Google Scholar Goes to School: The Presence of Google Scholar on College and University Web Sites

by Chris Neuhaus, Ellen Neuhaus, and Alan Asher

This study measured the degree of Google Scholar adoption within academia by analyzing the frequency of Google Scholar appearances on 948 campus and library Web sites, and by ascertaining the establishment of link resolution between Google Scholar and library resources. Results indicate a positive correlation between the implementation of Google Scholar link resolution and the degree of Google Scholar adoption.

Chris Neuhaus is Library Instruction Coordinator, University of Northern Iowa, USA <chris.neuhaus@uni.edu>; Ellen Neuhaus is Reference Librarian and Bibliographer, University of Northern Iowa, USA <ellen.neuhaus@uni.edu>; Alan Asher is Art and Music Librarian, University of Northern Iowa, USA <alan.asher@uni.edu>.

INTRODUCTION

Google Scholar has now been available to academia for more than two years. This free resource, with its very basic and familiar interface, could potentially serve as an alternative to metasearch engines such as MetaLib and WebFeat, and the more sophisticated and decidedly more expensive citation indexes Elsevier's Scopus and Thompson Scientific's Web of Knowledge. The reviews and critiques of Google Scholar have been, at best, mixed. The content, the search engine, the interface, and the citation counts of this product have all been criticized. Yet, despite these many published caveats and warnings of Google Scholar inadequacy, it is probable that academic scholars, attracted by the simplicity and familiarity of the Google interface, are now using Google Scholar in ever-increasing and substantial numbers. To what extent then has Google Scholar found a home on campus? How often does Google Scholar grace the Web sites of universities and colleges? Does the degree of Google Scholar utilization and adoption vary by type of academic institution? Is there a discernable relationship between promotion of Google Scholar and its prevalence and prominence within the campus Web space? To what degree is Google Scholar arrayed on the Web sites of library and information studies (LIS) programs?

This study summarizes reviews and research on Google Scholar from the past two years. New questions are raised concerning the degree and nature of Google Scholar as a presence within the framework of university and college research. As part of this study, a series of experiments, designed to determine the degree of Google Scholar utilization and adoption, were conducted and are here described, discussed, and followed by conjectures on the impact of Google Scholar and the possible directions for further inquiry.

Recent Studies and Critiques of Google Scholar

Research on and about Google Scholar has been confined by the sheer novelty of this search engine, to the past two years. Studies and discussions to date have focused on one or more of the following five questions: the relative strength of the Google Scholar search engine, the quality and quantity of Google Scholar "Cited by" entries relative to Web of Science and Scopus, the content of Google Scholar and the rate at which new content is added, the degree to which libraries are adopting and promoting Google Scholar, and the number of libraries

providing link resolver access from Google Scholar to their institutional resources.

The Google Scholar Search Engine and Interface

Soon after the release of Google Scholar, Peter Jacso served as the point-man for Google Scholar search capabilities and content analysis. Early into the release of Google Scholar, Jacso urged Google to improve its new product by incorporating the metadata provided by publisher's archives and by creating indexes to utilize this metadata. Jacso also encouraged Google to equip Google Scholar with a search interface that would offer pull-down menus for limiting searches to journal, publisher, and document type.[1] In a second critique of search capabilities, Jacso drew attention to the fact that the intrepid researcher is unable to use standard search options such as truncation and proximity while searching Google Scholar. Jacso noted that Google Scholar appears to have significant problems with Boolean operators and publication-year limits. This lack of reliable basic search options was compounded by the absence of browseable or searchable index fields for Author and Journal.[2]

Felter compared the Google Scholar interface with that of the search engine Scirus. Though Scirus clearly has the more sophisticated interface, Felter argued that most researchers preferred the simplicity of Google and would likely opt for Google Scholar over many more capable, but complicated, databases.[3] Henderson tested the search capabilities of Google Scholar and found a ranking bias toward older articles that had, as the result of the passage of time, been cited the greatest number of times. Henderson also lamented that Google Scholar lacked the standard Google search feature "Similar pages" and the "Did you mean:" feature for alternative spellings.[4] More recently, Golderman and Connolly applauded the compatibility of Google Scholar with bibliographic software such as Endnote and RefWorks, but faulted Google Scholar for failing to include search histories, alert services, and utilities for sorting, marking and saving results.[5]

The Relative Strength of Google Scholar as a Citation-Index

In their study of Google Scholar citations, Kousha and Thelwall found strong correlations between Google Scholar and ISI Web of Science for biology, computer science, and physics journals. They found more moderate, but still statistically significant, correlations between citations from Google Scholar and ISI for journals in sociology/psychology, education, chemistry, and economics. Kousha and Thelwall also demonstrated significant correlations between Google Scholar citation counts and ISI Journal Impact Factors.[6] Noruzi tested the citation counts for thirty-six frequently cited papers in the field of webometrics in both Google Scholar and Web of Science. Google Scholar identified more citations than Web of Science for all but three of these articles.[7] Bakkalbasi and Bauer compared citation counts in Google Scholar, Web of Science, and Scopus for 1985 and 2000 articles from the journal *JASIST*. They found that while Web of Science returned the most citation counts for 1985, Google Scholar tallied the highest citation counts for all *JASIST* articles published in 2000.[8] Jacso weighed-in during June 2005 plaintively noting that Google Scholar fell far short of two other free citation databases CiteSeer and eBizSearch, though Jacso conceded that Google Scholar might well be valued at institutions lacking the funds to support either Web of Science or Scopus.[9] In another published comparison of Google Scholar, Scopus, and Web of Science, Jacso investigated the relative coverage of the most heavily cited papers from the journal *Current Science*. Jacso found that for the sum total citation count of all thirty articles in question, both Web of Science and Scopus out-performed Google Scholar by a three to one margin.[10] Bakkalbasi, Bauer, Glover, and Wang analyzed the number of citing references for articles from the fields of oncology and condensed matter physics as generated by Web of Science, Scopus, and Google Scholar. They found that while Web of Science and Scopus did return more citing references than Google Scholar, Google Scholar returned the largest number of unique references. These authors concluded that no one of these three resources clearly outperformed the others and that a researcher relying on just one or even two of these resources might fail to find all references.[11] In their comparative bibliometric study of Web of Science, Scopus, and Google Scholar, Menho and Yang concluded that for the field of library and information science Google Scholar provides citations from a broader array of sources than either Scopus or Web of Science, though many of these additional sources come from low impact journals and conference proceedings.[12]

