

Using Text Analysis to Inform Outreach Efforts

J. GORDON DAINES III and CORY L. NIMER

Brigham Young University

Abstract: This article discusses research at Brigham Young University on use of text analysis tools to identify underserved programs and courses that could benefit from primary source instruction by curators in the L. Tom Perry Special Collections. It shows that text analysis is a useful initial screening methodology. Researchers initially chose to analyze course learning outcomes because of their accessibility. These course learning outcomes describe “the knowledge, skills, and attitudes that learners should have after successfully completing a learning experience or program.” The analysis identified a number of courses and programs that could potentially benefit from primary source instruction offered by the Perry Special Collections. Some of the results were expected, such as the relevance of primary source instruction for the history department, and some were quite surprising, such as the unexpected indication that German language courses on campus might benefit from primary source instruction. A similar text analysis on the syllabi for the identified courses was performed and the results compared with the text analysis of the course learning outcomes. The comparison underscored the utility of text analysis as an initial screening methodology.

Introduction

Instruction using primary sources has many benefits in an academic setting. Teaching with primary sources fosters critical thinking, historical empathy, and cultural competency as well as research and writing skills. While faculty generally understand the value of teaching with primary sources, they don't always know how to find primary sources that will be useful in their specific teaching contexts. They also struggle to fit primary sources into their courses and to design assignments that leverage those primary sources. The need to connect faculty and primary sources provides an important opportunity for special collections professionals.

As experts in using primary sources, special collections professionals can help faculty meet their pedagogical goals. One of the challenges facing special collections professionals is identifying faculty who could benefit from their help—some faculty self-identify, and others don't. This paper describes a project begun in 2019 to identify academic programs and courses at Brigham Young University that have been underserved by the L. Tom Perry Special Collections (hereafter Perry Special Collections). The project used Voyant Tools¹, a text analysis program, to examine a corpus of course learning outcomes and syllabi at Brigham Young University. Searching for terms related to archives and special collections, the analysis sought to identify departments and their respective programs that teach primary source literacy skills. A particular aim was to identify those not engaged with the Perry Special Collections. This paper focuses on the methodology used to identify these departments and programs and describes the benefits of using text analysis as an initial screening tool when doing learning outcomes and syllabus mining. It uses results from a pilot study done at Brigham Young University to demonstrate the usefulness of the methodology. While text mining is a useful initial screening tool, special collections professionals still need to do the heavy lifting of reaching out to departments and programs to work on integrating their holdings into the curricula of those departments and programs.

Problem Statement

¹ Voyant Tools, "Voyant: See Through Your Text," <http://voyant-tools.org/> (accessed October 4, 2021).

Syllabus mining has long been considered an important method for identifying learning objectives that have connections with library instruction. Butler and Calcagno reviewed the literature on syllabus mining and found that “even though the results of a syllabus study are not generalizable, the methods can be consistent and provide valuable information to an information literacy program.”² Most of the studies identified used a manual review of syllabi looking for “any mention of library use.” The recent promotion of the ACRL standards and the Information Literacy Framework have facilitated the identification of more specific concepts and assignments that fall within information literacy, allowing for syllabi review to be more systematic and accurate in identifying courses that would benefit from information literacy.³ Text analysis as a technique for identifying instruction opportunities for librarians was explored at San Diego State University. Librarians there found that a text-mining analysis of syllabi “shed light on the information needs of their campus communities. It also revealed gaps where the library could intervene and provide support, especially in the area of research support.”⁴ Special collections professionals have not taken full advantage of such reviews as a tool for identifying potential outreach opportunities.

Outreach programs are “essential in fostering the use of library collections and services.”⁵ They are particularly important for special collections programs attempting to connect their resources with faculty and students. At the outset of the 21st century Tamar Chute noted that “very little has been written about the different ways that college and university archives reach their constituents, i.e., faculty, students, administration, and staff.”⁶ She describes several different approaches that college and university archivists have used to try to connect their constituents with materials in their holdings. These approaches included exhibits, newspaper articles, presentations to university groups, and email blasts. A decade later Valerie A. Harris and Ann C. Weller described how a “common type of special collections outreach in the academic setting is that of special collections librarians working with professors to highlight relevant collections within special collections and to instruct students how to find, access, and use primary sources most productively.”⁷ They did not discuss how special collections librarians and archivists had connected with faculty in the first place. One of the perennial challenges facing special collections professionals is identifying which courses might benefit from the use of special collections materials in an in-depth, sustained way. Harris and Weller pointed out that it is easy for archivists to default to the “show and tell” presentation and that if archivists and special collections librarians want to grow their instruction programs, that they need to move beyond the “traditional activities of outreach such as collection ‘show and tells’ and exhibits.”⁸

