

Savvy searching Citation-enhanced indexing/abstracting databases

Péter Jacsó

The author

Péter Jacsó is Professor, University of Hawaii, Manoa, Hawaii, USA.

Keywords

Online databases, Abstracting services, Indexing services

Abstract

Provides an introductory look at what savvy users should know about the implications of having information about cited references in I/A records. Looks at the more sophisticated, link-enabled cited references and the novel citation scores in full-text collections, then discusses the alternatives for searching efficiently by elements of cited references: cited author, cited title, cited source and cited year in I/A databases and full-text archives.

Electronic access

The Emerald Research Register for this journal is available at
www.emeraldinsight.com/researchregister

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1468-4527.htm

Most indexing/abstracting (I/A) services have been enhancing their records with a variety of value-added information beyond the traditional bibliographic data elements quite steadily, but none of them is comparable to the potential provided by enhancing the records with cited references. For decades, apart from the pioneering Shepard's Citations series for case law, and to a limited extent the various patent databases, the Institute for Scientific Information (ISI) has been the only I/A service which included references to cited works in the records of journal articles. It made a mighty business of it by enabling savvy searchers to explore related publications. With the instant linking to the cited and citing documents made possible by the Web and the ever-growing full-text digital archives of journal collections, enhancing I/A records by citations is certain to become much more common in I/A databases not only to inform readers about the cited publications, but also to lead them directly to those publications, or at least to their I/A records. This has become common practice in the best full-text digital archives.

This part provides an introductory look at what savvy users should know about the implications of having information about cited references in I/A records. The second part looks at the more sophisticated, link-enabled cited references and the novel citation scores in full-text collections, while the third part discusses the alternatives for searching efficiently by elements of cited references: cited author, cited title, cited source and cited year in I/A databases and full-text archives.

Value-added information

I/A services, also known as secondary information services, have been competing with each other in adding more useful data elements to their records, which in a succinct form provide more information for the users about the primary items, and can be even used to filter the search. When UMI (now ProQuest) added information about the length of the primary documents (short, medium, long), Infotrac (now part of Gale) added the number of words in the original article to the record, then EBSCO enhanced records in some of its databases with a reading level indicator. Soon after InfoTrac assigned letter grades (A, B, C, etc.) to movie reviews, UMI countered with adding informative qualifiers to some of its subject terms, such as "movie review – favourable", "movie review – unfavourable", etc. in records about reviews of books, movies, theatrical performances and the like.

Similarly, when H.W. Wilson started to add information about the peer-reviewed status of the



journals (refereed vs non-refereed), some of its competitors started to add other journal qualifiers, such as “Peer Reviewed”, “Scholarly Journal”, “Trade Publication”, “Magazine”, while INSPEC went to the analytic level to indicate on an article-by-article basis if the topic is given – among others – a theoretical or practical treatment. There have been many similar enhancements in the best I/A databases, along with the flurry of pseudo high-tech enhancements by adding the e-mail address of the authors, the URL of the journal, the publisher – not that these were found useful, but rather because it was good public relations and easy. Of course, in careless hands these can turn into frustration, as happened with PsycINFO, which managed to mess up not only its own simple URL <http://www.apa.org> by using http://www.apa.org in 746 records, but also those of many other publishers in almost 12,000 records starting the URL with http://www, not to mention the tens of thousands of syntactically invalid journal URLs. And this is from a unit of the American Psychological Association which preaches in its Style Manual the importance of verifying your URLs!

Inclusion of references to cited publications

The above data elements are all useful enhancements but are dwarfed by the value added through the inclusion in the indexing/abstracting records of short bibliographic references to the articles, conference papers and/or books cited. Most of these are genuine citations, directly or indirectly related to the article at hand, and thus may lead the users to other publications relevant to their topic – without the need to formulate queries. In his classic article written almost 40 years ago, Garfield (1955) convincingly made the case for the benefits of using cited references as a complementary alternative to descriptor searching. His thoughts are mightily validated on the Web in general, and among the I/A databases in the Web of Science (WoS) system, a grown-up, virile version of his brain-child, the Science Citation Index (see Figure 1) in particular. Tenopir (2003) explains in her trademark style of clarity what makes citation searching so powerful in this age using WoS.

Although the title of the cited item is not included in the cited references (except for cited books), the citation-enhanced records add significant value for the information retrieval process compared with traditional I/A records, especially in those implementations where items in the citation list are hot-linked to their master record if available. They serve as

Figure 1 Citation enhanced record from Social Science Citation Index

TITLE: Journal impact factors and self-citations: Implications for psychology journals

AUTHOR(S): Anseel F (REPRINT); Duyck W; De Baene W; Brysbaert M
 JOURNAL: AMERICAN PSYCHOLOGIST, 2004, V59, N1 (JAN), P49-51

CITED REFERENCES:

NATURE, 2002, V415, P101
 *AM PSYCH ASS, 2001, PUBL MAN AM PSYCH AS
 ADAIR JG, 2003, V58, P15, AM PSYCHOL
 AKSNES DW, 2003, V56, P235, SCIENTOMETRICS
 BOOR M, 1982, V37, P975, AM PSYCHOL
 GOTTFREDSON SD, 1978, V33, P920, AM PSYCHOL
 LAWRENCE PA, 2003, V422, P259, NATURE
 MCGARTHY C, 2000, V5, P1, CURRENT RES SOCIAL P
 MOED HF, 1999, V46, P575, SCIENTOMETRICS

SocialSciSearch(R)(Dialog® File 7): (c)2004 Inst for Sci Info. All rights reserved.

mini-bibliographies, which sometimes are not that mini, consisting on the average of 30-40 cited references in social science articles.