Google Scholar Content and Rate of Updating

The collection management policies and practices of Google Scholar are something of a mystery. The content of Google Scholar, also an enigma, has been studied and guessed at since Google Scholar was first introduced. Jacso was one of the first to analyze the possible content of Google Scholar and to criticize Google Scholar for its obscurity, its inclusion of nonscholarly material, and its many omissions and redundancies.[13] Inviting others to test Google Scholar for themselves, Jacso created and then updated a Web site that allows the curious to compare search results of Google Scholar with search engines of publishers Annual Reviews, Blackwell, Institute of Physics, Nature Publishing Group, and Wiley Interscience.[14] Robinson and Wuseman found that for scientific literature, Google Scholar outperformed Ask.com, Google, and Yahoo! in terms of precision, recall, and retrieval of top ranked pages. However, Google performed poorly in these same tests for nonscientific literature.[15] Gardner and Eng utilized a standard search to test the content of Google Scholar against the content of PsycINFO, ERIC, and the ISI Social Science Citation Index. For the search <homeschooling OR "home schooling"> these authors found only modest to minimal overlap between Google Scholar and these three social science databases.[16] Investigations by Giustini and Barsky identified PubMed and nine scientific and medical publishers from the CrossRef Search pilot project as major content contributors to Google Scholar.[17] Walters compared Google Scholar to the databases Academic Search Elite, Ageline, ARticleFirst, GEOBASE, POPLINE, Social Sciences Index, and Social Sciences Citation Index. For a core list of 155 articles on later-life migration, published between 1990 and 2000, Google Scholar was found to index the greatest percentage of titles. Walters cautioned that despite this strong performance, roughly one third of all Google Scholar citations were incomplete, and that one third of all Google Scholar entries studied lacked abstracts.[18] In a comparison of Google Scholar with the Chemical Abstracts Service, Levine-Clark and Kraus found that Google Scholar returned more results than the Chemical Abstracts Service for topical searches, but that the reverse was true for chemical compound and

personal name searches.[19] In their study of Google Scholar content, C. Neuhaus, E. Neuhaus, Asher, and Wrede checked Google Scholar against random samples from forty-seven different databases. Google Scholar content strengths included coverage of the literature in the natural sciences, medicine, and computer sciences. Google Scholar content weaknesses included coverage of the literature in business, education, humanities, and the social and psychological sciences. In addition, Google Scholar was shown to have significant English language and publication date biases, particularly for material published prior to 1960.[20] On this same note, Noruzi found that Google Scholar did not index articles in either Persian or Chinese.[21] In contrast, for library and information science literature, Meho and Yang found that Google Scholar provided better coverage of international research than either Web of Science or Scopus for citations published after 1993.[22]

In one of the first qualitative studies of Google Scholar, Helms-Park, Radia, and Stapleton conducted an analysis of the literature retrieved from Google Scholar vs. literature retrieved from the university library catalog. The results of this study revealed no significant differences between library catalog derived literature and Google Scholar obtained literature as rated by three independent university instructors.[23]

Is Google Scholar keeping up with current research? Initial tests by Vine, in February 2005, indicated that Google Scholar was, at that time, more than a year behind in updating research found in PubMed.[24] Subsequent testing from April to July 2005 showed some improvement, but Google Scholar still suffered an updating lag of approximately fifteen weeks for new records held by PubMed and BioMed Central.[25] A follow-up study by Vine in January 2006 revealed that Google Scholar was nearly five months behind in uploading randomly selected clinical trials from PubMed.[26]

Library Promotion of Google Scholar

Mullen and Hartman analyzed the use and promotion of Google Scholar by 113 ARL libraries. In particular, the authors looked for the presence of Google Scholar on library home pages, library directories of indexes and abstracts, library research guides, and library directories of search engines. In addition, Mullen and Hartman looked for inclusion of Google Scholar within library catalogs. Results of this study, conducted during the summer of 2005, showed that only 6 percent of ARL libraries had cataloged Google Scholar or placed a direct link to Google Scholar on their home page. This study also reported that 24 percent of ARL libraries listed Google Scholar among their indexes and abstracts (alphabetical list of indexes) and 19.5 percent of ARL libraries listed Google Scholar among recommended Internet search engines.[27]

Institutional Link Resolver Access from Google Scholar

Oder, in April 2005, highlighted the progression of libraries providing link resolver access from Google Scholar to their online resources. In February 2005, there were twenty-five institutions experimenting with the provision of institutional access from Google Scholar to their full text databases using link resolvers such as SFX, Article Linker, and 1 Cate. Oder noted that by March 2005 Serials Solutions (Article Linker) was inviting roughly 200 of its clients to participate in this pilot project.[28] By May 2005, Young reported that there were over 100 campuses providing their faculty and students with access to online library resources via Google Scholar.[29]

Google Site Search and the Google Scholar Advertisement

A recent development involving the automatic generation of a hyperlinked Google Scholar advertisement by certain Google-powered site search engines caught the eyes of a number of scholars, information technologists, and librarians. Listserv discussions,[30] blogs,[31] and work group minutes[32] all mentioned that when the Public Service version of Google site search engine returns results, the message "Find academic research papers with Google Scholar" is produced at the bottom of the first page of search results. Though this Google Scholar marketing phenomenon is relatively recent, perhaps starting sometime in early 2005, this study included an analysis of the ubiquity of this particular product placement and its possible impact on campus-wide adoption of Google Scholar.

New Questions for Google Scholar

This study, conducted during March and April 2007, attempted to gauge the current degree of adoption of Google Scholar by universities and colleges. More specifically this study addressed the following questions:

1. How frequently does Google Scholar appear on the campus Web sites of universities and colleges?

2. How frequently does Google Scholar appear on the library Web sites of universities and colleges?

3. How often is Google used to power campus and library site searches?

4. Where Google is featured as the site search engine, how often does the accompanying hyperlink and advertisement "Find academic research papers with Google Scholar" appear?

5. How often does a direct link to Google Scholar appear on a campus or library home page?

6. How many libraries are cooperating with Google Scholar to provide access to their online resources through Google Scholar link resolvers?

7. How does the degree of Google Scholar adoption vary by type of academic institution as defined by broad Carnegie Classifications: Research Universities, Master's Colleges and Universities, Baccalaureate Colleges, and Associate's Colleges?

8. To what degree does Google Scholar appear on the Web sites of accredited LIS programs?

METHODOLOGY

To select the universities and colleges to be analyzed for their degree of Google Scholar utilization and adoption, this study utilized the 2005 Carnegie Classification of Institutions of Higher Education.[33] The 2005 Carnegie Classification classifies all accredited degree-granting colleges and universities in the United States using data from 2003 and 2004. Four of the basic Carnegie Classification categories were used to identify U.S. academic institutions:

- Research Universities=Research Institutions

- Master's Colleges and Universities=Master's Institutions

- Baccalaureate Colleges=Baccalaureate Institutions

- Associate's (primarily community colleges)=Associate's Institutions

A total of 948 universities and colleges, 132 Research Institutions, 256 Master's Institutions, 267 Baccalaureate Institutions, and 293 Associate's Institutions were analyzed in this study, with sample sizes based on known populations, a 95 percent confidence level, and a precision level of ±5 percent.[34] Each institution was randomly selected using the random number generator created by Random.org at http://www.random.org/nform.html (Appendix A – Sampled Institutions by Carnegie Classification Category – see online version of this article). If any institution or its library lacked a Web site, or if the library Web site pages were not assigned distinct and recognizable URLs, this institution was dropped from the sample and a new replacement institution was randomly selected from the same Carnegie Classification Category. In addition to these institutions, forty-five U.S.-based ALA-accredited LIS programs were also examined.

The campus and library Web pages for each of the 948 institutions were analyzed in the following manner:

a. The campus and library home pages were reviewed to determine the existence of a site search engine and whether the site search engine was powered by Google or by non-Google search software.

b. If a campus or library used Google to power its site search, the results returned from successful site searches were analyzed for the presence of the hyperlinked advertisement "Find academic research papers with Google Scholar."

c. For both institutional and library Web sites, a webometric analysis was conducted to determine the number of Web pages that either linked or referred to Google Scholar. This measurement was conducted using the phrase search "Google Scholar" within an Advanced Search of Google that was limited to primary campus or library domains.

