Finding Chemistry Information using Google Scholar:

A Comparison with Chemical Abstracts Service

[Accepted for publication in *Science and Technology Libraries*, Haworth Press]

Michael Levine-Clark

Collections Librarian

miclark@du.edu

303-871-3413

Fax: 303-871-2290

Joseph Kraus

Science Librarian

jokraus@du.edu

303-871-4586

Fax: 303-871-2290

Penrose Library

University of Denver

2150 E. Evans Ave.

Denver, CO 80208

Finding Chemistry Information using Google Scholar: A Comparison with Chemical Abstracts Service

ABSTRACT

Since its introduction in November 2004, Google Scholar has been the subject of considerable discussion among librarians. Though there has been much concern about the lack of transparency of the product, there has been relatively little direct comparison between Google Scholar and traditional library resources. This study compares Google Scholar and Chemical Abstracts Service (CAS) as resources for finding chemistry information. Of the 702 records found in six different searches, 65.1% were in Google Scholar and 45.1% were in CAS. Of these, 55.0% were unique to Google Scholar, 34.9% were unique to CAS, and 10.1% overlapped. When each record found was searched by title in the two databases, the figures change, with 79.5% in Google Scholar, 85.6% in CAS, and 65.1% overlapping. Based on this, researchers are more likely to find known published information through CAS than in Google Scholar. Results vary by type of search, type of resource, and date. For many types of searching, CAS performs significantly better than Google Scholar. This is especially true for searches on compounds or a personal name, both of which take advantage of advanced search features in CAS. For simple keyword searches, Google Scholar tends to perform better, most likely because Google Scholar searches through the full text of journal articles, while a keyword search through CAS only finds abstract and index terms.

Chemical Abstracts Service (CAS) has long been the major research tool for chemistry. It is a valuable resource, containing information back to 1907 from over 40,000 scientific journals, and has complex search features uniquely suited to chemistry information retrieval (About CAS). It is also very expensive, so for many libraries, a subscription to CAS is not an option. With that in mind, it seemed worthwhile to examine the relative utility of Chemical Abstracts Service, delivered via the SciFinder Scholar platform, and Google Scholar for locating chemistry-related information. Google Scholar, the free and much-hyped Google tool for locating scholarly information, could be a useful substitute for an expensive specialized database such as CAS. Five paired searches, representing different common types of chemistry search strategies, were conducted in each database, allowing a direct comparison of the two. Though the results of this comparison are particularly useful for chemistry research, it is expected that they will apply to Google Scholar in general.

ABOUT GOOGLE SCHOLAR AND CAS

Google Scholar was introduced in November 2004 as a free tool to search the scholarly Web. It is still in beta release 24 months later. According to the site's own information,

> Google Scholar provides a simple way to broadly search for scholarly literature. From one place, you can search across many disciplines and sources: peer-reviewed papers, theses, books, abstracts and articles, from academic publishers, professional societies, preprint repositories, universities and other scholarly organizations. Google Scholar helps you identify the most relevant research across the world of scholarly research.[1]

This relevancy is accomplished by taking into account "the full text of each article, the author, the publication in which the article appears, and how often the piece has been cited in other scholarly literature. The most relevant results will always appear on the first page."[2] There are flaws to this approach to relevancy, most notably the fact that recent articles, not yet cited in other literature but often most desirable from a researcher's perspective, get pushed lower on the list.[3]

Though there is a great deal of information on the Google Scholar information pages about how to search, how publishers can get their content indexed, and how libraries can use a link resolver with Google Scholar, there is no information about what content from publishers is already available or about exactly why search results sort the way they do. Marian Burright observes that, though results are ostensibly sorted by number of citations, this does not appear to be the case.[4]

Chemical Abstracts Service has been published since 1907 and indexes articles, patents, conference proceedings, and other documents related to chemistry. With over 25 million items indexed, CAS provides access to a wealth of scholarly material. It is also an extremely complex searching tool. CAS includes the CAS Registry,

> the largest substance identification tool in existence. When a chemical substance, newly encountered in the literature, is processed by CAS, its molecular structure diagram, systematic chemical name, molecular formula, and other identifying information are added to the Registry and it is assigned a unique CAS Registry Number. Registry now contains records for more than 30 million organic and inorganic substances and more than 58 million sequences.[5]

