Can Google Scholar Replace Your Database?

By Dave Schmitt, Engineering Librarian, UC San Diego

Why look at Google Scholar?

In the face of (A) widespread evidence that users turn to Google over traditional A&I databases and (B), budget cuts which force us to examine these same databases, I began to wonder just how complete Google Scholar is. Ultimately, I hoped to discover whether or not Google Scholar could be a viable replacement for the subject-specific databases to which we subscribe.

So I focused on answering this fairly simple question:

How many of the documents that researchers actually cite are indexed in Google Scholar?

Study Parameters:

To answer my question, I turned the school to which I am liaison, the Jacobs School of Engineering (JSOE) at UCSD. I searched for all dissertations and theses produced by JSOE in the year 2008 in ProQuest Dissertations & Theses. Most, if not all, of the JSOE theses in the year 2008 are electronic and located in this database. I found 157 theses (47 MS and 110 PhD) across five departments.

Method:

After locating the documents, every citation was searched for in Google Scholar. As I searched for the documents, I noted the type of document (journal article, website, etc.) and tallied whether or not it was findable. This method allowed me to answer the question of whether or not the cited literature had been indexed by Google, but also had the side effect of allowing me to see the difference in information needs between the various departments, and the citation habits of each.

Results Highlights:

- **95.8%** of all documents being cited can be found in Google Scholar
- Limiting only to document types usually found in A&I databases (Journal Articles, Conferences, Books, Tech Reports, Theses, Standards and Patents), this number jumps to **98.8%**
- Journal Articles and Books are most likely to be found in Google Scholar with **99.7%** included
- Only **11.9%** of cited websites are found in Google Scholar

Conclusion:

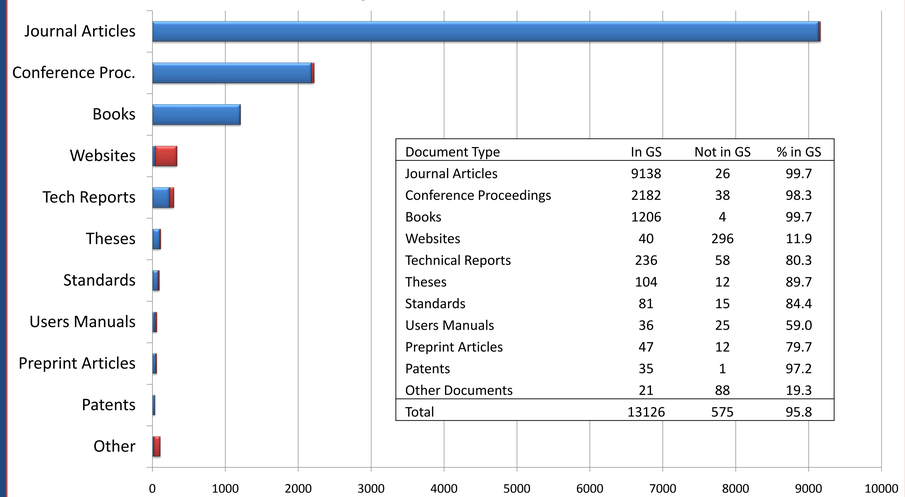
Purely seen in terms of indexing information, Google Scholar surpasses A&I databases. Benefits include:

- Wide range of recall across disciplines, as opposed to single-discipline databases
- Indexes all types of documents, not limited to journal articles and conference proceedings
- Easy to use and familiar
- It's free

What Google Scholar lacks is the dexterity of searching. Drawbacks:

- Doesn't include many detailed search fields, which could be useful in specialized subjects
- Difficult to refine searches
- The volume of results can be overwhelming

All Departments: 157 Theses



Document Type	In GS	Not in GS	% in GS
Journal Articles	9138	26	99.7
Conference Proceedings	2182	38	98.3
Books	1206	4	99.7
Websites	40	296	11.9
Technical Reports	236	58	80.3
Theses	104	12	89.7
Standards	81	15	84.4
Users Manuals	36	25	59.0
Preprint Articles	47	12	79.7
Patents	35	1	97.2
Other Documents	21	88	19.3
Total	13126	575	95.8

Other Document Types: News Articles (5/25), Class Resources (3/7), Unpublished Documents (1/7), Software Programs (8/0), Press Releases (2/5), Congressional Statements (2/2), Private Communications (0/25), MSDS (0/7), Catalogs (0/2), Maps (0/1), Circuit Designs (0/1), Unknown (0/6)

So what Isn't in Google Scholar?

Of the 575 citations I couldn't find in Google Scholar, there are three main categories:

1. Websites = 296 citations
2. "Non-traditional" documents, which you wouldn't expect to find in a database (private communications, catalogs, maps, circuit designs, class resources, news) = 125 citations
3. "Traditional" documents, such as Journal Articles, Conferences, Books, Tech Reports) = 154 citations

Of the traditional documents, standards and tech reports were the most likely not to be indexed. I suspect, though am not certain, this is due to the relatively non-systematic method in which agencies publish and distribute these documents.

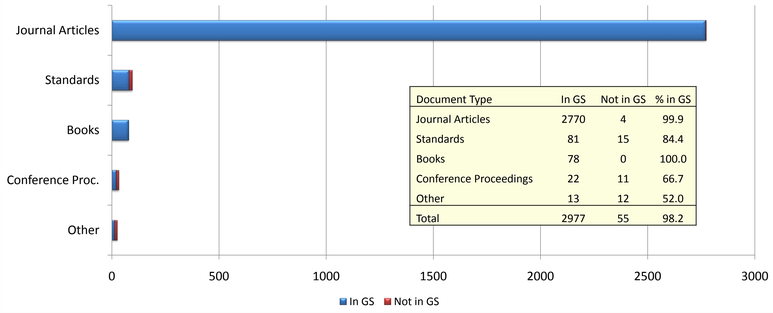
Of the 26 articles, most were in smaller publications or supplementary material.