In this study, only those pages initially returned by the Google search engine were tallied. Excluded from the tally were those pages that could be displayed by clicking on the option "In order to show you the most relevant results, we have omitted some entries very similar to the <number of items returned> already displayed. If you like, you can repeat the search with the omitted results included."

Institutional Web pages and their corresponding library Web pages were treated as exclusive entities. For each institution, the number of occurrences of Google Scholar on library Web sites was subtracted from the total number of occurrences of Google Scholar recorded for the entire institution.

Institutional Occurrences of Google Scholar=(occurrences of Google Scholar on all campus Web pages)−(occurrences of Google Scholar on all library Web pages).

d. Campus and library home pages were reviewed for the existence of a direct link to Google Scholar.

e. The establishment of link resolution from Google Scholar citations to restricted-access full-text articles licensed and offered by an institution's library was determined for each institution. This was done by typing the institutional name within the Library Links locator function of the Scholar Preferences feature of Google Scholar.

Though the scope and size of this study prevented the deployment of a fail-safe method for determining whether any given institution actually possessed link resolution software, an estimate was derived by utilizing the Library Links locator function of Google Scholar. In its current state, the Google Scholar Library Links function generates, by keyword search, a list of potential link resolvers for one or more institutions sharing a common keyword. When using the Library Links function, institutions that have established Google Scholar link resolution appear in bold with an active check box allowing for selection of the institution. For example, at the time of this writing, the Library Links search for "Cincinnati" results in active link resolution choices for both "University of Cincinnati (Find Full-Text at UC)" and "OhioLINK (Find it with OLinks)," both viable link options for the University of Cincinnati. Institutions with existing link resolution software that have not yet been linked to Google Scholar appear as nonbolded entries with inactive check boxes. Though the authors were unable to obtain a definitive statement from the creators of Google Scholar as to how complete or current their institutional link resolver database was during the time of this study, the premise for this study is that those institutions that failed to appear in a Library Links keyword search did not possess link resolver software.

f. The Web sites of all ALA accredited LIS programs offered in the United States were also analyzed as outlined in steps a through d.

RESULTS

The results of this study indicate that Research Institutions are linking and referring to Google Scholar much more frequently than smaller institutions from the other Carnegie classification categories. The average number of nonlibrary institutional Web pages containing references or links to Google Scholar was seen to vary from 189.73 for Research Institutions, to 4.09 for Master's Institutions, to 2.93 for Baccalaureate Institutions, and to 0.25 for Associate's Institutions (Fig. 1).

> "The results of this study indicate that Research Institutions are linking and referring to Google Scholar much more frequently than smaller institutions from the other Carnegie classification categories."

The presence of Google Scholar on library Web pages followed a similar pattern. The average number of library Web pages with references and links to Google Scholar was 41.72 for Research Institutions, 2.28 for Master's Institutions, 2.03 for Baccalaureate Institutions, and 0.72 for Associate's Institutions (Fig. 2).

This study found that ninety-six (73 percent) Research Institutions offered link resolver access from Google Scholar to their online library resources. This was in contrast to the eighty-five (33 percent) Master's Institutions, thirty-eight (14 percent) Baccalaureate Institutions, and seven (2 percent) Associate's Institutions that provided similar link resolver access from Google Scholar. Data obtained from the Library Links function of Google Scholar indicated that eight (6 percent) of the

Figure 1
Nonlibrary Institutional Web Pages Containing Google Scholar
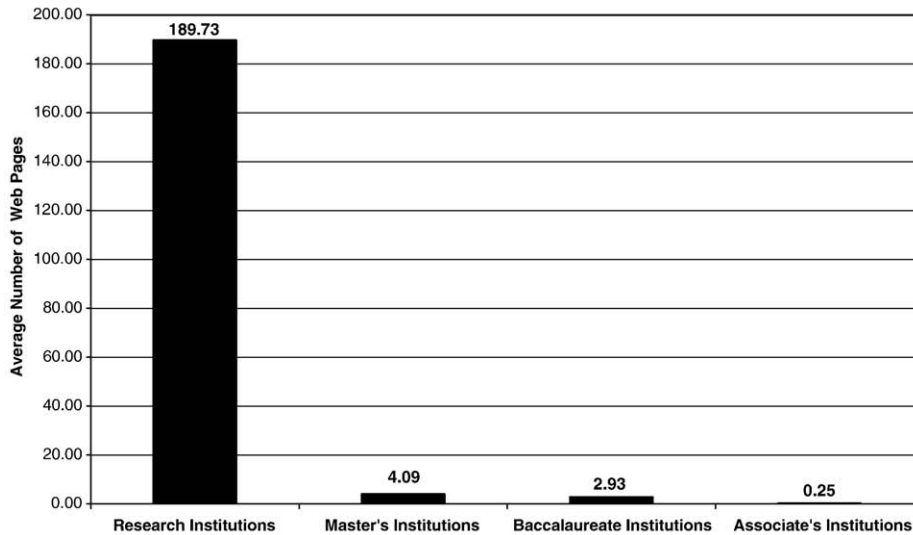
Research Institutions sampled possessed link resolver software, but had not established link resolution between Google Scholar and their library's online resources. Library Links data indicated that eight (3 percent) of the Master's Institutions sampled and three (1 percent) of the Baccalaureate Institutions sampled also possessed link resolver software without establishing a connection between their online resources and Google Scholar. The Library Links function data indicated that all of the sampled Associate's Institutions that owned link resolver software had established link resolution with Google Scholar (Fig. 3).

None of the institutions in this study provided direct access to Google Scholar from their campus home pages. However, seven (5 percent) of the library home pages of Research Institutions, two (1 percent) of the library home pages of Master's Institutions, and three (1 percent) of library home

pages of Baccalaureate Institutions featured direct links to Google Scholar.

For smaller institutions, the library serves as the primary promoter of Google Scholar. The percentage of pages containing Google Scholar that were attributed to the institution's library was highest for the Associate's Institutions at 74 percent. The library contribution to campus Web pages containing Google Scholar then decreased with increasing institutional size, with Baccalaureate Institutions at 41 percent, Master's Institutions at 36 percent, and Research Institutions at 18 percent.

A majority of institutions appear to utilize Google to power their site searches. Of those institutions included in this study, 97 (73 percent) Research Institutions, 176 (69 percent) Master's Institutions, 144 (54 percent) Baccalaureate Institutions, and 134 (46 percent) Associate's Institutions utilized a Google-

