6-1-2007

What ABOUT the book? Google-izing the Catalog with Tables of Contents

Angi Faiks
Macalester College

Follow this and additional works at: http://digitalcommons.macalester.edu/lib_pubs

Recommended Citation

What ABOUT the book? Google-izing the Catalog with Tables of Contents

Angi Faiks
Associate Director
Collection Management Team Leader
Dewitt Wallace Library
Macalester College
St. Paul MN

Amy Radermacher
Reference/Cataloging/Electronic Resources Librarian
Library Technology Center
Concordia University
St. Paul, MN

Amy Sheehan
Reference and Instruction Librarian
Bush Memorial Library
Hamline University
St. Paul, MN

Meeting User Expectations in the Age of Google

Whether Google's massive efforts to create the "world's largest digital library" (Thompson D01) make your heart race in anticipation or cause your blood to boil in exasperation, there is no question that the Google search engine and its many features—including Google Book Search—have forever changed the way that people expect to interact with information. These new developments have increased information seekers' expectations. Web users are already accustomed to finding information freely available online. Google's book digitization project, intended to digitize the full text of millions of books for online access does the same thing as regular Google, but with the full text of *books.* With this project, Google has further increased user expectations by allowing users' searches to delve deeply into the content of books and library collections in a way that is currently not possible using library catalogs of primarily print volumes. The data included in basic catalog records

is no longer enough information. Recent trends have led users to expect (1) easy, full-text access to books and other resources and (2) more information about each book. This paper addresses how one consortium of eight private academic libraries, Cooperating Libraries in Consortium (CLIC), addressed these expectations and thereby ensured their patrons a better search experience. The members of CLIC are Augsburg College, Bethel University, College of St. Catherine, Concordia University, St. Paul, Hamline University, Macalester College, Northwestern College, and University of St. Thomas.

The New Challenge: Full Text, Not Just Metadata

In this networked information age, there is little patience for separation between information about information and the information itself. In this sense, library catalogs no longer meet their users' needs because they typically contain only MARC records, metadata, information *about* the book. On the other hand, the Google Book project offers searchers the *actual* information, the actual book, the actual text: not just the meta, but the data. The actual information is offered in other information arenas as well. Amazon.com, for example, now offers the Search Inside!™ feature, which allows users to search on the content of the book's pages and even to read selected pages. Users have grown accustomed to such features. The University of California Libraries Bibliographic Task Force found that "users want to move easily/seamlessly from a citation ABOUT an item to the item itself. Discovery alone is [no longer] enough" (III.2.c). Users are also looking for information tools that save them time, such as tools that allow them to make their book selections while sitting at the computer at home, at a coffee shop, or at work. Users are no longer willing to wait until they physically have the item in hand to determine whether the content is applicable to their research or information needs. Anecdotally, librarians often hear questions and comments from patrons such as: Why are there no reviews attached to this book? I'd really like to glance at this section of the book to see if this is what I'm looking for. I can do these things on Amazon.com, so why can't I do them here? Enhancing the traditional catalog is a necessary next step. Karen Calhoun's report made a courageous call to "improve the user experience" by "enrich[ing] the catalog with services (e.g., 'more like this,' 'get it' options, new book lists, etc.), and data (cover art, reviews, TOCs)" (19). Libraries worldwide have begun to answer this call in a number of ways. Currently there are a range of options to choose from as library catalog and system vendors upgrade traditional library catalogs to act more like Google. Features like spell checking, sophisticated relevance ranking of results, and subject browsing are common add-ons to existing catalogs or features of next generation systems. One such library catalog product, Endeca Profind, is described by a customer as "provid[ing] the speed and flexibility of popular online search engines while capitalizing on existing catalog records. As a result, students, faculty, and researchers can now search and browse the [libraries'] collection as quickly and easily as searching and browsing the Web, while taking advantage of rich content and cutting-edge capabilities that no Web search engine can match" (NCSU Libraries News, par. 1). While these systems and features can improve the user's experience,

they do not give what Calhoun says is paramount for our systems to remain vital: useful information about the book.

