

## Finding and Counting Papers

“Those who count count” an old academic friend of mine used to say, and I have found myself often repeating this adage. But do they count right?

In a recent meeting of the ASME Journals chief editors, one of my colleagues mentioned that he personally recalculated his journal’s impact factor and found it to be at least three times higher than that reported by a citation service. He had several explanations of why this happened, and other editors chimed in. While the discussion was quite engaging, it became clear once again that the vagaries of data mining techniques are starting to rule our lives in many ways, including finding and counting our papers.

While data mining is done by machines, the rules that the machines operate under are made by humans and are subject to the usual human traits.

You can enjoy the many twists and turns these countings can take by reading the “h-index” talk stream in Wikipedia <http://en.wikipedia.org/wiki/Talk%3AH-index>. A discussion on citation tracking in the sciences can be found in an article cited in another Wikidedia article (Bakkalbasi et al. *Biomedical Digital Libraries* 2006 3:7 doi: 10.1186/1742-5581-3-7) cited in [http://en.wikipedia.org/wiki/Academic\\_journal](http://en.wikipedia.org/wiki/Academic_journal) (March 21, 2012). This article includes the comparison of features in the three citation services, Web of Science, Scopus, and Google Scholar, and summarizes their features in the following table:

	Web of science	Scopus	Google scholar
Indexing and abstracting	Yes	Yes	No
Years covered-journals	1900 to present (Science); 1956 to present (Social Science); 1975 to present (Arts and Humanities)	1966 to present for some journals, but many date back to 1996 to present	Not revealed
Years covered-citations	1900 to present	1996 to present	Not revealed
Fee-based	Yes	Yes	No
Contents	9300 journals (Science, Social Science, and Arts and Humanities)	15,000 journals (Science and Social Science)	Not revealed

The authors end their abstract with the following conclusion: “Our data indicate that the question of which tool provides the most complete set of citing literature may depend on the subject and publication year of a given article.” The article was published

in 2006 so perhaps data mining techniques are more definitive now. The h-index discussion cited above gives us some clues about the difficulties we face. Indeed, this Wikipedia discussion would crack me up except that too many counters may be deciding others’ lives and careers based on these numbers.

There are two areas where I think we can help ourselves in developing more accurate metrics for those who care to use them. One is to avoid “fragmentation” of citations. For example, substantively identical work may appear as a conference paper in archived proceedings, as a preprint and as a final official article. Depending on the search engine, you may get separate citations for all three with different numbers. If on top of this, your publisher has your paper on their site as the definitive source, you have your paper also posted as a stand-alone pdf on your own site (copyright issues aside), and your co-authors have it also on their sites, these separate URLs could be found as separate papers rather than instances of the same work. The suggestion here is that you use your publisher’s URL as the definitive source and put a live link to that even if you keep a local pdf, say, by inserting a “click here” for the pdf.

This strategy does not resolve the issue of how to site the work itself in the proper way for your readers. For example, a conference paper may precede its journal appearance by 2–3 years; citing only the journal version gives the wrong impression of when the work was completed; citing only the conference version gives the wrong impression about the archival quality of the work. In the end, you have to include both sources but as a single citation, adding words like “also appeared as....” This is also a good practice for a resume to avoid perceptions of “padding.”

The second area we can help ourselves is to be precise on the journal name we use. Our own journal is a challenge: It can be several combinations of a subset of the words “ASME Transactions Journal of Mechanical Design” and in a variety of sequences. My current understanding is that using the abbreviation “J. Mech. Des.” or “ASME J. Mech. Des.” would do the trick, and we should make a habit of using just that.

There is a continuing effort from ASME staff, the ASME Publications Committee, and all the editors of our sister ASME journals to press the citation services for a more consistent and thorough service. While this effort will likely continue for ever, we can help in two ways. Be aware of the issues and use simple strategies as I suggested above to help things along. Remind the counters, any chance we get, what all engineers know: All measured numbers have a distribution and unless you know what the distribution really is you use the numbers at your peril (and the peril of others).

**Panos Y. Papalambros**  
Editor