Figure 2
Library Web Pages Containing Google Scholar
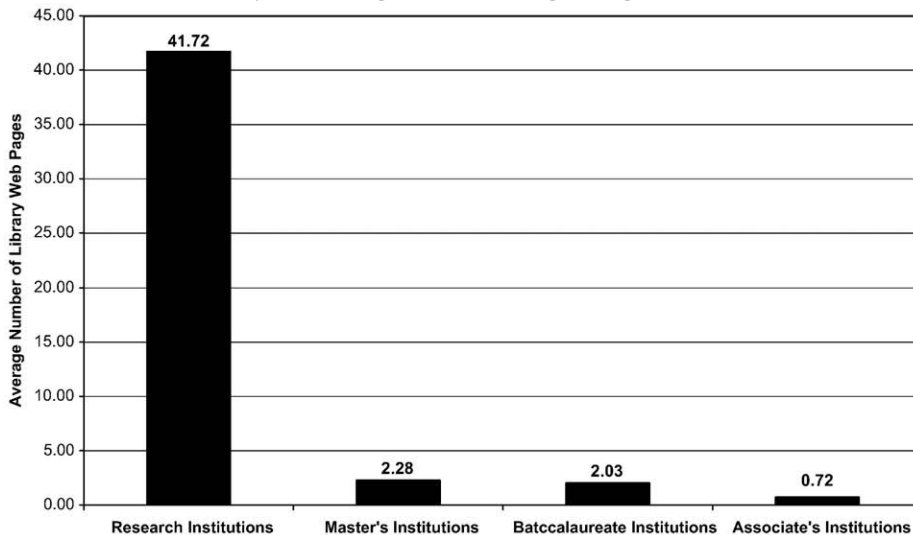
Figure 3
Institutions Providing Links from Google Scholar to Library Resources
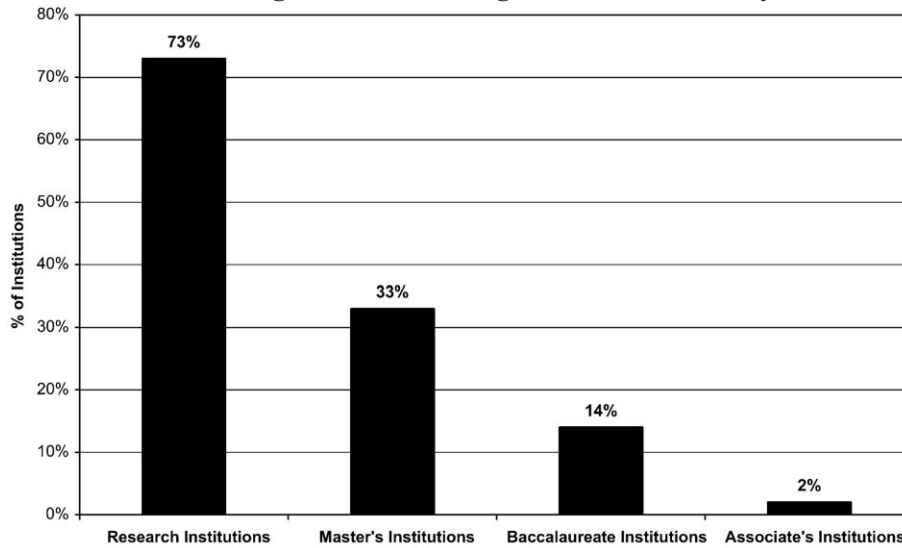
based site search. Many institutions have Google-based site search engines that provide the "Find Academic Research Papers with Google Scholar" hyperlinked advertisement at the end of the first page of generated search results. The number of institutions in this study that utilized Google-based site search engines which generated the "Find Academic Research Papers with Google Scholar" hyperlinked advertisement was 127 (50 percent) for Master's Institutions, 60 (45 percent) for Research Institutions, 102 (38 percent) for Baccalaureate Institutions, and 82 (28 percent) for Associate's Institutions (Fig. 4).

The majority of institutions included in this study possessed site-specific search engines. However, in many instances, the institution's library used the institution's site search engine instead of providing a search engine which searched only

library Web pages. Of the Research Institution libraries under consideration, 121 (92 percent) featured site search engines that limited the focus of the search to library Web pages. In contrast, only seventy-three (29 percent) Master's Institution libraries, forty-eight (18 percent) Baccalaureate Institution libraries, and just thirteen (4 percent) Associate's Institution libraries offered library site-specific search engines.

In this study, forty (30 percent) Research Institution libraries provided search engines with Google Scholar advertisements. This number decreased to thirty (12 percent) for Master's Institutions, eighteen (7 percent) for Baccalaureate Institutions, and two (1 percent) for Associate's Institutions (Fig. 5).

For the forty-five ALA accredited LIS programs sampled in this study, the average number of Web pages containing

Figure 4
Institutions with Site Searches That Advertise Google Scholar
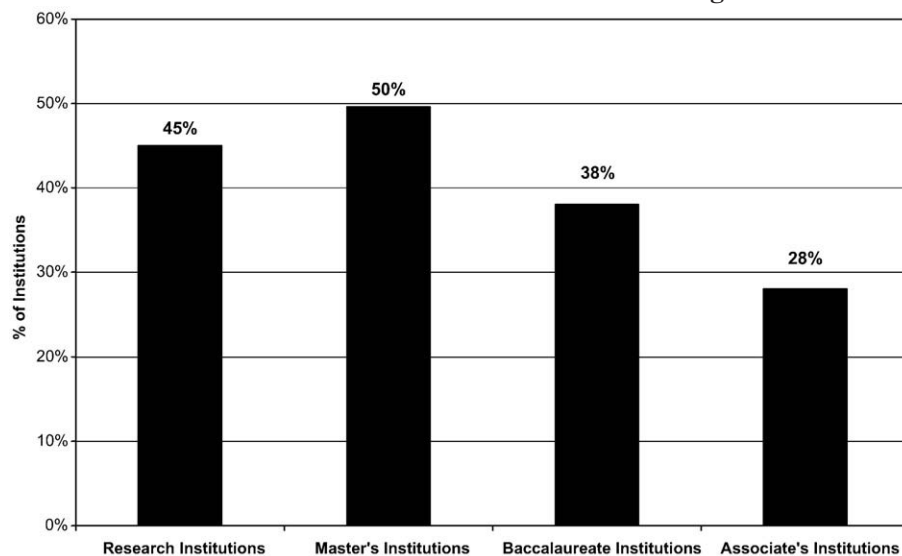
Figure 5
Libraries with Site Searches That Advertise Google Scholar
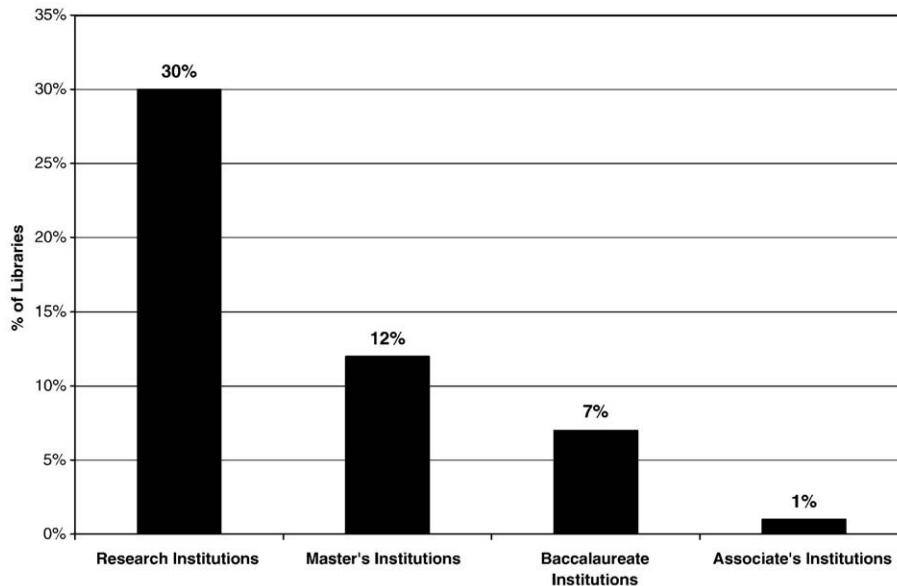
references and links to Google Scholar was 44.58. Fifty-eight percent of the forty-five programs provided site-specific search engines for their Web pages, and forty-two percent of all programs surveyed utilized Google for this purpose. Of these forty-five LIS programs, fifteen (33 percent) had site searches that generated "Find Academic Research Papers with Google Scholar" hyperlinked advertisements.

