

PART II: STATISTICS PERTAINING TO RECURRING PROBLEMS OF THREE MASS DIGITIZATION PROJECTS¹⁰

I. GOOGLE BOOK SEARCH (GBS)

1) NUMBERS OF TITLES AND VOLUMES SEARCHED AND SELECTED

Titles searched

Number of titles searched: 347

Pre-1923: 122

Post-1922: 225

Number of titles available in GBS

full text or preview: 229 (66% of all titles)

full text: 111 (32% of all titles)

preview: 118 (34.1% of all titles)

snippets: 45

no page digitizations available (including snippets): 118 (34% of all titles)

Number of pre-1923 titles available in GBS

full text or preview: 116 (95.1% of pre-1923 titles)

full text: 111 (91% of pre-1923 titles)

preview (for later editions): 5

snippets (for later editions): 2

no page digitizations available (including snippets): 6 (4.9% of pre-1923 titles)

Number of post-1922 titles available in GBS

preview: 113 (50.2% of post-1922 titles)

snippets: 44

no page digitizations available (including snippets): 112 (49.8% of post-1922 titles)

Notes: 1) Most of the GBS preview digitizations for post-1922 titles were published since 1981. Very few were published prior to 1941. Of 113 titles published since 1981, GBS provided preview digitizations for 84 titles, or 74.3%. Of the 81 titles published between 1941 and 1980, GBS provided preview digitizations for 27 titles, or 33%. Of the 31 titles published between 1923 and 1940, GBS provided preview digitizations for only 2 titles, or 6.5%.

2) For some post-1922 published volumes for which GBS has not obtained permission to make available to the public its digitizations, GBS provides “snippets” of text in response

¹⁰ As only one volume was reviewed in Early American Imprints, Series I: Evans, 1639-1800 (EAI), no statistics were gathered with regard to that project. Information concerning the EAI digitization may be found in Part I, entry 85.

to word searches—selected sentences in which search terms occur. Snippets can be of use to scholars, because they can alert users to works in which relevant information occurs. For other post-1922 volumes, GBS offers “no preview” listings, which have more limited use for researchers. Word searches for “no preview” listings do not retrieve any results from these books. For books offering “no preview” listings, GBS includes some metadata information: title, author, date, publisher, pages, subject, ISBN number, a paragraph summary, links for buying the book, with prices if available, where to borrow, reviews, references from web pages, and other editions. Mistakes occur even within these limited views, however. For example, for entry 10 (Hofstadter, Richard. *The Age of Reform: From Bryan to F. D. R.* New York: Vintage Books, 1955), for which GBS provides a “no preview” listing, under other editions, it lists *The Progressive Presidents* by John Morton Blum, a book that obviously should not be listed as an “other edition” of the Hofstadter book. If the information in the “no preview” listings was correct and comprehensive, it would be valuable. Because the information appears to be machine-generated, however, there is little chance in GBS accomplishing this. For example, for entry 10, the only reviews offered are an Article review in H-Diplo and one JSTOR review. In contrast, the HEB entry for this work includes links to six scholarly reviews. As scholars can find more correct and comprehensive bibliographic and reference information in sources other than GBS for these “no preview” works, I have not tallied figures for the number of volumes for which GBS includes this type of listing.

Volumes searched

Number of volumes searched: 608

Number of volumes searched available in GBS

full text or preview: 443 (72.9% of all volumes)

full text: 315 (51.8% of all volumes)

preview: 128 (21.1% of all volumes)

snippets: 56

no page digitizations available (including snippets): 165 (27.1% of all volumes)

Number of volumes reviewed

Total: 80

Pre-1923: 62

[entries 1, 3, 6, 8, 11, 13, 16, 21, 26, 28, 31, 36, 41, 43, 46, 48, 51, 53, 56, 58, 61, 63, 66, 68, 71, 73, 76, 78, 81, 83, 86, 88, 91, 93, 96, 98, 101, 103, 106, 108, 111, 113, 116, 118, 121, 123, 126, 128, 131, 133, 138, 141, 143, 146, 148, 151, 153, 156, 158, 161, 163, 198]

Post-1922: 18

[entries 18, 23, 33, 38, 136, 166, 168, 171, 173, 176, 178, 181, 183, 186, 188, 191, 193, 196]

2) ACCESS, DOWNLOADING, SHARING, AND NAVIGATION

Ease of finding specific volumes in multivolume works

Number of titles for which title or author/title searches did not retrieve desired volume: 8
(10% of all GBS volumes reviewed)

Pre-1923: 8 (12.9% of pre-1923 GBS volumes reviewed)

[entries 6, 8, 13, 46, 48, 56, 76, 143]

Post-1922: none

Finding specific volumes of multivolume works can take a lot of time. Title searches retrieve hits that often do not list the specific volume number of the hit in the bibliographic information offered. To learn the volume number of the hit, users must click on the hit and then find the title page. If that volume is not the one desired, a user must then click on a “More editions” link that retrieves a list of digitized volumes, usually with no indication of volume number in the bibliographic information offered. Users must click on each listing of a particular edition and page to the title page to find out if the link leads to the volume desired.

For example, a GBS title search for “Works of Alexander Hamilton” (entry 46) yielded four hits: one volume each from the 1904, 1851, 1810, and 2005 editions. The bibliographic information for the first three listings did not include the volume number of the specific hit. The information for the fourth did indicate the volume number, but stated “No preview available.” Clicking on each of the first three listings led to volume I of the particular edition. Clicking on the “More editions” links of the 1904 volume yielded approximately 100 hits for digitizations of specific volumes from a variety of editions; clicking on the “More editions” links of the 1851 volume retrieved 3 hits; the 1810 volume, 2 hits; and the 2005 volume, 12 hits. To find volumes 2-7 of the 7-volume 1850-51 edition, users must click not only on the 3 hits retrieved from the “More editions” links of the 1851 edition, but also on hits retrieved from the “More editions” link of the 1904 edition, a time-consuming hit-or-miss process. As some of the links for specific volumes do not lead to title pages, users must also spend time paging to title pages to find the volume numbers.

OPAC records

Number of volumes for which OPAC did not contain links for retrieval of records: 22
(27.5% of all GBS volumes reviewed)

Pre-1923: 15 (24.2% of pre-1923 GBS volumes reviewed)

[entries 6, 56, 61, 68, 73, 81, 88, 91, 108, 111, 126, 131, 138, 146, 158]

Post-1922: 7 (38.9% of post-1922 GBS volumes reviewed)

[entries 23, 173, 178, 181, 183, 188, 191]

I checked for links to GBS digitizations in Hollis Catalog (Harvard University Library) and Socrates (Stanford University’s Online Catalog). While Hollis Catalog links lead directly to GBS digitizations, the Socrates links lead to listings in Google Scholar, which often includes links to digitizations in GBS.

MARC records

MARC records were not available for GBS digitizations.

Downloading, searching PDF files, and copying text

In GBS, the PDF files that users can download are paginated according to the physical page rather than the image page. For example, in the PDF downloaded version of entry 6 (Miller, Samuel. *A Brief Retrospect of the Eighteenth Century. Part First; in Two Volumes: Containing a Sketch of the Improvements in Science, Arts, and Literature, During That Period.* 2 vols. New York: T. and J. Swords, 1803. [Vol. I]), page 31 in the text was listed as page 54 in the page indicator, because it was the 54th physical page of the book. If a user finds a reference in another text to page 31 of this book, he or she must make a calculation to find this page. Often the difference between the image page number and the physical page number changes as pages progress. In the case of entry 6, while the difference at page 31 was 23, by page 324, the difference is 17, as the physical page number that corresponds to 324 is 341. The change in difference can occur because of mistakes in scanning.

MLSB bypasses this problem because its downloaded files indicate the number of the image page in the page indicator along with the physical page listed in parenthesis; for example, “31 (54 of 563),” where 31 is the image page number, 54 the physical page number, and 563 the total number of physical pages.

Users cannot search for text in the GBS downloaded PDF file, unlike the MLSB downloaded PDF files.

Users can copy selections from the image, but not selections from the OCR reading of the text, as users can do with MLSB.

Sharing results

Users can share a “clip” from the text by selecting it and then sending it as an image or text to a blog. Users must sign in with a Google Account in order to do this. To access the clip, a receiver must have a Google Account.

Navigation

Unlike in MLSB, users have the option to input a desired page number in the page indicator in order to advance to a specific page. Users also can move page-by-page through clicking forward and backward buttons. Unlike in MLSB, users have the option of viewing double-page displays on one screen.

