Borgeson Paper Archive   Law Librarianship Program

2005

The Next Big Thing?: Google Scholar and Legal Scholarship

Shannon Malcolm

Follow this and additional works at: https://digitalcommons.law.uw.edu/law-lib_borgeson

Recommended Citation

THE NEXT BIG THING?: GOOGLE SCHOLAR AND LEGAL SCHOLARSHIP

Shannon L. Malcolm[1]

"While some look to the Internet as an innovative vehicle for communication, the court continues to warily and wearily view it largely as one large catalyst for rumour, innuendo, and misinformation."[2]

I. OVERVIEW

For better or worse, online research using electronic resources is rapidly displacing dependence upon traditional (i.e., printed) resources, especially for younger legal researchers.[3] It only makes sense, then, that more and more companies have moved to tap the market of academic research online; just as companies like LexisNexis seized the opportunity to target products to students years ago,[4] now Google is seeking to reach the academic community's Internet research needs with its latest offering: Google Scholar.

One critic has dismissed Google Scholar with the pronouncement that it "is an interesting and noble but less-than-groundbreaking contribution to research."[5] It is easy to be unimpressed by a new application of the now familiar search engine, especially an application still in its testing stages. I believe Google Scholar is not important for any groundbreaking contributions it has made so much as for the potential for groundbreaking developments it suggests. This paper attempts to analyze the ramifications of Google Scholar for law librarians and others interested in scholarly legal research, discussing how it works, how to use it, and how it stacks up against comparable research tools (including LegalTrac, Hein Online, Westlaw, and LexisNexis). An assessment of Google Scholar's problems, both real and imagined, is made, with suggestions for possible solutions to the former.

II. BACKGROUND

A. Whence It Came: The Brief History of Google Scholar

In 2004, Google released Google Scholar, a new Internet search engine designed to locate academic material online for researchers interested in scholarly and scientific literature.[6] Designed by Anurag Acharya, formerly a professor of computer science with the University of California at Santa Barbara, the engine "searches only research publications such as journal articles, books, preprints and technical reports, [ranking results] by means of algorithms[;] papers

---

[1] Shannon L. Malcolm is an intern for the Marian Gould Gallagher Law Library at the University of Washington in Seattle.

[2] St. Clair v. Johnny's Oyster & Shrimp, Inc., 76 F. Supp. 2d 773 (S.D. Tex. 1999).

[3] Thomas Keefe, *Teaching Legal Research from the Inside Out*, 97 L. LIBR. J. 117, 119-21 (2005).

[4] *Id.*

[5] Andrew Goodman, *Google Scholar vs. Real Scholarship*, TRAFFICK: INTERNET SEARCH ENLIGHTENMENT, *at* http://www.traffick.com/2004/11/google-scholar-vs-real-scholarship.asp (last visited Mar. 7, 2005).

[6] Declan Butler, *Science Searches Shift Up a Gear as Google Starts Scholar Search Engine*, 432 NATURE 423 (2004).

with many citations are generally ranked highest, and they get a further boost if they are referenced by highly cited articles."[7]

Google Scholar is working with seventeen institutions: Australian National University, Cornell University, Cranfield University, the European University Institute, Hong Kong University of Science and Technology, Indiana University-Purdue University at Indianapolis, Massachusetts Institute of Technology (whose DSpace software enabled the development of institutional electronic repositories in the first instance), Minho University, Ohio State University, the University of Arizona, the University of Calgary, the University of Oregon, the University of Parma, the University of Rochester, the University of Toronto, the University of Washington, and the University of Wisconsin.[8] These institutions have made their repositories of materials—an average of about one thousand documents per institution, many of them unpublished and therefore widely unknown outside of those institutions—available for indexing by Google Scholar, with the Online Computer Library Center ("OCLC") providing the interface between participating universities and Google Scholar.[9] Google Scholar also includes content from the Institute of Electrical and Electronics Engineers ("IEEE"), Association for Computing Machinery ("ACM"), OCLC's Worldcat, the NASA Astrophysics Data System, airXiv server, and PubMed; although Elsevier has refused to cooperate with the project, Google Scholar generates results of abstracts for over one million of that publisher's articles.[10] Yet no exhaustive, authoritative list can be presented of what resources are and are not accessible via Google Scholar, because the company will not disclose the size of its index or the number of publishers included therein, stating only that "most major scholarly publishers are included," but not specifying those that are so included.[11] This inscrutability is a major sticking point for Google Scholar's viability because it hinders' users' ability to assess what kinds of information the new tool is and is not able to locate.

B. How It Works: Functions of Google Scholar

Google Scholar's more familiar progenitor, Google, employs an algorithm called PageRank, which essentially "turns the Internet into a huge popularity contest" by ranking results largely "according to the number of times they are linked to by other websites."[12] In contrast, its developers claim that Google Scholar's relevance rankings consider an "article's author, the publication in which the article appeared, and how often it has been cited in scholarly literature."[13] Another distinguishing feature of Google Scholar is its inclusion of citations to materials not available online.[14] While the inclusion of such materials may be frustrating to users seeking instant gratification, it is in fact very valuable, since, as Acharya points out, "Not

---

[7] *Id.*

[8] Jeffrey R. Young, *Google Tests Search Engine for Colleges' Scholarly Materials*, CHRON. HIGHER EDUC., Apr. 23, 2004, at A36.

[9] *Id.*

[10] Krishnan Thiagarajan, *For the Scholar in You*, BUSINESSLINE, Dec. 6, 2004, at 1.

[11] *Id.*

[12] Chris Gaither, *Google Designs an Engine for Eggheads; With Its Scholar Research Software, the Web Giant Hopes to Tap into a New Source of Traffic—and Revenue*, L.A. TIMES, Dec. 27, 2004, at C1.

[13] Thiagarajan, *supra* note 8.

[14] *Id.*

knowing about a paper that is relevant to your work is much worse than actually knowing it exists but not being able to get to it immediately."[15]

To help researchers more efficiently obtain desired documents, even where they are not available online to the public at large, Google Scholar has implemented a feature whereby searchers can set preferences to display "institutional access links" for up to three participating organizations.[16] This feature facilitates easy hyper-linking directly to resources, such as commercial databases, which the searcher may access if affiliated with a subscribing institution.[17] It also helps to localize results in situations where searchers desire to do so, without sacrificing the wide-ranging capabilities Google Scholar promises; such options for customization are crucial for building effective search engines.[18] Notably lacking from Google Scholar is any kind of hierarchical arrangement which might accommodate browsing. However, Google has always geared its products more towards searching than browsing, and this has helped it to focus upon a simple interface.

Unlike Google, Google Scholar does not permit searchers to control for the document format of results retrieved (e.g., excluding—or including only—those with .html, .rtf, .pdf, .doc, or similar file extensions) nor does it allow control over the domain of results retrieved (e.g, excluding—or including only—sites from .org, .edu., or similar types of domains).[19] Searchers can, however, search for documents written by a particular author, published in a particular publication, or published in particular years (Google Scholar's interface uses the term "date" but in fact one can only limit searches to particular years or spans of years; thus, the term "date" implies a level of granularity not actually available).[20]

C. How It's Used: Google Scholar's Interface

Google Scholar defaults to the now familiar, simple interface of a blank field inviting users to enter search terms without further direction or explanation.[21] It does however offer an advanced search option, also serving up the usual suspects: fields that apply Boolean *ands* or *ors* to terms entered therein, that will search for an exact phrase, and that will exclude certain terms (thus accommodating searches such as "hip-hop NOT rabbit").[22] Users can choose whether they wish to retrieve any documents containing the sought terms, or only those documents containing

---

[15] Daniel Terdiman, *A Tool for Scholars Who Like to Dig Deep*, N.Y. TIMES, Nov. 25, 2004, at G6.

