

***Preprint version. Scheduled for publication in the 2014 CUNY Assessment Conference Proceedings.**

**Scholarly E-Book Use across Disciplines:
Content Analysis of Usage Reports and Search Terms**
Nisa Bakkalbasi (Assessment Coordinator)
Melissa Goertzen (E-Book Program Development Librarian)
Columbia University Libraries, New York, New York, USA

Abstract

Data collected through COUNTER usage statistics and the LibQUAL+ service quality assessment survey tell us that faculty, graduate students, and undergraduates value access to the growing e-book collection at Columbia University Libraries (CUL). While the aggregate results indicate that e-book use continues to increase, usage rates are not uniform across disciplines. Anecdotal evidence suggests that while e-book use has grown in the sciences and social sciences, scholars in the arts and humanities rely heavily on print books. Given the highly diverse research needs of the university community, CUL is keen to understand scholarly e-book usage in various disciplines.

In this study, we sought an innovative research method to understand e-book usage. This method utilizes data from two sources: readers' e-book search terms harvested by Google Analytics; and requested e-book titles provided by the COUNTER e-book usage reports. The data was analyzed using NVivo, a qualitative analysis software, to examine popular scholarly e-book topics and the correlation between search and delivery.

Introduction

Over the past decade, electronic books (e-books) have become increasingly popular in the academic community. In response to this demand, Columbia University Libraries (CUL) provides access to over two million e-books that support research, teaching, and learning activities across campus and within the wider scholarly community. As the collection continues to grow, CUL is developing a unique strategy and vision for e-book programs and initiatives. To achieve this goal, the Collection Development Department launched the E-Book Program Development Study in 2013. This ambitious assessment project centers on the collection of essential data to drive the development of policies related to e-book acquisition, discovery, and access.

During the same year, data collected through COUNTER usage statistics and the LibQUAL+ service quality assessment survey indicated that faculty, graduate students, and undergraduates value access to the growing e-book collection at CUL. While the aggregate results indicate that e-book use continues to increase, usage rates are not uniform across disciplines. Anecdotal evidence suggests that while e-book use has grown in the sciences and social sciences, scholars in the arts and humanities rely heavily on print books. Given the highly diverse research needs of the university community, we wanted to understand scholarly e-book usage in various disciplines.

The aim of this study is to better understand how scholarly e-books are used in various disciplines in teaching, learning, and scholarly pursuits through readily available data. This study seeks to gather data to drive the creation of best practices and policies to support the delivery of e-book collections and programs that facilitate research, teaching, and learning across campus and within the wider scholarly community.

Literature Review

Determining how e-books are used for scholarly purposes is a complex issue. The e-book landscape is evolving at a rapid pace and a wide range of factors, including business models, e-book formats, and platform functionality, impact how library clients discover and access e-books for research, teaching, and learning activities. It is more important than ever for librarians to understand when, how, and why clients use e-books in order to design services that meet existing needs.

Over the past several years, a number of studies were conducted to determine how e-book use differs across scholarly disciplines. Littman and Connaway (2004), Christianson (2006), Bailey (2006), and Kimball, Ives, and Jackson (2010) examined e-book use according to subject and all suggest that the highest usage rates were typically found in computers, technology, business, and the sciences. The lowest usage rates were most often discovered in the humanities and arts. This finding was consistent across academic institutions of various sizes, funding structure, and missions. Staiger (2012) discovered a trend that suggests a relationship between the currency of an e-book and its relevance to researchers, particularly in fields like business, computer science and technology. He attributed this finding to the fact that researchers in these disciplines have an acute need for current information.

A study by Levine-Clark (2007) suggests that there is no correlation between the awareness of e-book collections within disciplines and e-book usage rates. At the University of Denver, Levine-Clark conducted a survey that measured knowledge and usage of e-books in the humanities. In total, 2,067 faculty, students, incoming students, and alumni responded. The results indicated that 74.4 percent of humanists were aware of e-book collections available through the university. In all other disciplines, awareness ranged from 49 to 69 percent. However, humanists use e-books less often than scholars in other disciplines.