3) PHYSICAL CONDITION OF VOLUMES REVIEWED

Number of volumes with markings, folds, splotches, tears, stains, smudges, broken letters, or tape: 40 (50% of all GBS volumes reviewed)

Pre-1923: 40 (64.5% of pre-1923 GBS volumes reviewed)

[entries 1, 3, 8, 11, 13, 21, 26, 28, 31, 41, 43, 48, 53, 58, 63, 66, 68, 71, 81, 83, 86, 88, 91, 93, 96, 98, 103, 108, 111, 113, 116, 123, 128, 143, 148, 151, 153, 156, 161, 163]

Post-1922: none

Number of volumes with markings, folds, splotches, tears, stains, smudges, broken letters, or tape affecting 10 or more pages: 21 (26.25% of all GBS volumes reviewed)

Pre-1923: 21 (33.9% of pre-1923 GBS volumes reviewed)

[entries 11, 26, 43, 48, 58, 63, 71, 81, 83, 86, 88, 93, 96, 108, 111, 113, 123, 128, 148, 151, 163]

Post-1922: none

Number of volumes with markings, folds, splotches, tears, stains, smudges, broken letters, or tape affecting OCR: 31 (38.75% of all GBS volumes reviewed)

Pre-1923: 31 (50% of pre-1923 GBS volumes reviewed)

[entries 1, 3, 8, 11, 13, 21, 26, 28, 31, 41, 43, 48, 58, 63, 81, 83, 86, 88, 91, 93, 96, 108, 113, 116, 123, 128, 143, 148, 151, 161, 163]

Post-1922: none

Number of volumes with markings, folds, splotches, tears, stains, smudges, broken letters, or tape affecting OCR on 10 or more pages: 7 (8.75% of all GBS volumes reviewed)

Pre-1923: 7 (11.3% of pre-1923 GBS volumes reviewed)

[entries 43, 48, 58, 63, 123, 128, 151]

Post-1922: none

4) METADATA

Author, title, and publication listings

Number of volumes with incorrect author, title, and publication information: 3 (3.75% of all GBS volumes reviewed)

Pre-1923: 2 (3.2% of pre-1923 GBS volumes reviewed)

[entries 1, 8]

Post-1922: 1(5.6% of post-1922 GBS volumes reviewed)

[entry 196]

Contents links

Number of volumes with problematic contents links: 32 (40% of all GBS volumes reviewed)

Pre-1923: 25 (40.3% of pre-1923 GBS volumes reviewed)

[entries 6, 8, 13, 21, 26, 28, 36, 43, 46, 48, 53, 56, 58, 63, 71, 73, 76, 88, 101, 103, 121, 128, 148, 158, 198]

Post-1922: 7 (38.9% of post-1922 GBS volumes reviewed)

[entries 33, 38, 173, 183, 186, 188, 193]

Number of volumes with contents links to incorrect pages: 23 (28.75% of all GBS volumes reviewed)

Pre-1923: 17 (27.4% of pre-1923 GBS volumes reviewed)

[entries 6, 13, 26, 28, 36, 46, 53, 56, 58, 71, 73, 88, 101, 121, 148, 158, 198]

Post-1922: 6 (33.3% of post-1922 GBS volumes reviewed)

[entries 33, 38, 183, 186, 188, 193]

Number of volumes with incorrect or misspelled chapter listings in contents: 11 (13.75% of all GBS volumes reviewed)

Pre-1923: 9 (14.5% of pre-1923 GBS volumes reviewed)

[entries 8, 43, 48, 53, 63, 71, 73, 76, 128]

Post-1922: 2 (11.1% of post-1922 GBS volumes reviewed)

[entries 186, 193]

Number of volumes with contents links that are useless: 2 (2.5% of all GBS volumes reviewed)

Pre-1923: 1 (1.6% of pre-1923 GBS volumes reviewed)

[entry 198]

Post-1922: 1 (5.6% of post-1922 GBS volumes reviewed)

[entries 173]

Number of multivolume works with contents links that lead to incorrect volumes: 1 (1.25% of all GBS volumes reviewed)

Pre-1923: 1 (1.6% of pre-1923 GBS volumes reviewed)

[entry 46]

Post-1922: none

Number of volumes with contents links for an index where none exists: 2 (2.5% of all GBS volumes reviewed).

Pre-1923: 2 (3.2% of pre-1923 GBS volumes reviewed)

[entries 8, 21]

Post-1922: none

Index links

Number of volumes with incorrect index links: 1 (1.25% of all GBS volumes reviewed)

Pre-1923: 1 (1.6% of pre-1923 GBS volumes reviewed)

[entry 43]

Post-1922: none

Number of pages

Number of volumes with listings for number of pages different than that given in MARC records (LC catalog, Harvard Library's Hollis Catalog) or last page of text: 16 (20% of all GBS volumes reviewed)

Pre-1923: 4 (6.5% of pre-1923 GBS volumes reviewed)

[entries 3, 16, 28, 93]

Post-1922: 12 (66.7% of post-1922 GBS volumes reviewed)

[entries 18, 23, 33, 166, 173, 178, 181, 183, 188, 191, 193, 196]

In most, if not all, of these entries, the number of pages listed corresponded to the physical number of pages in the book, not the pagination as given by the publisher and included in authoritative sources.

Page indicators

Number of volumes with discrepancies between numbers listed in the page indicator and actual page numbers appearing in the text: 13 (16.25% of all GBS volumes reviewed)

Pre-1923: 11 (17.7% of pre-1923 GBS volumes reviewed)

[entries 46, 73, 83, 88, 101, 126, 141, 148, 151, 161, 184]

Post-1922: 2 (11.1% of post-1922 GBS volumes reviewed).

[entries 33, 166]

Subject headings

Number of volumes that used LC main subject headings but not the subheadings provided by LC: 23 (28.75% of all GBS volumes reviewed)

Pre-1923: 23 (37.1% of pre-1923 GBS volumes reviewed).

[entries 3, 6, 8, 16, 21, 41, 48, 51, 56, 71, 86, 93, 96, 111, 116, 131, 133, 138, 141, 143, 151, 161, 198]

Post-1922: none

A problem that such a practice can cause for scholars is exemplified in the following example. For entry 51 (Tyler, Moses Coit. *A History of American Literature*. Vol. I: 1607-1676. New York: G. P. Putnam's Sons, 1878), LC assigned the subject heading "American literature--Colonial period, ca. 1600-1775--History and criticism," while GBS assigned just "American literature." Clicking on the GBS link for "American literature" retrieved 1,173 books, listed in an order that is not explained by GBS. Users cannot order results differently. A scholar studying the history and criticism of colonial American literature will not be able to use the GBS results profitably because he or she will have no way to find out which of the 1,173 books deals with that topic.

In contrast, clicking on the subject heading link within the LC record for the Tyler book, one finds a list of LC subject headings beginning with "American literature--Colonial period, ca. 1600-1775--History and criticism" with links to relevant books pertaining to each heading. Clicking on the link for "American literature--Colonial period, ca. 1600-1775--History and criticism" retrieves 119 titles that a user could arrange by title, author

or creator, date (ascending), or date (descending). In addition to “American literature--Colonial period, ca. 1600-1775--History and criticism,” LC offers in the list the following headings with links provided to relevant books: “American literature--Colonial period, ca. 1600-1775--History and criticism—Congress”; “American literature--Colonial period, ca. 1600-1775--History and criticism—Periodicals”; “American literature--Colonial period, ca. 1600-1775--History and criticism—Theory, etc.” LC also offers many other headings beginning with the term “American literature--Colonial period, ca. 1600-1775.” LC’s subject heading links are organized according to transparent rules and standards that scholars may learn in order to help them find relevant works, and that professional librarians employ to answer reference queries. GBS’s methods of organizing digitizations are not openly accessible.

Number of volumes that either omitted some or all LC headings; used headings not derived from LC; used no subject headings even though LC provided some; or disjointed LC subject headings so that subdivisions became separate headings: 22 (27.5% of all GBS volumes reviewed)

Pre-1923: 11 (17.7% of pre-1923 GBS volumes reviewed)

[entries 28, 31, 58, 63, 73, 83, 101, 121, 156, 186, 193]

Post-1922: 11 (61.1% of post-1922 GBS volumes reviewed)

[entries 18, 166, 168, 173, 176, 178, 181, 183, 188, 191, 196]

For example, in entry 18 (Hollinger, David A. *Science, Jews, and Secular Culture: Studies in Mid-Twentieth-Century American Intellectual History*. Princeton: Princeton University Press, 1996), GBS used two of the five LC headings (“Jews—United States—Intellectual life” and “Secularism—United States—History—20th century”); augmented one (“Science—United States—20th century—History” was changed to “Science—History”); omitted two (“United States—Intellectual Life—20th century” and “United States—Ethnic relations”); added one general heading that seems of limited use because it is so broad (“History—United States—General”); and added one misleading heading (“Religion—Judaism—Rituals & Practices”), as the book addresses the influence of Jewish intellectuals on the secularization and de-Christianization of American culture, not than the rituals and practices of Judaism.