Cited references are hard to add (accurately)

In spite of the obvious benefits, other I/A databases of scholarly literature have not followed suit by adding citations to their records for three decades. In the mid-1990s the Current Index to Nursing and Allied Health Literature (CINAHL) database started to add cited references to its records. It is a valuable feature, and not only for nursing and allied health practitioners and librarians, but also for many other information professionals because CINAHL has very good coverage of key library and information science and technology journals. Unfortunately, in many of the CINAHL records the citations were added in an unusual manner. Spelling author and journal names correctly, using the right volume, issue and page numbers have been long sacrificed by some publishers at the altar of rigorous verification of adherence to their discombobulating preferences in one of the 900+home-grown citation styles, and by database producers at the altar of pumping more records at less cost in their databases by outsourced data entry operations. CINAHL's handling of cited author names added a novel flavour: abbreviating the surname of the authors to an initial while spelling out the first names and/or middle names. In the excerpt below Lawrence D is Lawrence Dowler, Steve L. is Steve Lawrence, Lee GC is C.L. Giles, and Hans R. is Hans Roes, not to mention that the place of the conference where Roes' paper was presented is not Prage but Prague, and the URL of the full-text paper is syntactically and semantically incorrect: it should have been

<http://drcwww.kub.nl/~roes/articles/ej_praag.htm>. The irony of the fact that this is an article about the lack of correct citation style on the Web is not lost on me. To CINAHL's credit such citation format is not characteristic of most of its records (although I could only test this by modest sampling), and semantically and syntactically incorrect citations are pervasive in every database (see Figure 2).

Citation enhancement takes a lot of time, money and care

No wonder that publishers of I/A services have shied away from including cited references, which have been growing substantially, especially in the social sciences. Adding citations to I/A records incurs extra costs. Adding them correctly and in a functionally optimal way incurs even more costs. The producers of the e-psyche database of Database Access Group have learned this the hard way. While they kept promising a rose garden, they kept digging their own grave by drastically cutting back on the inclusion of cited references, first eliminating references to books, then to pre-1970 works, quickly resorting to selecting (or rather, randomly picking) up to ten citations, irrespective of the number of citations in the articles (Jacsó, 2004). To put things into perspective, more than half of the article records in the various psychology subsets of Social Science Citation Index had more than 10 cited references, and 21 per cent had more than 30 cited references in March 2004. For articles published in 2003 these ratios grew close to 65 per cent and 42 per cent, respectively - which helps in gauging how much users of e-psyche have been short-changed in cited references.

Figure 2 Novel type of discombobulating citation style in CINAHL, citing authors by their surname initial and fully spelled out first name

Authors

Mehta U.

Title

On the World Wide Web: where are you going, where have you been?

Source

Internet Reference Services Quarterly. 5(1):51-66, 2000. (26 ref)

Abstract

An analysis of Web citations was performed to examine three aspects of scholarly and research communication: use of authoritative information, use of correct citation style, and availability of the cited sources for the purpose of verification. Results of the study indicate that a database or a repository of Web citations needs to be established.

Cited References

Lawrence D ed. Gateways to knowledge: the role of academic libraries in teaching, learning, and research. Cambridge, MA: MIT Press; 1997: 17-18.

Steve L, Lee GC. Searching the World Wide Web. Science 1998; 280: 98.

Peter L. The gateway library: teaching and researching in the global reference room, in Gateway Library.: 135-147.

Hans R. Electronic Journals: transforming the information cycle? Paper presented at the Internet Conference. Prague: 1994 June. Available: http://cwis.kub.nl/~dbi/users/roes/articles/ej_praag.htm (October 24, 1999).

Launching e-psyche had one advantage. The chance of possible competition by e-psyche spurred into action the producers of the PsycINFO database. They hurriedly started to enhance records with cited references in 2001. Impressively, by March 2004 there were 227,000 records with 9.8 million citations added to PsycINFO, an average of 43 citations per item. This is impressive even if in the psychology subsections of the Social Science Citation Index (dating back to 1972) there are more than twice as many records enhanced by cited references. However, the ISI records have lower average citations per record than PsycINFO because ISI does not process books for its traditional indexes. Books have an order of magnitude higher cited references on average than journal articles. The first 51 books added to PsycINFO with 2004 publication year averaged a whopping 625 cited references.

