TechMatters: A Tale of Two Book Search Tools

Krista Graham, Central Michigan University

It was the best of times, it was the worst of times, it was the age of Google Book Search, it was the time of Microsoft Live Search Books …"

With apologies to Charles Dickens

Chapter 1: A Brief History

The year was 2002, when a small group of Google staff members began investigating the feasibility of undertaking a massive project. Their hope was to digitize and index a copy of every book in the world. Realizing that such a huge project would take careful planning, they spent the next two years investigating other digitization projects that were already in progress, exploring and developing the necessary technology to safely speed up the digitization process, and approaching and lining up partner libraries and publishers.

In October 2004, with their investigations complete, Google announced their plans to launch "Google Print" in cooperation with a number of prominent publishers at the Frankfurt Book Fair in Germany. Two months later, they announced the companion "Google Print Library Project" in partnership with the University of Michigan, Stanford, Harvard, Oxford, and the New York Public Library indicating their intention to digitize the bulk of the collections at these prominent institutions.

Almost immediately controversy surrounded the project as librarians, publishers, the news media, and even the general public debated the implications of the project. In fact, lawsuits have been filed, such as by the Authors Guild, over whether Google has the right to scan and make available "snippets" of in-copyright works without explicit permission. While debate raged, Google continued to digitize, index, and make available materials supplied by its library and publisher partners. One year later, in 2005, the Google Print project was renamed "Google Book Search" in order to "more accurately reflect" how the tool was being used (http://books.google.com/googlebooks/newsviews/history.html).

Not to be outdone, Microsoft, one of Google's rivals in the search realm, announced in October of 2005 their plans to develop a book search tool to allow users to search, read, and download digital copies of "books, aca-demic materials, periodicals, and other print resources" (http://www.microsoft.com/presspass/press/2005/oct05/10-25MSNBookSearchPR.mspx). Like Google, Microsoft planned to work with libraries and publishers; however, their focus would be primarily on works in the public domain. On December 6, 2006, the beta version of "Microsoft Live Search Books" was released making direct comparisons between the two competing search tools possible.

Chapter 2: Two Philosophies

Although the two projects may seem incredibly similar on the surface, there is actually a significant underlying philosophical difference between them. Google has adopted an "opt-out" philosophy with regard to the digitization of materials. Their intention is to digitize and index all of the literature that they can obtain from their partner libraries and publishers. This includes works in the public domain, as well as those still under copyright protection.

As a result of this policy, a significant proportion of the works indexed in Google Book Search are not yet in the public domain and so cannot legally be made fully available without the copyright holder's permission. To address this issue, Google's approach is to vary the amount of information the user may access depending on the status of the work in question. There are three possibilities:

- For books that are out of print and no longer under copyright the user may view the entire work.

- For books under copyright by any of Google's partner publishers, users are presented with a "limited preview" of a length determined by the copyright holder, along with a link to purchase a copy.

- For all other materials, the user is presented with a "snippet view" which will display only three limited sections of text containing the user's search terms.

  ⇒ This last possibility only is applied by Google if there has not been an "opt-out"; in cases where the copyright holder requests removal of a book, or "opts-out" of the project, no preview or snippet will be made available at all.

In contrast, Microsoft has adopted an "opt-in" approach towards the inclusion of works in their search tool. The initial beta release includes only out-of-copyright books digitized from the collections of their partner libraries, which includes the University of California, the University of Toronto, and the British Library. As a result, there are currently no viewing restrictions on the materials retrieved via Live Search Books and all books are downloadable in their entirety in PDF format. The next phase of the project will include in-copyright works that are submitted for inclusion by the copyright holders who thereby "opt-in" to the program. Upon submission, copyright holders will be required to specify "preview rights" setting either a percent viewable, a number of viewable pages forward and backward from the search terms, or the use of contextual snippets for each work.

Chapter 3: The Search Experience

So, in Google Book Search (http://books.google.com) and Microsoft Live Search Books (http://books.live.com) we find two similar plans to digitize the world's literature, but with two different philosophical approaches. But in the end is one actually superior to the other? How does the search experience compare between the two? Is one easier to use? Does the other have more useful features? To find out, I used each tool to locate a copy of Charles Dickens' "A Tale of Two Cities." Here is what I found.

Both tools have a clean search interface consisting of a simple text search box (see Figure 1). Google provides two initial search options: search all books or limit to "full view" materials only. Currently, the Microsoft database contains only public domain works that are fully accessible so there is no need to limit the results. Hopefully, Microsoft will provide a similar option once they begin adding limited view materials to their project.

Figure 1: Home Pages for both services

In both tools, I took a basic search approach just as students typically do and simply typed the complete title without using any special search operators. In each case, the search algorithm retrieved a copy of the book I was seeking at, or near, the top of the search results. In Google, a full view version of the book was the first result, while in Microsoft, Dickens' book was the second on the results list.