“About this book” reference page

The “About this book” page contains the following reference sections: “Key words and phrases” (terminology used for pre-1923 volumes) and “Key terms” (terminology used for post-1922 volumes); “References from books”; “References from scholarly works”; “Popular passages”; and “Related books.” GBS employs “automated methods to analyze the book” and select material for these sections, according to its help page.¹¹ While some of the material selected may have use for scholars, serious researchers will want to rely on traditional methods of research and to remain skeptical of claims that terms listed as “key” are significant or that “Related books” do bear significant relation to the books to which they supposed are related.

¹¹ Google Book Search, “Google Book Search Help Center.” Available at <http://books.google.com/support/bin/answer.py?answer=53549>.

“Key words and phrases” and “Key terms”

Note: A number of problems dealing with the “Key words and phrases” section that were found in the initial Jan-Feb 2008 examination of digitizations had been corrected by March 2008, though others remained and some new ones were present. Because of this, I have revised my report accordingly and the information offered below reflects the situation in March 2008.

Choice of terms: It is often unclear why one term was chosen and another was not. For example, for entry 1 (Riley, Woodbridge. *American Thought from Puritanism to Pragmatism*. New York: H. Holt, 1915), the only personal name listed in “Key words and phrases” was Asa Gray. Clicking on the link for Asa Gray retrieved 10 pages. The following names, each with a greater number of pages listed in the book’s index than Gray’s, were not included in “Key terms and phrases”: Benjamin Franklin, Ralph Waldo Emerson, William James, Thomas Jefferson, and Charles Peirce.

For entry 156 (Giddings, Franklin Henry. *Democracy and Empire, with Studies of Their Psychological, Economic, and Moral Foundations*. New York: Macmillan, 1900), “Babylonia” was listed in “Key words and phrases.” “Babylonia” appears in conjunction with the word “Egypt” on four of the five pages in which the former term occurs; however, “Egypt,” which occurs on six pages in the book, was not listed in “Key words and phrases.”

Inconsequential terms: For entry 43 (Dewey, John. *Democracy and Education: An Introduction to the Philosophy of Education*. New York: Macmillan, 1916), the name “PAUL MONROE” was listed in “Key words and phrases.” Clicking on the name retrieved 3 hits, all of which led to pages prior to the title page. On two of these, Monroe’s name does not appear. Monroe was the editor of the series in which this book was published, and his name does appear on one page prior to the title page that lists all the books in the series. While Monroe’s name was included in “Key words and phrases” because of this reference, the name “Rousseau” was not included even though the book devotes 8 pages to a discussion of Rousseau’s writings on natural education.

Incomplete listings of pages: The links to “Key words and phrases” often are incomplete. In entry 1 example listed above, pages retrieved for Asa Gray did not include all the pages listed in the book’s index in which Gray was referenced because clicking on the link retrieved only pages in which the complete name “Asa Gray” occurred and not the pages in which only his last name appeared.

Number of volumes with “Key words and phrases” or “Key terms” problems: 31 (38.75% of all GBS volumes reviewed)

Pre-1923: 26 (41.9% of pre-1923 GBS volumes reviewed)

[entries 1, 11, 13, 16, 28, 31, 36, 43, 48, 58, 63, 81, 83, 91, 101, 106, 108, 118, 123, 126, 131, 138, 141, 143, 156, 163]

Post-1922: 5 (27.8% of post-1922 GBS volumes reviewed)

[entries 136, 173, 176, 178, 181]

Number of volumes with OCR misreadings listed as “Key words and phrases” or “Key terms”: 5 (6.25% of all GBS volumes reviewed)

Pre-1923: 5 (8.1% of pre-1923 GBS volumes reviewed)

[entries 36, 58, 63, 126, 131]

Post-1922: none

Number of volumes that include as “Key words and phrases” or “Key terms” terminology for which searching retrieved no occurrences: 2 (2.5% of all GBS volumes reviewed)

Pre-1923: 1 (1.6% of pre-1923 GBS volumes reviewed)

[entries 123]

Post-1922: 1 (5.6% of post-1922 GBS volumes reviewed)

[entry 136]

Number of volumes with word fragments listed as “Key words and phrases” or “Key terms”: 7 (8.75% of all GBS volumes reviewed)

Pre-1923: 5 (8.1% of pre-1923 GBS volumes reviewed)

[entries 11, 101, 108, 123, 126]

Post-1922: 2 (11.1% of post-1922 GBS volumes reviewed)

[entries 136, 173]

Number of volumes with name fragments listed as “Key words and phrases” or “Key terms”: 3 (3.75% of all GBS volumes reviewed)

Pre-1923: 1 (1.6% of pre-1923 GBS volumes reviewed)

[entry 28]

Post-1922: 2 (11.1% of post-1922 GBS volumes reviewed)

[entries 178, 181]

Number of volumes with source citations listed as “Key words and phrases” or “Key terms”: 3 (3.75% of all GBS volumes reviewed)

Pre-1923: 2 (2.5% of pre-1923 GBS volumes reviewed)

[entries 13, 143]

Post-1922: 1 (5.6% of post-1922 GBS volumes reviewed)

[entry 176]

Number of volumes in which terminology listed as “Key words and phrases” or “Key terms” was not significant (for reasons other than those noted above) or occurred on only 1 or 2 pages: 14 (13.75% of all GBS volumes reviewed)

Pre-1923: 14 (17.7% of pre-1923 GBS volumes reviewed)

[entries 16, 31, 36, 43, 48, 81, 83, 91, 106, 118, 138, 141, 143, 163]

Post-1922: none

“References from books”

GBS offers no rationale for the selection of books included in its “References from books” section. In addition, links in this section on occasion did not lead to relevant pages.

For entry 61 (Fiske, John. *American Political Ideas Viewed from the Standpoint of Universal History: Three Lectures Delivered at the Royal Institution of Great Britain in May 1880*. New York: Harper & Brothers, 1885), under “References from books,” the listing “The American Government, National and State - Page 54 by Burke Aaron Hinsdale - United States – 1891” was offered. Clicking on the link retrieved page 54, yet that page contained no reference to Fiske’s book. Fiske’s book was listed in Hinsdale’s on p. 18 as a reference, not on p. 54.

For entry 83 (Addams, Jane. *Democracy and Social Ethics*. New York: Macmillan, 1902), GBS included “References from books” links to 5 full-text digitizations of books and approximately 45 books for which GBS offered previews. In 2 cases of the full-text books, the link from the Addams book retrieved a page in which the Addams book was not referenced, although the book was referenced on a different page. A search in GBS on the title and author of the Addams book yielded 626 full-text digitizations in which the book was referenced and approximately 63 preview digitizations. Just why GBS selected only the approximately 50 books listed in the “References from books” section and not the more than 600 others that reference the Addams book was not explained.

For entry 98 (Adams, Henry. *The Education of Henry Adams: An Autobiography*. Boston: Houghton Mifflin; Cambridge: Riverside Press, 1918), GBS listed 22 books in which this book was noted. According to a GBS search, however, the book has been cited in 1,310 sources, of which 338 are full-text sources. The 22 books listed in “References from Books” are not significant studies of the autobiography or books that use it significantly. The fact that only 2 of the 22 books were not previews leads me to speculate that this selection of references was produced in part to advertise in-print books rather than to further scholarship.

GBS should demonstrate the rationale behind their principles of inclusion and exclusion of texts in their “References from books” (and other) sections in order for scholars to gain confidence that perusing the selections would be worth their time.

“References from scholarly works”

For entry 98 (Adams, Henry. *The Education of Henry Adams: An Autobiography*. Boston: Houghton Mifflin; Cambridge: Riverside Press, 1918), “References from scholarly works” included a link that led to an article entitled “*Elektron: Electrical Systems in Retrospect and Prospect*.” A search in this article under “Henry Adams” retrieved one hit with the note “Sorry, this page’s content is restricted.” For scholars, pursuing such a link is a time-waster.

For entry 113 (James, William. *Pragmatism: A New Name for Some Old Ways of Thinking; Popular Lectures on Philosophy*. New York: Longmans, Green, 1907),

“References from scholarly works” included links to 5 articles and a “Show more” link. Clicking on the latter led to the note, “Sorry, we didn't find any articles that cite 10545162373793139760.” Clicking on the link for the first article listed, one entitled “Rationale for a pragma-dialectical perspective,” led to a subscription database. The article abstract did not indicate that the James text was discussed in it. A scholar pursuing this source might very well pay for an article that has tangential relation to the James text, if that.