The CAS Registry allows for much more precise searching of substances than a keyword or subject search would allow. This feature along with the ability to search by chemical formula or a graphical representation of a chemical structure makes CAS an extremely powerful and very complex searching tool. An indicator of the complexity of SciFinder Scholar is the fact that a 235-page book is needed to explain it.[6]

LITERATURE REVIEW

Judging from the number of references at recent library conferences and the amount of press it has been receiving in the library literature, Google Scholar has been a significant concern for librarians since its launch.[7]

Recent articles have examined the degree of acceptance of Google Scholar by librarians, without offering much analysis of the quality of the resource. Laura Bowering Mullen and Karen A. Hartmen's study of the number of ARL members that have linked to Google Scholar from their websites shows a relative lack of enthusiasm for the product. Though they look at a variety of measures of acceptance of the product, the most telling is that of the 113 ARL institutions, only 24% have linked to Google Scholar from the alphabetical lists of electronic resources on their websites.[8] Maurice C. York shows that many libraries have treated Google Scholar differently than other resources, indicating it as a lesser source while linking to it on their websites.[9] Jill E. Grogg and Christine L. Ferguson include a useful section on the reactions of librarians to the product in their article on OpenURL linking with Google Scholar.[10] Martin Kesselman and Sarah Watstein summarize nicely some of the arguments for and against Google Scholar from a librarian's perspective.[11] Jeffrey Pomerantz notes that Google Scholar has tremendous potential for bringing users to the library.[12]

Reviews have been consistently critical, noting the lack of transparency in listing the sources indexed; the impossibility of ranking results by date; the poor quality of the advanced search feature; the uneven coverage; and the time-lag between coverage on publishers' websites and in Google Scholar.[13] This last is an unfair criticism, given that traditional library indexes are always delayed in covering recent publications. Rita Vine and Burright note that Google Scholar shows a significant delay in indexing material on PubMed.[14] In contrast, Chuck Hamaker and Brad Spry observe that as of December 2005, Blackwell nursing titles were indexed 2.5 months sooner on Google Scholar than in CINAHL.[15] Several reviewers consider Google Scholar inadequate for serious research but useful for cursory investigation.[16]

The few studies that have made in-depth comparisons of Google Scholar with other resources have mixed impressions of Google Scholar. In a comparison with three databases in the social sciences (PsycInfo, Social Science Citation Index, and ERIC), Susan Gardner and Susana Eng find that Google Scholar is less current than the other databases, but that the results found tend to be highly relevant.[17] Their search of Google Scholar only takes the first 100 results (of 2,200) into account, so conclusions about currency, given the fact that results are listed based on number of citations, are suspect. In analyzing overlap, they do use all 2,200 records, and show that the percentage of records overlapping is fairly high for PsycInfo (79%) but relatively low (under 42%) for the others. Jascó finds large gaps in coverage in Google Scholar in comparison to Science, Nature Publishing Group, and the Astrophysics Data System.[18]

Janice Adlington and Chris Benda's comparison of Google Scholar to five library databases as well as Scirus and Google shows mixed results. While they find problems with the display, authority control, and lack of transparency, they are impressed with "the quality and number of relevant results" and usability of Google Scholar.[19]

In several comparisons, however, Google Scholar comes out much better, especially for relatively recent information. In a study looking at alternatives to BIOSIS, the standard tool for finding biological information, D. Yvonne Jones observes that 27% of the 110 results found in a sample search in BIOSIS were also found in Google Scholar, but that 56% of the results from 1996 or later were present. And Google Scholar found 72 relevant papers and articles not found in BIOSIS, of which 49 were published in the last five years.[20]

In their comparative study of citation analysis tools, Kathleen Bauer and Nisa Bakkalbasi note that for articles published in the Journal of the American Society for Information Science and Technology in 2000, Google Scholar returns "a statistically significant larger mean number of citations than either Web of Science or Scopus."[21] This is in contrast to the results for 1985, for which Web of Science proves the best resource. Burright, however, considers Google Scholar to be "an inadequate search tool for citation data since it does not provide nearly accurate data on publication names nor explain how citation rates are calculated."[22]