What ABOUT the Book?

The questions our consortium tackled were how to make our books more discoverable and how to give enough information *about each book* so that its usefulness to the user would be easily discernable, without necessarily having the resource in hand. Even librarians realize that many authority-controlled subject headings do not meet this real need. Subject headings are helpful—in that they provide authority control—but they do not always match the words that people use intuitively and, thus, are not always revealed by keyword searches. For example, the subject heading used for fresco painting is actually "Mural painting and decoration, Italian." In this case, if a user performs a keyword search for "fresco," unless the term appears in the title, a book that has whole sections on fresco painting (as indicated in a summary or table of contents) would never be discovered. The keyword search was born to make subjects more accessible, but more metadata is needed for fruitful keyword searching. More metadata doesn't necessarily mean better results, though, unless the metadata can be searched in an efficient way. While users may associate the ability to search the actual content of a book with the idea of more precise results, without the addition of authority control, full-text searching results in only more results, not necessarily better results. The solution, then, seems to be this: a good searching mechanism should allow for a keyword search to be conducted first, but then allow for the discovery and addition of subject headings later in the searching process in order to shape the hits into a precise group of desired results. Library catalogs are capable of this; Google, however, is not. "If one accepts the premise that library collections have value, then library leaders must move swiftly to establish the catalog within the framework of online information discovery systems of all kinds" (Calhoun 7). This includes Google and other Google-like environments. "Because it is catalog data that has made collections accessible over time, to fail to define a strategic future for library catalogs places in jeopardy the legacy of the world's library collections themselves" (Calhoun 7). With big Google projects on the horizon and the web embedded so deeply in our culture, it is now quite possible to conceive of a time when traditional library patrons will visit Google first in order to search for books and then be pointed—from Google—to various options (local library, neighboring library, interlibrary loan, book store, direct web download) for acquiring those books. In fact, this *is* now possible—at least partially—with various OCLC linking options (Find in a Library and Worldcat.org) and OpenURL initiatives. While perhaps such a day is in the library's future, it is not yet a distinct reality. In the meantime, librarians must still consider their users' increasing expectations for information access. Until the catalog is revamped, until it can fully communicate and interact with other available information sources and other searching methods, how can libraries better serve their users' needs and expectations with their often limited financial resources? How can libraries continue to be a useful part of Ranganathan's "growing [library] organism" when information in the world outside of libraries is now

connected and delivered in ways never before thought possible? How can we continue the "legacy" of libraries (Ranganathan)?

Taking Steps Toward Better Service

While Google continues to push the possibilities of better book searching, libraries can move towards improved systems using lower cost and lower overhead options that can truly assist users in achieving a more fruitful and delightful library catalog search experience today. As a case in point, the Cooperating Libraries in Consortium (CLIC), a consortium of libraries of eight academic institutions, decided that one method of addressing the abovementioned concerns was to implement small, immediate changes in the catalog, changes that could dress the catalog in Google-like attire. One of these changes is to provide more text to search against, just as Google Book Search does. One option for achieving this is to add tables of contents (TOCs) and summary notes into bibliographic records for books, thereby offering additional, highly relevant search terms into a library's catalog. Readers might already be thinking "Oh, how mundane and outdated!" Compared to the cutting edge approach of Google and even many library system vendors, this idea may seem dull or passé. However, it is one way that libraries can provide users with more information about the books in their catalogs. The remainder of this paper will explain how the addition of tables of contents to the CLIC consortium's local catalog provided positive results, especially when weighed against the costs of incorporating them into the catalog.