It is another question that in most implementations of PsycINFO often only a fraction of the references could be included in the records due to size limitations of the MARC communications format (Jacsó, 2004). It has been used for distributing PsycINFO to database hosts until mid-2003, when delivery in XML format (which has no size limit) became available. Nine months later the majority of PsycINFO implementations are still created from the MARC records, missing a significant portion of cited references. APA should have prepared much better for this important project and the transition. Not even the MARC format limitation would explain why many of the records published in the past three years miss all the citations which appear in the article, including records for articles published in APA's own journals.

On the positive side, the cited references in PsycINFO are much more informative by virtue of including the title of cited journal articles in the records displayed, as is obvious if you compare the record for the same article in Figures 1 and 3.

Figure 3 Cited references with article titles in PsycINFO (albeit in redundant formatting in some records in Ovid's implementation)

Author

Anseel, Frederik; Duyck, Wouter; De Baene, Wouter; Brysbaert, Marc.

Title

Journal Impact Factors and Self-Citations: Implications for Psychology Journals.

Source

American Psychologist. Vol 59(1) Jan 2004, 49-51.
American Psychological Assn, US

Cited References

(1) Adair, J. G., & Vohra, N. (2003). The explosion of knowledge, references, and citations: Psychology's unique response to a crisis. *American Psychologist*, 58, 15-23. Adair, J. G., & Vohra, N. The explosion of knowledge, references, and citations: Psychology's unique response to a crisis. *American Psychologist*.
(2) Aksnes, D. W. (2003). A macro study of self-citation. *Scientometrics*, 56, 235-246. Aksnes, D. W. A macro study of self-citation. *Scientometrics*.

Figure 4

 Hide Cited References

Total Number: 9 Total Number Displayed: 9

1. Adair, J. G., & Vohra, N. (2003). The explosion of knowledge, references, and citations: Psychology's unique response to a crisis. *American Psychologist*, 58, 15-23. **PsycINFO No.:** 2003-02034-002
 2. Aksnes, D. W. (2003). A macro study of self-citation. *Scientometrics*, 56, 235-246.
 3. American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
 4. Boor, M. (1982). The citation impact factor: Another dubious index of journal quality. *American Psychologist*, 37, 975-977. **PsycINFO No.:** 1983-06523-001
 5. Errors in citation statistics: Opinion of the Nature editors. (2002, January 10). *Nature*, 415, 101.
 6. Gottfredson, S. D. (1978). Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments. *American Psychologist*, 33, 920-934. **PsycINFO No.:** 1979-24404-001
 7. Lawrence, P. A. (2003, March 20). The politics of publication. *Nature*, 422, 259-261.
 8. McGarty, C. (2000). The citation impact factor in social psychology: A bad statistic that encourages bad science? *Current Research in Social Psychology*, 5(1), 1-16. Retrieved October 16, 2003, from <http://www.uiowa.edu/~grpproc/crisp/crisp.5.1.htm>
 9. Moed, H. F., van Leeuwen, T. N., & Reedijk, J. (1999). Towards appropriate indicators of journal impact. *Scientometrics*, 46, 575-589.
-

This also has a significant advantage in searching for cited references. Unfortunately, the citations in Ovid's implementation sometimes inexplicably repeat the author's name, the article title and the journal title, making the citations a bit convoluted.

OCLC handles better the cited references in its implementation of PsycINFO, not only in terms of the layout, but also because of consistently linking the users to records of the cited articles if PsycINFO has a record. OCLC also adds unique value-added information (not shown here): the number of OCLC member libraries which subscribe(d) to the journal (see Figure 4).

Other implementations of PsycINFO, particularly by CSA-IDS and Thomson ISI (in its new role as an aggregator of third-party content via the Knowledge of Science platform), add even more teeth to the potential of citation enhancement of I/A and full-text

records. The next column will review those savvy features.

References

- Garfield, E. (1955), "Citation indexes for science: a new dimension in documentation through association of ideas", *Science*, Vol. 122 No. 3159, pp. 108-11, Open Access PDF version, available at: http://profiles.nlm.nih.gov/BB/A/1/X/K/_/bbaixk.pdf, Open Access HTML version, available at: [www.garfield.library.upenn.edu/papers/science_v122\(3159\)p108y1955.html](http://www.garfield.library.upenn.edu/papers/science_v122(3159)p108y1955.html)
- Jacsó, P. (2004), "Jacsó's cheers and jeers for 2004", *Information Today*, Vol. 21 No. 1, pp. 17, 19, 54, Open Access enhanced HTML version, available at: <http://hypatia.slis.hawaii.edu/~jacso/extra/infoday/cj-03/psy/psy.htm>
- Tenopir, C. (2003), "Online databases: the power of citation searching", *Library Journal*, Vol. 126, 1 November, pp. 39-40, Open Access HTML version, available at: www.libraryjournal.com/article/CA178212