

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



*Photo credit: M. Goertzen

Introduction

- ▶ Since 2010, there has been marked growth in Columbia University Libraries' (CUL) e-book programs
- ▶ The collection contains more than two million e-books

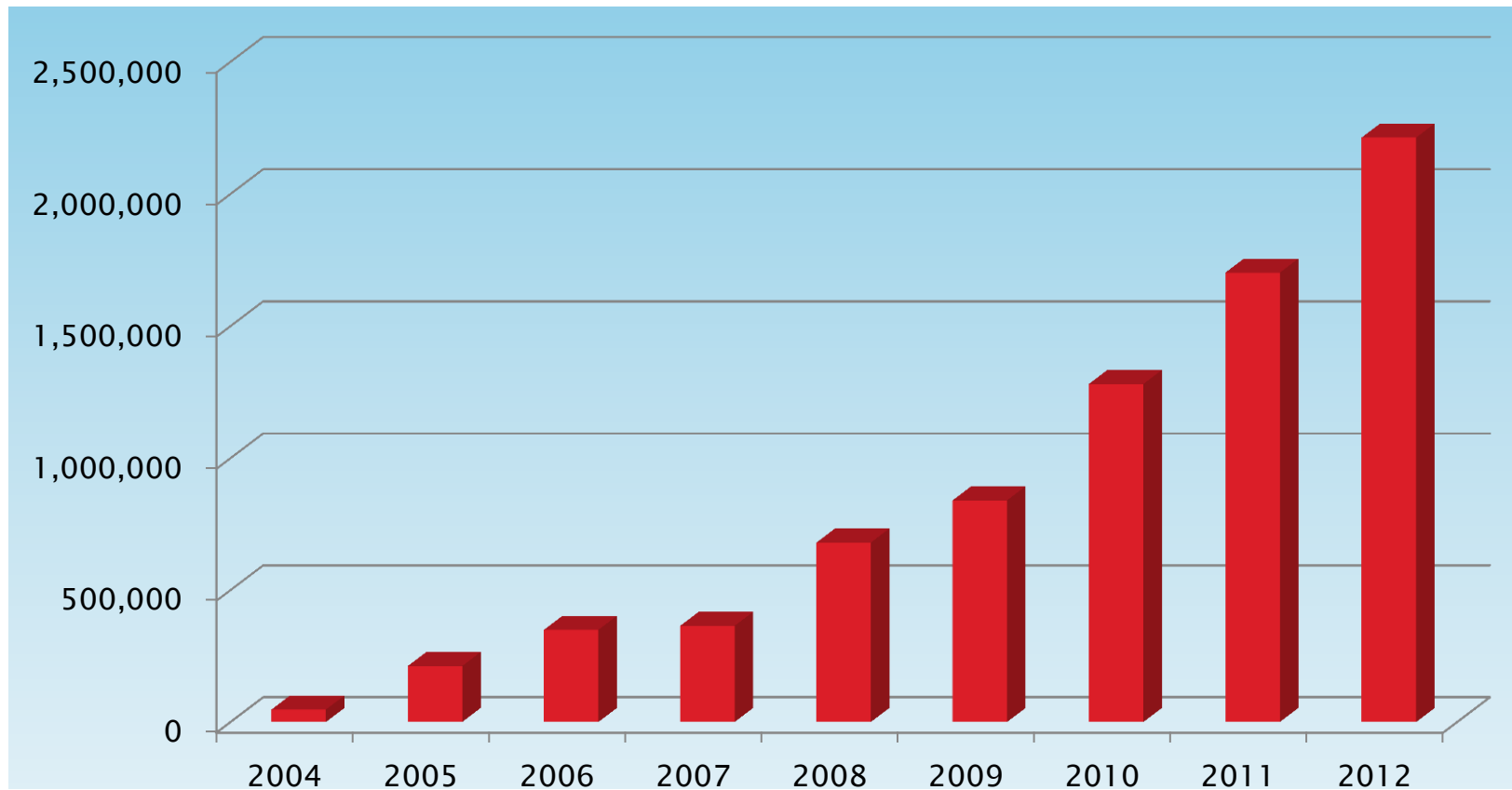
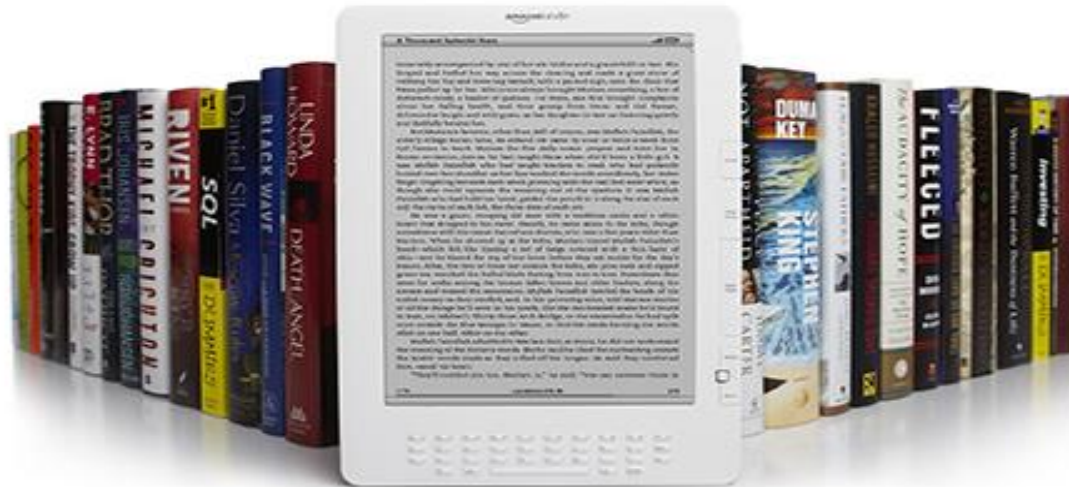