When the data gathered in this study were arranged and analyzed, a number of patterns and tendencies were revealed. Research Institutions that provided link resolver access from Google Scholar to their library resources had an average of 247.72 Web pages that linked or referred to Google Scholar. In contrast, Research Institutions that did not provide such link resolver access from Google Scholar had an average of only 35.08 Web pages with links or references to Google Scholar. This same trend was repeated among Master's, Baccalaureate, and Associate's Institutions. Master's Institutions that provided link resolver access from Google Scholar had an average of 6.29 Web pages that linked or referred to Google Scholar, while those schools without link resolver access to Google Scholar averaged only 2.95 such Web pages. Baccalaureate Institutions with Google Scholar link resolution averaged 13.79 Web pages referring to Google Scholar while Baccalaureate Institutions without averaged only 1.13 Web pages mentioning Google Scholar. Associate's Institutions providing link resolution from Google Scholar averaged 2.29 Web pages that linked or referred to Google Scholar and those without averaged on 0.20 references to Google Scholar (Fig. 6).

A similar pattern was observed for library Web pages. The average number of Research Institution library Web pages with references or links to Google Scholar was 53.54 pages for libraries with link resolver access to Google Scholar and 10.19 for libraries without. For Master's Institution libraries, the average number of library Web pages with references or links to Google Scholar was 4.06 for libraries with link resolver access to Google Scholar and just 1.39 Web pages for libraries without. Likewise Baccalaureate Institution libraries that provided link

resolver access to Google Scholar averaged 6.24 Web pages that referred to Google Scholar and those that did not averaged 1.33 Web pages with references to Google Scholar. Finally, Associate's Institution libraries with link resolver access to Google Scholar averaged 5.43 pages that referred to Google Scholar while those without averaged only 0.60 Web pages with a Google Scholar reference (Fig. 7).

For each of the eight Carnegie Class samples of institutions and libraries, tests of correlation were conducted to determine if these patterns indicated statistically significant relationships. Point-biserial correlations were calculated and t-tests of significance conducted on the relationship:

(number of Web pages linked to Google Scholar) vs.
(presence of link resolver access from Google Scholar to library resources)

with a null hypothesis: no correlation ($r = 0$), and the alternative hypothesis: there is a correlation ($r \neq 0$).

Statistically significant positive correlations were found between the presence of link resolver access from Google Scholar to library resources and the number of nonlibrary institutional Web pages that linked or referred to Google Scholar for all Carnegie Classifications.

Research Institutions: $r_{pb}(132) = +.17$, $p<.05$ and
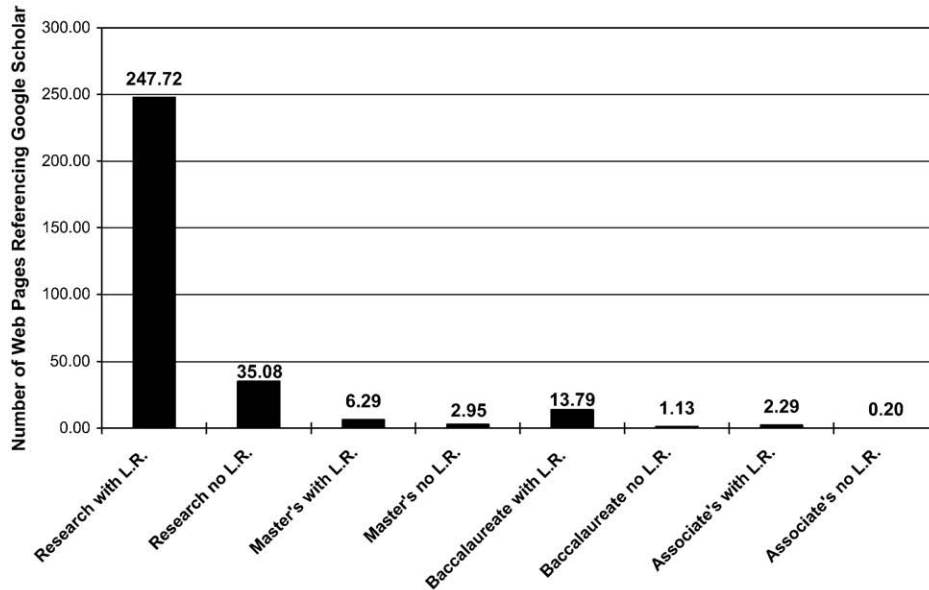 $t = 1.98$ (significant)

Master's Institutions: $r_{pb}(256) = +.16$, $p<.05$ and
  $t = 2.61$ (significant)

Baccalaureate Institutions: $r_{pb}(267) = +.35$, $p<.05$ and
  $t = 6.10$ (significant)

Associate's Institutions: $r_{pb}(293) = +.36$, $p<.05$ and
  $t = 6.61$ (significant)

Figure 6
Nonlibrary Institutional Web Pages Containing Google Scholar vs. Link Resolver Access to Google Scholar (L.R.)

For libraries of these institutions, statistically significant positive correlations were also found between the presence of link revolver access from Google Scholar to library resources and the number of library Web pages that linked or referred to Google Scholar for all but the Research Institution libraries.

Research Libraries: $r_{pb}(132) = +.06$, $p<.05$ and
  $t = 0.67$ (not significant)

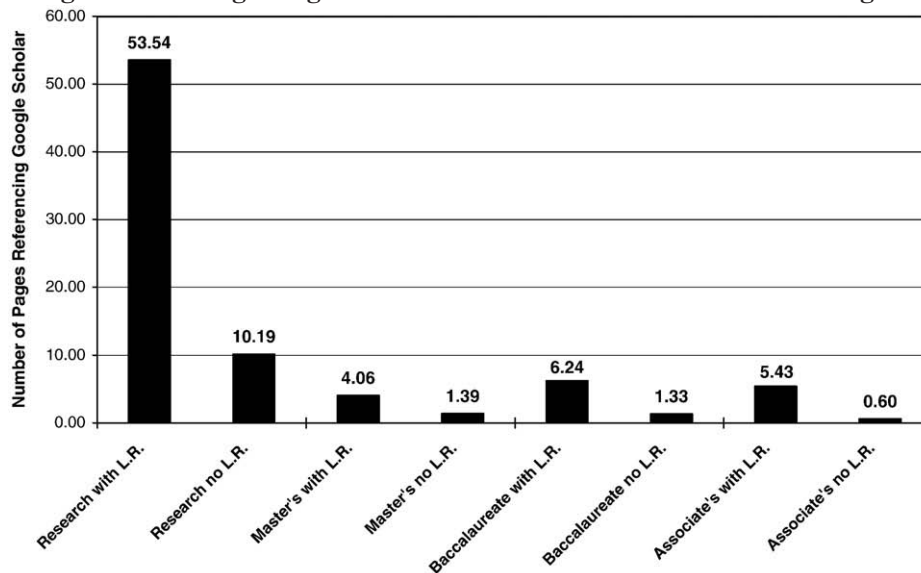Master's Libraries: $r_{pb}(256) = +.25$, $p<.05$ and
  $t = 4.14$ (significant)