[16] This feature remains a pilot project; currently, participating institutions include: American University of Beirut, Bates College Brandeis University, California Institute of Technology, Canterbury Christ Church University College, Colgate University, College of Saint Benedict at Saint John's University, Consejo Superior de Investigaciones Cientificas, Duke University, Emory University, Hampshire College, Harvard University, Houston Academy of Medicine at the Texas Medical Center, Los Alamos National Laboratory, Massachusetts Institute of Technology, Stanford University, Technische Universität München, Universiteit van Amsterdam, University of Calgary, University of California, University of Iowa, University of Manitoba, University of Michigan, University of Oregon, University of Pennsylvania, Vrije Universiteit Brussel, Washington University in St. Louis, and Yale University. *Google Scholar Beta Preferences*, at http://www.scholar.google.com/scholar_preferences (last visited Mar. 18, 2005).

[17] *Id.*

[18] Michael S. Malone, *Yahoo!: The Invisible Giant Turns 10*, WIRED, Mar. 2005, at 120, 123.

[19] Faisal al Yafai & Steven Morris, *Mars Shines but Hamlet is Out of Question*, GUARDIAN, Nov. 22, 2004, at 7.

[20] *Google Scholar Beta Advanced Search*, at http://www.scholar.google.com/advanced_scholar_search (last visited Mar. 7, 2005).

[21] *Google Scholar Beta*, at http://www.scholar.google.com/ (last visited Mar. 22, 2005).

[22] *Google Scholar Beta Advanced Scholar Search*, at http://www.scholar.google.com/advanced_scholar_search (last visited Mar. 7, 2005).

the terms in their titles. Users can also choose how many results are to be displayed per page (ten, twenty, thirty, fifty, or one hundred).[23]

The *Advanced Search Tips* page is praiseworthy. It explains proper usage of common Boolean operators, including how to search for terms which would otherwise be identified as stop words and thus ignored by the search engine.[24] It also includes practical tips to assist novice users in searching more effectively by, for example, pointing out that publications may have variant spellings or abbreviations and reminding users that "information may be incomplete or even incorrect; many preprints, for instance, don't say where (or even whether) the article was ultimately published."[25] This statement is reassuring in that, while it demonstrates Google Scholar's shortcomings, its candor allays any concerns about developers' frankness in acknowledging those shortcomings.

Results are displayed with prefatory descriptors. If a result is a book, the word "book" appears in brackets before the result. Searchers interested in a book can follow one of two hyperlinks: "Library Search" allows users to determine what libraries hold the book (similarly to a search of Worldcat), and "Web Search" takes researchers to sites online where the book may be purchased.[26] A bracketed indicator similarly marks citations to a document in order to differentiate them from the document itself. Thus, a search for "Janis Johnston" returns a result for an article by her[27] not itself available via Google Scholar[28] but which is cited by a document that is available via Google Scholar.[29] Users can jump to the citing document via a hyperlink. This feature serves two purposes: it helps users locate material not necessarily available online (via the free Web, at any rate), just as the inclusion of books mentioned earlier does; it also assists with the familiar technique of citation chaining, a technique whereby the documents cited in a known, relevant source are sought out as additional, potentially useful documents, themselves containing still more citations to potentially useful material, and so on.[30]

III. RAMIFICATIONS FOR LEGAL SCHOLARSHIP

A. A View from the Trenches: The Legal Researcher's Perspective

The most obvious advantage Google Scholar has over analogous tools such as the four mentioned earlier (Hein Online, LexisNexis, Westlaw, and LegalTrac) is its cost: as of now, it is available free to the public, and since it provides access to citations for "documents currently

---

[23] *Id.*

[24] *Google Scholar Beta: Advanced Scholar Search Tips, at*

http://www.scholar.google.com/scholar/refinesearch.html (last visited Mar. 22, 2005).

[25] *Id.*

[26] *Google Scholar Beta: About Google Scholar, at* http://www.scholar.google.com/scholar/about.html (last visited Mar. 26, 2005).

[27] Janis L. Johnston, *Outsourcing: New Name for an Old Practice,* 88 L. Libr. J. 128 (1996).

[28] *Law Library Journal* is not available on the free web prior to volume 92 (2000). American Association of Law Libraries, *Products and Publications: Law Library Journal,* at http://www.aall.org/products/pub_journal.asp (last visited Mar. 25, 2005).

[29] Robert S. Martin et al., *The Impact of Outsourcing and Privatization on Library Services and Management: A Study for the American Library Association* (2000), *at*

https://www.ala.org/ala/oif/iftoolkits/outsourcing/outsourcing_doc.pdf (last visited Mar. 25, 2005).

[30] *See* M. S. Dalton & L. Charnigo, *Historians and Their Information Sources,* 65 COLLEGE & RESEARCH LIBR. 400-425 (2004), R. Delgadillo & B. P. Lynch, *Future Historians: Their Quest for Information* 60 COLLEGE & RESEARCH LIBR., 245-259 (1999).

held behind subscription pay walls"[31] it can be useful for locating resources for procurement via resource-sharing channels (e.g., inter-library loans) even if a researcher's affiliated institution does not subscribe to a particular pay service, since it provides the citation information necessary to facilitate such resource sharing. Nevertheless, there is truth to the cliché that one gets what one pays for; even though Google Scholar is currently free, it is inferior to many commercially available analogues, as a recent review by Péter Jasco points out:

> The citedness [sic] scores of documents displayed in the results lists have great potential for choosing the most promising articles and books on a subject, but they are often inflated. The prominent display of the citedness scores could help the scholars and practitioners whose libraries don't have access to the best citation-based systems, such as Web of Science and Scopus, or to the smartest implementations of citation-enhanced abstracting indexing databases, like some on CSA and EBSCO. [Other] open-access services . . . such as CiteBase, Research Index and RePEc/LogEc . . . handle citing and cited references and citedness scores much better . . . . [Google Scholar] picked up . . . many redundant and irrelevant pages and ignored a few million full-text scholarly papers and/or their citation/abstract records.[32]

Jasco goes on to point out that the crux of the problem lies in Google Scholar's treatment of the items in the archives used for the project "as any of the zillions of unstructured pages on the Web" and explains how Google Scholar compares negatively to myriad commercial services.[33]

However, neither Jasco's review nor any of the others available so far addresses Google Scholar's specific utility for legal researchers, nor compares it to analogous specialty legal researching tools. The resulting gap in the literature is addressed hereafter by a comparison between Google Scholar and comparable tools used to research legal scholarship.

Google Scholar retrieves only scholarly material, which is to say, for legal researchers, secondary material. It cannot (nor does it purport to) compete with the all-in-one, integrated resources provided by industry giants LexisNexis and Westlaw. Even scholarly legal research more often than not entails primary materials such as cases, statutes, and regulations, and Google Scholar is not the place to go for these materials. Although the increasing availability of such materials via the free Web—from governmental (both state and federal) and non-profit clearinghouse sites[34]—makes locating them online less problematic than it used to be, the value added to them by commercial providers remains all but indispensable; we remain a very long way off indeed from being able to dismiss the utility of a tool like the *United States Code Service*, available from LexisNexis, simply because the *United States Code* is available online for free from any number of sources. The annotations that facilitate cross-referencing to related parts of the code, regulations, case law, and secondary sources; the indexing, and tools like popular names tables all provide utility unmatched by even the best free resources (be they electronic or printed) known to me. Nevertheless, if open access tools like Google Scholar prove

---

[31] Andrew Albanese, *Google Launches Scholarly Search Service*, 251 PUBLISHERS WEEKLY 13 (2004).

[32] Péter Jasco, *Google Scholar Beta*, Péter's Digital Reference Shelf, *at* http://www.galegroup.com/servlet//reference/archive/200412/googlescholar.html (last visited Mar. 7, 2005).