“Popular passages”

Often quoted “Popular passages” were from texts that the author had quoted, not the author’s own words. Often listed page numbers were incorrect, as were the number of books listed that purported to include the quote.

For entry 63 (Ross, Edward Alsworth. *Social Control: A Survey of the Foundations of Order*. New York: Macmillan, 1901), a quote in “Popular passages” was from the New Testament, but GBS did not indicate this information in its citation. The fact that GBS included a link to “958 books from 1774-2007” in which the quote appeared is useless for scholars interested in the Ross text. The fact that clicking on the “958 books” link led only to 160 “Limited previews” indicated that “Popular passages” may function more as a tool for marketing than one for scholarship. In addition, while GBS stated that the quote appeared on p. 375 in the Ross book, it actually appears on p. 373.

For entry 98 (Adams, Henry. *The Education of Henry Adams: An Autobiography*. Boston: Houghton Mifflin; Cambridge: Riverside Press, 1918), under “Popular passages,” the first quote listed was purported to have appeared in 69 books from 1889-2007, yet clicking on the link led to 48 books, nine of which were editions of this very book.

For entry 156 (Giddings, Franklin Henry. *Democracy and Empire, with Studies of Their Psychological, Economic, and Moral Foundations*. New York: Macmillan, 1900), in “Popular passages,” GBS quoted 10 passages, but these did not appear on the pages from the book that were retrieved.

“Related books”

This section also appears to be a marketing tool; however, often the books listed have little or no apparent connection to the text under consideration.

For entry 26 (Robinson, James Harvey. *The New History: Essays Illustrating the Modern Historical Outlook*. New York: Macmillan, 1922), “Related books” included links to a “Snippet view” of Toni Morrison’s novel, *Song of Solomon*, a “Snippet view” of Tom Wolfe’s novel, *The Right Stuff*, a “Limited preview” to James Joyce’s *Ulysses*, and a “Limited preview” to Thomas Pynchon’s *Gravity’s Rainbow*, among other irrelevant texts. Robinson’s text, a manifesto advocating a new approach to the discipline of history, has nothing to do with the novels listed as “Related books.” Books in the field that do

discuss Robinson's, such as John Higham's *History: Professional Scholarship in America* and Peter Novick's *That Noble Dream: The "Objectivity Question" and the American Historical Profession* were not listed as "Related Books" by GBS.

For entry 198 (*The New Encyclopedia of Social Reform*. Ed. William D. P. Bliss and Rudolph M. Binder. New edition. Funk & Wagnalls, 1908), "Related books" listed James H. Billington, *Fire in the Minds of Men: Origins of the Revolutionary Faith*, a book bearing limited relation to the thrust of the encyclopedia.

5) QUALITY OF SCANS

Number of volumes rated "poor": 21 (26.25% of all GBS volumes reviewed)

Pre-1923: 20 (32.3% of pre-1923 GBS volumes reviewed)

[entries 26, 31, 43, 46, 56, 66, 73, 88, 101, 108, 118, 121, 131, 141, 143, 146, 148, 151, 161, 198]

Post-1922: 1 (5.6% of post-1922 GBS volumes reviewed)

[entry 196]

Missing pages

Number of volumes with missing pages: 13 (16.25% of all GBS volumes reviewed)

Pre-1923: 13 (21% of pre-1923 GBS volumes reviewed)

[entries 26, 43, 56, 73, 88, 101, 118, 121, 141, 146, 148, 151, 161]

Post-1922: none

Pages out-of-order

Number of volumes with pages displayed out-of-order: 4 (5% of all GBS volumes reviewed)

Pre-1923: 4 (6.5% of pre-1923 GBS volumes reviewed)

[entries 46, 56, 131, 141]

Post-1922: none

Pages appearing more than once

Number of volumes with one or more pages displayed more than one time: 8 (10% of all GBS volumes reviewed)

Pre-1923: 8 (12.9 % of pre-1923 GBS volumes reviewed)

[entries 26, 73, 83, 101, 141, 148, 151, 161]

Post-1922: none

Illegible pages

Number of volumes with one or more illegible pages: 6 (7.5% of all GBS volumes reviewed)

Pre-1923: 6 (9.7% of pre-1923 GBS volumes reviewed)
[entries 31, 56, 66, 73, 88, 101]
Post-1922: none

Pages with portions cut off or obscured

Number of volumes with portions of one or more page cut off or obscured: 12 (15% of all GBS volumes reviewed)

Pre-1923: 12 (19.4% of pre-1923 GBS volumes reviewed)
[entries 46, 56, 58, 73, 76, 101, 108, 131, 141, 148, 151, 198]
Post-1922: none

Pages with distorted lines

Number of volumes with distorted lines: 2 (2.5% of all GBS volumes reviewed)

Pre-1923: 1 (1.6% of pre-1923 GBS volumes reviewed)
[entry 126]
Post-1922: 1 (5.6% of post-1922 GBS volumes reviewed)
[entry 168]

Light, dark, or blurry pages, affecting OCR readings adversely

Number of volumes with light, dark, or blurry pages, parts of pages, or letters, affecting OCR readings adversely: 13 (16.25% of all GBS volumes reviewed)

Pre-1923: 13 (21% of pre-1923 GBS volumes reviewed)
[entries 46, 48, 51, 53, 56, 58, 63, 68, 73, 83, 88, 101, 158]
Post-1922: none

Scanned image quality

Number of volumes in which letters were not sharp: 20 (25% of all GBS volumes reviewed)

Pre-1923: 9 (14.5% of pre-1923 GBS volumes reviewed)
[entries 98, 103, 121, 133, 138, 153, 158, 193]
Post-1922: 11 (61.1% of post-1922 GBS volumes reviewed)
[entries 18, 33, 38, 136, 166, 168, 171, 173, 176, 178, 183]

6) ACCURACY OF WORD SEARCHING WITHIN TEXTS

30 hit limit

GBS retrieves a maximum of 30 hits of pages in which a searched word or phrase occurs, even if the word or phrase occurs on more than 30 pages.

For example, for entry 1 (Riley, Woodbridge. *American Thought from Puritanism to Pragmatism*. New York: H. Holt, 1915), a search for “idealism” retrieved 30 hits of pages, while a search for the same word in a digitization of the same book in MLSB retrieved 59 hits of pages. There is no rationale given for the particular selection of hits offered by GBS.

Truncation of search terms

HEB allows users to truncate search terms with the use of an asterisk. This type of truncation for searching would be useful for scholars if it was incorporated into GBS.

For example, in entry 188 (McClay, Wilfred M. *The Masterless: Self and Society in Modern America*. Chapel Hill: University of North Carolina Press, 1994), a search for “Lazarsfeld” does not yield p. 254, which mentions “Paul Lazarsfeld’s Bureau of Applied Social Research.” The index for the book does include p. 254 under “Lazarsfeld.” It would save scholars time to be able to truncate in order to retrieve one list of pages for terms and their derivatives.

OCR misreadings

Number of volumes in which OCR reading was affected adversely: 71 (88.75% of all GBS volumes reviewed)

Pre-1923: 57 (91.9% of pre-1923 GBS volumes reviewed)

[entries 1, 3, 6, 8, 11, 13, 16, 21, 26, 28, 31, 36, 41, 43, 46, 48, 51, 53, 56, 58, 61, 63, 66, 68, 71, 73, 76, 78, 81, 83, 86, 88, 91, 93, 96, 98, 101, 108, 111, 113, 116, 118, 121, 123, 126, 128, 131, 138, 141, 143, 146, 148, 151, 158, 161, 163, 198]

Post-1922: 14 (77.8% of post-1922 GBS volumes reviewed)

[entries 23, 33, 136, 166, 168, 171, 173, 176, 181, 183, 186, 188, 193, 196]

OCR misreadings due to footnote indicators

Number of volumes in which searching was inaccurate because footnote indicator were read by OCR as parts of words: 2 (2.5% of all GBS volumes reviewed)

Pre-1923: 1 (1.6% of pre-1923 GBS volumes reviewed)

[entry 93]

Post-1922: 1 (5.6% of post-1922 GBS volumes reviewed)

[entry 196]

OCR misreadings due to run-together words

Number of volumes in which searching was inaccurate because OCR read separate words as one word: 2 (2.5% of all GBS volumes reviewed)

Pre-1923: 2 (3.2% of pre-1923 GBS volumes reviewed)

[entries 81, 98]

Post-1922: none

OCR misreadings due to italicized words

Number of volumes in which searching was inaccurate because OCR read italicized words incorrectly: 14 (13.75% of all GBS volumes reviewed)

Pre-1923: 10 (16.1% of pre-1923 GBS volumes reviewed)

[entries 3, 8, 31, 36, 53, 61, 66, 71, 118, 138]

Post-1922: 4 (22.2% of post-1922 GBS volumes reviewed)