Baccalaureate Libraries: $r_{pb}(267) = +.27$, $p<.05$ and
  $t = 4.49$ (significant)

Associate's Libraries: $r_{pb}(293) = +.17$, $p<.05$ and
  $t = 3.01$ (significant)

The relationship between the presence of the Google site search advertisement "Find academic research papers with Google Scholar" and the number of Web pages that linked or referred to Google Scholar was also investigated. While
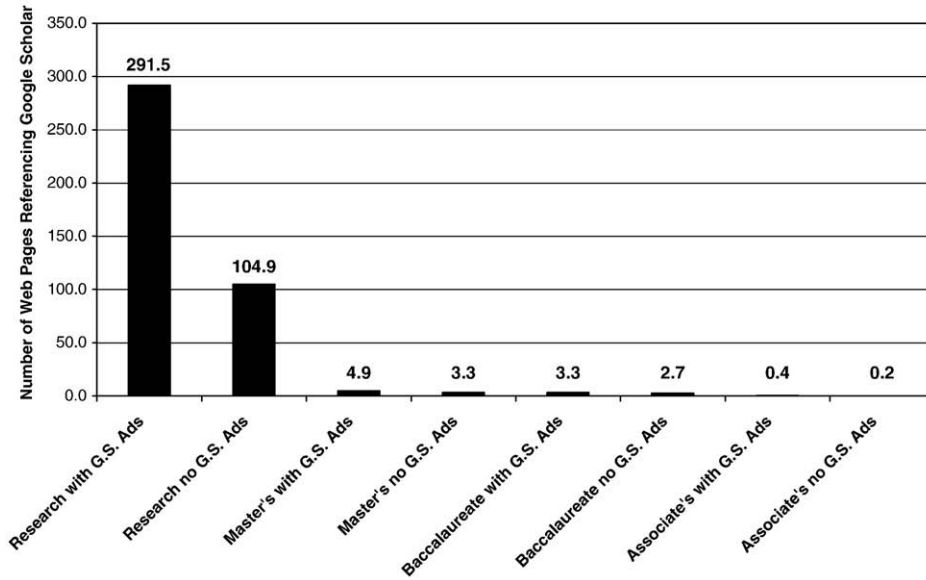
Figure 7
Library Web Pages Containing Google Scholar vs. Link Resolver Access to Google Scholar (L.R.)

Figure 8
Nonlibrary Institutional Web Pages Containing Google Scholar vs. Google Scholar Advertisements (G.S. Ads)

only one correlation proved statistically significant for these advertisements,

Master's Libraries: $r_{pb}(256) = +.19$, $p<.05$ and

$t = 3.07$ (significant),

the following pattern was noticed for all Carnegie Classes: institutions and libraries that utilized Google site search engines that generated Google Scholar advertisements had more Web pages pointing to Google Scholar than those schools and libraries without such advertisements (Figs. 8 and 9).

WEIGHTED RESULTS—A CAUTIONARY TALE

To determine whether the number of Google Scholar appearances for a given campus was simply a function of Web site size, this study sought to determine the relative prevalence of Google Scholar when compared to both the total number of nonlibrary institutional Web pages and the total number of

Figure 9
Library Web Pages Containing Google Scholar vs. Google Scholar Advertisements (G.S. Ads)
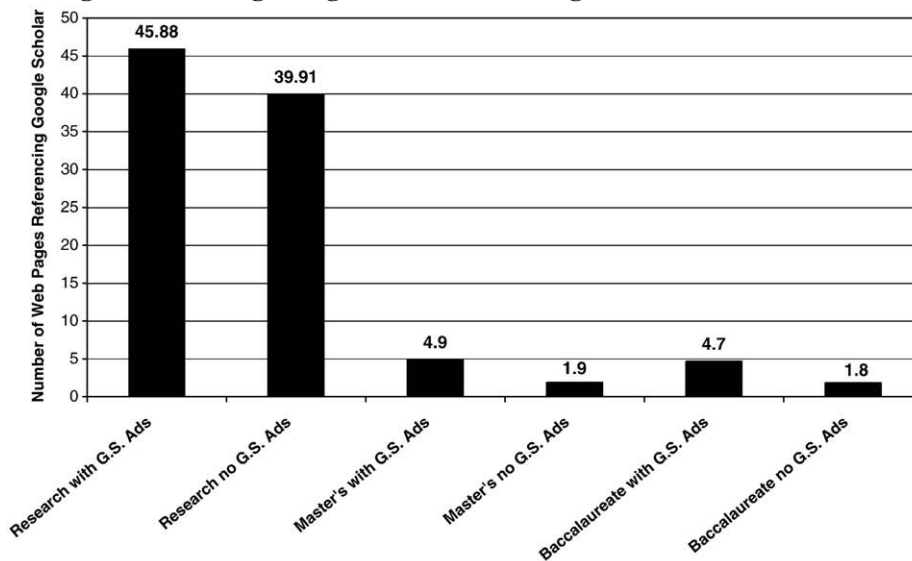
Table 1
Weighted Nonlibrary Institutional Web Pages Containing Google Scholar vs. Presence of Link Resolver Access to Google Scholar (L.R.)

Link Resolver Access to Google Scholar	Percent of Total Institution Web Pages
Research with L.R.	0.062 percent
Research Institution no L.R.	0.013 percent
Master's with L.R.	0.016 percent
Master's no L.R.	0.009 percent
Baccalaureate with L.R.	0.019 percent
Baccalaureate no L.R.	0.013 percent
Associate's with L.R.	0.013 percent
Associate's no L.R.	0.007 percent

library Web pages. These weighted statistics were thus calculated as:

Weighted Institutional Occurrences of Google Scholar

= (Nonlibrary Institutional Occurrences of Google Scholar)

÷(Weighted Library Occurrences of Google Scholar

= (Library Occurrences of Google Scholar)

÷(Total Number of Library Institutional Web Pages)

Weighted institutional occurrences of Google Scholar for Research Institutions averaged 0.048 percent of all institution Web pages. Baccalaureate Institutions averaged 0.014 percent, Master's Institutions 0.011 percent, and Associate's Institutions 0.007 percent. Thus not only did Research Institutions have by far the greatest average number of Web pages that refer to Google Scholar, Research Institutions also appear to have a greater percentage of their total Web pages linking to or referencing Google Scholar.

The weighted library occurrences of Google Scholar followed a very different pattern. For Baccalaureate Libraries, on average, 1.200 percent of all library pages contained a reference to Google Scholar. For Master's Libraries, this

Table 2
Weighted Library Web Pages Containing Google Scholar vs. Presence of Link Resolver Access to Google Scholar (L.R.)

Link Resolver Access to Google Scholar	Percent of Total Library Web Pages
Research with L.R.	0.620 percent
Research no L.R.	0.507 percent
Master's with L.R.	1.143 percent
Master's no L.R.	0.529 percent
Baccalaureate with L.R.	1.459 percent
Baccalaureate no L.R.	1.157 percent
Associate's with L.R.	0.752 percent
Associate's no L.R.	0.668 percent

Table 3
Weighted Nonlibrary Institutional Web Pages Containing Google Scholar vs. Presence of Google Scholar Advertisements (G.S. Ads)

Presence of Google Scholar Advertisements	Percent of Total Institution Web Pages
Research with G.S. Ads	0.074 percent
Research no G.S. Ads	0.027 percent
Master's with G.S. Ads	0.011 percent
Master's no G.S. Ads	0.012 percent
Baccalaureate with G.S. Ads	0.017 percent
Baccalaureate no G.S. Ads	0.012 percent
Associate's with G.S. Ads	0.007 percent
Associate's no G.S. Ads	0.007 percent

average was 0.733 percent, for Associate's Libraries 0.670 percent, and for Research Libraries 0.670 percent.