² Kathleen Butler and Theresa Calcagno, “Syllabus Mining for Information Literacy Instruction: A Scoping Review,” *Evidence Based Library & Information Practice* 15, no. 4 (2020): 101.

³ *Ibid.*, 100.

⁴ Keven M. Jeffrey et al., “Digging in the Mines: Mining Course Syllabi in Search of the Library,” *Evidence Based Library & Information Practice* 12, no. 1 (2017): 83.

⁵ Barbara Blummer and Jeffrey M. Kenton, “Academic Libraries’ Outreach Efforts: Identifying Themes in the Literature,” *Public Services Quarterly* 15, no. 3 (2019): 179.

⁶ Tamar G. Chute, “Selling the College and University Archives: Current Outreach Perspectives,” *Archival Issues* 25, no. 1/2 (2000): 34.

⁷ Valerie A. Harris and Ann C. Weller, “Use of Special Collections as an Opportunity for Outreach in the Academic Library,” *Journal of Library Administration* 52, no. 3-4 (2012): 295.

⁸ *Ibid.*, 302.

Moving beyond traditional outreach activities is an important, emergent theme over the last decade as special collections professionals focus more on instruction. Bryan Bance posits that academic archives need to align themselves “with an emerging engagement-based instructional paradigm” and that they need to insert themselves more “fully into the teaching function of the university.”⁹ This activist approach to outreach is necessary because faculty are not often talking to special collections professionals about sources. Doris Malkmus found that faculty “find out about online primary sources in informal and unsystematic ways—equally often through e-mail, professional publications, online browsing, and word of mouth.” Malkmus also found that rarely did librarians (or special collections professionals) play a significant role in helping faculty find primary sources.¹⁰

The literature makes clear the need to be more proactive in identifying courses and faculty that would benefit from the use of primary sources. Curators in the Perry Special Collections decided to try text mining as a way to identify faculty to approach about utilizing primary sources held by the department. They decided to look at learning outcomes to see if they could be text mined for information that would direct them to faculty across different programs that could be better served by Special Collections. Preliminary results from this stage led the research team to apply text mining to course syllabi to see if findings differed. This allowed us to compare the utility of these two different approaches to identifying faculty to approach about primary source instruction.

Methodology

In early 2019, the Special Collections Outreach Committee was tasked with identifying university departments and programs that would benefit from expanded use of primary sources. The committee created a project team comprised of the authors. The project team eventually grew to include two student research assistants. The project team was familiar with text analysis from a previous project examining the *American Archivist*¹¹ and wondered if text analysis could be used as an initial screening tool to identify underserved programs and faculty. Curators could then approach these programs and faculty about utilizing Special Collections materials as part of their curriculum.

The project team initially decided to analyze course learning outcomes aggregated by department because of their accessibility and the fact that the learning outcomes site had a complete list of learning outcomes for every course broken out by department. At Brigham Young University academic departments and programs are asked to formulate expected learning outcomes. Learning outcomes describe “the knowledge, skills, and attitudes that learners should have after successfully completing a learning experience or program.”¹²

The authors worked with two student research assistants to compile relevant learning outcomes from the university’s website and convert them into a useable textual format. The text was then analyzed using

⁹ Bryan Bance, “Outreach in the Academic Community: Enhancing the Teaching Role of University Archives,” MA thesis, University of Manitoba/University of Winnipeg, 2012, 3.

¹⁰ Doris Malkmus, ““Old Stuff” for New Teaching Methods: Outreach to History Faculty Teaching with Primary Sources,” *portal: Libraries and the Academy* 10, no. 4 (2010): 424.