[33] *Id.*

[34] *See, e.g.*, Access Indiana, *Law and Justice*, *at* http://www.in.gov/ai/law/ (last visited Mar. 26, 2005), Cornell University Law School, *Legal Information Institute*, *at* http://www.law.cornell.edu/ (last visited Mar. 26, 2005), *Findlaw*, *at* http://www.findlaw.com/ (last visited Mar. 26, 2005), Municipal Research and Services Center, *LegalWA.org*, at http://www.legalwa.org/ (last visited Mar. 26, 2005), U.S. Government Printing Office, *GPO Access*, *at* http://www.gpoaccess.gov/index.html (last visited Mar. 25, 2005).

effective and popular, they can be used in conjunction with other online tools that retrieve primary material to facilitate research slightly less conveniently, but significantly less expensively, than is possible via commercial services. Google Scholar is a tool designed for scholars, as that community is most broadly understood. It is worth observing that legal scholars are a somewhat odd lot, who embody the needs of both the erudite academic theorist and the practicing professional. Perhaps a later, more specialized incarnation of Google Scholar (or a competitor) could address this fact and provide specialized hybrid tools better designed to serve those needs by "paving the way for similar services targeting other reseach-heavy professions[, such as] Google Lawyer."[35]

B. A Test Drive: Comparing Google Scholar to Legal Scholarship Research Tools

In order to better understand Google Scholar's applications for scholarly legal research, I sought to compare it to four leading tools for locating legal scholarship: Hein Online, LegalTrac, LexisNexis, and Westlaw.[36]

To see how Google Scholar compared to competitors, I searched for scholarship produced by particular authors about a particular field in which they had recognized expertise. I selected scholars from a variety of institutions, both those participating in the Google Scholar project and those not participating. I attempted to select scholars from a variety of fields of legal scholarship. I also selected scholars for whom I could easily obtain bibliographies, either via *curriculum vitae* or otherwise. I decided to focus on Google Scholar's searching capabilities, as opposed to its browsing capabilities, because that is its strength. The searches were undertaken using each of the five tools—Google Scholar, Hein Online, LegalTrac, LexisNexis, and Westlaw. These searches were not intended to be any kind of rigorous statistical analysis of the systems in question. Rather, they represent an attempt to explore the differences among the systems via empirical, albeit anecdotal, observations about how they compared to each other.

The single most salient observation I made in my experiments with Google Scholar is how few documents it retrieved in comparison to what specialty legal research analogues did. Comparing Google Scholar's retrievals to the *curriculum vitae* of scholars showed that it failed to retrieve the entire *oeuvre* of any of the scholars used for the testing. This finding was true regardless of whether the scholars were affiliated with an institution participating in the Google Scholar project (i.e., one of the seventeen making their archival materials available to Google Scholar's crawlers). Not only did it not retrieve many of the documents the scholars are known

---

[35] Gaither, *supra* note 12.

[36] Hein Online is a subscription service that includes some 500 legal periodicals (primarily law reviews and journals) in .pdf format, and accommodating full-text searching. LegalTrac is a subscription service providing indexing to some 1,300 legal periodicals (law reviews, law journals, and bar publications, as well as "law-related articles from over 1,000 additional business and general interest titles") published from 1980 to the present; 109 titles are available in full-text, but only citations and sometimes abstracts are available for the others, precluding full-text searches. Westlaw and LexisNexis provide full-text articles from legal periodicals in .html format; they contain some 600 and some 550 titles, respectively; coverage begins in the early 1990s for most titles in both systems. Barbara Monroe, *Georgetown Law Library Research Guides: Using Articles for Legal and Non-Legal Research, at* http://www.ll. georgetown.edu/guides/articles.cfm (last visited Apr. 10, 2005). Comparisons with the University of Washington's Current Index to Legal Periodicals ("CILP") were decided against because CILP is primarily a current awareness tool, not an archival database of the sort chosen for comparisons. *Marian Gould Gallagher Law Library, Current Index to Legal Periodicals, at* http://lib.law.washington.edu/cilp/cilp.html (last visited Mar. 26, 2005). The idea behind this decision is that Google Scholar purports to accommodate scholars proactively seeking information, and that not necessarily current.

to have authored, at times the results Google Scholar did retrieve were downright bizarre. For example, Nancy Hardin Rogers, Dean of the Moritz College of Law at Ohio State University (which is a participating institution) and a recognized authority in the area of mediation and other types of alternative dispute resolution, has authored some twenty-one documents.[37] Keyword searching in Google Scholar retrieved only six of these, two of which were available online in full-text (the other four were books). Three additional results generated by this search could not be confirmed as being actually authored by Rogers or not, since details for two were available only to subscribers of a pay service unavailable to me, and a third led to a dead link. Imagine my confusion when I discovered that, while keyword searching generated these meager results, a focused search for material by Rogers, using the "author" field, generated no results! It is certainly no surprise that a broader keyword search retrieves results that a focused search will not when one is casting about without the goal of retrieving specific documents. But when, armed with the author and title of a document previously retrieved by a broad keyword search, one later implements a focused search for that document (as I did), it ought to be easily retrieved. In all likelihood, this anomaly represents little more than some sort of technical glitch which will be remedied when Google Scholar moves beyond its current Beta incarnation. I nevertheless briefly wondered, tongue planted firmly in cheek, whether perhaps the folks at Google have a diabolical plan to increase users' dependence upon their famously simple interfaces,[38] to the exclusion of more advanced searching capabilities. In short, Google Scholar is both under-inclusive and over-inclusive depending upon what kinds of searches one constructs; broader keyword searches retrieve significant amounts of chaff with the wheat, yet more carefully-constructed searches of specific meta-data fields may retrieve no wheat at all.

Another major weakness of Google Scholar when compared to the specialized legal research tools is the number of redundant results it generates. This characteristic can lead to frustrating and time-consuming work to separate the wheat from the chaff. Imagine searching for works by a particular author or about a particular topic and receiving, say, twenty-nine results only to find, upon reading over them, that there were in fact only ten unique results. The time spent discovering this disparity is wasted time for busy legal scholars, be they professors, students, or librarians. And the greater the numbers involved, the more vexing the phenomenon becomes. This problem should be remedied with an application allowing users to eliminate duplicate results.

Google Scholar seems to function better when processing broad, unfocused (some might say ham-fisted) searches than do the other tools, since it returned relatively more relevant results for full-text keyword searches (as opposed to field-specific searches) than did its counterparts. For instance, while generally fewer results came up in Google Scholar than in other tools (in both keyword searches and those seeking terms in specific fields), those that were returned were likely to be written by an author whose name was used as a keyword or discuss a topic used as a keyword. While results for such broad keyword searches in the other four tools returned more results (many more for LexisNexis and Westlaw, predictably enough), those results were mostly for documents containing citations to the work of the author whose name was entered as a keyword. This phenomenon comports with the Google company's fame for developing tools

---

[37] *Dean Rogers' C.V.*, at http://moritzlaw.osu.edu/faculty/facultyprofiles/rogers.html (last visited Mar. 24, 2005).

[38] *See* Mary J. Koshollek, *"Google" Your Way to Better Web Searching: Now You Can Create Specialized Searches for Customized Results Using Google, the World's Biggest Search Engine Database*, 76 WIS. LAWYER 32 (2003) ("Since its inception, Google's corporate philosophy has been to present a clean, simple interface, free from the ubiquitous ads that plague other search sites.").

which return uncannily relevant results for even poorly constructed searches. Thus, perhaps Google Scholar will continue its progenitor's legacy as the first stop for novices, providing good ready reference that can be followed up by more refined searching using more sophisticated tools if it proves unavailing: "We know most scholarly researchers start with Google[; w]hen we are honest with ourselves, we admit that most librarians start with Google. It's a fine place to play with words, look for links, and focus a search before heading for the 'quality stuff' . . . ."[39]

I have already mentioned the advantage of one-stop shopping that LexisNexis and Westlaw provide over Google Scholar (for a price). Another advantage these two services, along with Hein Online, offer over the newcomer is the proportion of documents available in full-text online. While Google Scholar does provide access to more full-text items than LegalTrac, the latter's *raison d'etre* is its indexing power. Yet Google Scholar actually outperformed LegalTrac inasmuch as it also recovered more citations to documents than the latter in the searches I ran. It was not the case that Google Scholar retrieved everything fitting a given search parameter that LegalTrac did and then some; the results returned by the two tools were not particularly duplicative of each other. Nevertheless, among the four tools to which I compared Google Scholar, LegalTrac was hand over fist given the strongest run for its money, and that in a very real sense, since LegalTrac is a pay service and Google Scholar is free. Google Scholar tends to return more results than LegalTrac, and more of them in full-text. Both tools retrieved similarly relevant material, but of different types, as I will explain directly.