The most ambitious study to date, by Chris Neuhaus, et al., compares Google Scholar to 47 databases from a range of disciplines. This study shows that Google Scholar performs better for scientific and medical than for humanities or social sciences research and has better coverage of English-language than non-English and relatively current than older material. For a random sample of records from CAS tested in Google Scholar, the authors found 60% overlap. Though useful, this study does not take into

account records found in Google Scholar but not in the other databases. Nor does it factor in differences in search results, concentrating instead on the simple presence of records in Google Scholar that have already been identified in the other databases.[23]

METHODOLOGY

Five searches were conducted over a one-month period in Google Scholar and CAS. These included two topical searches (hypericin and epr and pulsed epr and biology and enzymes), two searches on compounds (hydroxybutyranilide and myrigalone), and one on a personal name (Andrei Kutateladze). The lists of resources returned by these searches were compared, giving a quantitative measure of the relative quality of the two databases in terms of search results.

Searches were conducted in the best way possible for the given database. The personal-name search, for instance, was entered as an author search in each database. In CAS, this was a simple process, with a search on "Kutateladze, A" easily grouping the various versions of the name together. In Google Scholar, however, there were multiple ways of searching. Author searches alone were ineffective, so a more complex search strategy was employed: "author:Andrei OR author:AG OR author:A (and) author:Kutateladze," limited to the category of "chemistry and materials science." See figure 1 for details on the search strategies attempted. The remaining searches were entered as keyword searches in Google Scholar as were the topical searches in CAS. The compound searches were conducted using CAS Registry numbers in CAS, while the name of the compound was entered as a keyword in Google Scholar.

Every item returned by the searches was assigned a resource type (article, book, dissertation, etc.). In general, no attempt was made to decide whether an article was the official version. As long as a citation was available, it was counted as an article for purposes of this study. Thus, articles include those found on the official publisher website (whether full-text or not), postprints, and citations. In some cases, the results in Google Scholar were unclear (generally broken links or incomplete citations). These findings were indicated as problems of some sort. Thirty-seven records were returned in Google Scholar for the wrong Kutateladze and have not been counted in the results.

Other than notations of problems, no attempt was made to judge the quality of results. The presence of keywords in the text and/or citation was assumed to be enough. There was also no attempt made to count the number of full-text results in Google Scholar since these sometimes depend on local subscriptions and because the comparative database, CAS, does not provide full-text access.

A second comparison was made by searching each database individually for each title initially found in only one database. The resulting list of titles allows a comparison of the content of the two databases while addressing the quality of the indexing and abstracting of each. Each of these titles was checked to determine whether all of the keywords were present. In the search for hydroxybutyranilide, for example, the initial searches yielded 135 separate records. Of these, 29 were in Google Scholar and 107 in CAS. Searches by keywords in the title upped the number to 57 in Google Scholar and 135 in CAS.

RESULTS

The five searches returned a total of 702 records, of which 457 (65.1%) were in Google Scholar and 316 (45.1%) were in CAS. Of these, 386 (55.0%) were unique to Google Scholar, 245 (34.9%) were unique to CAS, and 71 (10.1%) overlapped between

the two. Based on the searches done, Google Scholar returns larger sets of results. However, when the titles found in these initial searches were searched on keywords from the titles in the databases in which they had not initially been located, the totals increased to 558 (79.5%) in Google Scholar and 601 (85.6%) in CAS. The degree of overlap (65.0%) is a little higher than the 60% in the Neuhaus (2006) study.

CAS, while not as successful a tool for keyword searching in terms of the numbers of records found, is a little better in terms of overall numbers of records available. Part of the difference in the searching may be due to the fact that Google Scholar searches the full text of materials in some cases, making it more likely that search terms will be present. The records found in CAS, based on terms in the citations, abstracts, and index terms, are likely more focused on the topics in question than those in Google Scholar. Based on a measure of quantity, Google Scholar is a superior searching tool. Based on a measure of quality, CAS may be superior.

Of the records found, 564 were for journal articles. Of these, the format for which one typically searches CAS, 387 (68.6%) were found in Google Scholar, while 248 (44.0%) were in CAS. Seventy-one (12.6%) of these were in both sources, with 177 (31.4%) just in CAS and 316 (56.0%) in Google Scholar alone. Based on these results, it seems that a searcher is more likely to find desired articles in Google Scholar than in CAS.

