

# “Have You Searched Google Yet?” Using Google as a Discovery Tool for Cataloging

Jennifer Lang

*Electronic Resources Cataloger  
Princeton University Library  
Princeton, New Jersey 08544-2098*

The use of the Internet by catalogers as a tool for locating cataloging documentation and other information is not a new development; articles and surveys have been published on the topic for at least a decade. (For further reading on catalogers' use of the Internet for cataloging, see Dunsire; Landesman and Oberg; Leroy and Thomas; Long; and Russell.) Not a day passes when I don't consult one or more of the many reliable websites I've discovered through the use of Internet search engines. This article is not intended to provide a comparison of search engines since most search engines have similar features, and each has its strengths and weaknesses (see Notess, “Search Engine Features Chart”). The focus of this article is Google and how some of its features can be used to assist in the cataloging process. I prefer Google because it was the first commercial search engine to provide links to cached or archived webpages (Olsen), a feature I find extremely useful for two reasons: 1) it allows me to view earlier versions of webpages, including webpages that are no longer available on the Internet, and 2) many times, I am able to access Google's cached version of a webpage much more quickly than the actual webpage itself. This caching feature, along with Google's translation tool and its ability to limit searches in a variety of ways, have made Google an essential addition to my arsenal of cataloging tools.

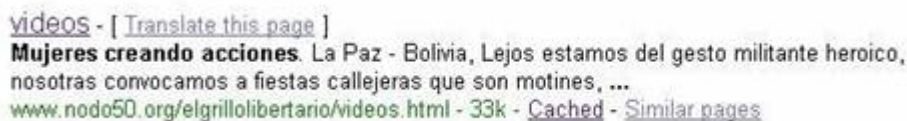
As the nature of the material I catalog has changed over time—I catalog electronic resources exclusively now—so has the nature of the information I seek on the Internet. Many of the CD-ROMs and DVDs I catalog are accompanied by little, if any, documentation. Often a jewel case has no insert and the disc itself has only a handwritten title. Normally, this doesn't present too much of a problem because I can determine a disc's content by loading it on my computer. Sometimes, however, the content is in a language that I don't know well or I encounter a DVD that won't play on my computer at all. Recently, I received a DVD to catalog that had virtually no information (Figure 1).

Figure 1. DVD with very little information



The disc would not play in my computer; the jewel case insert consisted of one sheet of paper that appeared to have been printed on an inkjet printer and cut to fit the case; there was a small sheet of paper with additional information taped to the back of the case insert; and the disc itself had no label. A Google search on the DVD title quickly led me to what appeared to be a catalog of videotapes and DVDs. I based this assumption on the “hints” provided on the Google search results page (see “Search Results Page”): the title of the webpage, with an excerpt that included my search term in bold; the place of production, which matched the information on the little sheet of paper in the jewel case; and a few words from the description of the DVD (Figure 2).

Figure 2. Google search retrieves webpage of catalog of videos and DVDs



By clicking the link labeled “Translate this page,” I used Google’s translation tool to translate the page and confirm at a glance that my search had indeed retrieved a catalog with useful information (Figure 3).

Figure 3. Translation of retrieved webpage



The Google search on the DVD title also led me to the website for the Museo de Arte Contemporáneo de Castilla y León, where I discovered that the DVD is part of the museum's collection and located more information about the production. Although I had enough information about the DVD to complete the cataloging at this point, my "cataloger's curiosity" prompted me to search for more information about the organization responsible for the production. I performed another Google search for the name of the organization, Mujeres Creando, and the first entry on the Google search results page pointed to their website. Again, I used Google's translation tool to view the page in English, and this led to a Google search on the English name of the organization, which revealed articles and other websites related to the group and its activities. Finally, I knew I had more than enough information to create a catalog record that was rich with descriptive information. A few extra minutes with Google helped me locate quality information I could use in my description of the DVD. Library users can only benefit from having the additional information available to them in the catalog record. In this case, not only will users know what the DVD is about, they will also learn something about the group that produced it.

In addition to cataloging "traditional" electronic resources such as DVDs, I catalog many statistical data sets that the library obtains through individual purchases or subscriptions. Because the data files themselves are numeric and have no information about the surveys from which the data were collected, I must rely on accompanying codebooks (either electronic or print) that contain information about the purpose of the surveys, how and by whom the data were collected, and the methodologies used. Sometimes the data sets are accompanied by very complete documentation, but other times I must be creative in obtaining descriptive information. When I encounter a data set for which there is little usable information, I use Google to search for additional information on the Internet.

