Title: Google Book Search coverage of core clinical textbooks

E. Diane Johnson, J. Otto Lottes Health Sciences Library, University of Missouri, Columbia, Missouri USA

Abstract
Rationale: Google has entered into book scanning agreements with prestigious libraries in eight countries, and search results reveal evidence of participation by commercial publishers as well. Since every word in the text is searchable, Google Book Search could become a powerful tool for finding precise answers to highly specific clinical questions. Although the exact size and coverage of Google Book Search will likely remain a trade secret, medical librarians need to assess its coverage and currency to determine its role in answering clinical questions.

Objective: To determine the extent of coverage of core medical textbooks in Google Book Search.

Methods: Titles and ISBNs for the most recent editions of the 87 minimal core clinical titles on the 2003 Brandon-Hill List were searched in Google Book Search to see if each is available. Titles that were listed were further classified as how much of the text can be viewed: full view, limited preview, snippet view, or no preview available.

Results: All 87 of the minimal core titles were listed in Google Book Search, with the most recent edition listed for all but 3 titles. Of the 84 current editions, 64% of the books were searchable, with 37 of the books providing a limited preview display, and 17 books showing a snippet display. The remaining 30 books had no preview available and were not searchable.

Conclusions: With 64% of the most current editions of 84 major clinical texts fully searchable, Google Book Search provides librarians with a new and powerful search tool for locating information in clinical textbooks. The “Find in a Library” link on the Google Book Search display provides a new way to leverage and expose libraries’ print textbook collections.

Introduction
Google has entered into book scanning agreements with prestigious libraries in eight countries, and search results reveal evidence of participation by commercial publishers as well. Since every word in the text is searchable, Google Book Search could become a powerful tool for finding precise answers to highly specific clinical questions. Although the exact size and coverage of Google Book Search will likely remain a trade secret, medical librarians need to
assess its coverage and currency to determine its role in answering clinical questions.

Google’s library partners include some of the world’s largest research libraries. Countries with participating libraries include Germany, Belgium, Japan, France, Spain, the United Kingdom, and Switzerland. Partner libraries in the United States include Harvard University, Stanford University, New York Public Library, University of Michigan, and many more. (1)

Each library has a different agreement with Google dictating which books it allows to be digitized. For example, the Oxford University agreement calls for digitizing more than 1 million books from the Bodleian Library over a three year period. The agreement covers only public domain books – books for which the copyright has expired; most published before 1920. Two copies of each book will be created: One fully indexed and searchable through Google, and the other linked from the library catalog. (2) In contrast, the University of Michigan agreement targets digitizing all of the print books in their collection; all would be searchable on Google, but the amount of full text displayed would be dictated by copyright. In addition, the publisher or rights holder can request that an item scanned at a library be removed. (3)

Google is also reaching out to publishers and other rights holders through its Google Books Partner Program. Participants can upload their content and track activity on the pages for the books they submit. (4)

**Google Book Search vs Google Search**

Before delving into the specifics of Google Book Search, it’s important to clarify that Google Book Search returns different results than the regular Google Search box. When you go to books.google.com and do a Google Book Search, you are “directly searching the text of over 7 million books”. (5) According to a member of the Google Book Search Online Team, these search results are only retrieved on regular Google searches at www.google.com “when we think it fits the query”. (5)

**Display and search**

Google Book Search presents books in a variety of ways(6), presumably based upon the agreement with the provider of the scanned material. Books may be fully available, where every word in the text can be searched, viewed, and printed, and the entire book is available for download. This option is most common with old books in the public domain and is called “full view”. Another option is “limited preview”. Every word in the full text of these books can be searched, but only portions of the text can be displayed. Typically, the reader can display several pages of the full text before reaching the allowable display limit. Certain pages or images may be omitted. However, on these books Google will list the page number(s) containing the search terms even if the pages can’t be displayed. If you can locate a paper copy of the book, you will know exactly where to find your search terms. Furthermore, by clicking on “find this book in a library”, the searcher can locate the nearest copy. Enter a country or a city to discover how many miles
you are from the closest library known to have a copy of the book. Find this book in a library uses the OCLC WorldCat database, which contains well over 100 million items in the collections of more than 71,000 participating libraries. While most of the libraries are in the United States, a total of 112 countries and territories worldwide participate in OCLC WorldCat. As of early 2009, WorldCat contains 76.2 million holdings from non-U.S. libraries. This year, plans call for adding more than 250 million additional holdings until the national libraries of 40 countries are represented.