When considering the potential impact of the presence or absence of link resolver access from Google Scholar to library resources, both weighted institutional occurrences of Google Scholar and weighted library occurrences of Google Scholar behaved in a similar pattern. All institutions and libraries with link resolver access to Google Scholar displayed higher weighted occurrences of Google Scholar references than those without link resolver access from Google Scholar to library resources (Tables 1 and 2).

When considering the presence or absence of the advertisement "Find Academic Research Papers with Google Scholar," the pattern becomes a bit murkier. For Research, Master's, and Associate's institutions and their corresponding libraries, the presence of this advertisement corresponded to greater weighted institutional and weighted library occurrences of Google Scholar references. However, the converse was true for Master's institutions and their libraries (Tables 3 and 4).

While the number of occurrences of Google Scholar for all institutions in this study was recorded by direct tally of the search results, the significant size and scope of most of these institutions and libraries necessitated an alternative approach for recording the total number of institutional and library Web pages (Table 5).

Table 4
Weighted Library Web Pages Containing Google Scholar vs. Presence of Google Scholar Advertisements (G.S. Ads)

Presence of Google Scholar Advertisements	Percent of Total Library Web Pages
Research with G.S. Ads	0.965 percent
Research no G.S. Ads	0.426 percent
Master's with G.S. Ads	0.466 percent
Master's no G.S. Ads	0.766 percent
Baccalaureate with G.S. Ads	1.433 percent
Baccalaureate no G.S. Ads	1.183 percent

Table 5
Average Number of Nonlibrary Institution Web Pages and Library Web Pages

	Nonlibrary Institution	Library
Research Institutions	365,920	26,625
Master's Institutions	35,910	1473
Baccalaureate Institutions	17,375	351
Associate's Institutions	6835	84

To ascertain an estimate of the size of the institutional and library Web spaces under consideration, this study relied on numbers returned by the Google Statistics Bar. However, Google cautions in its GoogleGuide, that the Statistics Bar "Describes your search, includes the number of results on the current results page and an estimate of the total number of results.… This estimate is unreliable."[35] Furthermore, throughout the duration of this study, the authors monitored the daily variation in reported results by the Google Statistics Bar for their institution, the University of Northern Iowa. The total number of institution Web pages recorded by the Google Statistics Bar for a domain name search "uni.edu" ranged from 127,000 to 106,000. The total number of library Web pages recorded by the Google Statistics Bar for the domain name search "library.uni.edu" during this time period also varied from a value as high as 5510 to a value as low as 4230. Thus, though many of the weighted results for this study roughly parallel many of the same findings for nonweighted results, the imprecise nature of the Google Statistics Bar discourages the drawing of definitive conclusions from these weighted statistics.

DISCUSSION

If the establishment of a link resolver from Google Scholar to licensed online university resources can be taken as a clear sign of institutional Google Scholar buy-in, then Google Scholar has found a home on the campuses of the majority (73 percent) of Research Institutions. A statistically significant positive correlation between the establishment of Google Scholar link resolvers and the increased presence of institutional Web pages referring to Google Scholar was demonstrated for all institution classes within this study.

At larger institutions, it is the Web sites of faculty and students, and not library Web sites, that account for the presence of Google Scholar on campus Web sites (with 82 percent, 64 percent, and 59 percent of all campus Web pages containing Google Scholar being attributed to nonlibrary Web sources of Research Institutions, Master's Institutions, and Baccalaureate Institutions, respectively).

Research, Master's, and Baccalaureate institutions and libraries that provided Google site search engines that generated the advertisement "Find academic research papers with Google Scholar" did show higher averages of Web pages referring to Google Scholar. However, with the exception of Master's libraries, no statistically significant positive correlation could be shown linking these Google Scholar advertisements to an increased presence of Google Scholar on institutional and library Web pages.

The authors will readily concede that the presence of link resolvers from Google Scholar to library resources, Google

Scholar Web page references, and Google Scholar advertisements are indirect measures of Google Scholar adoption. Still to be answered are important questions such as:

- How many academic faculty and students use Google Scholar and how often do they use it?
- How does Google Scholar utilization compare with the frequency of use for library-provided metasearch engines or individual proprietary databases?

As to the question of relative utilization of metasearch vs. Google Scholar, there is already a testimonial on record. In his February 2007 *Library Journal* article "(Meta)search Like Google," Jonathan Rochkind noted that at Johns Hopkins University "Google Scholar has become the largest single source of links to our link resolver product."[36] A survey of Google Scholar link resolver utilization, conducted among systems librarians of institutions providing Google Scholar link resolvers, might well document a significant trend, and could be considered as yet another measure of Google Scholar adoption on university and college campuses.

YET MORE QUESTIONS

For academic libraries, Google Scholar offers new opportunities and probably presents a few problems. Google Scholar also raises more questions than it answers:

- Should Google Scholar link resolution be established on campus?
- How well is Google Scholar link resolution working?
- What is the general quality of Google Scholar records and the material to which these Google Scholar records refer?
- How reliable and thorough are the Google Scholar "Cited by" references?
- How might libraries work with the architects of Google Scholar to improve the quality of this index?
- If a library is unable to afford Scopus or Web of Knowledge, can Google Scholar serve to fill this void?
- If a library is only able to subscribe to a few online full text resources, does Google Scholar point to enough free open-access journals to satisfy library patrons?
- Will Google Scholar place an increased burden on inter-library loan systems?
- How should libraries incorporate Google Scholar into their array of online resources and their services such as reference and instruction?
- Will Google Scholar remain free?
- Will Google Scholar go away?

CONCLUSION

Google Scholar is clearly making inroads into the info-structure of academia. At the time of this writing, 73 percent of all Research Institutions and one third of all Master's Institutions provide Google Scholar-mediated link resolution from Google Scholar records to licensed full-text library resources. Roughly half of all Research and Master's Institutions and about one out of every three Baccalaureate and Associate's Institutions utilize

Google-powered campus site search engines that advertise Google Scholar to all who search the academic Web space. Perhaps Google and Google Scholar are taking the library world by storm, washing away much that is good and cherished. Perhaps Google Scholar is simply a new wave on which we all might enjoy a long and exhilarating ride. Certainly Google Scholar remains a force to be reckoned with and a phenomenon and resource that bears further investigation and continued study.

> **"At the time of this writing 73 percent of all Research Institutions and one third of all Master's Institutions provide Google Scholar mediated link resolution from Google Scholar records to licensed full-text library resources."**

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.acalib.2007.11.009.