[entries 166, 168, 181, 188]

OCR misreadings in footnotes

Number of volumes in which searching was inaccurate because OCR read words in footnotes incorrectly: 24 (30% of all GBS volumes reviewed)

Pre-1923: 20 (32.3% of pre-1923 GBS volumes reviewed)

[entries 6, 8, 13, 16, 21, 28, 36, 46, 48, 53, 63, 66, 71, 81, 88, 91, 96, 118, 138, 151]

Post-1922: 4 (22.2 % of post-1922 GBS volumes reviewed)

[entries 176, 183, 186, 196]

OCR misreadings due to words in all caps

Number of volumes in which searching was inaccurate because OCR read words in all caps incorrectly: 10 (12.5% of all GBS volumes reviewed)

Pre-1923: 10 (16.1 % of pre-1923 GBS volumes reviewed)

[entries 3, 8, 11, 13, 46, 86, 91, 111, 121, 198]

Post-1922: none

OCR misreadings due to words with capitalized first letters

Number of volumes in which searching was inaccurate because OCR read words in which the first letter was capitalized incorrectly: 5 (6.25% of all GBS volumes reviewed)

Pre-1923: 5 (8.1 % of pre-1923 GBS volumes reviewed)

[entries 31, 61, 113, 116, 118]

Post-1922: none

OCR misreadings due to words in languages other than English

Number of volumes in which searching was inaccurate because OCR read words in languages other than English incorrectly: 7 (8.75% of all GBS volumes reviewed)

Pre-1923: 4 (6.5% of pre-1923 GBS volumes reviewed)

[entries 13, 66, 71, 111]

Post-1922: 3 (16.7% of post-1922 GBS volumes reviewed)

[entries 166, 168, 188]

OCR misreadings in Index pages

Number of volumes in which searching was inaccurate in Index pages: 4 (5% of all GBS volumes reviewed)

Pre-1923: 4 (6.5% of pre-1923 GBS volumes reviewed)

[entries 71, 91, 113, 138]

Post-1922: none

OCR misreadings due to hyphenation between lines

Number of volumes in which searching was inaccurate because OCR read hyphenated words between lines as separate words: 12 (15% of all GBS volumes reviewed)

Pre-1923: 7 (11.3% of pre-1923 GBS volumes reviewed)

[entries 11, 21, 96, 101, 108, 123, 126]

Post-1922: 5 (27.8% of post-1922 GBS volumes reviewed)

[entries 136, 171, 173, 186, 193]

OCR misreadings due to causes other than those surveyed above

Number of volumes in which searching was inaccurate because OCR misread words for reasons other than those surveyed above: 19 (23.75% of all GBS volumes reviewed)

Pre-1923: 13 (21% of pre-1923 GBS volumes reviewed)

[entries 41, 46, 73, 78, 96, 98, 116, 121, 126, 138, 141, 198, 201]

Post-1922: 6 (33.3% of post-1922 GBS volumes reviewed)

[entries 23, 33, 168, 183, 186, 196]

II. MICROSOFT LIVE SEARCH BOOKS (MLSB)

1) NUMBERS OF TITLES AND VOLUMES SEARCHED AND SELECTED

Titles searched

Number of titles searched: 347

Pre-1923: 122

Post-1922: 225

Number of titles available in MLSB

full text or preview: 106 (30.5% of all titles)

full text: 75 (21.6% of all titles)

preview: 31 (8.9% of all titles)

no page digitizations available: 241 (69.4% of all titles)

Number of pre-1923 titles available in MLSB

full text or preview: 76 (62.3% of pre-1923 titles)

full text: 75 (61.5% of pre-1923 titles)

preview (for later edition): 1

no digitizations available: 46 (37.7% of pre-1923 titles)

Number of post-1922 titles available in MLSB

preview: 30 (13.3% of post-1922 titles)

no digitizations available: 195 (86.7% of post-1922 titles)

Note: Most of the MLSB preview digitizations for post-1922 titles were published since 1981. Very few were published prior to 1981. Of 113 titles published since 1981, MLSB provided preview digitizations for 27 titles, or 23.9%. Of the 112 titles published between 1923 and 1980, MLSB provided preview digitizations for only 3 titles, or 2.7%.

Volumes searched

Number of volumes searched: 608

Number of volumes searched available in MLSB

full text or preview: 253 (41.6% of all volumes)

full text: 225 (37% of all volumes)

preview: 28 (4.6% of all volumes)

no page digitizations available: 355 (58.4% of all volumes)

Number of volumes reviewed

Total: 80

Pre-1923: 55

[entries 2, 4, 9, 12, 17, 19, 22, 32, 37, 39, 44, 47, 49, 57, 62, 64, 69, 72, 77, 79, 82, 87, 89, 92, 94, 97, 99, 102, 104, 107, 112, 114, 117, 147, 149, 152, 154, 157, 159, 162, 164, 167, 169, 172, 174, 177, 179, 182, 184, 187, 189, 192, 194, 197, 199]

Post 1922: 25

[entries 7, 14, 24, 27, 29, 34, 42, 52, 54, 59, 67, 74, 84, 109, 119, 122, 124, 127, 129, 132, 134, 137, 139, 142, 144]

2) ACCESS, DOWNLOADING, SHARING, AND NAVIGATION

In late January 2008, after reviewing volumes for one month, I found that I could no longer advance to the next page of a digitization in MLSB. I tried reloading the site and received an error message. Thereafter, I no longer could access MLSB at all on my home computer without receiving a similar error message, though I was able to access the site on a public computer terminal at a nearby library and continue the examination of digitizations. I eventually found a way to again access MLSB on my home computer by first searching for a book title in the main page of Microsoft Live Search (of which MLSB is a part), clicking on a "see also" link on the retrieved results page listing websites that mention the title, and finally clicking on a "Books" link on the subsequent page retrieved. This roundabout method worked only for a few weeks, after which I again began to receive error messages, though I still was able to access the site from a library's public computer terminal. I communicated my problem to a browser support site and received some suggestions, none of which resolved the problem. Two people viewing my communication wrote to say that they had encountered the same problem. I had a web chat with a volunteer from a browser service, who also duplicated the problem on his or her terminal. Whatever was causing the problem, thus, was not uniquely situated in my home computer. In fact, I connected another computer to my DSL line and the problem remained. In March 2008, during the final stages of the project, I tried again to access the site and this time succeeded. Whatever the reason that I temporarily was unable to access MLSB, or that I am now, perhaps temporarily, able to access the site, the incident highlights the need in the interest of open scholarship for all non-subscription mass digitization sites to remain accessible to all users at all times. Lack of access without explanation led to wasted time.

Ease of finding specific volumes in multivolume works

Number of titles for which title or author/title searches did not retrieve desired volume: 9
(11.25% of all MLSB volumes reviewed)

Pre-1923: 9 (16.4% of pre-1923 MLSB volumes reviewed)

[entries 12, 19, 22, 49, 69, 97, 114, 147, 162]

Post-1922: none

Finding specific volumes of multivolume works can take a lot of time. Title searches retrieve hits that often do not list volumes numbers. Users must click on each listing and page to the title page before finding out if the choice is the volume desired.

For example, an MLSB keyword search for “Works of John C. Calhoun” (see entry 162) yielded 50 hits. The first listing, *Correspondence of John C. Calhoun*, was not part of the “Works,” yet was retrieved because the search phrase occurred within that book. Hits 2 through 8, 10-11, and 13 were all volumes of the “Works” set, yet the listings did not indicate which volume numbers corresponded to the hits.

An MLSB keyword search for “Henry Adams” AND “History of the United States” (see entry 49) yielded 212 hits, most of which were works that were not volumes from Adams’s history, but works in which the title was mentioned. The hits for volumes of Adams’s work were: 3-6, 14-16, 18, 21-22, 24, 26, 32, 37-38, 41-42, 46, 49, 51, 53, 59, 68, 74, 83, 99, and 120. For most of these hits, the bibliographic information did not include volume numbers. In MLSB, as in GBS, users have no way to order hits for relevance.

OPAC records

OPAC records did not contain links for retrieval of MLSB digitizations.

MARC records

MARC records were not available for GBS digitizations.

Downloading, searching PDF files, and copying text

Unlike in GBS, PDF files downloaded from MLSB include the image page number in the page indicator, with the physical page number in parenthesis. For example “31 (54 of 563),” where 31 is the image page number, 54 the physical page number, and 563 the total number of physical pages.

Unlike in GBS, users can search for text in the downloaded PDF file from MLSB.

Users can copy selections from the image and selections from the OCR reading of downloaded text. In GBS, users can copy selections from the downloaded image only.

Sharing results

MLSB has no options for sharing results as in GBS.

