Can Google Scholar Replace Your Database?

By Dave Schmitt, Engineering Librarian, UC San Diego

Can Google Scholar Replace Your Database?

By Dave Schmitt, Engineering Librarian, UC San Diego

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.

Study Parameters:

To answer my question, I turned the school to which I am liaison, the Jacobs School of Engineering (JSE) at UCSD. I searched for all dissertations and theses produced by JSE in the year 2008 in ProQuest Dissertations & Theses. Most, if not all, of the JSE 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

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?