¹¹ J. Gordon Daines III, Cory L. Nimer, and Jacob R. Lee, “Exploring the American Archivist: Corpus Analysis Tools and the Professional Literature,” *Journal of Contemporary Archival Studies* 5, Article 3 (2018), accessed September 28, 2021, <https://elischolar.library.yale.edu/jcas/vol5/iss1/3>

¹² Brigham Young University, “Expected Learning Outcomes,” accessed September 28, 2021, <https://learningoutcomes.byu.edu/>

Voyant Tools to identify programs and courses using language in their learning outcomes that indicated that they might benefit from use of Special Collections materials.

A list of key terms related to libraries, archives, and special collections was developed to aid in the analysis. The initial list was based on the terms used by Jeffrey et al.¹³, and was comprised of the following: archiv*, librar*, reference, interlibrary, database, eBook, article, journal*, cit*, bibliography, research*, source, “primary source*”, scholarly, history, genealogy, book, APA, MLA*, Chicago, Turabian, “special collections,” and catalog. The results of text analysis using the initial list identified a number of STEM programs that were involved with research and library usage. However, while STEM courses can be research heavy, the type of research that they do does not typically involve the use of Special Collections resources. The initial term list was subsequently revised to be more specific to archives and special collections as it became clear the initial list was not providing the desired results. The adjusted list focused on the terms archiv*, librar*, and research*.

Preliminary analysis of course learning outcomes led to the expected result that many academic programs and faculty were already partnering with the Perry Special Collections. However, the analysis also identified several programs and faculty that were not working with the Perry Special Collections who were teaching primary source literacy skills—at least according to their course learning outcomes. As such, the authors conducted a follow-up analysis of syllabi to see if similar results were achieved. A similar process was followed for compiling relevant syllabi and converting them into a useable textual format. The text was then analyzed using Voyant Tools to identify programs and faculty using language in their syllabi that indicated that they might benefit from use of Special Collections materials. The review of the syllabi focused only on the term archiv*, librar*, and research*

Results

The initial review looked at both the total occurrences of terms from our key term list in a corpus of program learning outcomes, and then specifically at our shorter adjusted term list. Among the most prominent of the programs in the total occurrences list were Family History—Genealogy BA (123 terms), History BA (65 terms), German Linguistics BA (28 terms), German Studies BA (25 terms), and English BA (24 terms). With the narrowed term list the ranking of programs by overall occurrences varied, though many of these program remained near the top (see Table 1).

Program	archiv*	librar*	research*
Family History—Genealogy BA	5	6	14
German Linguistics BA	2	1	11
German Lit, Film, Culture BA	2	1	10
German Studies BA	2	1	10
German Teaching BA	2	1	10

Table 1. Occurrence of Key Terms in Program Learning Outcomes

Based on these preliminary results, we determined to use the seven programs with the highest term frequencies to conduct a text analysis of their course learning outcomes and another based on course syllabi. Corpora of each type of document were aggregated for the bachelor’s degree programs in Family History—Genealogy; History; German Linguistics; German Lit, Film, Culture; German Studies; German Teaching; and English.

¹³ Jeffrey et al., “Digging in the Mines: Mining Course Syllabi in Search of the Library,” 76.

Querying terms in each set of corpora we found similar patterns emerged for each, despite differences in corpus size and document type. The syllabi corpus held 2,474,809 total words and 28,407 unique word forms. The course learning outcomes corpus included 55,182 total words, with 2,720 unique word forms. Within each of these, occurrence counts and relative frequency measures were obtained for each of the terms in the more comprehensive term list (see Table 2). Generally, terms related to specific skills or standards, such as “interlibrary” or “Turabian” were less likely to appear in the course learning outcome documents than in course syllabi, while general concepts or institutions like “research*” or “librar*” remained. This tendency toward conceptual language led to greater relative visibility of archival terms, with more frequent use of the terms “archiv*” (0.0001016 freq./mill. for syllabi, and 0.0001974 freq./mill. in learning outcomes) and “primary source*” (0.0002727 freq./mill. for syllabi, and 0.0018589 freq./mill. for learning outcomes).