A number of studies have been conducted to understand how e-books are used for research, teaching, and learning activities. Shelburne (2009) conducted a large scale survey to learn about e-book usage patterns at the University of Illinois. In total, 1,547 responses were received. The results indicated that 78 percent of e-book use was intended for research purposes, 56 percent for study, 2 percent for teaching, and 2 percent for other purposes.

Levine-Clark (2007) found that library users typically “use rather than read” e-books. Typically, the format is viewed as a convenient source that provides quick reference for scholarly endeavors. Results from a survey of 2,067 faculty, students, incoming students, and alumni

indicated that 56 percent of respondents use e-books to read a chapter or article within a book, and 36 percent typically read a single entry or several pages.

Noorhidawati and Gibb (2008) and Berg, Hoffman, and Dawson (2010) suggest that e-books are primarily used for quick reference, limited reading, and citation checks as opposed to extended reading and research. In other cases, e-books serve as a convenient means to preview a text; students and faculty members peruse the e-version to gain a sense of the information, biases, or arguments presented in a scholarly monograph. If it is useful for their research purpose, a print version is often requested for extended reading.

A literature review by Staiger (2012) compared the results of two dozen studies regarding e-book usage by members of the academic community. Findings suggested that “academic users typically search e-books for discrete bits of information, a behavior summed up by the formula ‘use rather than read’” (p. 355). In general, members of the academic community do not immerse themselves in e-books for extended periods of time to examine entire arguments. Instead, they view e-books as “convenient sources from which to extract information for their scholarly endeavors” (p. 357). Essentially, e-books provide a means for power browsing. They allow users to preview a book without leaving their work stations, and then locate the print copy if the information is relevant to their studies (p. 358). A literature review by Ashcroft (2011) uncovered similar trends. Statistics showed that on average, “53.5 percent of students and 58.6 percent of teachers dipped in and out of several chapters, whereas very low percentages read the whole book – 5.5 percent of students and 7.1 percent of teachers” (p. 401).

E-book Collection at CUL

CUL is one of the top five academic research library systems in North America and serves a community of over 3,750 faculty members and 26,000 full-time students at the Morningside Campus and Medical Center. The collections are housed across 21 campus libraries and include over 12 million volumes, 160,000 current journals and serials, and an extensive collection of manuscripts, rare books, microforms, maps, and audiovisual materials. In 2004, CUL began purchasing e-books in an experimental capacity. Due to the positive reception by faculty and students, the Library continued to grow e-book holdings to support research, teaching, and learning activities across campus. Currently, CUL provides access to over two million titles.

CUL offers e-books through subscriptions packages (e.g. Knovel, Ebrary, Safari) as well as individually purchased titles. The Library also licenses e-books through publishers’ packages, including Cambridge University Press, Oxford University Press, Springer, and Wiley. Over the past several years, CUL has partnered with a number of academic and research institutions through consortial groups to investigate business models for shared e-book purchasing, including the Manhattan Research Library Initiative (MaRLI), 2CUL, Knowledge Unlatched (KU) and the North East Research Libraries (NERL) Consortium.

Methodology

Before discussing the methodology in detail, it is worth mentioning that our initial thought was to create a survey to gather information about e-book use across disciplines. However, two key factors influenced our assessment strategy and motivated us to tap into existing data sources rather than developing a survey instrument. First, during our initial consultations, it became apparent that using a low-overhead data collection technique that would allow us to systematically collect information over time would be most appropriate for this project. Due to our interest in continuously monitoring our user base in an ever-changing e-book landscape, reliance on readily available, continuous, and accurate data was an important factor in creating an effective and sustainable assessment plan.

Second, as survey participation rates have declined, survey research has experienced significant challenges that impact its use in library assessment plans. Participating in a survey to provide thoughtful and reflective feedback requires time and effort from respondents. The quality of the data begins to deteriorate when potential respondents do not make the effort to submit a completed survey or leave the survey incomplete. Surveys are of little, or no use, if the response rate is low or the data is inaccurate. Based on the low response rates from a recent survey, and in an attempt to avoid survey fatigue, we investigated alternative approaches of data collection.