CLIC catalog results for "homelessness and criminal and law"

CLIC catalog full record view with TOC enhancement

The idea of adding access points such as tables of contents to library catalog records, while not a new one, has certainly been a topic of interest since the very first attempts at collection description. As Dinkins and Kirkland report in their literature review of tables of contents, "[t]he usefulness of adding point of access information to bibliographic records has been the subject of numerous articles and studies" (60). The authors go on to assert that "most studies agree that increasing access points in an online catalog record results in greater retrieval" (60). If this is the case, then libraries should be interested in inserting meaningful additional points of access however they can. By increasing access points on which users can search, both Google and library vendors are hoping to increase use of materials, to ensure that Ranganathan's first law ("Books are for use") remains true. If adding tables of contents (a relatively affordable method of catalog enhancements) can help libraries achieve this goal, then there are few reasons not to consider doing so. In addition, using tables of contents incorporates both the positive aspects of a library catalog (authority control and precise searches, for example) and the positive aspects of many web tools (the ability to see some content without having to handle the actual book).

Enriching the Catalog with TOCs: Effects on Usage and Users

In 2003, CLIC decided to purchase a set of tables of contents (TOCs) and, for some records, tables of contents with summary notes from Blackwell's Book Services; add them to the records of CLICnet, CLIC's shared catalog; and then monitor their