Some of the titles in Google Book Search are available in “snippet view”, which allows viewing of a few sentences surrounding the search terms entered. As with limited preview titles, snippet view will show the page numbers containing the search terms entered, so that once you’ve got a paper copy of the book, you can quickly locate your search terms in the relevant portions of the text.

Finally, some of the titles included in Google Book Search have no preview capability at all. There also is no option to search inside these books. It appears that “no preview” titles are included on Google Book Search results only if all search terms are found in the title of the book, or in access points typically found at the level of indexing provided by a catalog record - that is, the complete text is not searchable.

Google Book Search provides an advanced search capability which supports author, title, ISBN and subject searching. Results can be limited by publication date. Limits are also available for 45 different languages. It is also possible to restrict search results to display only full view and/or limited preview titles.

**Methodology**

In this study, 87 core clinical medical textbooks were searched in Google Book Search to determine the level of search and display available. With respect to full text searchability, “Search inside the book” functionality was available for any books available in full view, limited preview, or snippet view.

The clinical textbooks searched were the minimal core titles on the Brandon Hill list selected list of print books for the small medical library, last published in 2003. Only books in clinical subject areas were examined for this study. Books listed under topics such as anatomy, dictionaries/directories, ethics, medical informatics, statistics and other non-clinical areas were omitted from the study. Within the clinical areas, the most recent edition of all of the titles that carried a “minimal core” designation were included in the study. WorldCat was used to determine the most recent edition of each book. It also served as a source of ISBNs. Each book was searched both by title and by all ISBNs associated with the most recent edition. If the current edition was not found in Google Book Search, the search was expanded to previous editions, and the latest edition found in Google Book Search was recorded. Searches were conducted in the first quarter of 2009.
Results

All 87 of the minimal core clinical titles were listed in Google Book Search, with the most recent edition listed for all but 3 titles. None of the 87 books had a full view of the latest edition available. However, 37 of the 84 current edition books (44%) had a limited preview, and an additional 17 (20%) included a snippet view. The remaining 30 (36%) had no preview available. There were no patterns discerned about which subjects had fuller displays than others. Some publishers were more liberal in the amount of text they were willing to display than others, although even this wasn’t completely consistent for a given publisher.

Surprisingly, many of the books contained more than one listing in Google Book Search for the current edition, and the level of search and display varied even for the same edition of a single book. For example, the latest edition of Scriven’s Metabolic and Molecular Bases of Inherited Disease was listed 8 times. Seven of the entries had no preview available; one entry had a snippet view.

For the purposes of Figure 1a, each of the 84 current edition titles was only counted once, and was recorded under the most liberal level of access available for the latest edition. So for example Scriven’s Metabolic and Molecular Bases of Inherited Disease was counted with the snippet view books.

Quite often when the same edition had varying levels of access, e.g. no preview and snippet view, the snippet view came from a library copy - most often from the University of Michigan. Unlike some of the other libraries, which only allowed Google to scan material in the public domain, Michigan agreed to let Google scan everything in their print collection. Ten of the books had both “no preview” and snippet views available.

Five translations were found for current editions; two in Spanish, two in Chinese, and one in Portuguese. No previews were available for four of the books. The remaining one, in Spanish, had a limited preview available.