When the total number of articles for each topic available in each database is analyzed, the results are quite different. Searches by title for all of the 564 articles indicated that 482 (85.5%) were in Google Scholar, while 521 (92.4%) were in CAS. 438 (77.7%) were in both sources, with 83 (14.7%) in CAS alone and 44 (7.8%) in Google

Scholar alone. Based just on these results, it seems that a searcher is slightly more likely to find known information in CAS than in Google Scholar.

There were 138 non-journal results found, of which 68 (49.3%) were in CAS and 70 (50.7%) were in Google Scholar. Of those amounts, 54 patents were found just in CAS. Conference proceedings, with 12 in CAS and 11 in Google Scholar, were found in good amounts in both databases. Twenty-six records that were problems of some sort were found in Google Scholar. An additional 20 records found in Google Scholar with incomplete citations were not classified as problems because they were also found in CAS and categorized according to the CAS record. See Tables 1 and 2 for a complete breakdown of other types of resources found. For non-journal, non-patent, and non-conference literature, of which there were only a small number of items, Google Scholar is superior.

Of course, these broad results do not take into account variables such as date of publication or subject of the search. These will be addressed in more detail below. Nor do they take into account relevancy. For the purposes of this study, the assumption has been made that presence of all keywords in a given record indicates a certain degree of relevancy to the topic. In the case of the author search when the wrong author (same last name and different first name) was located in Google Scholar, that record was discarded. And when the resulting record is unclear in any way, it is indicated as a problem. Otherwise no attempt has been made to determine relevancy.

For aggregate results, Google Scholar performs better than CAS. This is generally true for results by date as well. For the 170 pre-1989 titles, coverage is significantly

better in CAS. After 1990, Google Scholar performs better. See table 3 for a detailed breakdown of search results by date.

When results are compared by date for content available in each database, Google Scholar and CAS perform roughly equally. For pre-1994 titles, CAS performs better. After that, the results are roughly evenly spread between the two databases. See table 4 for a breakdown of coverage by date.

The three types of searches present very different results. For the two topical searches (hypericin and epr and pulsed epr and biology and enzymes), 429 total records were found. Of these, 84.1% were found in Google Scholar versus 19.8% in CAS. Though there is a range of results within these searches, it is clear that for topical searches, Google Scholar outperforms CAS.

For the personal name search (Andrei Kutateladze), 93 records were found, with 97.8% in CAS and 52.7% in Google Scholar. For the two searches on compounds (hydroxybutyranilide and myrigalone), there were 180 total records found, of which 77.8% were in CAS and 26.1% were in Google Scholar. There is a moderate disparity between the two, with the hydroxybutyranilide search returning a larger percentage of records in CAS and a lower percentage in Google Scholar than does the search for myrigalone. Despite the differences, it is clear that scholars searching for information about compounds will do better with CAS than Google Scholar as they will for name searches. See tables 5 and 6 for a more detailed listing of search differences.

These figures do not include records found post-search by checking the other database for each title found. Adding those records allows for a more accurate reflection of database content, and reflects some of the inadequacies in searching. For the topical searches, 96.3% of all records were in Google Scholar versus 78.3% in CAS; for the compound searches, 96.7% were in CAS and 51.1% were in Google Scholar; and for the name search 97.8% were in CAS with 57.0% in Google Scholar. See tables 7 and 8 for a comparison of records available in each database for the three types of search.

Some of the differences between the search types can be explained by the types of sources found. For instance, in the searches on compounds, 54 patents were found, all in CAS. Removing these from the equation brings the amount of records found in each database closer together, with 68.3% of these records in CAS and 37.3% in Google Scholar. Looking at articles only, of the 122 found in the compounds search, 69.7% were in CAS and 36.1% were in Google Scholar. Of the 359 articles found in the topical searches, 82.2% were in Google Scholar with 22.6% in CAS. And with the name search, of the 83 articles found, 98.8% were in CAS compared to 57.8% in Google Scholar. It is clear from this sample that searchers looking for information by name or compound will be more successful in CAS while those searching on a topic will have more success in Google Scholar.

The differences are not as extreme when the articles not found in the initial searches are added to the topical searches. In this case, 96.4% of the articles were in Google Scholar with 88.9% in CAS. For the two other types of searches the results still vary significantly in favor of CAS, with 97.5% of the articles from the compounds searches in CAS and 68.9% in Google Scholar with 94.57% in CAS. And with the name search, of the 83 articles found, 100.0% were in CAS compared to 62.7% in Google Scholar. Though the likelihood of there being articles in Google Scholar is about even

for the topical searches, for the other subject areas, it is more likely that articles will be in CAS.