Navigation

Unlike in GBS, users do not have the option to input a desired page number in the page indicator. Users can move page-by-page through clicking forward and backward buttons, move to the first physical page and the last physical page through clicking buttons, and move a marker to specific places on a vertical ruler to the right of the page image, but the

ruler is not marked by page numbers, only by unmarked hash marks. Unlike in GBS, users do not have the option of viewing double-page displays on one screen.

3) PHYSICAL CONDITION OF VOLUMES REVIEWED

Number of volumes with markings, folds, splotches, tears, stains, smudges, broken letters, or tape: 43 (53.75% of all MLSB volumes reviewed)

Pre-1923: 43 (78.2% of pre-1923 MLSB volumes reviewed)

[entries 2, 4, 9, 12, 17, 19, 32, 37, 39, 42, 44, 47, 49, 57, 62, 64, 69, 72, 77, 79, 87, 92, 94, 99, 102, 112, 117, 147, 149, 152, 154, 157, 159, 162, 164, 172, 174, 177, 179, 184, 187, 189, 197]

Post-1922: none

Number of volumes with markings, folds, splotches, tears, stains, smudges, broken letters, or tape affecting 10 or more pages: 21 (26.25% of all MLSB volumes reviewed)

Pre-1923: 21 (38.2% of pre-1923 MLSB volumes reviewed)

[entries 2, 12, 47, 49, 57, 62, 69, 72, 77, 87, 94, 117, 147, 149, 159, 162, 172, 177, 179, 187, 189]

Post-1922: none

Number of volumes with markings, folds, splotches, tears, stains, smudges, broken letters, or tape affecting OCR: 30 (37.5% of all MLSB volumes reviewed)

Pre-1923: 29 (52.7% of pre-1923 MLSB volumes reviewed)

[entries 2, 4, 9, 12, 17, 19, 32, 44, 47, 49, 62, 64, 69, 72, 77, 87, 92, 94, 117, 147, 152, 154, 157, 159, 162, 172, 177, 187, 189]

Post-1922: 1 (4% of post-1922 MLSB volumes reviewed)

[entry 42]

Number of volumes with markings, folds, splotches, tears, stains, smudges, broken letters, or tape affecting OCR on 10 or more pages: 3 (3.75% of all MLSB volumes reviewed)

Pre-1923: 3 (5.5% of pre-1923 MLSB volumes reviewed)

[entries 72, 162, 187]

Post-1922: none

4) METADATA

Author, title, and publication listings

Number of volumes with incorrect author, title, or publication information: 3 (3.75% of all MLSB volumes reviewed)

Pre-1923: 3 (5.5% of pre-1923 MLSB volumes reviewed)

[entries 9, 82, 102]

Post-1922: none

The titles of volumes in the bibliographic information to the left of the displayed volume systematically were cut-off in mid-word or otherwise truncated due to space limitations. The full title, however, could be found in a small box appearing on the screen as the cursor is placed on the title.

[examples: entries 14, 17, 19, 22, 27, 32, 34]

Contents links

Number of volumes with problematic contents links: 17 (21.25% of all MLSB volumes reviewed)

Pre-1923: 15 (27.2% of pre-1923 MLSB volumes reviewed)

[entries 9, 12, 17, 22, 32, 37, 39, 82, 87, 89, 94, 99, 102, 112, 157]

Post-1922: 2 (8% of post-1922 MLSB volumes reviewed)

[entries 27, 67]

Number of volumes with contents links to incorrect pages: 14 (17.5% of all MLSB volumes reviewed)

Pre-1923: 12 (21.8% of pre-1923 MLSB volumes reviewed)

[entries 17, 22, 32, 37, 82, 87, 89, 94, 99, 102, 112, 157]

Post-1922: 2 (8% of post-1922 MLSB volumes reviewed)

[entries 27, 67]

Number of multi-volume works with contents links that lead to incorrect volumes: 2 (2.5% of all MLSB volumes reviewed)

Pre-1923: 2 (3.6% of pre-1923 MLSB volumes reviewed)

[entries 9, 12]

Post-1922: none

Number of volumes with contents links to numbers other than page numbers (dates, section numbers, Roman numerals) that were useless: 3 (3.75% of all MLSB volumes reviewed)

Pre-1923: 3 (5.5% of pre-1923 MLSB volumes reviewed)

[entries 39, 89, 102]

Post-1922: none

Index links

Number of volumes with incorrect index links: 1 (1.25% of all MLSB volumes reviewed)

Pre-1923: 1 (1.8% of pre-1923 MLSB volumes reviewed)

[entry 19]

Post-1922: none

Number of pages

Number of volumes with listings for number of pages different than that given in MARC records (LC catalog, Harvard Library's Hollis Catalog) or last page of text: 75 (93.75% of all MLSB volumes reviewed)

Pre-1923: 52 (94.5% of pre-1923 MLSB volumes reviewed)

[entries 2, 4, 9, 12, 17, 19, 32, 37, 39, 47, 49, 57, 62, 64, 69, 72, 77, 79, 82, 87, 89, 92, 94, 97, 99, 104, 107, 112, 114, 117, 147, 149, 152, 154, 157, 159, 162, 164, 167, 169, 172, 174, 177, 179, 182, 184, 187, 189, 192, 194, 197, 199]

Post-1922: 23 (92% of post-1922 MLSB volumes reviewed)

[entries 7, 14, 27, 29, 34, 42, 52, 54, 59, 67, 74, 109, 119, 122, 124, 127, 129, 132, 134, 137, 139, 142, 144]

Page indicators

Number of volumes with discrepancy between page indicator and page number in text: 3 (3.75% of all MLSB volumes reviewed)

Pre-1923: 3 (5.5% of pre-1923 MLSB volumes reviewed)

[entries 184, 189, 192]

Post-1922: none

Subject headings

After a user performs a basic search for a book (the only search available in MLSB), results appear on the left side of the screen, with titles of books, a small thumbnail image, and limited additional information, usually including the author, year of publication, number of page numbers, percentage of pages viewable, and quoted text in which the word search appears. When a user places the cursor over the entry on the left side, a larger image appears on the right side with a limited amount of bibliographic information including the title, author, subject headings, publisher, and publication date. Unlike in GBS, the subject headings are not hyperlinked to other digitizations. Users can search subjects in the basic search option to retrieve digitizations of other books with the same subjects, but cannot order the results list according to author, title, or date, as in library catalogs.

Number of volumes that use LC subject headings, without subdivisions: 2 (2.5% of all MLSB volumes reviewed)

Pre-1923: 1 (1.8% of pre-1923 MLSB volumes reviewed)

[entry 184]

Post-1922: 1 (1.25% of post-1922 MLSB volumes reviewed)

[entry 59]

Number of volumes that use complete LC subject headings, but truncate them by number of characters: 5 (6.25% of all MLSB volumes reviewed)

Pre-1923: 5 (6.25% of pre-1923 MLSB volumes reviewed)

[entries 9, 69, 167, 182, 194]

Post-1922: none

Note: While the LC subject headings were truncated in MLSB displays, searching for LC-assigned headings that had been left off of displays did retrieve works in question. For example, for entry 9 (Franklin, Benjamin. *The Works of Benjamin Franklin; Containing Several Political and Historical Tracts Not Included in Any Former Edition, and Many Letters, Official and Private Not Hitherto Published; with Notes and a Life of the Author by Jared Sparks*. 10 vols. Boston, Hilliard, Gray, Co., 1840. [Vol. I]), MLSB reproduces the first two subjects listed by LC (“United States—Foreign Relations—1775-1783”; “United States—Politics and government—To 1775”), with an ellipsis following, thus leaving out two remaining LC subjects (“United States—Politics and government—1775-1783”; “United States—Politics and government—1783-1789”). While a basic search for the latter two subjects, did retrieve this title, users who do not consult other library catalogs for a text’s subject headings will not be aware of the omitted headings and thus will not be able to find other relevant texts bearing the same or similar headings.

Number of volumes that either augmented LC subject headings; used some LC headings, omitted others, and used headings not derived from LC; used no subject headings even though LC provided some; or disjointed LC subject headings so that subdivisions became separate headings: 30 (37.5% of all MLSB volumes reviewed)

Pre-1923: 11 (20% of pre-1923 MLSB volumes reviewed)

[entries 12, 17, 37, 39, 49, 77, 87, 149, 172, 177, 189]

Post-1922: 19 (76% of post-1922 MLSB volumes reviewed)

[entries 7, 24, 34, 42, 52, 54, 59, 67, 84, 109, 119, 122, 127, 129, 134, 137, 139, 142, 144]

Note: Often the non-LC headings appear to be headings applied to the books for marketing purposes. For example, for entry 24 (Capper, Charles. *Margaret Fuller: An American Romantic Life*. Vol. 1, *The Private Years*. New York : Oxford University Press, 1992), LC lists 3 subjects: “Fuller, Margaret, 1810-1850”; “Authors, American—19th century—Biography”; and “Feminists—United States—Biography.” MLSB lists only one subject: “BIOGRAPHY AND AUTOBIOGRAPHY / Women,” a classification designation used by Powell’s Books, among other book sellers.