Term	Syllabi		Course Learning Outcomes	
	Count	Relative Frequency	Count	Relative Frequency
archiv*	252	0.0001016	12	0.0001974
librar*	622	0.0002509	38	0.0006251
research*	5072	0.0101337	616	0.0020458
reference	361	0.0001456	2	0.0000329
interlibrary	12	0.0000048	0	0.0000000
database	126	0.0000508	8	0.0000000
eBook	26	0.0000105	0	0.0000000
journal*	1085	0.0004376	7	0.0001152
cit*	1856	0.0007486	109	0.0017931
bibliography*	485	0.0001956	16	0.0002632
source	2474	0.0009979	143	0.0023525
“primary source*”	1256	0.0002727	182	0.0018589
scholarly	310	0.0001250	20	0.0003290
history	5297	0.0021366	502	0.0082583
genealogy	304	0.0001226	0	0.0000000
book	2595	0.0007212	12	0.0000329
APA	36	0.0000145	0	0.0000000
MLA*	320	0.0001291	60	0.0009870
Chicago	176	0.0000710	0	0.0000000
Turabian	140	0.0000565	0	0.0000000
“special collections”	54	0.0000218	0	0.0000000
catalog	296	0.0001194	0	0.0000000

Table 2. Comparison of Count and Relative Frequency of Terms in Syllabi and Course Learning Outcomes

When looking at the occurrence of terms from our narrowed list within each corpus independently, we found differences between terms and across degree programs. Looking at the course learning outcomes corpus (see Fig. 1), results from programs within the same department tended to be similar, with the general task-based term research* being much the same between History and Family History—Genealogy (approximately 0.014 freq./mill.), or at a lower rate across the German language undergraduate majors in the Department of German and Russian (approximately 0.004 freq./mill.). At the same time, the institution-based terms of librar* and archiv* appeared at low rates across all degree programs. In the case

of German-language programs, *archiv** did not appear at all, while unexpectedly the term *librar** did not appear in course learning outcomes for the English major.

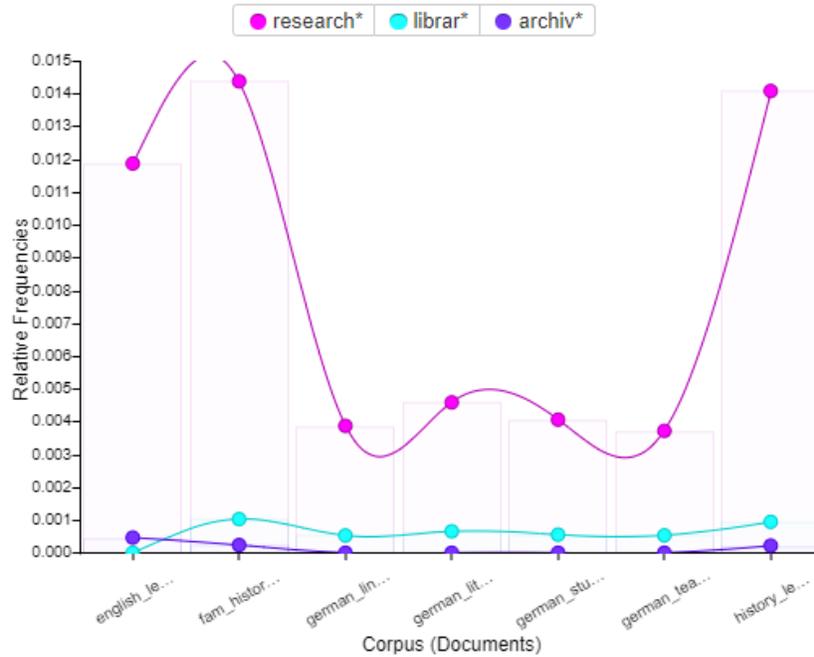


Figure 1. Relative Frequency of Terms in Course Learning Outcomes, by Degree Program

Reviewing the same terms in our corpus of course syllabi, we immediately noted a similar “W”-shaped curve when using the terms from our narrowed term list (see Fig. 2). However, there were significant differences in the scale of the relative frequencies for some programs. For Family History—Genealogy and History, in particular, “*research**” appeared at a rate of 0.0032 freq./mill. in syllabi compared to the earlier 0.014 freq./mill. for learning outcomes. These differences reflected the overall decline in frequency seen between the corpora, though the segmentation by major suggested that these reductions may occur proportionally.