CONCLUSIONS

Google Scholar has been greeted with a mixture of enthusiasm (for its potential) and criticism (for its lack of transparency and some of its flaws in searching). Most of those who have studied Google Scholar have examined its use for finding a pre-defined set of information, in general not comparing it to other resources. Relatively few studies have compared Google Scholar with other databases.

This comparison of Google Scholar and CAS for retrieval of chemistry information shows that Google Scholar returns more results, even for journal literature, than does CAS. The results vary by type of search, with topical searches more likely to be successful in Google Scholar and compound and personal-name searches more likely to be successful in CAS.

Part of the advantage for Google Scholar in this comparison of searching is that it often searches the full text of resources. When comparing the number of records present, the two databases are more equal. Though there are many legitimate concerns about Google Scholar, it seems that it is just as useful as the traditional tool for finding many sorts of chemistry information.

The SciFinder Scholar platform for searching CAS has much more sophisticated search functionality than Google Scholar, with many chemistry-specific tools available, such as the ability to search on a graphic representation of a chemical structure, on a molecular formula, or on the chemical registry number. Thus, for some of the ways that chemists find information, Google Scholar, even if the information is available in the database, is an insufficient tool.

CAS is still a more valuable tool for searching for information about compounds, and because it better standardizes names, it seems to be a more useful tool for personal name searches, but Google Scholar can be a worthwhile substitute in many cases. Researchers who do not have access to CAS can confidently make use of Google Scholar to find information on chemistry topics. Those with CAS can use Google Scholar to expand the range of information available to them.

---

[1] About Google Scholar, http://scholar.google.com/scholar/about.html (accessed October 26, 2006).

[2] Ibid.

[3] Dean Giustini and Eugene Barsky, "A Look at Google Scholar, PubMed, and Scirus: Comparisons and Recommendations," The Journal of the Canadian Health Libraries Association/Journal de l'Association des Bibliotheques de la Santé du Canada 26, no. 3 (Summer 2005): 86.

[4] Marian Burright, "Google Scholar – Science & Technology," Issues in Science and Technology Librarianship 45 (Winter 2006), http://www.istl.org/06-winter/databases2.html (accessed October 30, 2006).

[5] About CAS, http://www.cas.org/about.html (accessed October 26, 2006).

[6] D.D. Ridley, Information Retrieval: SciFinder and SciFinder Scholar. Chicester: John Wiley & Sons, 2002.

[7] Recent searches on LISA: Library and Information Science Abstracts and Library, Information Science & Technology Abstracts turned up 60 and 138 results respectively (October 30, 2006).

[8] Laura Bowering Mullen and Karen A. Hartman, "Google Scholar and the Library Web Site: The Early Response by ARL Libraries," College & Research Libraries 67, no. 2 (March 2006): 106-122.

[9] Maurice C. York, "Calling the Scholars Home: Google Scholar as a Tool for Rediscovering the Academic Library," Internet Reference Services Quarterly 10 no. 3/4 (2005): 117-133.

[10] Jill E. Grogg and Christine L. Ferguson, "OpenURL Linking with Google SCHOLAR," Searcher 13, no. 9 (October 2005): 39-46.

[11] Martin Kesselman and Sarah Barbara Watstein, "Google Scholar and Libraries: Point/Counterpoint," Reference Services Review 33, no. 4 (2005): 380-387.

[12] Jeffrey Pomerantz, "Google Scholar and 100 Percent Availability of Information," Information Technology and Libraries 25, no. 2 (2006): 52-56.