Of the 38 conferences and technical reports, many were of local nature, either held at or published by UC San Diego.

UCSD Libraries

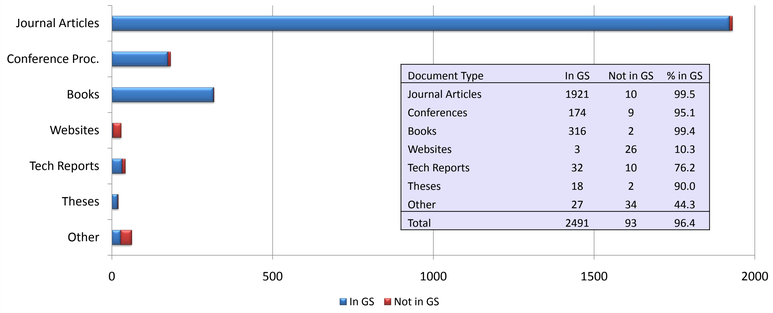
Science & Engineering Library

BioEngineering: 25 Theses



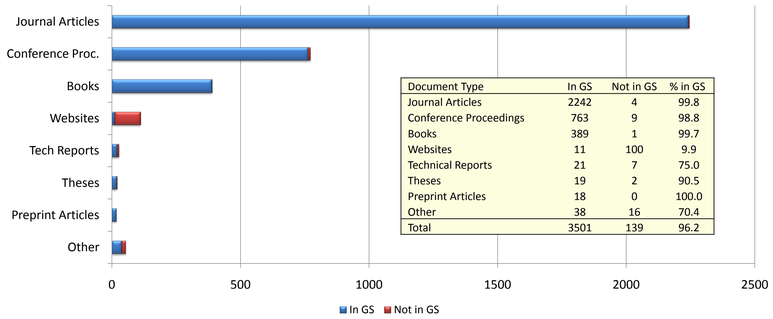
Document Type	In GS	Not in GS	% in GS
Journal Articles	2770	4	99.9
Standards	81	15	84.4
Books	78	0	100.0
Conference Proceedings	22	11	66.7
Other	13	12	52.0
Total	2977	55	98.2

Mechanical and Aerospace Engineering: 29 Theses



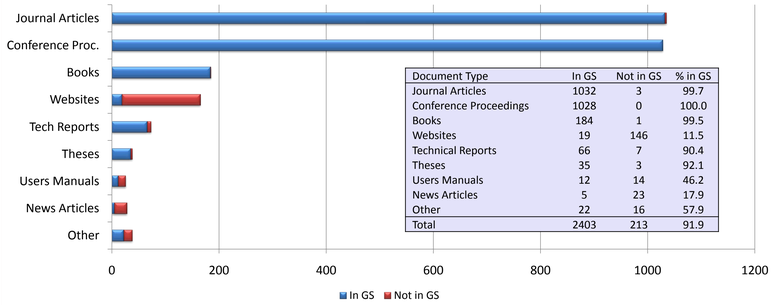
Document Type	In GS	Not in GS	% in GS
Journal Articles	1921	10	99.5
Conferences	174	9	95.1
Books	316	2	99.4
Websites	3	26	10.3
Tech Reports	32	10	76.2
Theses	18	2	90.0
Other	27	34	44.3
Total	2491	93	96.4

Electrical and Computer Engineering: 39 Theses



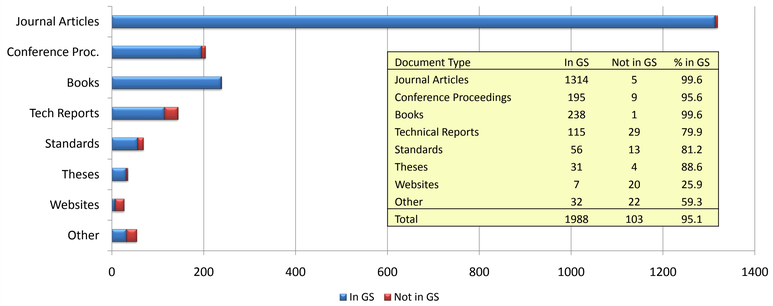
Document Type	In GS	Not in GS	% in GS
Journal Articles	2242	4	99.8
Conference Proceedings	763	9	98.8
Books	389	1	99.7
Websites	11	100	9.9
Technical Reports	21	7	75.0
Theses	19	2	90.5
Preprint Articles	18	0	100.0
Other	38	16	70.4
Total	3501	139	96.2

Computer Science and Engineering: 42 Theses



Document Type	In GS	Not in GS	% in GS
Journal Articles	1032	3	99.7
Conference Proceedings	1028	0	100.0
Books	184	1	99.5
Websites	19	146	11.5
Technical Reports	66	7	90.4
Theses	35	3	92.1
Users Manuals	12	14	46.2
News Articles	5	23	17.9
Other	22	16	57.9
Total	2403	213	91.9

Structural Engineering: 22 Theses



Document Type	In GS	Not in GS	% in GS
Journal Articles	1314	5	99.6
Conference Proceedings	195	9	95.6
Books	238	1	99.6
Technical Reports	115	29	79.9
Standards	56	13	81.2
Theses	31	4	88.6
Websites	7	20	25.9
Other	32	22	59.3
Total	1988	103	95.1