impact. At the time of CLIC's initial project, Blackwell's would enhance batches of records at the standard price of $1.05 per record for a single institution and $2.10 per record for consortium. We hoped that the additions of these records would (1) increase circulation of the books with enhanced records; (2) provide users with more detail about the contents of a given book so they could retrieve resources that are truly on target; and (3) allow more "free flow" (i.e. keyword) searching of the catalog, much like Google.   Devising a study on usage of titles that have been enhanced with TOCs is inherently complicated. There are so many factors that can impact circulation that being able to prove that TOCs were the main driver of a change in circulation is challenging, if not entirely impossible. In addition, both increased and decreased circulation of a book with a TOC can be considered a positive outcome of adding a TOC. In other words, if a TOC provides enough data for a patron to check out a book, that is a positive outcome. However, if a TOC provides enough data for a patron not to check out a book that is irrelevant to their information need, that is also positive outcome. The ideal way to find out how users are reacting to TOCs would be to ask them. Eventually we would like to design a user survey to gauge reactions to TOCs and thus further understanding of the impact of adding this data. Understanding these complications, we decided to go forward in capturing and analyzing data to see what might be revealed about the impact of added TOCs on circulation statistics. In gathering data, we had a choice of either A) comparing our enhanced record set's usage to itself before enhancement or B) comparing the enhanced set to a set of completely different unenhanced records. For our first round of analysis, we chose to compare the record set before and after enhancement. However, we believe that in order to have a more complete understanding of the impact of TOCs, we must also compare enhanced records with unenhanced ones and are therefore considering a study that will reveal data in this manner as well.   To begin with, we looked at circulation of the books in the set before and after enhancement. TOCs were added in November of 2004. We chose to enhance a random sample of records for monographs published between 2000 and 2005 that did not already have a table of contents (MARC 21 field 505) or a contents note (MARC 21 field 520). We did not specify which types of titles to enhance since our TOC vendor states that they focus on highly academic, wide-distribution titles and exclude titles that would not benefit from an added TOC (Blackwell's Book Services 1).
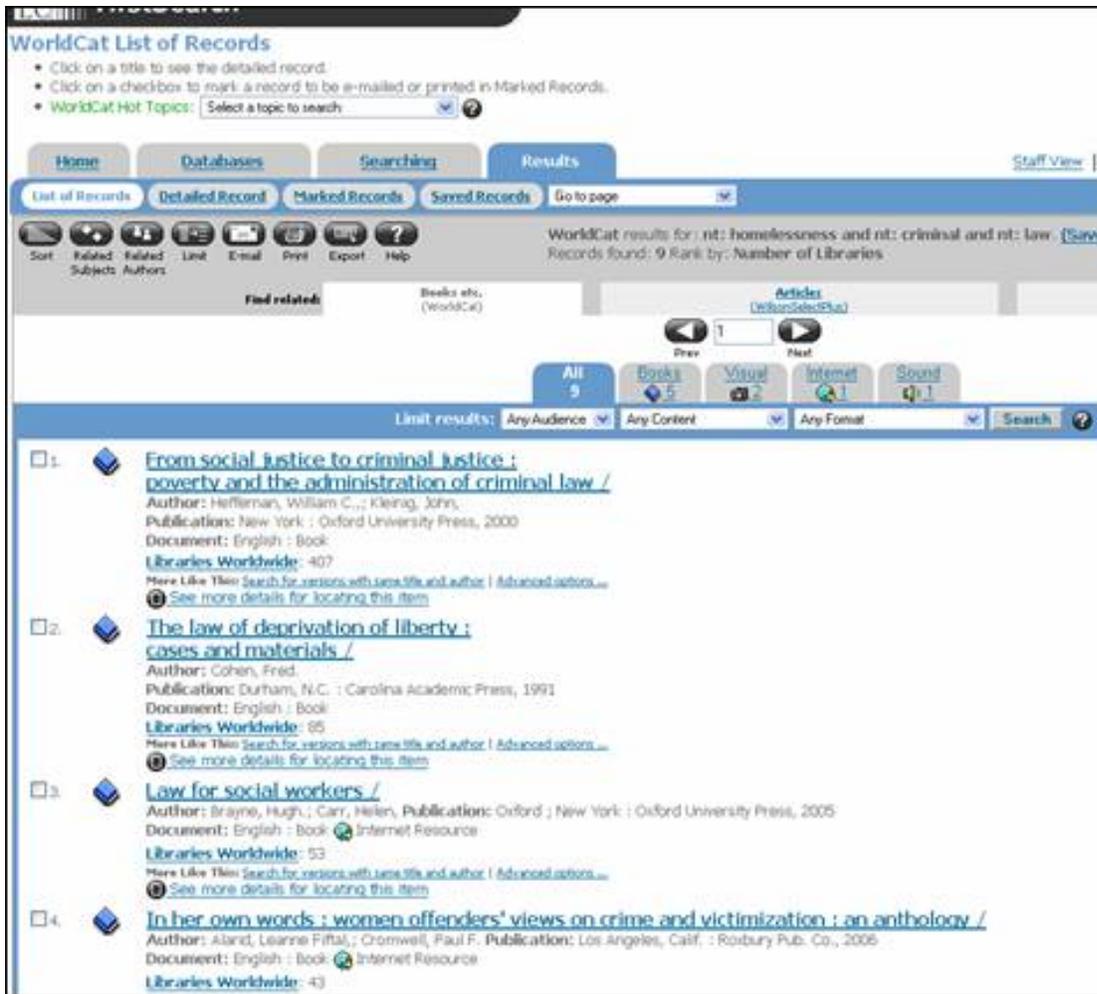
After the enhancement process was completed, the committee began tracking usage of titles with enhanced records. We were pleased—and even a bit surprised—to find that after adding TOCs/Summary Notes, circulation rose significantly more than expected. To gather our data, we compared usage of the books with enhanced records in 2004-2005 to usage of the same books from the previous year, 2003-2004. The percentage increase in circulation after TOCs/Summary Notes were added was 20.40%. This increase is certainly positive and may offer some indication that enhancing records can increase usage of material. As mentioned earlier, TOCs can also help users determine that a given title is *not* useful to their information need. However, data on the times a user found an enhanced record and chose not to check out the title would be extremely difficult—if not impossible—to gather strictly from usage data.

We agree that more data gathering and analysis must be done to explore additional factors in our TOC analysis. For example, we are curious whether certain subjects benefit more from TOCs and whether TOCs affect usage of interlibrary loan. User surveys would also contribute to a better understanding of how TOCs impact the actions of those who find them. Further research in this area is certainly warranted. Regardless of data, however, we believe that there is enough evidence in the literature and in our collective library experience to intuit that the more useful added data is, the more useful our catalogs will be. TOCs can help users know more about a book than is currently possible in most catalogs. In addition, future Google-like enhancements to OPACs, such as relevance ranking, will work best if the data within the records serve to meaningfully influence term frequency (the number of times a term appears in the record/document) and inverse document frequency (the importance of the word within the entire database) (Schneider). Until our catalogs search full text, a TOC can provide targeted search terms that can make relevance ranking more accurate. And finally, the value of TOCs may be in their ability to prevent disappointment by fruitless actions such as needlessly checking out or making interlibrary loan requests for unhelpful books.