[13] See, for example, Burright; Laura B. Cohen, "Finding Scholarly Content on the Web: From Google Scholar to RSS Feeds," Choice: Current Reviews for Academic Libraries 42, Special Issue (2005): 7-17; Giustini and Barsky, 85-89; Peter Jascó, "Savvy Searching: Google Scholar: the Pros and Cons," Online Information Review 29, no. 2 (2005): 208-214; Martin Myhill, "The Advisor Reviews . . . Google Scholar," The Charleston Advisor 6, no. 4 (April 2005), http://www.charlestonco.com/review.cfm?id=225 (accessed 23 may, 2006); Greg R. Notess, "Scholarly Web Searching: Google Scholar and Scirus," Online 29, no. 4

(July/August 2005): 39-41; Mick O'Leary, "Google Scholar," Information Today 22, no. 7 (July/August 2005): 35, 39; Kathryn Skhal and Rita Vine, "Google Scholar," Journal of the Medical Library Association 94, no. 1 (January 2006): 97-99 Roy Tennant, "Google, the Naked Emperor," Library Journal 130, no. 13 (August 2005): 29; and Tennant, "Is Metasearching Dead?," Library Journal 130, no. 12 (July 2005): 28.

[14] Rita Vine, "Google Scholar Gets Better at Indexing PubMed Content, but It's Still Several Months Behind," SiteLines, http://www.workingfaster.com/sitelines/archives/2006_01.html#000365 (accessed May 23, 2006); Burright.

[15] Chuck Hamaker and Brad Spry, "Key Issue: Google Scholar," Serials 18, no. 1 (March 2005): 71.

[16] See, for example, O'Leary; Skhal and Vine; and Giustini and Barsky.

[17] Susan Gardner and Susanna Eng, "Gaga Over Google? Scholar in the Social Sciences," Library Hi Tech News 22, no. 8(2005): 42-45.

[18] Jascó.

[19] Janice Adlington and Chris Benda, "Checking Under the Hood: Evaluating Google Scholar for Reference Use," Internet Reference Services Quarterly 10, no. 3/4 (2005):135-148.

[20] D. Yvonne Jones, "Biology Article Retrieval from Various Databases: Making Good Choices with Limited Resources," Issues in Science & Technology Librarianship 44 (Fall 2005), http://www.istl.org/05-fall/refereed.html (accessed May 2, 2006).

[21] Kathleen Bauer and Nisa Bakkalbasi, "An Examination of Citation Counts in a New Scholarly Communication Environment," D-Lib Magazine 11, no. 9 (September 2005), http://www.dlib.org/dlib/september05/bauer/09bauer.html (accessed May 1, 2006).

[22] Burright.

[23] Chris Neuhaus, et al., "The Depth and Breadth of Google Scholar: An Empirical Study," portal: Libraries and the Academy 6, no. 2 (2006): 127:141.

Table 1. Types of resources found in initial search

Resource Type	*Google Scholar*	*CAS*	Both
Journal article (n=564)	387 (68.6%)	248 (44.0%)	71 (12.6%)
Patent (n=54)	0 (0.0%)	54 (100.0%)	0 (0.0%)
Problem (n=26)	26 (100.0%)	0 (0.0%)	0 (0.0%)
Conference proceedings (n=23)	11 (47.8%)	12 (52.2%)	0 (0.0%)
Book (n=21)	20 (95.2%)	1 (4.8%)	0 (0.0%)
Dissertation (n=9)	9 (100.0%)	0 (0.0%)	0 (0.0%)
Other (n=5)	4 (80.0%)	1 (20.0%)	0 (0.0%)
Total (n=702)	457 (65.1%)	316 (45.1%)	71 (10.1%)

Table 2. Types of resources available after further searching on keywords in titles

Resource Type	*Google Scholar*	*CAS*	Both
Journal article (n=564)	482 (85.5%)	521 (92.4%)	438 (77.7%)
Patent (n=54)	4 (7.4%)	54 (100.0%)	4 (7.4%)
Problem (n=26)	26 (100.0%)	0 (0.0%)	0 (0.0%)
Conference proceedings (n=23)	11 (47.8%)	12 (52.2%)	0 (0.0%)
Book (n=21)	21 (100.0%)	7 (33.3%)	7 (33.3%)
Dissertation (n=9)	9 (100.0%)	5 (55.6%)	5 (55.6%)
Other (n=5)	5 (100.0%)	2 (40.0%)	2 (40.0%)
Total (n=702)	558 (79.5%)	601 (85.6%)	456 (65.0%)