Notes and References

1. Peter Jacso, "Google Scholar Beta *Peter's Digital Reference Shelf* (December, 2004): Available: http://www.gale.com/servlet/HTMLFileServlet?imprint=9999&region=7&fileName=/reference/archive/200412/googlescholar.html (July 24, 2006).
2. Peter Jacso, "Side-by-Side2: Native Search Engines vs Google Scholar," (2005). Available: http://www2.hawaii.edu/~jacso/scholarly/side-by-side2.htm (July 24, 2006).
3. Laura M. Felter, "Google Scholar, Scirus, and the Scholarly Search revolution *Searcher* 13 (February, 2005): 43–48. Available: http://www.scirus.com/press/pdf/searcher_reprint.pdf (July 23, 2006).
4. Jim Henderson, "Google Scholar: A Source for Clinicians? *CMAJ: Canadian Medical Association Journal* 172 (June 7, 2005): 1549–1550.
5. Gail Golderman and Bruce Connolly, "Who Cited This? *Library Journal* 132 (Winter, 2007 Net Connect): 18–26.
6. Kayvan Kousha and Mike Thelwall, "Google Scholar Citations and Google Web/URL Citations: A Multi-Discipline Exploratory Analysis," *Proceedings International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting, Nancy (France)* (June 5, 2006). Available: http://eprints.rclis.org/archive/00006416/ (July 22, 2006).
7. Alireza Noruzi, "Google Scholar: The New Generation of Citation Indexes," *Libri* 55 (February 11, 2006): 170–180. Available: http://eprints.rclis.org/archive/00005542/01/Google_Scholar,_The_New_Generation_of_Citation_Indexes.pdf (July 23, 2006).
8. Nisa Bakkalbasi and Kathleen Bauer, "An Examination of Citation Counts in a New Scholarly Communication Environment," *D-Lib Magazine,* 11 (September, 2005): 1–10. Available: http://www.dlib.org/dlib/september05/bauer/09bauer.html (July 23, 2006).
9. Peter Jacso, "Google Scholar (Redux)," *Peter's Digital Reference Shelf,* (June, 2005). Available: http://www.galegroup.com/servlet/HTMLFileServlet?imprint=9999&region=7&fileName=/reference/archive/200506/google.html (July 24, 2006).
10. Peter Jacso, "As We May Search-Comparison of Major Features of the Web of Science, Scopus, and Google Scholar Citation-Based and Citation-Enhanced Databases," *Current Science,* 89 (November 10, 2005): 1537–1547.
11. Nisa Bakkalbasi, Kathleen Bauer, Janis Glover, and Lei Wang, "Three Options for Citation Tracking: Google Scholar, Scopus and Web of Science," *Biomedical Digital Libraries,* 3 (June 29, 2006). Available: http://www.bio-diglib.com/content/3/1/7 (July 22, 2006).
12. Lokman Meho and Kiduk Yang, "A New Era in Citation and Bibliometric Analyses: Web of Science, Scopus, and Google Scholar," (Accepted for publication in the *Journal of the American Society for Information Science and Technology*). Available: http://arxiv.org/ftp/cs/papers/0612/0612132.pdf (April 30, 2007).
13. Peter Jacso, 2004.
14. Peter Jacso, 2005. "Side-by-Side2".
15. Mary Robinson and Judith Wusteman, "Putting Google Scholar to the Test: A Preliminary Study," *Program: Electronic Library and Information Systems,* 41 (2007): 71–80.
16. Susan Gardner and Susanna Eng, "Gaga Over Google? Scholar in the Social Sciences," *Library Hi Tech News,* 22 (September, 2005): 42–45.
17. Dean Giustini and Eugene Barsky, "A Look at Google Scholar, PubMed, and Scirus: Comparisons and Recommendations," *Journal of the Canadian Health Libraries Association/Journal de l'Association des Bibliothèques de la Santé du Canada,* 26 (Summer, 2005): 85–89 Available: http://weblogs.elearning.ubc.ca/googlescholar/giustini_barsky.pdf (July 22, 2006).
18. William Walters, "Google Scholar Coverage of a Multidisciplinary Field," *Information Processing & Management*, 43 (2007): 1121–1132.
19. Michael Levine-Clark and Joseph Kraus, "Finding Chemistry Information Using Google Scholar: A Comparison with Chemical Abstracts Service," (Accepted for publication in *Science and Technology Libraries*). Available: http://eprints.rclis.org/archive/00008633/01/Levine-Clark_Kraus.pdf (April 20, 2007).
20. Chris Neuhaus, Ellen Neuhaus, Alan Asher, and Clint Wrede, "The Depth and Breadth of Google Scholar: An Empirical Study," *Portal: Libraries and the Academy,* 6 (April, 2006): 127–141.
21. Alireza Noruzi.
22. Lokman Meho and Kiduk Yang.
23. Rena Helms-Park, Pavlina Radia, and Paul Stapleton, "A Preliminary Assessment of Google Scholar as a Source of EAP Students' Research Materials," *The Internet and Higher Education,* 10 (2007): 65–76.
24. Rita Vine, "Google Scholar is a Full Year Late Indexing PubMed Content," *Site-Lines: Ideas About Web Searching,* (February 23, 2005). Available: http://www.workingfaster.com/sitelines/archives/2005_02.html#000282 (July 22, 2006).
25. Chris Neuhaus, Ellen Neuhaus, Alan Asher, and Clint Wrede.
26. Rita Vine, "Google Scholar Gets Better at Indexing PubMed Content, but It's Still Several Months Behind," *Site-Lines: Ideas About Web Searching,* (January 9, 2006). Available: http://www.workingfaster.com/sitelines/archives/2006_01.html#000365 (July 22, 2006).
27. Laura Bowering Mullen and Karen A. Hartman, "Google Scholar and the Library Web Site: The Early Response by ARL Libraries," *College & Research Libraries,* 67 (March, 2006): 106–122.
28. Norman Oder, "Google Scholar Links with Libs," *Library Journal,* 130 (April 15, 2005): 17–18.
29. Jeffrey Young, "More than 100 Colleges Work with Google to Speed Campus Users to Library Resources," *The Chronicle of Higher Education,* 50 (May 20, 2005): A30.
30. Bernie Sloan, "Creating a Google-Based Site Search Capability," *Web4lib,* (November 2, 2005). Available: http://lists.webjunction.org/wjlists/web4lib/2005-November/038917.html (July 24, 2006).
31. Michael Holman, "Plagiarpolicogony: Homan Said Let there be a New Plagiarism Policy, and There Was, and Several Students Cried in his Office, but Overall it was Good," *A Personal Journal About Teaching the Bible and Ancient Near Eastern History/Theology/Religion/Archaeology to University,* (April 27, 2005). Available: http://michaelhoman.blogspot.com/2005/04/plagiarpolicogony-homan-said-let-there.html (July 24, 2006).

32. Electronic Resources Work Group, University of Illinois Libraries, *Meeting Summary* (May 25, 2005). Available: http://www.library.uiuc.edu/administration/collections/committees/erwg/erwg_minutes_5_2005.doc (July 24, 2006).

33. Carnegie Foundation for the Advancement of Teaching, *Basic Classification Description.* Available: http://www.carnegiefoundation.org/classifications/index.asp?key=791 (February 25, 2007).

34. Taro Yamane, *Elementary Sampling Theory* (Englewood Cliffs, NJ: Prentice-Hall, 1967).

35. Google, "Results Page," *GoogleGuide: Making Searching Even Easier.* Available: http://www.googleguide.com/results_page.html (April 27, 2007).

36. Jonathan Rochkind, "(Meta)search Like Google," *Library Journal,* 132 (February 15, 2007): 28–30.