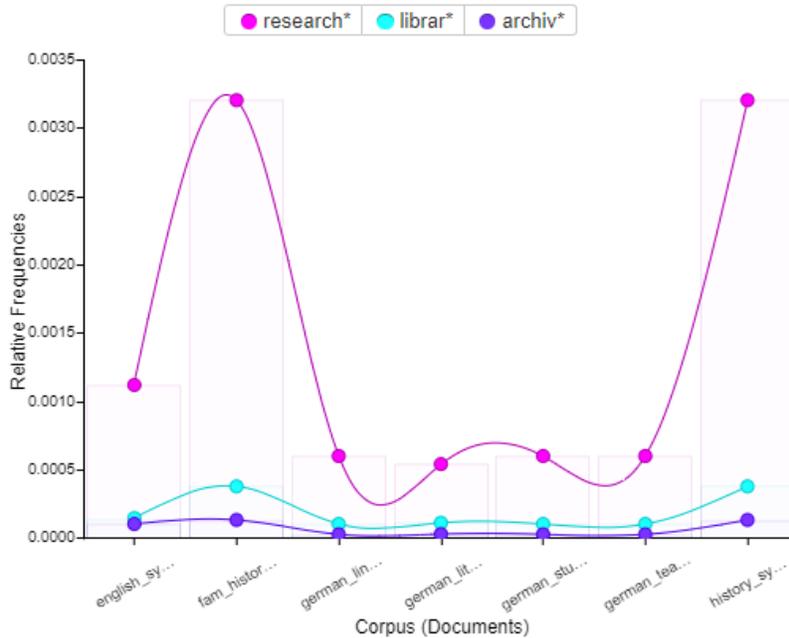


Figure 2. Relative Frequency of Terms in Course Syllabi, by Degree Program

A final area of investigation was to use text analysis to explore the scaffolding used within undergraduate programs based on our term list. Comparing our narrow list of terms, we found, in general, the greatest occurrence of these terms in 200-level courses, declining in the third year of study. In 400-level courses there was a renewed focus on research, but the relative frequency of the terms “librar*” and “archiv*” continued to decline over the 300-level counts (see Fig. 3).

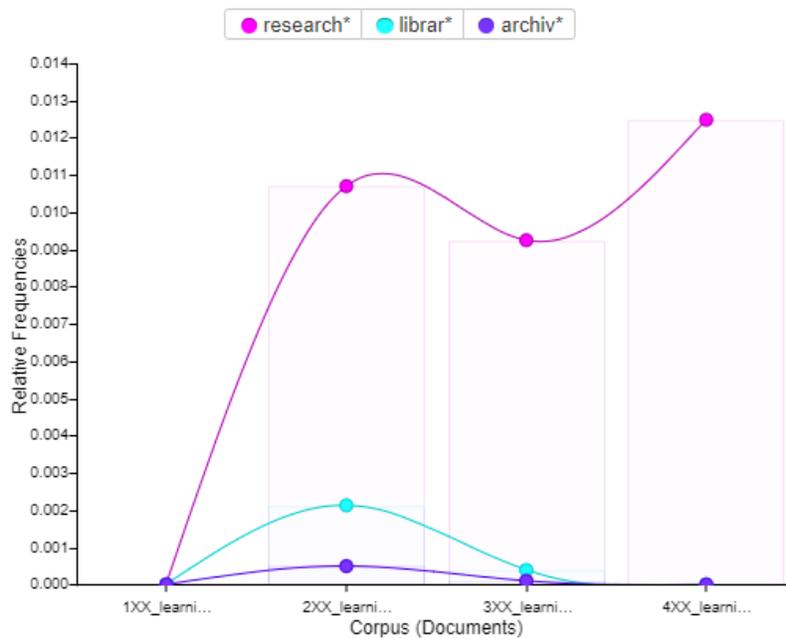


Figure 3. Relative Frequency of Terms in Learning Outcomes, by Course Level

Findings

In examining the text analysis results of syllabi and course learning outcomes within a sample of seven undergraduate programs, we identified four primary findings:

- A broader range of skill or task-based terms appear in course syllabi than in course learning outcomes
- Course learning outcomes tend to utilize more conceptual language resulting in higher relative frequency of conceptual terms
- Conceptual terms generally appeared proportionally between the syllabi and the course learning outcome documents resulting in similar trends for both document types
- Course learning outcomes directly addressing the conceptual terms examined were seen most frequently in 200-level courses

These findings suggest that conducting a textual analysis of course learning outcome documentation was an adequate substitute for the more involved review of all the content of syllabi when focusing on conceptual terms. A comparison of the relative frequency trends between learning outcomes and syllabi resulted in similar trendlines, both when aggregated by academic program and by course level. The analysis also provided useful information on the sequencing of instruction within individual program and generally across the selected departments, identifying second-year courses as being more focused on introducing students to libraries and archives.

Conclusion

This paper examines the utility of using text analysis to aid in the identification of academic departments and programs that could potentially benefit from deeper integration with special collections. It describes a methodology in which the project team initially performed text analysis on course learning outcomes to see where outreach possibilities existed. Project members were surprised by the results of the textual analysis of course learning outcomes. While some of the results were expected, revealing programs and faculty already partnering with the Perry Special Collections; other results were not expected and showed that there were programs and faculty on campus whose course learning outcomes indicated that they were teaching primary source literacy skills. Given the results, the project team decided to conduct a textual analysis of course syllabi to compare the outcome with the analysis of learning outcomes.

The results of the comparison of the textual analyses of course learning outcomes and syllabi suggests that a textual analysis of course learning outcomes is a reasonable initial screening step when preparing to do a deeper syllabus review. Our project showed that text analysis is a viable tool for helping special collections professionals identify programs and faculty that they can reach out to about deeper integration. While text mining of course learning outcomes is a useful initial screening tool, special collections professionals still need to carefully mine all the content of syllabi to better understand which programs and faculty would be most likely to be receptive to outreach efforts. Special Collections professionals then need to do the work of reaching out to identified programs and faculty to work on integrating their holdings into the courses offered by those programs and faculty.

In addition to exploring course learning outcomes, the project also provided insights into the scaffolding of primary source instruction in the overall curriculum. Our analysis found that second-year courses were the most likely to include content regarding libraries and archives, providing a foundation for increasing expectations regarding research in capstone courses. By better understanding the sequence of instruction within departments and programs, outreach to faculty can be further targeted to promote greater integration of special collections and archives in their curriculum.

Resources

- Bance, Bryan. "Outreach in the Academic Community: Enhancing the Teaching Role of University Archives" (MA thesis, University of Manitoba/University of Winnipeg, 2012).
- Blummer, Barbara and Jeffrey M. Kenton. "Academic Libraries' Outreach Efforts: Identifying Themes in the Literature." *Public Services Quarterly* 15, no. 3 (2019): 179-204.
- Brigham Young University, "Expected Learning Outcomes," accessed September 28, 2021, <https://learningoutcomes.byu.edu/>
- Butler, Kathleen and Theresa Calcagno. "Syllabus Mining for Information Literacy Instruction: A Scoping Review." *Evidence Based Library & Information Practice* 15, no. 4 (2020): 83-104.
- Chute, Tamar G. "Selling the College and University Archives: Current Outreach Perspectives." *Archival Issues* 25 no. 1/2 (2000): 33-48.
- Daines, J. Gordon III, Cory L. Nimer, and Jacob R. Lee. "Exploring the American Archivist: Corpus Analysis Tools and the Professional Literature." *Journal of Contemporary Archival Studies* 5, Article 3 (2018), accessed September 28, 2021, <https://elischolar.library.yale.edu/jcas/vol5/iss1/3>
- Harris, Valerie A. and Ann C. Weller. "Use of Special Collections as an Opportunity for Outreach in the Academic Library." *Journal of Library Administration* 52, no. 3-4 (2012): 294-303.
- Jeffrey, Keven M. et al. "Digging in the Mines: Mining Course Syllabi in Search of the Library." *Evidence Based Library & Information Practice* 12, no. 1 (2017): 72-84.
- Malkmus, Doris. "'Old Stuff' for New Teaching Methods: Outreach to History Faculty Teaching with Primary Sources." *portal: Libraries and the Academy* 10, no. 4 (2010): 413-435.