In this study, we sought an innovative research method to understand e-book usage. This method utilizes data from two sources: readers' e-book search terms harvested by Google Analytics; and requested e-book titles provided by the COUNTER e-book usage reports. The data sets present CUL with an accurate, continuous, and objective picture of e-book use.

The study covers the period from January 1, 2013 through December 31, 2013. It is worth noting that CLIO became the default discovery tool for the library at the beginning of June 2013. Thus, searches tracked by Google Analytics prior to June 2013 are limited. We included eight major e-book platforms in the study (i.e. Springer, Wiley, Oxford University Press, Elsevier, EBSCO, Ebrary, Cambridge University Press, and Safari Books Online) to ensure e-books were included from all three major disciplines, namely humanities, social sciences, and sciences.

For the indicated time period, we exported all search terms limited by format to e-books from our Google Analytics account. After data clean-up and formatting, requested e-book titles from COUNTER reports and e-book search terms from our Google Analytics account were loaded into the qualitative analysis software, NVivo to identify frequently used words and explore recurring patterns. Then, we performed text analysis to generate word frequency tables and word clouds for each of the frequency sets to graphically display how each of the collections, at least in terms of the titles used, covers a different sector of the e-book platform universe.

Findings and discussion

The most frequently repeated search word was "history," which was entered 526 times into the search field to search for e-books. It was followed by the word "theory" (entered 378 times). The

most frequently requested e-book title word was “edition” (repeated 3,284 times), followed by the word “volume” (repeated 2,306 times). In the preliminary analysis, we refrained from adding words such as “edition,” “volume,” and “2nd” to a stop list, as we determined they might shed a special light on what was being searched and delivered in some instances.

Table 14 lists the top 25 most frequently repeated search words and requested title words. We found an overlap of 60% (15 words) in both lists, indicating a correlation between search and delivery of e-books. The words that are present in both lists are reported in *italics* (see Table 1).

Table 1. Most frequently repeated search and requested title words

Rank	Search terms			Requested title words		
	Word	Length	Count	Word	Length	Count
1	<i>history</i>	7	526	edition	7	3284
2	<i>theory</i>	6	378	volume	6	2306
3	<i>social</i>	6	368	<i>history</i>	7	1949
4	introduction	12	359	<i>theory</i>	6	1777
5	<i>new</i>	3	358	<i>new</i>	3	1730
6	<i>analysis</i>	8	326	<i>american</i>	8	1689
7	<i>american</i>	8	309	<i>analysis</i>	8	1651
8	<i>handbook</i>	8	303	advances	8	1577
9	human	5	281	systems	7	1558
10	<i>research</i>	8	281	culture	7	1552
11	<i>health</i>	6	265	studies	7	1532
12	<i>world</i>	5	227	<i>world</i>	5	1510
13	<i>modern</i>	6	223	<i>guide</i>	5	1502
14	<i>guide</i>	5	219	<i>social</i>	6	1479
15	law	3	211	<i>handbook</i>	8	1468
16	medicine	8	207	applications	12	1412
17	<i>management</i>	10	198	<i>politics</i>	8	1367
18	rights	6	193	<i>science</i>	7	1365
19	war	3	191	<i>modern</i>	6	1230
20	<i>development</i>	11	188	<i>research</i>	8	1198
21	art	3	186	<i>development</i>	11	1196
22	<i>science</i>	7	183	international	13	1196
23	<i>politics</i>	8	181	<i>management</i>	10	1126
24	design	6	176	<i>health</i>	6	1107
25	political	9	172	global	6	1034

The expected role of a book title is to provide a compact summary of the book and help the reader identify typical content of the book. The prominence of “history” in both lists was an interesting reflection on the kinds of works being used, as were the terms “handbook,” “guide,” and “manual.” The high frequency of these words leads us to believe that users were searching for broad topics, reference works, or other collections of instructions, all of which are intended to provide ready reference. These results mirror a number of findings mentioned in the literature review, namely by Levine-Clark (2007), Shelburne (2009) and Staiger (2012), who suggest that e-books are used to read chapters or articles for study purposes.

