How Data Librarians Support their Whole Institution

*Providing Cross-Campus and Cross-Disciplinary support
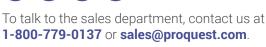at the University of Toronto.*

Kara Handren is a Data Librarian at the University of Toronto. Handren provides support for data discovery and analysis across the University's St. George Campus.

About the University

Founded in 1827, The University of Toronto is one of the top research-intensive universities in the world. More than 620,000 graduates advance U of T's impact on communities across the globe. Bringing together thought leaders from various backgrounds and disciplines to collaborate on the world's most critical challenges, the University of Toronto has transformed how communities think, work, and conduct relationships with one another for more than a century.

For over a decade, researchers at U of T have been involved in text mining and collaborating across disciplines. Currently, the university is experiencing a significant growth in the number of these collaborative projects, along with text analysis techniques being applied by students in traditionally data light fields. These text mining shifts have driven a need to build additional supports, as an increase in demand has developed. Internally, text analysis is being used by faculty to update information and fine tune data for undiscovered links between ongoing medical research projects that facilitate future funding requests. Some libraries on campus are also using text mining to better inform their collection development policies moving forward.

The Map and Data Library (MDL) at U of T acquires and maintains data collections, and assists its users with finding data and maps, ranging from historical collections to more recent government-conducted surveys. The Map & Data Library also offers software support for tools used in data cleaning and analysis, guidance on statistical analysis techniques, help with research data management, advice on best practices and software for data visualization, and assistance with geographic information systems (GIS) tools. This support is offered to staff, faculty, students, and members of the public in the form of workshops, online tutorials, and on demand in-person reference and consultation services.

Identifying the Problem

Kara Handren, Data Librarian, at the University of Toronto shares her experiences and contribution to these growing text analysis needs. Handren provides support for data discovery and analysis across the University's St. George Campus.

Handren explained that as demand grew for text and data mining licensed collections, major difficulties and hurdles remained when it came to getting licenses cleared from each publisher. These difficulties were frequently encountered by library staff over the course of consultations with patrons. In some cases, researchers had the tools they needed to conduct their research but needed guidance about how to access content and gain the appropriate rights for text and data mining. In other cases, instruction on those tools, such as programming languages or text analysis techniques, was also required. Handren shared that while the library was able to provide this foundational support, assist patrons with questions about appropriate types and analysis for their research questions, and navigate licensing restrictions, accessing content was a drawn-out process that was handled on a case-by-case basis. "For example, a researcher may have secured funding and a well-thought-out research process, including a trained model, but would be held up for months due to the time required to negotiate and purchase access to a particular corpus. These roadblocks were particularly visible in the case of newspaper articles, which were frequently requested," says Handren.

"patrons seeking to mine content across a large breadth of newspaper articles... can be directed to TDM Studio without requiring any additional licensing negotiations"

How TDM Studio Solved the Problem

In 2019, the Mills Harris Library at the Rotman School of Management at U of T was approached by a Faculty member in Economics. This faculty member was interested in analyzing word frequencies in housing ads posted in the New York Times Sunday edition over a twenty-year period. The director of Milt Harris reached out to the wider U of T library community for possible databases of this newspaper with cleared text mining rights, and a colleague who works closely with ProQuest, part of Clarivate, was able to flag TDM Studio as a tool of potential interest.

TDM Studio is a text and data mining solution for research across disciplines. It offers exclusive access to rights-cleared content from ProQuest subscriptions, including the New York Times, one of 3,000 other news titles, and millions of journal articles, primary sources, and dissertations and theses.

> "After some initial conversations, U of T began a pilot of this tool in December 2019, to better evaluate its capabilities for future use. In May of 2020, the director of Milt Harris was invited to be a development partner on the advisory board of TDM Studio. This successful pilot, alongside the work on the advisory board, led to U of T signing a 1-year trial contract in November 2020, and a full three year contract in May of 2021", says Handren.

> Handren explained that as a result of this license, patrons seeking to mine content across a large breadth of newspaper articles, from a wide geographic range, can be directed to TDM Studio without requiring any additional licensing negotiations. There is no longer a need to wait for clearance and the Map & Data Library, in partnership with Milt Harris and other libraries on campus, is able to focus on researcher support and fill in any skill gaps, such as learning R or Python or choosing appropriate text analysis techniques.

By using TDM Studio, researchers save time on analysis by quickly creating datasets and visuals of content relationships across ProQuest resources. Now being used by the whole U of T campus, TDM Studio supports data literacy and research across disciplines for researchers with or without coding skills. It not only enhances research but can also elevate teaching and learning across campus with solutions designed to support varied skill levels in research approaches and methodologies.

Benefits Across Different Users

At the University of Toronto, the development environment in TDM Studio, known as a Workbench, has been used successfully in many projects. Many projects are implemented by expert researchers that have the skills to work in the tool independently. Some with beginner level experience often reach out to the Map & Data Library to support them. A few examples are listed below.

- An undergraduate student used the tool to analyze newspaper articles focusing on robotics for theme and emotions over a twenty-year period, and then used the analysis derived from TDM Studio to compare her findings to related data in public opinion polls.

- A PhD student used the tool to begin to examine how gang activity has been represented and reported across regional newspapers in Central America through successive changes in government.

- And a librarian at a specialized library on campus has used the tool to analyze subject metadata attached to acquisitions and chart these over time, to see how their collections development policies have been keeping up with scholarly publications.

The TDM Studio Workbench allows access to the full text as well as all metadata, so that code can be written against it. A virtual machine with ample storage and high computing capacity is leveraged through a Jupyter Notebook. For those that code in Python and R, the Workbench provides the flexibility for any workflow and includes sample scripts to help researchers get started. For those without coding skills there are three Visualizations that easily run some of the most popular analyses: topic modeling, geographic analysis, and sentiment analysis. By interacting with these Visualizations, a researcher who does not know how to code can uncover trends and patterns across mass quantities of information they could not have detected by just reading a few articles alone.

"TDM Studio removed barriers and encourages researchers in non-traditional areas to explore text mining avenues within their fields of research."

Conclusion

The library at the University of Toronto sought out to become known as the support and guidance destination on text and data mining for the entire campus. U of T has developed workshops to support this tool and strengthened their Python support.

To date, the Map & Data Library has supported eight research projects in using TDM Studio. The tool has also been referenced in untold numbers of other consultations as a potential solution for research needs. There are currently 39 projects using the tool at the University of Toronto – note that the library is not involved in all cases, as more experienced researchers are often able to get started on their own.

As part of the University of Toronto Libraries, the Map & Data Library measures its success by the quality of service that it is able to provide to the University community. Ensuring access to collections and tools that are in high demand or of high benefit to that community is key to this success. Handren shared, "By providing programmatic access to highly valued collections licensed by the University, robust support services and continued maintenance and development, TDM Studio removed barriers and encourages researchers in non-traditional areas to explore text mining avenues within their fields of research. As a result, the Map & Data Library, in partnership with Milt Harris and with other libraries and partners on campus, can better support these researchers moving forward."

About TDM Studio

ProQuest's workflow solution for text and data mining is designed for research, teaching and learning. TDM Studio provides access to sought-after content including current and historical newspapers, primary sources, scholarly journals, and dissertations and theses. It empowers researchers, students and faculty to analyze documents by uncovering connections and patterns that lead to career-defining discoveries.

Learn more at www.proquest.com/go/tdm-studio
or contact your ProQuest representative today.

ProQuest®

Part of **Clarivate**