Table 3. Search results by date

Date	*Google Scholar*	*CAS*	Both
pre-1969 (n=45)	16 (35.6%)	29 (64.5%)	0 (0.0%)
1970-1979 (n=34)	12 (35.3%)	22 (64.7%)	0 (0.0%)
1980-1989 (n=91)	27 (29.7%)	71 (78.0%)	7 (7.7%)
1990-1994 (n=81)	59 (72.8%)	28 (34.6%)	20 (24.7%)
1995-1999 (n=120)	80 (66.7%)	55 (45.8%)	15 (12.5%)
2000 (n=42)	35 (83.3%)	10 (23.8%)	3 (7.1%)
2001 (n=51)	43 (84.3%)	16 (31.4%)	8 (15.7%)
2002 (n=45)	40 (88.9%)	8 (17.8%)	3 (6.7%)
2003 (n=65)	48 (73.8%)	24 (36.9%)	7 (10.8%)
2004 (n=42)	30 (71.4%)	14 (33.3%)	2 (4.8%)
2005 (n=50)	38 (76.0%)	18 (36.0%)	6 (12.0%)
2006 (n=21)	14 (66.7%)	8 (38.1%)	1 (4.8%)
no date (n=15)	15 (100.0%)	0 (0.0%)	0 (0.0%)

Table 4. Records available by date after further searching on keywords in titles

Date	*Google Scholar*	*CAS*	Both
pre-1969 (n=45)	20 (44.4%)	45 (100.0%)	20 (44.4%)
1970-1979 (n=34)	23 (67.6%)	34 (100.0%)	23 (67.6%)
1980-1989 (n=91)	44 (48.4%)	85 (93.4%)	38 (41.8%)
1990-1994 (n=81)	70 (86.4%)	78 (96.3%)	67 (82.7%)
1995-1999 (n=120)	99 (82.5%)	105 (87.5%)	84 (70.0%)
2000 (n=42)	38 (90.5%)	37 (88.1%)	33 (78.6%)
2001 (n=51)	48 (94.1%)	44 (86.3%)	41 (80.4%)
2002 (n=45)	43 (95.6%)	36 (80.0%)	34 (75.6%)
2003 (n=65)	57 (87.7%)	58 (89.2%)	49 (75.4%)
2004 (n=42)	38 (90.5%)	25 (59.5%)	34 (81.0%)
2005 (n=50)	47 (94.0%)	37 (74.0%)	34 (68.0%)
2006 (n=21)	16 (76.2%)	18 (85.7%)	13 (61.9%)
no date (n=15)	15 (100.0%)	0 (0.0%)	0 (0.0%)

Table 5. Comparison of records found in three types of searches (larger figures in bold for emphasis)

Search Type	*Google Scholar*	*CAS*	Both
Topical (n=429)	**361 (84.1%)**	85 (19.8%)	17 (4.0%)
Compound (n=180)	47 (26.1%)	**140 (77.8%)**	7 (3.9%)
Name (n=93)	49 (52.7%)	**91 (97.8%)**	47 (50.5%)

Table 6. Comparison of records found in four searches (two topical and two compound)

Search	*Google Scholar*	*CAS*	Both
Hypericin and EPR (n=119)	**109 (91.6%)**	16 (13.4%)	6 (5.0%)
Pulsed EPR (n=310)	**252 (81.3%)**	69 (22.3%)	11 (3.5%)
Hydroxybutyranilide (n=135)	29 (21.5%)	**107 (79.2%)**	1 (0.7%)
Myrigalone (n=45)	18 (40.0%)	**33 (73.3%)**	6 (13.3%)

Table 7. Comparison of records available for three types of searches after further analysis

Search Type	*Google Scholar*	*CAS*	Both
Topical (n=429)	**413 (96.3%)**	336 (78.3%)	320 (74.6%)
Compound (n=180)	92 (51.1%)	**174 (96.7%)**	86 (47.8%)
Name (n=93)	53 (57.0%)	**91 (97.8%)**	51 (54.8%)

Table 8. Comparison of records available for four searches (two topical and two compound) after further analysis

Search	*Google Scholar*	*CAS*	Both
Hypericin and EPR (n=119)	**115 (96.6%)**	88 (73.9%)	84 (70.6%)
Pulsed EPR (n=310)	**298 (96.1%)**	248 (80.0%)	236 (76.1%)
Hydroxybutyranilide (n=135)	57 (42.2%)	**135 (100.0%)**	57 (42.2%)
Myrigalone (n=45)	35 (77.8%)	**39 (86.7%)**	29 (64.4%)