When we evaluated the word clouds, which are graphic representations of word frequencies for the e-book search terms and requested titles, a similar trend emerged (see Figures 1, 2, 3, and 4). For instance, Figures 1 and 2 show the frequencies of all requested e-book titles and search terms. Words like “history,” “edition,” “volume,” “introduction,” and “theory” are situated at the center of the clouds, meaning that they have the highest frequency.

Next, we examined the word clouds generated for each of the major platforms included in the study. For the purpose of this paper, we explored the preliminary results for the Ebrary platform (see Figure 3) and the Springer platform (see Figure 4). Again the results pointed towards broad topics that could be used for reference purposes. For instance, the most frequently repeated title words for the Ebrary platform are “volume” and “history,” and the most frequently repeated title words for Springer are “systems,” and “theory.”

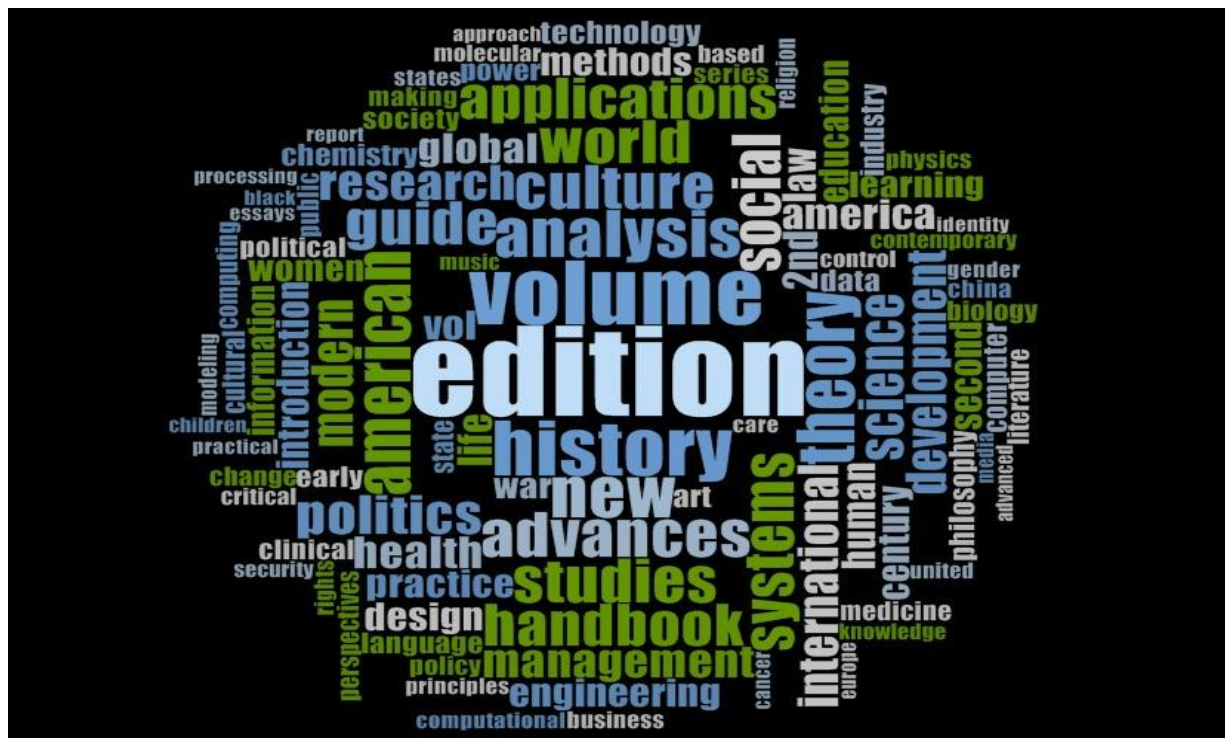


Figure 1. Word cloud for requested e-book titles.