The Challenges

Our consortium struggled with the tradeoffs that come from outsourcing TOC enhancement. We wanted quality records with quality information, but would the publisher-provided TOCs and content notes meet our standards of quality? Overall, our libraries were supportive of the idea of purchasing TOCs: they realized that we did not have the people power to enhance current records ourselves, much less the records of any past purchases. Further proof to us that the project was a worthwhile one was the fact that the Library of Congress, too, had recognized the importance of added contextual access points in records as evidenced by their current work on adding additional TOCs (Byrum). We could benefit from our TOC project and the Library of Congress's efforts by sending in our lists of records to Blackwell's only after MARC records had already been loaded into CLICnet. This way we could ensure that any TOC work done by the Library of Congress would already be reflected in our catalog and therefore would not need to be duplicated by an enhancement by Blackwell.   Money was, of course, a large part of the consortium's decision-making conversation as we wondered if enhancing a record for the consortium price of $2.10 each was worth the potential benefit to users. We eventually agreed that it was, and that notion was reinforced by the usage data and anecdotal feedback that we gathered. As one library staff member put it, we spend several thousands of dollars on books each year; adding an additional $2.10 to the cost of each book to help ensure and promote discovery and use is worth it.   MARC record placement was also discussed at length. We had the option of placing the TOCs and summary notes in two different places. The different placements interacted differently with our OPAC, and we had to consider issues of display consistency (especially with records containing TOCs not loaded by Blackwell's) and adherence to standards.   Prior to implementation, library staff members were also concerned with summary notes being taken from book jacket descriptions that contained flowery or effusive terms in

the publisher's description of the book rather than unbiased coverage. Some argued we would we be moving towards less objective descriptions when including publisher-provided summaries, as opposed to those selected by or edited by a library cataloger. Others argued for the benefits of publisher descriptions as the first step in opening the traditional library catalog to reader comments and other Web 2.0 features.   To address this concern, our current protocol asks for enhancements for records that are currently without contents notes (MARC 21 field 505) or summary notes (MARC 21 field 520). We accept only enhancements with 505s, some of which also include 520s. This way, each enhancement includes a table of contents, and some enhancements have the additional information included in a summary note (but never just a summary note).   While there are still issues that CLIC needs to work out—for example, what do we do about seemingly publisher-biased 520 notes, and how do we feel about records that contain contents without any true contextual terms—the project seems to be a success, and offers a point of discussion that helps our organization examine where we need to grow and simultaneously where we should "let go" in order to continue to be a part of the "living organism" of the information world.   A larger question also remains regarding the benefit of adding TOCs even though we can already conceive of a day when libraries, much like Google Book Search, will search against the full text of our holdings. The reality of such a scenario is not, however, a negative for TOC addition. We propound that the value of searching tables of contents will remain— even in a full text environment. While being able to search for terms against every word in a book is remarkable and certainly useful in many situations (looking for specific quotes or using the search as an index tool, for example), one can easily see how it could also be overwhelming. Even as "full text" catalog searching is developing, being able to focus on TOC metadata specifically is and will remain valuable.   To illustrate this point, we searched CLICnet, OCLC WorldCat (a database of combined bibliographic records from libraries worldwide), and Google Book Search for "homelessness and criminal and law." By limiting the search in our catalog to a keyword search, we found three hits for relevant titles that were only discovered because these search terms were found in the Tables of Contents. In OCLC WorldCat we were able to limit our search to only contents notes, resulting in five hits for books, all relevant to our search. Finally, this search in Google Books resulted in 777 results. While many of the books in the initial pages of the Google Books results look promising, 777 results may not be considered useful or valuable by all users. Perhaps the best case scenario in any system or catalog is to offer the richness of full text searching in addition to the option of limiting searches strictly to TOC metadata for increased search precision and relevant results. Since most catalogs today do not offer the former, there is increased incentive for libraries to offer the latter.