For entry 52 (Fitzhugh, George. *Cannibals All! or, Slaves without Masters*. Edited by C. Vann Woodward. Cambridge, Belknap Press of Harvard University Press, 1960), LC lists “Slavery—United States”; “Slavery—Justification”; and “Working class,” while MLSB lists “HISTORY / General,” a classification designation used by Amazon.com.

5) QUALITY OF SCANS

Number of volumes rated “poor”: 2 (2.5% of all MLSB volumes reviewed)

Pre-1923: 2 (3.6% of pre-1923 MLSB volumes reviewed)

[entries 99, 192]

Post-1922: none

Missing pages

Number of volumes with missing pages: 2 (2.5% of all MLSB volumes reviewed)

Pre-1923: 2 (3.6% of pre-1923 MLSB volumes reviewed)

[entries 17, 99]

Post-1922: none

Pages appearing more than once

Number of volumes with one or more pages displayed more than one time: 1 (1.25% of all MLSB volumes reviewed)

Pre-1923: 1 (1.8% of pre-1923 MLSB volumes reviewed)

[entries 17]

Post-1922: none

Pages with portions cut off

Number of volumes with portions of one or more page cut off: 2 (2.5% of all MLSB volumes reviewed)

Pre-1923: 2 (3.6% of pre-1923 MLSB volumes reviewed)

[entries 49, 147]

Post-1922: none

Light, dark, or blurry pages, affecting OCR readings adversely

Number of volumes with light, dark, or blurry pages, parts of pages, or letters, affecting OCR readings adversely: 8 (10% of all MLSB volumes reviewed)

Pre-1923: 7 (12.7% of pre-1923 MLSB volumes reviewed)

[entries 47, 49, 72, 92, 184, 194, 199]

Post-1922: 1 (4% of post-1922 MLSB volumes reviewed)

[entry 7]

Scanned image quality

Number of volumes in which letters were not sharp: 4 (5% of all MLSB volumes reviewed)

Pre-1923: 2 (3.6% of pre-1923 MLSB volumes reviewed)

[entries 177, 187]

Post-1922: 2 (8% of post-1922 MLSB volumes reviewed)

[entries 122, 139]

6) ACCURACY OF WORD SEARCHING WITHIN TEXTS

Sorting results of word searches by “relevance”

Searching of words and phrases purports to sort results according to relevance, but MLSB does not indicate the way that relevance is determined.

For entry 127 (Bremer, Francis J. *John Winthrop: America's Forgotten Founding Father*. New York: Oxford University Press, 2003), a search for "arminianism" placed p. 95 next to last in a search results list of 12 pages in which the term occurred even though the term was first introduced on p. 95 and its historical context was explained on that page. The second entry in the list of relevant pages was a page from the notes in which the title of a book that includes the word "arminianism" was listed; this note pertained to p. 95. It is unclear why the text of p. 95 was considered by MLSB less relevant than a citation pertaining to that page with regard to the subject matter discussed on that page.

For entry 199 (*The Encyclopedia of Social Reform*. Ed. William D. P. Bliss. New York: Funk & Wagnalls, 1897), a search for "Atkinson" retrieved 30 pages. The entry in the encyclopedia for "Atkinson, Edward," certainly the most relevant of the listed results, appeared 19th in the list of 30. A search for "socialism" retrieved 344 pages; the entry in the encyclopedia for "Socialism" was 8th in the list. A search for "individualism" retrieved 84 pages; the entry for "Individualism" was 3rd in the list. The results list did not indicate which of the pages listed included the actual encyclopedia entries for the terms searched.

It would be more relevant to scholars to list the pages in which search terms occur in page number order, as is the case with HEB. Authors often develop their discussions within narratives. By ordering search results by page number order, users can easily find the first usage of a search term and then click on subsequent pages in order to learn how an author builds his or her discussion of the term.

Basic searches only allowed

Unlike GBS and HEB, MLSB provides only basic searching capabilities. A basic search for "The Works of Jonathan Edwards" retrieved 160 results, which included works with that title and works in which that title is mentioned.

Truncation of search terms

HEB allows users to truncate search terms with the use of an asterisk. This type of truncation for searching would be useful for scholars if it was incorporated into MLSB.

OCR misreadings

Number of volumes in which OCR reading was affected adversely: 75 (93.75% of all MLSB volumes reviewed)

Pre-1923: 53 (96.4% of pre-1923 MLSB volumes reviewed)

[entries 2, 4, 9, 12, 17, 19, 32, 37, 39, 44, 47, 49, 62, 64, 69, 72, 77, 79, 82, 87, 89, 92, 94, 97, 99, 102, 104, 107, 112, 114, 117, 147, 149, 152, 154, 157, 159, 162, 164, 167, 169, 172, 174, 177, 179, 182, 184, 187, 189, 192, 194, 197, 199]

Post-1922: 22 (88% of post-1922 MLSB volumes reviewed)
[entries 7, 14, 27, 29, 34, 42, 52, 54, 59, 67, 74, 84, 119, 122, 127, 129, 132,
134, 137, 139, 142, 144]

OCR misreading of dates

Number of volumes in which searching was inaccurate because OCR misread the numeral “1” as “l” in dates: 1 (1.25% of all MLSB volumes reviewed)

Pre-1923: none

Post-1922: 1 (4% of post-1922 MLSB volumes reviewed)
[entry 14]

OCR misreadings due to footnote indicators

Number of volumes in which searching was inaccurate because footnote indicator were read by OCR as parts of words: 1 (1.25% of all MLSB volumes reviewed)

Pre-1923: none

Post-1922: 1 (4% of post-1922 MLSB volumes reviewed)
[entry 122]

OCR misreadings due to run-together words

Number of volumes in which searching was inaccurate because OCR read separate words as one word: 3 (3.75% of all MLSB volumes reviewed)

Pre-1923: none

Post-1922: 3 (12% of post-1922 MLSB volumes reviewed)
[entries 29, 122, 127]

OCR misreadings due to italicized words

Number of volumes in which searching was inaccurate because OCR read italicized words incorrectly: 19 (23.75% of all MLSB volumes reviewed)

Pre-1923: 12 (21.8% of pre-1923 MLSB volumes reviewed)

[entries 19, 64, 69, 87, 92, 94, 112, 149, 174, 182, 189, 199]

Post-1922: 7 (28% of post-1922 MLSB volumes reviewed)
[entries 52, 54, 67, 119, 129, 134, 139]

OCR misreadings in footnotes

Number of volumes in which searching was inaccurate because OCR read words in footnotes incorrectly: 29 (36.25% of all MLSB volumes reviewed)

Pre-1923: 27 (49.1% of pre-1923 MLSB volumes reviewed)

[entries 19, 32, 37, 44, 47, 64, 69, 79, 82, 87, 89, 92, 104, 107, 112, 114, 147,
154, 157, 162, 167, 174, 177, 179, 187, 192, 194]

Post-1922: 2 (8% of post-1922 MLSB volumes reviewed)
[entries 7, 67]

OCR misreadings due to words in all caps

Number of volumes in which searching was inaccurate because OCR read words in all caps incorrectly: 4 (5% of all MLSB volumes reviewed)

Pre-1923: 4 (7.3% of pre-1923 MLSB volumes reviewed)

[entries 99, 107, 172, 192]

Post-1922: none

OCR misreadings due to words with capitalized first letters

Number of volumes in which searching was inaccurate because OCR read words in which the first letter was capitalized incorrectly: 5 (6.25% of all MLSB volumes reviewed)

Pre-1923: 5 (9.1% of pre-1923 MLSB volumes reviewed)

[entries 79, 162, 167, 169, 194]

Post-1922: none

OCR misreadings due to words in languages other than English

Number of volumes in which searching was inaccurate because OCR read words in languages other than English incorrectly: 14 (17.5% of all MLSB volumes reviewed)

Pre-1923: 13 (23.6% of pre-1923 MLSB volumes reviewed)

[entries 64, 69, 77, 87, 92, 94, 97, 107, 149, 172, 174, 179, 197]

Post-1922: 1 (4% of post-1922 MLSB volumes reviewed)

[entry 119]

OCR misreadings in Index pages

Number of volumes in which searching was inaccurate in Index pages: 22 (27.5% of all MLSB volumes reviewed)

Pre-1923: 11 (20% of pre-1923 MLSB volumes reviewed)