Table 1. E-book acquisition at Columbia University Libraries

E-Book Program Development Study

- **Method:** Complete an **assessment** aimed at gathering essential data to drive the development of policies related to e-book development programs.
- **Intended outcome:** The results provide **best practices** for stakeholders as they collaborate on the development of e-book programs.



LibQUAL+ Service Quality Assessment Survey

- Distributed across campus in October 2013
- 3,793 respondents (17% response rate)
- Participants commented on satisfaction with print and electronic resources
- Results indicate that faculty and students value access to e-book collections at CUL
- E-book usage rates are not uniform across disciplines



Study Objectives

1. To understand how e-books are **discovered and used** in various disciplines for research, teaching, and learning purposes.
2. To gather data that will drive the development of **e-book polices and best practices**.
3. To provide recommendations that will support **e-book collection development efforts** across campus.

Assessment Strategy

Key Consideration:

The e-book landscape is evolving at a rapid pace

- Use a low-overhead data collection technique that allows for the systematic collection of information over time
- Reliance on readily available, continuous, and accurate data
- Sustainability



Assessment Strategy

Initial Assessment Idea: Develop a survey

Challenges:

1. Survey participation rates have declined
2. Survey research has experienced significant challenges that impact its use in library assessment plans
3. The quality of data deteriorates when respondents do not submit a completed survey



Methodology

Data sources:

1. Readers' e-book search terms harvested by Google Analytics
2. Requested e-book titles provided by COUNTER reports
 - Springer
 - Wiley
 - Oxford University Press
 - Elsevier
 - EBSCO
 - Ebrary
 - Cambridge University Press
 - Safari



Methodology

- E-book titles from COUNTER reports and search terms from Google Analytics were loaded into NVivo
- Identification of frequently used words
- Text analysis was used to generate word frequency tables and word clouds



Results: Word Frequency

- Most frequently repeated search words were “history” and “theory”
- Most frequently repeated e-book title words were “edition” and “volume”
- Discovery of a 60% overlap between search terms and e-book titles



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

Table 2. Most frequently repeated search and requested title words

Results: Word Frequency

- Expected role of a title is to provide a compact summary of the book content
- The high frequency of “handbook” and “guide” suggests that users were searching for reference works or collections of instructions