The comparable, though not identical, results generated by Google Scholar and LegalTrac beg the question "Right, then, in what ways did they differ?" The answer is simple: They differ in exactly the ways one would expect. Google Scholar does a much better job at locating books, as well as pre-prints, conference materials, white papers, and similar ostensibly unpublished materials than do LegalTrac, Hein Online, LexisNexis, or Westlaw.[40] Google Scholar's location of books is nothing impressive to anyone who's familiar with Worldcat, of course; but the location of heretofore secreted materials in nebulous, ostensibly unpublished states is quite nice.

Another thing Google Scholar shines at is location of materials in resources which are unknown and arcane (and hence neither searched in nor discovered) from the legal researcher's point of view. I am a great believer in the value of specialization, but as legal scholarship becomes increasingly interdisciplinary,[41] the need for legal scholars and the information professionals who assist them to locate material outside the universe of traditional legal resources is likely to increase. Often, Google Scholar returns results for material available via services such as Blackwell-Synergy and SpringerLink, which are unfamiliar to many legal scholars, but which could prove useful to them, and which are often available via subscriptions through their affiliated institutions and therefore may as well be utilized. A perennial problem with online versions of material originally published in print is the failure of online interfaces to

---

[39] Stephen Abram, *Google Scholar: Thin End of the Wedge?*, INFORMATION OUTLOOK, Jan. 2005, *at* http://www.findarticles.com/p/articles/mi_m0FWE/is_1_9/ai_n10300800.

[40] This finding is predictable, since those tools have well-defined databases primarily consisting of articles published in traditional periodicals.

[41] This trend is best exemplified by the proliferation of the "law and (philosophy, economics, literature, etc.)" traditions and work in developing areas such as health law and technology law which require expertise and material from other areas in addition to the law. *See* James G. Milles, *Leaky Boundaries and the Decline of the Autonomous Law School Library*, 96 L. LIBR. J. 387 (2004).

properly display documents involving graphics,[42] but Google Scholar often retrieved such documents in their entirety, including any accompanying graphics. Of the other legal research tools tested, only Hein Online, with its handy .pdf format, has this capacity; LegalTrac contains relatively few full-text documents, and the predominant .html format employed by LexisNeis and Westlaw mean that users of those services are unable to view charts, tables, diagrams, and such which may have been included in an article as it was originally published. In addition, because of the participation of institutions from multiple nations, Google Scholar may be more likely than the other four tools to locate material from outside the United States, given the limited foreign material contained in their databases.[43]

One of Google Scholar's most striking limitations is that it does not enable users to search by subject, (or abstract) as does LegalTrac. Indeed, this limitation was in part responsible for the clumsiness of the tests and comparisons I was able to make. In addition to a capacity to search for specific domains and document types, it would be desirable if users could search by subject,[44] title, or type of work (e.g., article, book, white paper, conference proceeding, etc.). Since Google Scholar currently remains in testing stages, it is to be hoped that other searchable fields will be added to it as it develops.

IV. LOOKING AHEAD

A. Google Scholar's Potentially Problematic Impact upon Unwary Researchers

Paula Hane recently pointed out that "Google is a huge brand for good reason—it's good for general purposes, but it's still frustrating for scholarly research."[45] Hane goes on to discuss the prevalent ignorance about focused, topical tools, even among scholars themselves and other ostensibly expert researchers, and the possibility that Google Scholar may become commonly

---

[42] Steven J. Bell, *Is More Always Better?: When Quality is the Goal, Access to Everything May Not Be the User's Best Bet*, 34 AM. LIBR. 44, 46 (2003).

[43] Anthony Paonita, *Tech Trends Hands On: Googling is Good for You*, N.Y. L. J. (July 2, 2001) ("Believe it or not, everyone does not write in English, and there is useful information out there in such languages as French, German, Italian and Scandinavian languages [sic].")

[44] Some might argue that keyword searching is rendering subject searching obsolete, but a strong argument can be made for the continuing value of indexing, abstracting, and the assignment of subject terminology that can result, despite the power of full-text searching. The reasoning is that a document may or may not actually contain those terms that best describe its subject (consider an article about the phenomenon of absentee parents, latchkey children, and the sociological ramifications thereof which nevertheless includes neither the terms "family" nor "sociology." While experienced searchers, such as librarians, may overcome the problems these kinds of situations present by anticipating them and formulating clever searches accordingly, such solutions may require lengthy and detailed reference interviews of the sort impetuous and harried students and faculty can be loathe to entertain, and even then problems may persist. For instance, what of the researcher who fails to find relevant material and gives up without consulting an expert researcher? Systems should be designed with such possible problems (and solutions to them) in mind; subject classification is precisely such a solution and ought not to be dismissed out of hand. Homonymy presents another obstacle to relying solely upon full-text keyword searching. Walt Crawford, *The Card Catalog and Other Digital Controversies: What's Obsolete and What's Not in the Age of Information*, 30 AMER. LIBR. 53, 56 (1999). I believe that properly constructed nested, Boolean searches can significantly, though not completely, alleviate both problems, but such a discussion is beyond the scope of this paper.

[45] Paula J. Hane, *Getting "Scirus About Scholarly Content*, NEWSLINK *at* http://www.infotoday.com/newslink/newslink0412.htm. (quoting Ammy Vogtlander). In the interest of full disclosure, note that Vogtlander is general manager of Scirus, an Elsevier service competing with Google Scholar. Note also that her observation is nevertheless absolutely true.

used even when better, more appropriate tools might fit the bill.[46] Two aspects of legal scholarship may alleviate this potential problem. Legal scholars, primarily faculty at law schools, are often more involved with the analysis of material than location thereof; much of their research is performed not by these professors themselves but by research assistants (usually law students) and law librarians.[47] The former are all too likely to fall into the trap of relying upon Google Scholar and overlooking more effective tools, while (it is to be hoped!) the latter are especially unlikely to do so. Ironically, the more obscure a given specialized research tool is, the more arcane the materials it contains, the more likely it is to be the most effective means of locating and retrieving those materials—and the more likely it is to be overlooked by relatively inexpert users like professors and law students. After all, every professor or student of law worthy of the title is familiar (if not necessarily adroit) with LexisNexis and Westlaw. But how many are familiar with Pacer, Worldcat, the Derwent World Patents Index, Law Library Microform Consortium Digital, or Securities Mosaic, much less the specialized resources outside the legal field which may nevertheless be critical in light of the aforementioned growth of interdisciplinary scholarship?

This dilemma is indicative of the growing need to emphasize not mere location of materials, but proper verification that those materials are authoritative and reliable.[48] Even if Google Scholar and its ilk make locating information easier, concern must shift to ensuring that information is accurate; law librarians and other legal research educators need to teach both students and faculty how to evaluate what they find, especially since "many students [have] almost unquestioning faith in information derived from the Web, no matter the source."[49] Google Scholar, despite its tendency to be both under-inclusive and over-inclusive, is most attractive to students, especially the large number of students whose penchants for procrastination lead to a demand for materials in a hurry. In addition, more and more professors are not to be outdone by their pupils in the realm of procrastination and impatience, any popularly received notions to the contrary notwithstanding. The temptation to rely upon Google for a quick fix to last-minute research may be exacerbated by Google Scholar, which purports to be a more respectable and therefore reliable tool: "Google has made us complacent in a certain way, [but] the reality is that . . . search engine returns are incomplete for research purposes."[50] Indeed, Google Scholar will be more popular with students than faculty precisely because of its shortcomings, since earnest scholars (i.e., responsible, properly educated faculty) will hesitate to use it because of them. It is therefore imperative for students to understand that, just as Google, while wonderful, is no panacea, neither is Google Scholar, any hype to the contrary notwithstanding.[51] Furthermore, as younger, more technologically savvy persons join faculties, they too must be warned about Google Scholar's shortcomings.

---

[46] *Id.*

[47] Ed Edmonds, *The University of St. Thomas Law Library: A New Library for a New Era in Legal Education,* TRENDS IN L. LIBR. MGMT. & TECH., Apr. 26, 2002, at 5.

[48] Nicholas Pengelley, *I Google Therefore I Am,* 28 CAN. LAW LIBR. 66 (2003).