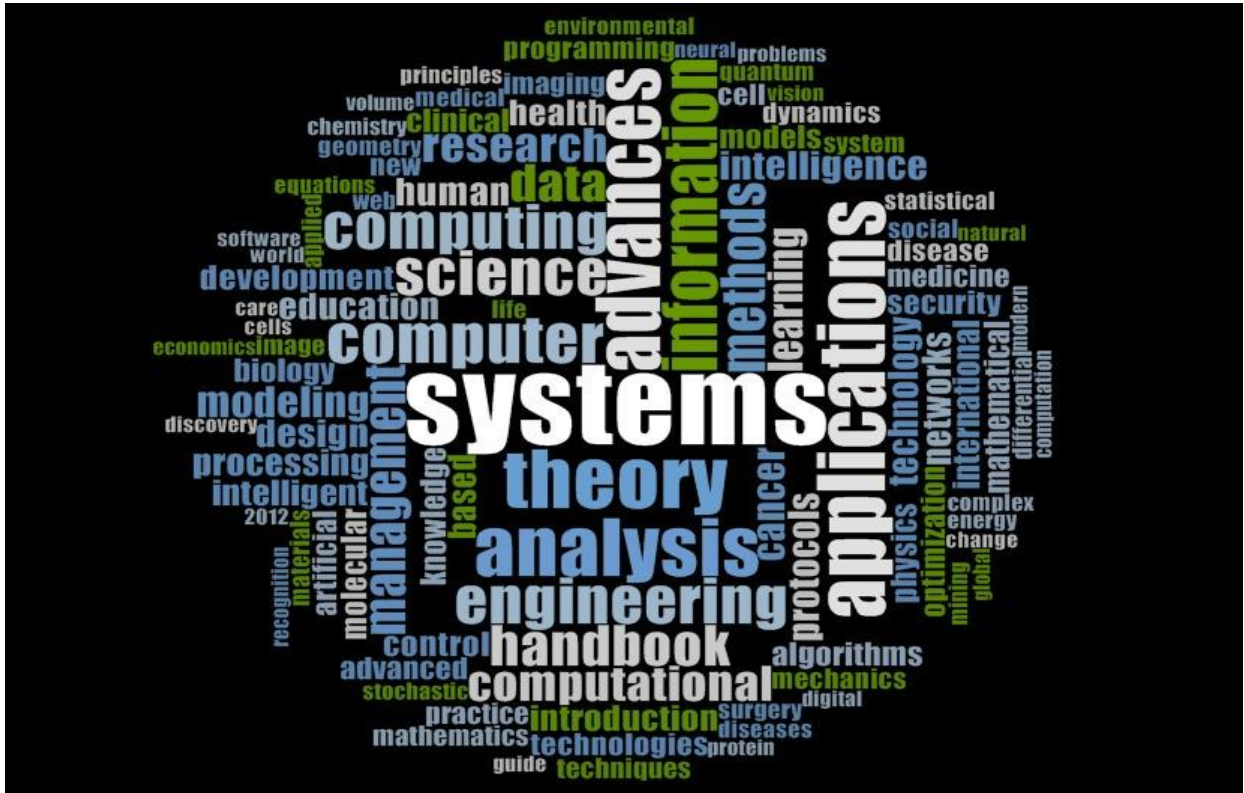


Figure 4. Word cloud requested e-book titles from the Springer e-book collection.

To analyze our findings in greater depth, we turned to open-ended comments collected through the 2013 LibQUAL+ service quality assessment survey. Comments relating to the e-book collection indicated that many users access e-books to read course materials. Both undergraduate and masters-level students expressed an interest in greater access to course readings in electronic format. For instance, an undergraduate computer science major said that “all of the Core texts should be available from the library digitally!” Another undergraduate studying public affairs wrote, “please provide more copies of course textbooks or enable electronic copies.” A doctoral student in the social sciences said that e-books available as PDF files are most convenient because “I want to be able to flip through the whole book without having to log back in.” These comments are consistent with our findings that the e-book collection is widely used across major disciplines to support instruction and learning.