One such data set, available through the Roper Center for Public Opinion Research, has very little information on the Roper Center website or in the accompanying codebook. The website indicates that the name of the survey firm is "Smith Research Center" (Figure 4), and this information matches the information in the accompanying printed codebook.

Figure 4. Roper Center website for data set

<b>TITLE</b>	<b>Adult Jewish Israel Opinion of the Mideast Peace Process</b> [Study# ISUSIA1992-92104]
<b>SURVEY FIRM</b>	Smith Research Center
<b>SURVEY SPONSOR</b>	United States Information Agency [USIA]

Both the website and codebook provide information about the survey dates, questions asked, etc., and the codebook also includes a copy of a memo from someone associated with the firm, Hanoch Smith. It also shows that the survey firm is located in Jerusalem. This information by itself is sufficient for creating a catalog record for the data set and establishing a name authority record for the firm, since one doesn't exist in the Library of Congress name authority file. Nevertheless,

perfectionism—or perhaps “cataloger's instinct”—led me to the Internet to see if I could find additional information about the survey firm. I knew a phrase search of “Smith Research Center “by itself would probably result in too many hits, so I added the name “Hanoch” to my search to narrow the search results. By performing a Google search with these search terms, I was able to locate several online references to what appeared to be the same firm that conducted the survey:

[Reverbiage: 'hanoch smith' news feed - stories from NPR, National ...](#)  
Madeleine Brand speaks with **Hanoch Smith**, director of the Israeli public opinion polling firm **Smith Research Center**, about the response to Sharon's ...  
[www.reverbiage.com/find/hanoch-smith - 7k - Supplemental Result - Cached - Similar pages](#)

[j - As immigrants flex muscles, Israeli campaign shifts tactics](#)  
“The key community in this election is the Russians,” said **Hanoch Smith**, founder of **Smith Research Center**, an Israeli polling institute. ...  
[www.jewishsf.com/content/2-0-/module/displaystory/story\\_id/11225/edition\\_id/215/format/html/displaystory.html - 21k - Cached - Similar pages](#)

[JSTOR: The Intifada and Political Discourse in Israel](#)  
For example, a March 1989 survey conducted for the New York Times by the **Hanoch Smith Research Center** in Jerusalem found that 54 percent of all Israelis ...  
[links.jstor.org/sici?sici=0377-919X\(199024\)19%3A2%3C43%3ATIAPDI%3E2.0.CO%3B2-O - Similar pages](#)

The Google search results page not only revealed several links that referenced Hanoch Smith as founder and director of Smith Research Center, it also pointed to a page that had a slightly different form of the firm's name than what appears in the codebook and on the Roper Center webpage. Because it is important that authority records be as accurate as possible and include any variations of a name for cross-reference purposes, I needed to investigate this discrepancy further. I already knew that the company was known as both “Smith Research Center “and “Hanoch Smith Research Center,” so my next task was to determine if the name of the company had changed or if there were just slight variations in the name. The ideal way to accomplish this would have been to find the website for the company, if one existed, or to locate an email address for Hanoch Smith and email him.

None of the webpages had contact information for the company, so I used another Google feature to limit my search even further. I knew from the codebook that the firm was in Jerusalem, so I did a Google search with the `site:` operator, which limits results by domain. By using the abbreviation for Israel (il), I was able to limit the results to only those websites that had “il” as the last element of the domain name. (Using the `site:` operator is different from using the `inurl:` operator, which would retrieve websites in Israel but also sites with “il” anywhere in the URL [e.g., pages with “il” in the URL because they had something to do with the state of Illinois]). Adding “`site:.il`” to my existing search resulted in no hits, so I removed the quotation marks surrounding Smith Research Center and found a reference to yet another variation on the company's name, Hanoch and Rafi Smith Research and Consulting:

[More Israelis hooked on Internet by Judy Siegel-Itzkovich May 6 ...](#)

The survey, by **Hanoch** and Rafi **Smith Research** and Consulting, polled a representative sample of 500 Israeli adults, with a 4% margin of error. ...