In both tools, retrieving a copy of the book is quite straightforward. By simply clicking on the link the first page of the book is displayed. Both tools provide links to easily download a full version (if out of copyright) in PDF format, and in both cases the quality of the digitized materials is fairly good.

Google provides two additional linking options: "Table of Contents" which leads to the contents page as one would expect, and "About this book" which provides a number of additional functions including links to purchase a copy from an online bookseller, find a copy in a nearby library via Open WorldCat, locate other digitized editions of the work, as well as "related books" such as criticism. Microsoft lacks these additional options (see Figure 2).

"My name is [insert guy's name] and I'm the librarian at [community college]. We don't have any formal classes for our students, but I was thinking we could maybe add a one-unit module to the freshmen English classes and teach library skills to them."

"My name is [insert woman's name] and I work in the school library at [small town]. We just started requiring the students to do a set of worksheets. It's like a work-book. They all have to use the library to answer ques-tions, like what is an index, or how to look up a Roman god in the OPAC."

"I'm [insert guy's name] and I work at the county library. We don't have anything formal, but we like to show stu-dents our reference books and special encyclopedias, be-cause they just use the computers and don't ever want to use books. And, they are really expensive."

It went on like this for half an hour. I felt like the host of a talk show that was going sour.

"What do you do at the university to teach students library skills?" someone asked.

I answered politely. Succinctly. Hearing the words as they came out of my mouth.

"You know," I said. "I had this very same conversation 30 years ago."

"Really?"

"Yep," I said, realizing that I was older than their parents(!)

"Well, what happened? Did you come up with anything? Were you able to make a change in the curriculum, or de-velop any common instruction?"

"Nope," I said with a sigh, "but that doesn't mean we can't try again!" And, I passed around the plate of cookies.

TechMatters...Continued from page 7)

Figure 2: Results Pages for both services
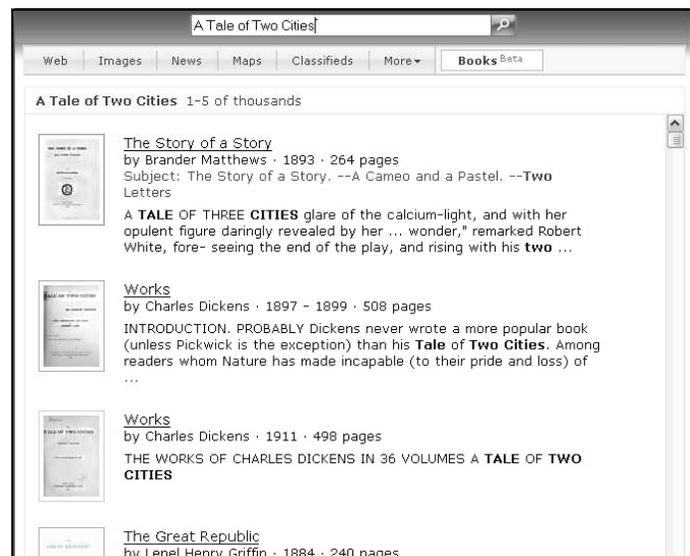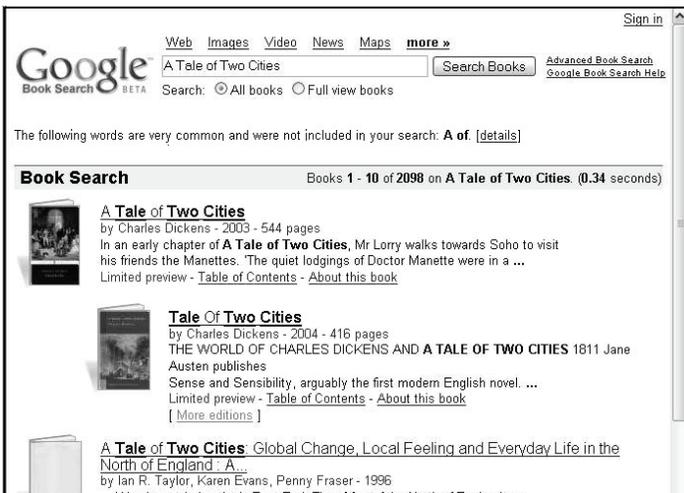
One notable weakness of the Microsoft search tool is the absence of an Advanced Search feature to facilitate searching for a specific edition of a work. In contrast, Google's Advanced Search screen allows users to enter the title, author, publisher, date of publication, and/or ISBN which makes finding a specific work significantly easier.

In the end, Google currently seems to have the clear edge over Microsoft in the book search realm. Google

Book Search is the more full-featured search tool, and cer-tainly their database contains a larger collection of materi-als. Of course, the Microsoft tool is very new and undoubt-edly they will continue to develop their product to make it a more competitive option. As I've said before, competition in the search domain can only benefit our students and li-brary users as companies will be pushed to continually im-prove and expand their Book Search programs.