[49] *Id.*

[50] Molly McDonough, *In Google We Trust?: Critics Question How Much Judges, Lawyers Should rely on Internet Search Results,* 90 A.B.A. J. 30 (2004) (internal quotes omitted).

[51] Another problem of relying too much on electronic tools like Google Scholar is the loss of context that can occur. Judy Meadows & Kay Todd, *Our Question—Your Answers,* 13 PERSPECTIVES: TEACHING LEGAL RESEARCH AND WRITING 113 (2005) ("Without [the big] picture . . . legal research becomes a matter of fact patterns and key words instead of legal principles[;] information can easily be missed by using only electronic resources.") (internal quotes omitted).

Neither my own explorations of Google Scholar nor those in the literature indicate problems of inaccuracy or unreliability in the records it retrieves; rather, its shortcomings lie in its incompleteness, which Acharya himself has acknowledged while pointing out that Google Scholar merely harvests what is available, stating that "data providers do not always provide full information. We are not able to distinguish the adequate records at this point . . . ."[52] Again, Acharya and his colleagues' honesty is laudable, but their recalcitrance to reveal details about how Google Scholar functions (understandable though it may be to ensure protection of intellectual property) in many ways stonewalls effective analysis.

For instance, while "Google Scholar gathers bibliographical data from many sources, including automatically extracting it from text and citations"[53] it is unclear whether the engine makes allowances for textual variations which represent the same concept. Does its software recognise that the Mississippi Law Journal and the Miss. L.J. are one and the same? Apparently it does not, given the admonitions for caution that appear in the interface and my own experiments. Indeed, searching for "Miss LJ"[54] in Google Scholar retrieves only six results, none of which have anything to do with the periodical thus abbreviated, and most of which involve an acknowledgement of gratitude to an apparently unmarried woman whose first two initials are "L.J."; only "Mississippi Law Journal" yields relevant results. Consistency is lacking, however, as searches for "Harv. L. Rev.", "Ga. L. Rev." and "Geo. L.J." retrieve results from and referring to those legal periodicals. In many ways, the more I explore and experiment with Google Scholar, the greater my uncertainty about exactly how it operates. Jasco expressed similar frustrations, putting me in good company, but boding ill for the engine's efficacy in its current state.

B. Google Scholar's Potential (Inevitable?) Commercialization

There is no guarantee that Google Scholar will remain free. Google currently generates some three billion dollars annually from sales of advertisements displayed in conjunction with certain keywords, and such advertisements might be worth even more to marketers wishing to realize the potential for targeting niche demographics that tools like Google Scholar offer.[55] Indeed, according to its most recent earnings statement, "Google reported record revenues of $1.032 billion for the quarter ended December 31, 2004, up to 101% year over year."[56] This company is not in business to serve the public interest.[57]

Given what Google Scholar offers in its present incarnation, any decision to transform it to a pay service seems doomed to failure, because it simply does not offer enough to justify payment. However, as a recent article points out, "Google defines itself by its technology[; it]

---

[52] Barbara Quint, *Google Scholar Focuses on Research-Quality Content*, NEWSBREAKS (November 22, 2004) *at* http://www.infotoday.com/newsbreaks/nb041122-1.shtml

[53] *Google Scholar Beta: Advanced Scholar Search Tips, supra* note 24.

[54] Inserting periods where appropriate returns identical results, as Google Scholar ignores them, a characteristic I actually prefer, just as I prefer Westlaw's independence of punctuation requirements to LexisNexis' less forgiving interface.

[55] Gaither, *supra* note 12.

[56] Google Investor Relations, *Google Announces Record Revenues for Fourth Quarter and Fiscal Year 2004*, Feb. 1, 2005, *at* http://investor.google.com/releases/20050201.html (last visited Mar. 27, 2005).

[57] Nor is it the nefarious, mustachio-twirling villain some may think it to be; among other things, it is currently part of a consortium of companies working towards making laptop computers available to schoolchildren for $100 per unit by the end of 2006. Joseph Jacobson, *What if Every Kid Had a Computer?: The MIT Media Lab's Plan for a $100 Laptop*, WIRED, Apr. 2005, at 81.

invents for markets that don't exist."[58] If Google Scholar addresses the problems identified in the literature so far (and its employees are no doubt busily toiling toward that end), it may well become a force to be reckoned with, especially in regard to the invisible Web[59] and the ostensibly unpublished material it excels at locating. If fees are then implemented, libraries will have to decide whether to add yet another electronic resource to the already lengthy rolls of ever more expensive subscriptions.[60]

The other avenue Google Scholar may take, should it thrive, is to incorporate advertising, as has its progenitor.[61] Of especial concern is the degree to which the commercial interests of advertisers and other information providers to whom Google Scholar (or any similar purveyor of searching technologies, for that matter) is in some way financially beholden may influence the functions of the tool, especially its relevance rankings. After all, basing a document's potential relevance and utility upon the amount of times others have cited it makes some sense, but basing that determination upon how much money its creators or purveyors command is silly and unhelpful at best, despicable and obstructive at worst. Such concerns are magnified by the power of a Web site's PageRank to affect its visibility, given Google's dominance of the field.[62] Currently, the Google company does not sell higher rankings, so "Web sites cannot determine how they are ranked, or even if they show up on Google's search engine in the first place."[63] But these rankings do have value, even though they cannot be purchased directly, and businesses have taken advantage of that fact. One such business, SearchKing, offered a service called PR Ad Network ("PRAN"), which located Web sites ranked highly by Google to assist advertisers interested in maximizing their exposure by placing advertisements on such sites; SearchKing's fees were based in part upon the ranking Google assigns to the Web page where clients' advertizing is placed.[64]

From early in 2001 to the middle of 2002, SearchKing's PageRank declined significantly, and PRAN's PageRank was eliminated outright, prompting SearchKing to sue Google for tortious interference with contractual relations.[65] SearchKing responded to Google's claim that

---

[58] Malone, *supra* note 18.

[59] The invisible, or deep, Web is that portion of the Web commonly "overlooked by conventional search engines . . . . . It is the fastest growing category of new information sources on the Web. Most of this overlooked material resides in subject databases that require individualized searching" (e.g., records in a databases like ERIC) "or constitutes information that is dynamically generated for a particular enquiry" (i.e., data generated in real-time as it is retrieved) Jane Devine & Francine Egger-Sider, *Beyond Google: The Invisible Web in the Academic Library*, 30 J. ACAD. LIBRARIANSHIP 265-69 (2004).

[60] For an excellent discussion of the financial difficulties facing libraries in maintaining both printed and electronic resources, see KENDALL F. SVENGALIS, *A Brief History of Legal Publishing, in* LEGAL INFORMATION BUYER'S GUIDE AND REFERENCE MANUAL 7-18 (2004).

[61] Of course, I do not mean to suggest that Google Scholar will either become a pay service or incorporate advertisements; it may be that elements of both pay service and advertising will find their way into Google Scholar. I do conjecture with great confidence, however, that it will not remain both free and free of advertisement indefinitely.

[62] Note that here I use the term PageRank to refer to a site's ranking by Google's algorithmic system, not to the system of ranking itself. The technology's creators are themselves to blame for this somewhat confusing homonymy of nomenclature. SERGEY BRIN & LAWRENCE PAGE, THE ANATOMY OF A LARGE-SCALE HYPERTEXTUAL SEARCH ENGINE 4, *available at* www.stanford.edu/class/cs240/readings/google.pdf (last visited Apr. 11, 2005).

[63] Eric J. Sinrod, *Google's PageRank Deemed Protected Speech*, LEGAL TIMES, June 23, 2003, at 21.