[www.cji.co.il/cji-n252.txt](http://www.cji.co.il/cji-n252.txt) - 3k - [Cached](#) - [Similar pages](#)

The search results page also included a link to what appeared to be an online directory of company names that included the names Hanoch and Rafi Smith. Another Google feature I take advantage of frequently is its cached versions of webpages, which are “snapshots” of pages indexed by Google and stored for backup purposes and which have search terms highlighted (see “Cached Pages”). By clicking the link for the cached version of the webpage, I was able to quickly locate the directory entry by finding the highlighted name of Hanoch Smith on the page:

4567	<a href="#">Hanoch &amp; Rafi Smith Consulting &amp; Research Ltd.</a>	03-6726522	03-6726562	<b>Hanoch Smith</b>
------	--	------------	------------	---------------------

Once I found this directory entry, which included an email address for Hanoch Smith, I was able to contact him in order to verify his firm's name for the authority record, which made the record more accurate and useful because it included all of the various names under which his firm is known.

My experiences using Google are not unique. For this article, I designed a very informal online survey aimed at determining how other catalogers use Google in their work and advertised it via several cataloging-related discussion lists, including AUTOCAT (for issues related to cataloging and authority control in libraries), NGC4LIB (Next Generation Catalogs for Libraries, for cataloging and catalog functionality), SERIALST (Serials in Libraries Discussion Forum, for various aspects of serials processing in libraries), OCLC-CAT (for librarians using OCLC's cataloging service), and OLAC-LIST (OnLine Audiovisual Catalogers, for catalogers of audiovisual materials). The survey was conducted between October 4 and October 10, 2006, and during that time, 571 catalogers answered all or most of the survey questions. The survey consisted of nine questions, six of which were general in nature: type of institution, number of years cataloging, age, gender, MLS degree or not, and type(s) of material cataloged. While the survey is hardly scientific—it was not comprehensive, not all respondents answered all questions, the survey sample was not random, etc.—after summarizing the results, I was able to make some interesting observations that affirmed my own choice to use Google as a cataloging tool. The results of the first six questions are summarized in below.

### 1. At what type of institution do you work?

	Response Percent	Response Total
Academic	64.8%	367
Public	17.8%	101
School	1.1%	6
Other*	16.3%	92

*\*Covers all types of special libraries not included in the first three categories.*

### 2. How many years have you been cataloging?

	Response Percent	Response Total
0-4	17.7%	100
5-10	26.5%	150
10-15	17.3%	98
15-20	12.5%	77
20+	26%	147

### 3. What is your age?

	Response Percent	Response Total
20-25	1.6%	9
26-30	6.3%	35
31-35	11.1%	62
36-40	13.4%	75
41-45	12.5%	70
46-50	15.9%	89
51-55	18.2%	102

56-60	13.6%	76
Over 60	7.3%	41

**4. Are you...**

	Response Percent	Response Total
Male	19.6%	109
Female	80.4%	448

**5. Do you have an MLS degree?**

	Response Percent	Response Total
Yes	79.6%	448
No	20.4%	115

**6. What type(s) of material do you catalog primarily?\***

	Response Percent	Response Total
Monographs	84.2%	475
Serials	46.6%	263
Electronic resources (e.g., CD-ROMs and websites)	54.8%	309
Audiovisual	59.2%	334
Other**	29.8%	168

*\*Respondents were instructed to select all categories of materials that applied.*

*\*\*Includes archival materials and manuscripts, curriculum materials, digital images, government documents, maps, and microforms.*

The final three questions, listed below, allowed for open-ended answers (see Appendix for selected comments).

**7. If you DO use Google as a tool to assist you with cataloging, in what ways do you use it? If you DON'T use Google as a cataloging tool, why not?**

8. Google has strengths and weaknesses. In terms of your work as a cataloger, in what areas do you find Google to be most useful? What weaknesses have you observed in Google (related to its usefulness as a tool for cataloging)?

9. Is there anything else you would like to add about using Google as a cataloging tool? If so, please use the space provided below.

Responses to these questions indicated that most catalogers who use Google as a cataloging tool use it in much the same way I described earlier. I was surprised to discover that there is little difference in use versus non-use of Google based on age or years of cataloging experience. Of the respondents under the age of 40, almost 95% reported using Google in their cataloging, while a little over 85% of the catalogers over 40 reported the same. In terms of number of years of cataloging experience, the difference was even less: roughly 87% of catalogers with up to 20 years of cataloging experience use Google in some way in their work. Surprisingly, this percentage was slightly higher for the catalogers with over 20 years of experience—almost 91% of this group use Google in their work. The difference in use of Google as a cataloging tool between catalogers in academic and non-academic libraries was negligible: 87% of the respondents in academic libraries reported using Google, while 89% of catalogers in other types of libraries reported the same. I'm not sure what accounts for the lack of difference; perhaps it has something to do with the nature of cataloging. Unlike public services librarian, we are not necessarily looking for information as a basis for research; rather, we are generally trying to verify discrete pieces of information to help us craft a bibliographic description. Many respondents consider Google important for authority work, and others reported using Google for the same reasons I do, including its translation and caching functions.