OCLC WorldCat results for "homelessness and criminal and law"

What ABOUT the book? Google-izing the Catalog with Tables of Contents, Angi Faiks, Amy Radermacher, Amy Sheehan, *Library Philosophy and Practice 2007 (June), LPP Special Issue on Libraries and Google*

9

Google Book Search Results for "homelessness and criminal and law"

Currently, Google is a large proponent of the change being introduced into our library world. CLIC's TOC project offered the consortium the impetus for discussion about the catalog, its features, its needs for enhancement, and its value in our organization. Library staff were given the opportunity to share their concerns and opinions about adding TOCs via a consortium-wide survey. Many of the topics of this survey were related to larger topics surrounding the Googlization of information seeking. Because of this discussion and others that have taken place in various consortium meetings, CLIC hosted a more formal discussion of the ILS and the future of its catalog (and perhaps by extension, *the* catalog). Our consortium recognizes the continued need for enhanced catalog records in addition to making more radical changes to our library systems and services to meet the needs of users today. Pursuing a TOC/Summary Note addition project was the first step in the direction of increased information access and retrieval for a relatively reasonable price. At the time of writing, we continue to enhance thousands of records with TOCs twice annually. Adding TOCs continues to be an effective method of fulfilling the users' increased expectations while simultaneously maintaining the integrity of the catalog that

librarians have known and trusted for years, another way to modernize the laws of Ranganathan. *Acknowledgements:* The authors would like to thank Steve Waage, Amy Reinhold and John McDonald for their contributions to this project.

Works Cited

Blackwell's Book Services."Table of Contents Enrichment Guide: Simply Accessible." 5 May 2005. 31 Oct. 2006. http://www.blackwells.com/downloads/TOCEnrichment.pdf>.

Brin, Sergey and Lawrence Page. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *Stanford University InfoLab Computer Science Department Conference Paper.* 1998. 7 Oct. 2006. <http://infolab.stanford.edu/~backrub/google.html#data>.

Byrum, John D. "Machine-generated Contents Notes." *Library of Congress, Bibliographic Enrichment Advisory Team.* 16 July 2005. 17 Oct. 2006. <http://www.loc.gov/catdir/beat/mg505.html>.

Calhoun, Karen. "The Changing Nature of the Catalog and its Integration with Other Discovery Tool." *Final Report.* 17 March 2006. 13 Oct. 2006. <http://www.loc.gov/catdir/calhoun-report-final.pdf>.

California. University of California Libraries Bibliographic Services Task Force. "Rethinking How We Provide Bibliographic Services for the University of California." *Final Report,* Dec. 2005. 13 Oct. 2006. <http://libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf>.

Dinkins, Debbi and Laura N. Kirkland. "It's What's Inside That Counts: Adding Content Notes to Bibliographic Records and Its Impact on Circulation." *College and Undergraduate Libraries* 13.1 (2006): 59-71.

North Carolina. North Carolina State University Libraries. "NCSU Libraries Unveils Revolutionary, Endeca-Powered Online Catalog." *NCSU Libraries News.* 12 Jan. 2006. 15 Aug. 2006. <http://www.lib.ncsu.edu/news/libraries.php?p=1998&more=1>.

Ranganathan, S. R. *Five Laws of Library Science.* Madras: Madras Library Association; London: Blunt, 1957.

Schneider, Karen G. "How OPACs Suck, Part 1: Relevance Rank (Or the Lack of It)." Online Posting. 13 March 2006. *ALA TechSource.* 25 Oct. 2006. <http://www.techsource.ala.org/blog/2006/03/how-opacs-suck-part-1-relevance-rank-or-the-lack-of-it.html>.

Thomson, Bob. "Search Me? Google Wants to Digitize Every Book. Publishers Say Read the Fine Print First." *The Washington Post.* 13 Aug. 2006: D01. 12 Sept. 2006 <http://www.washingtonpost.com/wp-dyn/content/article/2006/08/12/AR2006081200886.html>