[64] *Id.*

[65] Search King Inc. v. Google Technologies, Inc., No. CIV-02-1457-M, 2003 WL 21464568, at *1 (W. D. Okla. May 27, 2003).

assignment of PageRanks was expression protected by the First Amendment by pointing to Google's claims that its ranking system is "mechanical," "honest," and "objective," arguing that Google could not "have it both ways, professing the objectivity of the PageRank system on one hand, and relying on the subjective nature of the system in order to avoid tort liability on the other"; the court held that "there is no conceivable way to prove that the relative significance assigned to a given web site is false."[66]

Thus, although Google's motion to dismiss the complaint was granted, they won a kind of Pyrrhic victory. In essence, Google admitted, and the court's ruling affirmed, that PageRank and similar systems are indeed subjective. Users were explicitly put on notice that the purveyors of search technologies are wholly within their rights to exercise that subjectivity in assigning relevance rankings, and librarians have a professional responsibility to ensure users appreciate the consequences of this fact. The issues raised by this case, and the prescient ruling of the court, should make clear the need for a healthy distrust of automated systems for determining relevance. Where incentives exist for altering how such relevance values are assigned in order to accommodate advertisers or thwart competitors, we ought to be doubly wary. I wish to be absolutely clear that I do not mean to suggest that Google has or will engage in such dubious behavior. In fact, it may be argued that the steps Google took to circumvent SearchKing in the first place represent a commitment to prevent companies' gaming the system to manipulate search results for their benefit. It could also be argued (as SearchKing contended) that those same steps were merely an attempt by Google to keep other entrepreneurs from getting a piece of the action.

C. Privacy Concerns?

Libraries have traditionally supported the idea that "users have a right to be free from any unreasonable intrusion into or surveillance of their lawful library use."[67] On its face, such a view correlates to disapproval of the collection of information about users' research habits. A number of such privacy concerns have been raised about Google's search engine technologies, "ranging from how long Google's cookies linger on a user's hard drive before expiring to what the company might be doing with personal data to the genuinely conspiratorial."[68] These concerns are largely unwarranted, as Google's own privacy policy makes clear that its cookies collect only "limited non-personally identifying information your browser makes available whenever you visit a website. This log information includes your Internet Protocol address, browser type, browser language, the date and time of your query and one or more cookies that may uniquely identify your browser."[69] The policy goes on to explain that any personally identifying information the company does acquire is released to third parties only when required by law or with one's consent, and that only aggregate data are used for marketing and product development purposes.[70] Furthermore, search engines' cookies can record only "that particular browser software, on a particular computer, made a request. A cookie does [enable engines to] potentially

---

[66] *Id.* at *3, *4 (internal quotes omitted).

[67] American Library Association, *Privacy: An Interpretation of the Library Bill of Rights, at* http://www.ala.org/ala/oif/statementspols/statementsif/interpretations/privacy.htm (last visited Apr. 17, 2005).

[68] Dave Gussow, *Personal Tech: Google Under Fire for Privacy,* ST. PETERSBURG TIMES, Apr. 14, 2003, at 1E.

[69] Google Privacy Center, *Google Privacy Policy,* at http://www.google.com/privacy.html (last visited Apr. 17, 2005).

[70] *Id.*

see all requests made by that particular browser software, over time[, but it] doesn't know who [made the request]; all it knows is an anonymous number."[71]

Some might contend that these practices are all well and good, but that the very collection of any information about users' practices is undesirable, regardless of its security. For instance, the policy of the law library where I work is succinct, clear, and leaves little room for potential abuse: "We protect the privacy of all library users including names, addresses, telephone numbers, social security numbers, books checked out, and questions asked. We will not give any information except to comply with a search warrant or a court order."[72] Not only is such information as we do maintain permanently (i.e., borrower records) kept confidential to the greatest extent possible, but information about usage (e.g., what books a given patron has checked out over time) is not even stored. Once a patron returns an item and it is discharged by our automated system, no record is maintained indicating that item was ever charged out to that specific patron.[73] It is worth noting, of course, that law libraries raise special concerns about protecting the confidentiality of patrons' research practices.[74]

At any rate, none of the information collected by the Google company is any different from that collected by any number of other search engines; Google Scholar's arrival on the scene raises no issues about users' privacy beyond those arising from use of any generic Internet search engine.[75] Furthermore, simple measures can be taken to by-pass any collection of data about usage, none of which affect the tools' efficacy; these include disabling cookies and employing an "anonymizing [sic] tool," such as a proxy server.[76] Thus, privacy and confidentiality of users' information gathering habits will remain quite sound (or at least become no more unsound than it is now) even if Google Scholar becomes a popular scholarly research tool.

Readers who still have concerns about Google Scholar users' privacy are invited to consider Westlaw's. Among other things, that company collects the following:

> your Internet protocol address, what kind of browser or computer you use, the number of links you click within the site, the state or country from which you accessed the site, the date and time of your visit, the name of your Internet service provider, the Web page you linked to our site from, [and] the pages you viewed on the site.[77]

LexisNexis' policy is similar.[78] It is an open secret among law librarians that the marketing

---

[71] Danny Sullivan, *Search Privacy at Google and Other Search Engines*, SEARCH ENGINE WATCH, *at* http://searchenginewatch.com/sereport/article.php/2189531 (last visited Apr. 17, 2005).

[72] Marian Gould Gallagher Law Library, *Privacy Policy*, *at* http://lib.law.washington.edu/dir/webprivacy.html (last visited Apr. 17, 2005).

[73] This explanation oversimplifies things slightly; very recently discharged items can be linked to the last patron who used them, for instance, but for all practical purposes such information is entirely ephemeral and irretrievable, unless patrons opt in to a password-protected feature enabling them to maintain a "reading history" for their own purposes. Larisa Bosma, *My Law Library*, LAW LIBRARY NEWS, at http://lib.law.washington.edu/news/2004/Oct4.html (last visited Apr. 17, 2005).

[74] Mary Whisner, *Library Records are Confidential*, LAW LIBRARY NEWS, *at* http://lib.law.washington.edu/news/2005/Jan10.html (last visited Apr. 17, 2005) (citing American Association of Law Libraries, *Ethical Principles*, *at* http://www.aallnet.org/about/ policy_ethics.asp (last visited Apr. 17, 2005).

[75] Danny Sullivan, *Google and the Big Brother Nomination*, SEARCH ENGINE WATCH, *at* http://searchenginewatch.com/sereport/article.php/2175251 (last visited Apr. 17, 2005).

[76] *Id.*

[77] Thomson West, *West Privacy Policy*, *at* http://west.thomson.com/privacy/Default.asp (last visited Apr. 17, 2005).

[78] LexisNexis, *Company Information: Privacy Notice*, *at* http://www.lexisnexis.com/terms/ privacy/ (last visited Apr. 17, 2005).

purposes alluded to in both companies' policies entail (not necessarily desirable) price modifications based upon customers' usage patterns. For example, by collecting detailed information about, say, the types and quantities of databases accessed by a law library operating under a flat-rate contract for its services, a vendor can more accurately tailor the pricing of such flat-rate plans to reflect actual usage (i.e., to be more like what users might be charged if they were paying *a la carte*): "Firms entering into such contracts should . . . be warned to control costs since higher usage will be factored into the next fixed-rate proposal and/or contract."[79] Such automated data collection thereby increases efficiency by ensuring that consumers bear the appropriate costs of the goods they consume. This phenomenon occurs because flat rates disproportionately "benefit those who would be prepared to pay more (those who, for example, require the information most urgently . . .) but systematically exclude those who cannot or will not meet the price demanded. Price discrimination overcomes these inefficiencies by tailoring availability more closely to demand."[80] Automating the data collection necessary to make price discrimination feasible also creates efficiencies by decreasing transaction costs (a significant reason for implementation of uniform pricing in the first place).[81] While economically optimal, then, such practices are not always favored by law librarians and other consumers of information, because the benefits realized accrue to the providers of that information—hardly a recipe for merriment and cheer in this time of shrinking budgets and rising prices.[82] Thus, whatever Google Scholar's owners do with any data it collects about users' habits, its policies are certainly no more intrusive or disturbing than its competitors, and they certainly have not impacted law libraries' budgets the way the major vendors' practices have (not yet anyway).

D. Hope for the Future: Some Suggestions for Improvement

1. Functionality

Unlike humanities scholars, such as historians, for whom the serendipitous discovery of tangentially related materials plays a major role in the research process,[83] legal scholars often seek very particular information from specific sources. This characteristic of legal research makes progressively focused searching important, and the best specialized legal research tools accommodate that importance (witness the hierarchically arranged databases of LexisNexis and Westlaw).