The catalogers who reported not using Google had a wide variety of reasons ranging from distrust to not considering Google to be a reliable source of information. Some simply preferred other search engines. Others reported never having considered using Google before seeing my survey, and some disagreed with calling Google a cataloging "tool." Several catalogers were in agreement that Google sometimes returns too many results to be effective, and there were even some questions about the effectiveness of Google's relevancy ranking, which is based on the number of pages that link back to a particular page ("Our Search: Google Technology"). Nevertheless, most respondents agreed that Google's main strengths are its ease of use and its ability to reduce the amount of time required to find information on the Internet. Many find Google to be a good complement to the other cataloging tools they use and report using it on a daily basis.

Thus it seems that most catalogers continue to find the Internet a rich source of information. Google in particular has several features that catalogers find useful. Information that used to be difficult if not impossible to find, especially given the unstable nature of websites, is quite easily found using Google's cached pages. In addition, Google's translation tool and other features make it a search engine worth exploring as a supplement to "traditional" cataloging reference sources.

## Appendix: Selected Comments from Survey Participants

"I'm not a 'Google-prude' and get weary of librarians who basically imply (or even outright state) that Google is beneath them. It is simply another tool we use. They should learn to accept it (or at least, tolerate it) rather than being such snobs and acting like anyone who doesn't disdain the use of Google has stepped into something smelly!"

"It never occurred to me to use Google. I have access to LCSH and Class Plus online, as well as the LC authority headings through OCLC. I can't think offhand what Google would do for me that these sources can't."

"Haven't tried it yet for cataloging. Will look at it more now that you've recommended it."

"I do not use it because I do not trust it."

"I use the 'translate this page' function to translate reviews/abstracts of books in foreign languages."

"I prefer Google to other search tools because it usually provides a cached link, so if the site no longer exists, often there is a cached copy that I use to find the information."

"I think Google has become an essential tool in doing NACO work. Being able to find new or confirming information without having to make phone calls or write to authors has facilitated the creation of name authority records enormously." (NACO is the Name Authority Cooperative of the Program for Cooperative Cataloging; see <http://www.loc.gov/catdir/pcc/naco/naco.html>)

## Works Cited

"Cached Pages." *Google Guide*. 2 Jan 2006. 6 Nov. 2006.  
<[http://www.googleguide.com/cached\\_pages.html](http://www.googleguide.com/cached_pages.html)>.

Dunsire, Gordon. "The Internet as a Tool for Cataloging and Classification, a View from the UK." *Journal of Internet Cataloging*, 2.3-4 (2000): 187-95.

Landesman, Betty and Steve Oberg. "What's in It for Us? Internet Use in Technical Services." *Serials Librarian*, 28.3-4 (1996): 317-23.

Leroy, Sarah Yoder and Suzanne Leffard Thomas. "Impact of Web Access on Cataloging." *Cataloging & Classification Quarterly*, 38.2 (2004): 7-16.

Long, Chris Evin. "The Internet's Value to Catalogers: Results of a Survey." *Cataloging & Classification Quarterly*, 23.3-4 (1997): 65-74.

Notess, Greg R. "Review of Google." *Search Engine Showdown*. 10 Oct. 2006. 4 Nov. 2006 <<http://searchengineshowdown.com/features/google/review.html>>.

Notess, Greg R. "Search Engine Features Chart." *Search Engine Showdown*. 17 Sept. 2006. 6 Nov. 2006 <<http://searchengineshowdown.com/features/>>.

Olsen, Stephanie. "Google Cache Raises Copyright Concerns." *CNET News.com*. 9 July 2003 . 6 Nov. 2006 <[http://news.com.com/2100-1038\\_3-1024234.html](http://news.com.com/2100-1038_3-1024234.html)>.

"Our Search: Google Technology." *Google*. 6 Nov. 2006. <<http://www.google.com/technology/>>.

Russell, Beth M. "Using the Web for Name Authority Work." *Library Resources and Technical Services*, 45.2 (Apr. 2001): 73-79.

"Search Results Page." *Google*. 5 Nov. 2006 <<http://www.google.com/support/bin/static.py?page=searchguides.html&ctx=results>>.

Sullivan, Danny. "Major Search Engines and Directories." *SearchEngineWatch*. 4 March 2004. 4 Nov. 2006 <<http://searchenginewatch.com/showPage.html?page=2156221>>.