[entries 39, 49, 62, 87, 97, 114, 149, 152, 159, 167, 184]

Post-1922: 11 (44% of post-1922 MLSB volumes reviewed)

[entries 27, 34, 42, 67, 74, 84, 127, 132, 134, 137, 144]

OCR misreadings due to hyphenation between lines

Number of volumes in which searching was inaccurate because OCR read hyphenated words between lines as separate words: 39 (48.75% of all MLSB volumes reviewed)

Pre-1923: 26 (47.3% of pre-1923 MLSB volumes reviewed)

[entries 4, 32, 37, 39, 44, 62, 79, 82, 87, 89, 92, 97, 102, 104, 117, 149, 152, 154, 167, 177, 179, 182, 184, 187, 197, 199]

Post-1922: 13 (52% of post-1922 MLSB volumes reviewed)

[entries 14, 29, 34, 59, 84, 119, 127, 129, 132, 134, 137, 142, 144]

OCR misreadings due to causes other than those surveyed above

Number of volumes in which searching was inaccurate because OCR misread words for reasons other than those surveyed above: 26 (32.5% of all MLSB volumes reviewed)

Pre-1923: 17 (30.9% of pre-1923 MLSB volumes reviewed)

[entries 2, 12, 19, 37, 47, 77, 94, 112, 114, 164, 177, 182, 184, 187, 189, 194, 197]

Post-1922: 9 (36% of post-1922 MLSB volumes reviewed)

[entries 14, 27, 34, 52, 127, 129, 134, 142, 144]

III. ACLS HUMANITIES E-BOOK (HEB)

1) NUMBERS OF TITLES AND VOLUMES SEARCHED AND SELECTED

Titles searched

Number of titles searched: 347

Pre-1923: 122

Post-1922: 225

Number of titles searched available in HEB

full text: 41 (11.8% of all titles)

no page digitizations available: 306 (88.2% of all titles)

Number of pre-1923 titles searched available in HEB

full text: none

no digitizations available: 122 (100%)

Number of post-1922 titles searched available in HEB

full text: 41 (18.2% of post-1922 titles)

no digitizations available: 184 (81.8% of post-1922 titles)

Note: Of 158 titles published since 1981, HEB provided digitizations for 38 titles, or 24.1%. Of the 67 titles published between 1923 and 1980, HEB provided digitizations for only 3 titles, or 4.5%.

Number of volumes reviewed

Total: 39 (all post-1922)

[entries 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200]

2) ACCESS, DOWNLOADING, SHARING, AND NAVIGATION

Ease of finding specific volumes in multivolume works

Number of titles for which title or author/title searches did not retrieve desired volume:
none.

OPAC records

Number of volumes for which OPAC did not contain links for retrieval of records: none

MARC records

According to the HEB website, “MARC records are available free to subscribers. Once a subscription is set up, the contacts at the subscribing institution will be sent download instructions. The records are MARC21, USMARC records prepared by the University of Michigan Libraries.”

Downloading, searching PDF files, and copying text.

Users cannot download text from HEB. Users can print individual pages.

Sharing results

HEB has no options for sharing results as in GBS.

3) PHYSICAL CONDITION OF VOLUMES REVIEWED

Number of volumes with markings, folds, splotches, tears, stains, smudges, broken letters, or tape: 4 (10.3% of all HEB volumes reviewed).
[entries 30, 40, 45, 125]

Number of volumes with markings, folds, splotches, tears, stains, smudges, broken letters, or tape affecting 10 or more pages: 1 (2.6% of all HEB volumes reviewed).
[entry 40]

Number of volumes with markings, folds, splotches, tears, stains, smudges, broken letters, or tape affecting OCR: 4 (10.3% of all HEB volumes reviewed).
[entries 30, 40, 45, 125]

Number of volumes with markings, folds, splotches, tears, stains, smudges, broken letters, or tape affecting OCR on 10 or more pages: 1 (2.6% of all HEB volumes reviewed).
[entry 40]

4) METADATA

Author, title, and publication listings

Number of volumes with incorrect author, title, and publication information: 1 (2.6% of all HEB volumes reviewed).
[entry 150]

Contents links

Number of volumes with problematic contents links: none

Index links

Number of volumes with incorrect index links: none

Number of pages

Number of volumes with listings for number of pages different than that given in MARC records (LC catalog, Harvard Library's Hollis Catalog) or last page of text: none

Page indicators

Number of volumes with discrepancy between page indicator and page number in text: 1
(2.6% of all HEB volumes reviewed).

[entry 200]

Note: See also Appendix B. In two entries (204C, 205C), all pages were mislabeled in the page indicator, which resulted in problematic searching.

Subject headings

Number of volumes that used LC main subject headings without subdivisions provided by LC: none

5) QUALITY OF SCANS

Number of volumes rated "poor": none.

Missing pages

Number of volumes with missing pages: none.

Pages out-of-order

Number of volumes with pages displayed out-of-order: none.

Pages appearing more than once

Number of volumes with one or more pages displayed more than one time: none.

Illegible pages

Number of volumes with one or more illegible pages: none.

Pages with portions cut off or obscured

Number of volumes with portions of one or more page cut off or obscured: none.

Pages with distorted lines

Number of volumes with distorted lines: none.

Light, dark, or blurry pages, affecting OCR readings adversely

Number of volumes with light, dark, or blurry pages, parts of pages, or letters, affecting OCR readings adversely: 1 (2.6% of all HEB volumes reviewed)
[entries 35]

Scanned image quality

Number of volumes in which letters were not sharp: 5 (12.8% of all HEB volumes reviewed)
[entries 30, 50, 55, 140, 150]

6) ACCURACY OF WORD SEARCHING WITHIN TEXTS

Each search result indicated one more occurrence than was listed.

For example, in entry 5 (Kloppenber, James T. *Uncertain Victory: Social Democracy and Progressivism in European and American Thought, 1870-1920*. New York: Oxford University Press, 1988), a search for the phrase “American exceptionalism” cited “1 to 2 of 2 matching terms in this item,” then listed only one occurrence, not two. A search for “Henry George” cited “1 to 14 of 14 matching terms,” but listed only 13 pages in which “Henry George” appeared, and within those 13 pages, 15 occurrences.

OCR misreadings

Number of volumes in which OCR reading was affected adversely: 31 (79.5% of all HEB volumes reviewed)
[entries 5, 15, 20, 25, 30, 35, 40, 45, 50, 55, 70, 75, 80, 90, 95, 105, 110, 115, 120, 125, 130, 135, 140, 155, 160, 165, 180, 185, 190, 195, 200]

OCR misreading of dates

Number of volumes in which searching was inaccurate because OCR misread the numeral “1” as “I” in dates: 19 (48.7% of all HEB volumes reviewed).
[entries 5, 15, 25, 30, 35, 40, 45, 50, 55, 70, 75, 95, 130, 155, 160, 165, 180, 185, 190]

OCR misreadings due to footnote indicators

Number of volumes in which searching was inaccurate because footnote indicator were read by OCR as parts of words: 3 (7.7% of all HEB volumes reviewed).

[entries 90, 160, 195]

OCR misreadings due to run-together words

Number of volumes in which searching was inaccurate because OCR read separate words as one word: 8 (20.5% of all HEB volumes reviewed).

[entries 20, 30, 80, 95, 105, 110, 115, 120]

OCR misreadings due to italicized words

Number of volumes in which searching was inaccurate because OCR read italicized words incorrectly: 5 (12.8% of all HEB volumes reviewed).

[entries 35, 55, 135, 140, 200]

OCR misreadings in footnotes

Number of volumes in which searching was inaccurate because OCR read words in footnotes incorrectly: 1 (2.6% of all HEB volumes reviewed).

[entry 20]

OCR misreadings due to words in all caps

Number of volumes in which searching was inaccurate because OCR read words in all caps incorrectly: none

OCR misreadings due to words with capitalized first letters

Number of volumes in which searching was inaccurate because OCR read words in which the first letter was capitalized incorrectly: none

OCR misreadings due to words in languages other than English

Number of volumes in which searching was inaccurate because OCR read words in languages other than English incorrectly: none

OCR misreadings in Index pages

Number of volumes in which searching was inaccurate in Index pages: 1 (2.6% of all HEB volumes reviewed).

[entry 25]

OCR misreadings due to hyphenation between lines

Number of volumes in which searching was inaccurate because OCR read hyphenated words between lines as separate words: 5 (12.8% of all HEB volumes reviewed)
[entries 20, 25, 35, 55, 180]

OCR misreadings due to causes other than those surveyed above

Number of volumes in which searching was inaccurate because OCR misread words for reasons other than those surveyed above: 4 (10.3% of all HEB volumes reviewed).
[entries 35, 155, 165, 195]