Conclusions

Running search terms and requested title words through a text analysis tool reveals new ideas and concepts relating to e-book use, and reaffirms certain findings that we discovered through the LibQUAL+ service quality survey. The preliminary text analysis of search terms and requested title words was useful in gaining insight into the nature of e-book use across disciplines, including broad topic (e.g. history), academic level of use (e.g. introductory), and genre/type (e.g. reference).

It is challenging to deduce reader intent from word frequencies, as text data remain widely open for interpretation. However, responses to open-ended questions from the most recent LibQUAL+ survey are consistent with our findings that e-book collections are widely used across all major disciplines to support instruction and learning. User sentiments from the LibQUAL+ survey mirror a number of findings mentioned in the literature review, namely by Levine-Clark (2007) and Shelburne (2009), who suggest that e-books are used primarily to read chapters or articles for study purposes.

The ability to analyze word frequencies allows us to dig deeper and think about the many usage patterns that we wouldn't otherwise observe. While relying on a text analysis tool for these sorts of conclusions feels a bit nebulous, future work could clarify and extend present findings. Next, we plan to dig deeper into the text data by running *exact match* and *stemmed word* queries for those titles with 50 or more uses included in large platforms such as Springer, Ebrary, and EBSCO. Our preliminary analysis convinced us that words like "edition," "volume," and "2d" should be added to the stop list. They appear high in some e-book collections, and not at all in others, which may point to differences in the way databases formulate their titles as opposed to differences in the content of e-book collections. We will carry out formal statistical analysis to investigate the rank correlation and measure the relationship between search terms and e-book titles to assess the significance of the relationship between them.

References

- Ashcroft, L. (2011). E-books in libraries: An overview of the current situation. *Library Management*, 32(6), 398-407.
- Bailey, T. P. (2006, January). Electronic book usage at a master's level 1 university: A longitudinal study. *Journal of Academic Librarianship*, 32(1), 52 – 59.
- Berg, A. S., Hoffmann, K., & Dawson, D. (2010, November). Not on the same page: Undergraduates' information retrieval in electronic and print books. *The Journal of Academic Librarianship*, 36(6), 518-525. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0099133310002168>
- Christianson, M. (2005, December). Patterns of use of electronic books. *Library Collections, Acquisitions, & Technical Services*, 29(4), 351-363. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1464905506000042>
- Levine-Clark, M. (2007). Electronic books and the humanities: A survey at the University of Denver. *Collection Building*, 26(1), 7-14. Retrieved from <http://search.proquest.com/docview/219889113?accountid=10226>

- Littman, J., & Connaway, L. S. (2004). A circulation analysis of print books and e-books in an academic research library. *Library Resources and Technical Services*, 48, 256-262. Retrieved from <http://www.oclc.org/content/dam/research/publications/library/2004/littman-connaway-duke.pdf?urlm=162892>
- Mays, A. (2014, January). Biz of acq —PDA, e-books, print books usage and expenditures: Knowledge ecosystem remix. *Against the Grain*, 25(6), 61-64.
- Noorhidawati, A., & Gibb, F. (2008). How students use e-books – reading or referring? *Malaysian Journal of Library and Information Science*, 13(2), 1-14. Retrieved from <http://strathprints.strath.ac.uk/19896/>
- Shelburne, W.A. (2009, June). E-book usage in an academic library: User attitudes and behaviors. *Library Collections, Acquisitions, and Technical Services*, 33(2009), 59-72.
- Staiger, J. (2012, June). How e-books are used: A literature review of the e-book studies conducted from 2006 to 2011. *Reference & User Services Quarterly*, 51(4), 355 – 365. Retrieved from <http://rusa.metapress.com/content/u084729074582u77/fulltext.pdf>