In Google Book Search, the complete text is searchable for any book with full view, limited preview, or snippet view. Searching is provided across the entire collection, as well as via a “search inside this book” box. While 36% of the current edition titles in the study had no preview available and therefore were not fulltext searchable, the remaining 64% were fulltext searchable. (These were the books that had either a snippet view or a limited preview; 20% and 44% respectively.)
In several cases, the searching proved to be somewhat inconsistent. Retrieval would vary depending upon whether the title or the ISBN was searched. Searching the ISBN for Swartz's Textbook of Physical Diagnosis: History and Examination retrieved two listings for the 2005 5th edition: one with no preview, and another with a snippet view and accompanying search capability. Searching by title instead of ISBN did not retrieve the snippet view on the 5\textsuperscript{th} edition, only a snippet view of the 2002 edition.

As mentioned earlier, multiple records appeared for a single edition of the same textbook. In some cases, these described different editions, such as
the two records found for the paperback and hardcover editions of the *Oxford Textbook of Palliative Medicine*. Other times, there was a separate record for each volume of a multivolume text (such as *Nathan & Oski’s Hematology of Infancy and Childhood*).

With full text searchable for 64% of the recent editions of these core clinical texts, what are the implications of Google Book Search for health sciences libraries?

**Increased searchability**

For items available in snippet view, limited preview, or full view, every word in the text is searchable, making it possible to look for answers to highly specific clinical questions like these:

- Transthoracic vs transesophageal echocardiography in right sided endocarditis
- Fat solubility of oxycodone
- The vitreous in infant heat stroke

Having every word in the text searchable also means that cited references are searchable. This can be a mixed blessing, since sometimes most of the search results come from the cited references rather than the text itself. On the other hand, the searchability of cited references in Google Books has proven to be an asset in citation verification.

The search capabilities are not very sophisticated. The advanced book search screen supports the use of one phrase, one Boolean OR statement, one Boolean AND statement, and one Boolean NOT statement. All the normal pitfalls of natural language searching apply. The specific words you use, and the way you combine them, impact your search results.

There is a subject search available, providing a few controlled vocabulary search terms per book. This appears to search the subject headings from the cataloging-in-publication statement, and perhaps also from the subject headings on WorldCat catalog record, although it's difficult to tell for sure.

In addition to the ability to search for words or phrases in a specific book, it is also possible to create a custom collection of hand selected books, and search only in those titles. This functionality comes from a companion Google service called “My Library”. It is free like other Google services, but it does require registration. It also has a batch upload function whereby you can simultaneously add a group of ISBNs to My Library. In this way, you could upload ISBNs corresponding to a certain collection of books in your library - such as the core textbook collection - thereby making it searchable. Remember that no search is available on “no preview” records; but all items available in either snippet view, limited preview or full view can be fully searched in this way.

**Other uses**

In addition to searchability, Google Book Search gives us a convenient way to quote a source when responding to a reference question. You can
create and share persistent links to searches, to specific books, or to a specific page in the book.(13) However, Google Books does not support selecting, copying and pasting portions of fulltext into email messages, Word documents, etc (except for books available in full view). Typically we provide links to pertinent passages and a link to the library catalog record showing the location of the book in our collection.

Furthermore, Google Book Search provides links to book reviews, which can be used to inform purchase decisions.

While this study has shown that a fair number of current clinical textbooks are full text searchable in Google Book Search, a lot of the search hits are for old books, or previous editions of current textbooks. However, it’s possible to focus on new material by supplying a date range when searching on the advanced search screen.

The Google Books Proposed Settlement

In 2005, the Authors Guild and the Association of American Publishers filed a class action lawsuit against Google in the United States District Court of New York, claiming copyright infringement. On October 28, 2008, the parties announced that a settlement had been reached.(14) The settlement still requires court approval by the presiding judge, and there is still much speculation as to what this settlement will mean for the future of Google Book Search and for libraries both within and outside the United States.(15) By and large, the settlement focuses on in-copyright books that are not commercially available. It does not apply to any books first published after January 5, 2009. Some observers believe that for in-print books, the displays may disappear.(16) It’s not clear at this point if the page numbers where the search terms were found will remain. In any event, the focus of the settlement has to do with the display functions rather than the search functions of Google Book Search. With the full text of over 7 million books already scanned and searchable, it would seem that the “genie is out of the bottle” where search is concerned, giving librarians the capability to search their book collections at a level of granularity far beyond the capabilities of other search tools.