Results: Word Clouds

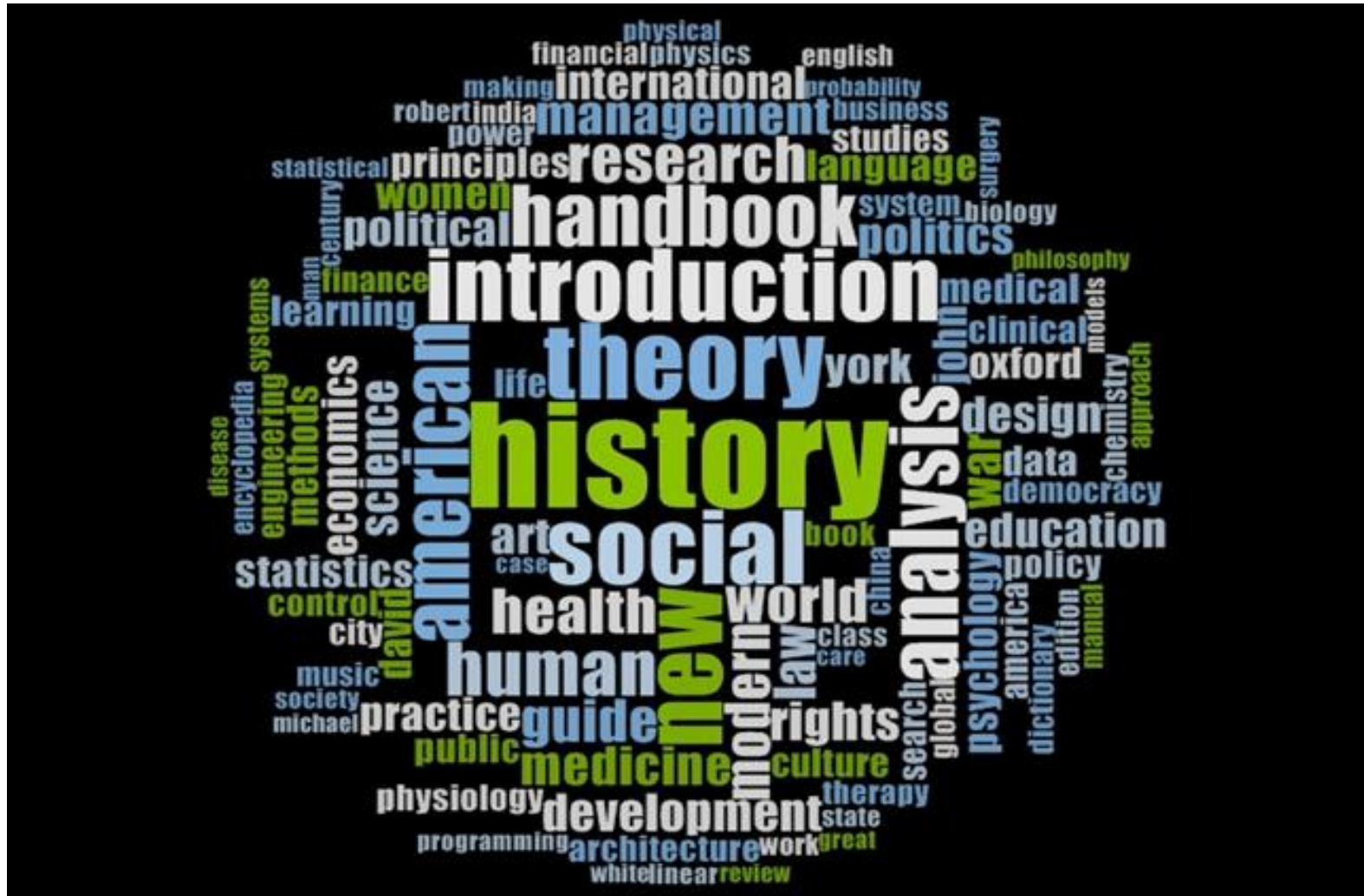


Figure 2. Word cloud for search terms harvested by Google Analytics

Results: Word Clouds

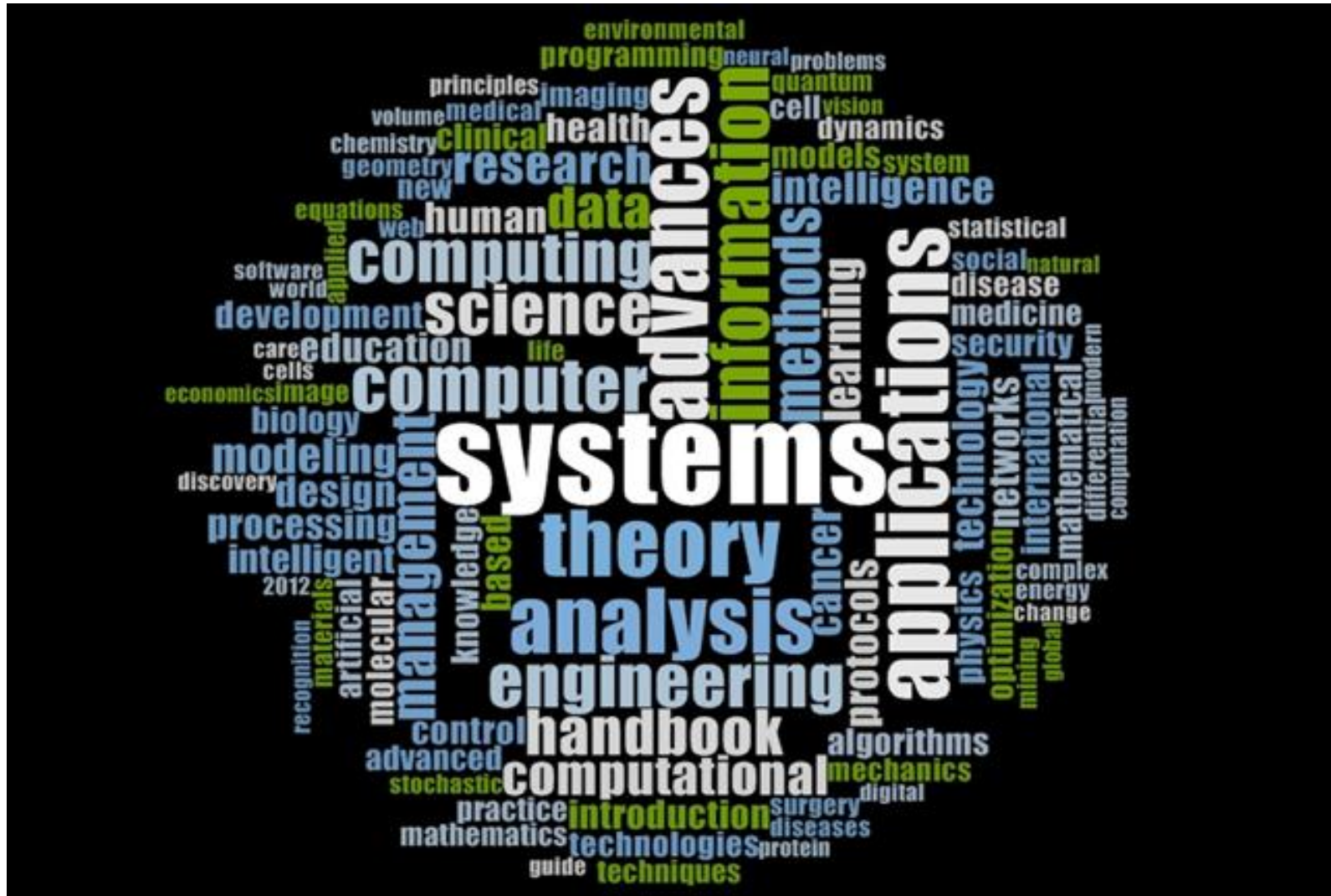


Figure 2. Word cloud for requested e-book titles from Springer

Conclusions

1. Text analysis of search terms and requested title words provided **insight into content of books**: broad topic (e.g. history), academic level of use (e.g. undergraduate/introductory), and genre/type: reference).
2. LibQUAL+ survey results are consistent with the finding that e-book collections are widely used across all major disciplines to **support instruction and learning**.
3. The ability to analyze word frequencies allows us to go deeper and think about the many **patterns** that we wouldn't otherwise notice.

Future Work

1. Dig deeper into the text data by running *exact match* and *stemmed word* queries for titles with 50 or more uses
2. Preliminary analysis suggests that the words like “edition”, “volume”, “2d” should be added to the stop list
3. Carry out a formal statistical analysis to investigate rank correlation and measure the relationship between search terms and e-book titles to assess the significance of the relationship between them.

Thank You

Nisa Bakkalbasi

Assessment Coordinator
hnb2119@columbia.edu

Melissa Goertzen

E-Book Program Development Librarian
mjg2227@columbia.edu