To be more useful for legal researchers, Google Scholar might consider incorporating similar capabilities. Researchers needing articles from bar publications may not be interested in those appearing in law journals or unpublished white papers. To accommodate those searchers wanting to search only specific subsets of data, Google Scholar might consider what have been called search zones in the literature: "The creation of search zones—pockets of more homogenous content—reduces the apples-and-oranges effect and allows users to focus their

---

[79] Kendall F. Svengalis, *Computer-Assisted Legal Research, in* LEGAL INFORMATION BUYER'S GUIDE AND REFERENCE MANUAL 137, 148 (2004).

[80] Paul Ganley, *Digital Copyright and the New Creative Dynamics*, 12 INT'L J. L. & INFORMATION TECH. 282, 289 (2004) (internal quotations omitted).

[81] *Id.* at 290.

[82] *See* Hugh Spitzer, Law Libraries Losing to "Double Inflation", WASH. STATE BAR NEWS, Nov. 1980, at 31; Herbert S. White, White Papers: The Tragic Cost of Being "Reasonable", LIBR. J., Feb. 15, 1991, at 166.

[83] W. M Duff & C. A. Johnson, *Accidentally Found on Purpose: Information-seeking Behavior of Historians in Archives*, 72 LIBR. Q. 472-496.

searches."[84] Arranging such subdivisions hierarchically provides even more options for users, and the groupings necessary to produce them can be created automatically via document similarity: "algorithms convert . . . document[s] into . . . a query . . . . Stop words [are] stripped out of the . . . document, leaving a useful set of semantically rich terms that, ideally, represent the document well. These terms are then converted into a query that should retrieve similar results."[85] One search engine, Teoma, currently creates these kinds of subdivisions and hierarchies by organizing results into "naturally occurring communities" via its Refine feature.[86] Thus, searching "maritime law" with Teoma offers the user a choice of several categories of results, including: admiralty law, U.S. maritime law, international maritime law, shipping law, salvage laws, and maritime liens.[87] This granularity more closely approximates the kind of specificity legal scholars require, whether they are expert professors who know what they want and cannot waste time weeding through irrelevant results, or novice students who need collocation and classification systems like this to guide them to resources on point. Search engines such as Google Scholar should consider incorporating these kinds of tools if they are to appeal to legal scholars.

As mentioned earlier, the popularity contest upon which the basic Google search engine bases relevance is unsuited to scholarly research. However, there is some value in the popularity contest approach, even for scholarly research, if executed properly. Again, Teoma provides a great example, outdoing Google Scholar procedurally if not substantively. To rank a web page's relevance, "Teoma . . . first identifies other sites on the same topic, then analyzes how often those sites link to the page. Teoma calls this 'subject specific popularity.' [T]he best [sites] for lawyers are likely to be the ones that lawyers as a group most often link to."[88] We might extend this principle to any specialization. Thus, legal scholars are likely to find useful those search results that other legal scholars have found useful, family law scholars those their peers have found useful, and so forth. Google Scholar—and any competitors' products seeking to serve scholarly researchers—will achieve a breakthrough if they develop systems which enable users to retrieve results based upon their peers' linkages and citations, as Teoma has. By knowing what others in their field have viewed or linked to, users will be closer to realizing citation chaining on the Internet, much as tools like Web of Science enable them to do for defined databases. After all, incorporating a hyperlink to a given site is really a form of citation.

2. Interface

Many shortcomings of Google Scholar's interface have already been identified; only a few more will be mentioned here. A natural extension of the "institutional access links"[89] feature would be to include clear and forthright identification of those results which will be unavailable absent paid access to a subscription database. Even better might be an option to exclude such

---

[84] Peter Morville & Louis Rosenfeld, *Search Systems, in* INFORMATION ARCHITECTURE FOR THE WORLD WIDE WEB 132, 137 (2d ed. 2002).

[85] *Id.*

[86] About Teoma, *at* http://sp.teoma.com/docs/teoma/about/searchwithauthority.html

[87] Teoma, *at* www.teoma.com (search conducted Mar. 15 2005).

[88] Robert J. Ambrogi, *Search Engine Showdown: The Reigning Best of Breed is Google, but a Handful of Upstarts Are laying Claim to the Title,* 168 N.J. L. J. 34 (2002).

[89] *See supra* notes 16-17 and accompanying text.

results altogether; while it is desirable for users to be able to access all indexed documents by Google Scholar, often one will want only to deal with those readily available.[90]

More options must also be made available for refining searches. Allowing users to limit results to those featuring sought terms in their titles is a great start, but there is much room for improvement. The effectiveness of search systems such as Google Scholar is commonly measured in terms of recall and precision; the former "is the proportion of relevant documents that are retrieved" and the latter "the proportion of retrieved documents that are relevant."[91] Because of the secrecy surrounding what resources are currently indexed by and included in the Google Scholar system, it is not possible to meaningfully assess its recall. It certainly did not retrieve all of the documents written by the legal scholars whose *curriculum vitae* I reviewed. While this failure could be either due to the unfound items' not being indexed in the first instance or because of a flaw in the searching software itself, past experience and the proven, successful track record of Google's searching algorithms suggests the problem lies with incomplete indexing, even for those sources which Google Scholar claims to have indexed in their entirety.[92] Given the lack of information available about these matters, I can only make the observation and hope the powers that be take note of its implications.

E. Ramifications for (Law) Libraries

It has been argued that "[b]ecause research processes and patrons' usage have shifted more and more to electronic rather than print sources, libraries no longer need to purchase and house extensive print sets to support their main mission to users."[93] Google Scholar represents an undeniable move toward increasing the viability of such ideas about digital access, storage, and preservation via the DSpace archiving of faculty writings by their affiliated institutions which underpins it; regardless of its current imperfections, Google Scholar's functionality and interface represent a workable prototype for systematic access to such materials.

If anyone is up to the task of creating a comprehensive tool for locating scholarly material online, the folks at Google ought to be. They index over three billion Web documents, (their closest competitors can claim just about half as many), and it is estimated that searching through this many documents by hand would take an individual some six thousand years.[94] Despite these impressive figures, it has been pointed out that "Google Scholar doesn't contain anything that isn't already available for free over the Web from thousands of public and academic libraries all around the world."[95] However, libraries have been somewhat ineffective at providing a strong case for the superior quality and comprehensiveness of preexisting services

---

[90] Compare the criteria used by some libraries for determining whether to include websites in their online catalogues: "[O]n our 'not' list are URLs that require an access fee, since patrons accessing items through our catalog shouldn't find themselves blocked at a host's front end. However, we do link to some free items housed on sites that charge users for other menu choices." James R. Veatch, *Insourcing the Web*, 30 AMER. LIBR. 64, 66 (1999).

[91] W.B. Croft, *An Overview of Information Systems*, INFORMATION TECHNOLOGY: RESEARCH AND DEVELOPMENT 73, 84-85(1982).

[92] Jasco, *supra* note 32.

[93] Kevin M. Marmion & Richard J. Spinelli, *The Changing Role of Law Library Vendors: The William S. Hein & Co. Perspective*, 21 LEGAL REFERENCE SERVICES Q. 301, 305 (2002).

[94] Robert J. Ambrogi, *The Best Search Engines: For Fact Research, Check Out These New Tools*, LEGAL TIMES, July 15, 2002, at 31.

[95] Chuck Richard & David Curle, *A Quick Take on Google's New Google Scholar*, OUTSELL NOW, Nov. 18, 2004, *at* http://now.outsellinc.com/now/2004/11/a_quick_take_on.html.

and facilities in the wake of the hype over Google, and, now, Google Scholar—a competitor that (ostensibly) hits even closer to home, although progress is being made as libraries better recognize the importance of public relations to their survival in the face of growing (albeit often misplaced) faith in the Internet to meet more and more information needs.[96] If Google Scholar attains the kind of popularity with researchers that its progenitor has with the populace at large[97] it will become even more important for law librarians and other information professionals to educate users about the importance of a balanced approach to research and the hazards of reliance on any one resource.

