their platforms, services, and products. Librarians must become informed customers and users of those platforms, services, and products. Perhaps most importantly, librarians need to prepare college and university graduates to be informed citizens and to develop fulfilling and useful professional lives in a world infused with big data, machine learning, and artificial intelligence.

Notes
2. I also hope to use my privileged position to focus on diversifying my profession, but that is a topic for another column. However, the two issues are related. We must ensure that incorporating artificial intelligence and machine learning into librarianship does not reinforce existing bias and privilege.
5. For example, the National Endowment for the Humanities awarded a grant (PAA-256249-17) to Carnegie Mellon University to develop image-identification tools and techniques to expedite description and improve access to the Charles “Tenie” Harris Archive of African American Life in Pittsburgh, a large photographic collection.
6. For example, Stanford University and the National Library of Norway have co-sponsored the first two Fantastic Futures conferences.

Jonathan Miller (jm30@williams.edu) is Director of Libraries at Williams College.

© 2020 Jonathan Miller. The text of this article is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License.