The "need to put . . . emphasis on providing access to as much information as possible and leave the decision making of what information is useful or not up to readers"[98] is central to the philosophy underlying projects like Google Scholar. Lest my previous discussion of the potential pitfalls of Google Scholar for the unwary left any room for doubt, let me now be explicit that I do not subscribe to this philosophy wholesale myself. However, there are benefits to Google Scholar, including simultaneous access and use of information by multiple users in multiple locations, as well as the potential to fight the unsettling trend for librarians to buy "the same content produced by their own scholars—but at outrageous prices."[99] The DSpace concept undergirding Google Scholar, along with similar ventures such as the Scholarly Publishing and Academic Resources Coalition ("SPARC")[100] represent a significant step toward alleviating the latter problem. Persistent financial concerns about libraries' budgets, especially in relation to scholarly periodicals, underscore how important it is to support the development of such alternatives.[101]

Academics are not paid for the articles they write in the direct fashion that, say, journalists are. One can certainly argue, and I would agree, that the work they do to produce writings is expected as a part of their employment by universities (and that their salaries therefore, at least in part, compensate them for their writings), but that is neither here nor there for the purposes of this discussion; the point is that they are not paid by the periodicals in which their work appears. It would therefore be financially beneficial to everyone except for publishers if these articles were simply made available for free online, where they could be retrieved by tools like Google Scholar. This arrangement would not only save money, it would advance the presumably desirable goal of making knowledge more widely available because those who might otherwise be denied access to some work or another because, say, their library does not subscribe to the journal in which it would have appeared. Of course, Internet access is crucial to the viability of such a model, but providing access to the Internet seems likely to be less expensive

---

[96] *Id.* See also Penny A. Hazelton, *How Much of Your Print Collection is Really on WESTLAW or LEXIS/NEXIS?*, 18 LEGAL REFERENCE. SERVICES Q. 3 (1999) (demonstrating only some 7% of a large academic law library's resources were available on these two commercial databases).

[97] Some 150 million searches are run on Google every day. Michael B. Reddy, *Google Hacks: 100 Industrial Strength Tips and Tools*, 22 LEGAL INFO. ALERT. 12 (2003).

[98] Anthony W. Ferguson, *Digital Library Selection: Maximum Access, Not Buying the Best Titles; Libraries Should Become Full-Text Amazon.com's*, 31 J. LIBR. ADMIN. 27, 28 (2000).

[99] *Id.* at 35.

[100] SPARC describes itself as "an alliance of academic and research libraries and organizations working to correct market dysfunctions in the scholarly publishing system [by facilitating] he emergence off systems that capitalize on the networked environment to disseminate research[,] expand competition and support open access" *SPARC, at* http://www.arl.org/sparc/ (last visited Apr. 15, 2005).

[101] The thorny nexus of collections development, budgetary constraints, and digital materials is compellingly discussed in Michael Chiorazzi, *Books, Bytes, Bricks, and Bodies: Thinking About collection Use in Academic Law Libraries*, 21 LEGAL REFERENCE SERVICES Q. 1 (2002).

and troublesome an undertaking for libraries than maintaining subscriptions to hundreds of periodicals and paying the overhead needed to acquire, process, and store those periodicals. Monographs, for which authors—including academics—are normally paid in the form of royalties, do not fit neatly into such a model, but then, they are not at this time generally retrieved from online sources, printed, and read in the way that articles have come to be. Even so, Google Scholar, or a similar tool, suggests a potentially promising method for at least locating monographs. (*Quaere*: Could such tools transform the Internet into a kind of *uber*-catalogue of the future?)

Some (perhaps many) may argue that Google Scholar represents not a trend that may alleviate the current oligopolistic conditions of traditional publishing by making information more readily accessible online from its authors, but merely a changing of the guard as libraries and their patrons become beholden to companies such as Google rather than traditional publishers. That may or may not be; the present expenses and complications of online information services suggest that such a prediction is not wholly preposterous. In any case, we should remember that Google Scholar represents a paradigm shift from companies which collects, organizes, and then distributes information—whether by via traditional printing or online services—to companies which provide tools for locating information created and stored independently of those companies' actions. Such a shift is a positive thing, in my opinion, because, among other things, it may reduce the phony elitism established by traditional editorial processes. An article may come to be widely accepted or respected on its own merits more often in the future, without consideration for whether it appeared in the *Harvard Law Review* or the *Florida Coastal Law Review*.[102] It also may help to empower individuals, via technology, to better organize and locate information for themselves:

> Ultimately no human brain, no planet full of human brains, can possibly catalog the dark, expanding ocean of data we spew. In a future of information auto-organized by folksonomy, we may not even have words for the kinds of sorting that will be going on; like mathematical proofs with 30,000 steps, they may be beyond comprhension. But they'll enable searches that are vast and eerily powerful.[103]

Folksonomy refers to the process whereby users tag the information they encounter using "machines that can automate at least some of what it takes to classify information and . . . social software that makes users willing to do at least some of the work for nothing"; it is not currently useful for locating a specific piece of information in and of itself, but progress may be made toward that end in time.[104]

---

[102] I realize my contempt for the Byzantine jockeying for prestige that accompanies modern scholarship probably places me in a minority. I nevertheless stand by it. Those who speak of this or that article being more worthy of praise or attention because of the publication in which it appeared (which may have as much to do with the timing of its completion or its author's personal connections as with the significance of its content) remind me of a statement made by an especially perspicacious musician about his work: "[N]ot the individuals in the band, but the name Pink Floyd is worth millions of pounds. The name is probably worth one million sales of an album, any album we put out. Even if we just coughed, a million people will have ordered it simply because of the name." Nick Sedgewick, *A Rambling Conversation with Roger Waters Concerning All This and That, in* SHINE ON 58, 59 (1992).

[103] Bruce Sterling, *Order Out of Chaos: What's the Best Way to Tag, Bag, and Sort Data? Give It to the Unorganized Masses,* WIRED, April 2005, at 83.

[104] *Id.*

V. CONCLUSION

Everywhere I turn, it seems—in conversations with colleagues, in reference interviews with students and faculty, in the literature of the profession, and even in popular media—a debate seems to be raging about the relative merits of digital and concrete media, about the (f?)utility of libraries versus that of the Internet, about whether information ought to be organized this way or that; the old and the new. Too much of the rhetoric seeks to establish a false dichotomy. We are currently experiencing an exciting period of transition comparable to that brought about by the industrial revolution, and the agricultural revolution before it. Tired polemic and the polarization that ensues is silly; instead, we must come to cope with what is, not damning new technologies as inherently undesirable, but rather recognizing them for the mere tools they are, with an accompanying capacity to solve or create problems. Google Scholar, and the technology and theories underlying it, represent such a tool, and one we librarians cannot dismiss out of hand.

Google Scholar is not yet comparable to Hein Online, LexisNexis or Westlaw for either full-text retrieval or comprehensive coverage. But, in its current incarnation, it seeks neither to be a specialised tool for researching legal scholarship nor for full-text retrieval of indexed documents, as these three services do. Tellingly, my admittedly limited tests indicate it equals, and in significant ways exceeds, the capabilities of LegalTrac, an observation made more significant by the similar, basic aims it shares with this service: the indexing of scholarly literature. To be sure, Google Scholar is still undergoing birthing pains: its interface could stand improvement, and the content it currently indexes is not nearly as comprehensive as it might be. But the former is a consequence of Google Scholar's relative novelty, and the latter of that novelty as welll as copyright and other legal concerns; neither has much to do with technological limitations. So long as scholars remain dedicated to the advancement of knowledge above financial profit (by no means a foregone thing), I for one am far more comfortable with whatever deals need be stricken between academic institutions and companies like Google than I am with the current state of the publishing industry (particularly the legal publishing industry) and its effect on libraries and their patrons. Furthermore, the idea that tools like Google Scholar will ever replace librarians and other information professionals is patently absurd. We will always have a role to play, whether it be in guiding end-users through efficient use of tools or designing and improving those tools to minimise the need for such guidance. We must not abandon that role, regardless of how it manifests itself, and we certainly must not pooh-pooh the technologies which may alter it out of any petty fear of change or entrenched antiquarianism.