The emergence of competitors to the *Science Citation Index* and the *Web of Science*

Dana L. Roth

California Institute of Technology Library System, Pasadena, CA 91125, USA

Keywords: Citation indexing, *Science Citation Index*, *Web of Science.*

EUGENE Garfield's seminal paper 'Citation indexes for science'[1] and his subsequent response[2] to Schoenbach[3] provided the intellectual foundation for the wide variety of citation indexing products emerging as competitors to what has become the *Web of Science (WoS)*[4]. As Garfield succinctly pointed out, subject or classified indexes can, as a practical matter, only provide 'subject' access at the article level. Citation indexing, however, takes advantage of an author's indexing of parts of the article and has the potential to obviate many of the problems scientists experience with indexing terminology and nomenclature.

Furthermore, Garfield suggests that, 'it is impossible for any one person (e.g. the indexer) to anticipate all the thought processes of a user' … and 'by using authors' references in compiling the citation index, we are in reality utilizing an army of indexers, for every time an author makes a reference he is in effect indexing that work from his point of view'[1].

In this regard, Garfield also alluded to an often overlooked analogy between subject and citation indexing and *Chemical Abstracts* and *Beilstein*. Subject indexes, including the *Chemical Abstracts General Subject Index* (CA GSI), generally index articles at the level at which they are written. For example, a comprehensive review article on 'Dissociation constants and limiting conductances of organic acids in water' only has index terms for 'dissociation constants' and 'acids, properties'. While, as Garfield noted, 'in a certain sense a citation index is not very different from a compendium like *Beilstein*'[1] which would index each compound described in the article and relate them to the physical property reported.

Fortunately, Garfield's seminal article 'Citation indexes for science' not only led to the *Science Citation Index* and the *WoS*, but also to the development of subject retrieval, based on cited references, in a plethora of other services. In contrast with these other services, however, the *WoS* uniquely offers 'author/source' citation searching which, with some diligence, also uniquely retrieves mis-cited or in-press references.

e-mail: dzrlib@library.caltech.edu

Services currently offering cited reference searching include:

- *Chemical Abstracts*/*SciFinder*/*SciFinder Scholar*
- *NASA Astrophysics Data System Abstract Service*
- Amazon.com's 'Search Inside this Book' program
- *Scopus*
- *Scitation*/*Spin Web*
- *PROLA* (*Physical Review* Online Archive)
- Citation Bridge (US Patents)
- US Patent and Trademark Office
- *Google Scholar*
- *Optics InfoBase*
- *CiteSeer*
- *Science Direct*
- *PsycINFO*
- *IEEE Xplore*
- *Spires HEP*
- IOP (Institute of Physics)
- *CrossRef*

Chemical Abracts/SciFinder/SciFinder Scholar (http://www.cas.org/casdb.html)

Chemical Abstracts (*CA*) offers broad subject coverage, extending to the chemical aspects of astronomy, biology, education, engineering, economics, geology, history, mathematics, medicine, and physics. In addition, *CA*'s uniquely extensive format coverage includes articles from journals, preprints (2000+) and regularly published conference proceedings (73%), articles from one-time or first-time conference proceedings (7%), dissertations (2%), technical reports (1%), patents (16%) and edited research monograph chapters (1%).

The electronic version of *CA* on STN, *SciFinder* and *SciFinder Scholar* all offer cited reference searching. First announced in early 2000, cited references are included in CAplus records from (i) non-patent documents published in the Roman alphabet since 1997, and (ii) examiner citations from basic patents from the US, EPO, WIPO, and German patent offices, starting in 1999 … and from the UK and French patent offices beginning in 2003. These 'cited references' are algorithmically linked to their CAplus records. Thus, each CAplus record (since the early 1900s) is linked to CAplus records (beginning in 1997) that cite it correctly. The 'cited reference'

searching feature is available for a single reference or set of references first via the 'Get Related' button in *SciFinder Scholar*, and then by clicking 'Get Citing References'.

This feature is similar to the 'Times Cited' link appearing on the full record for individual articles in *WoS*, which retrieves 'exact' citation matches but not citations with minor errors. There are some examples (*J. Chem. Soc. Chem. Commun./Chem. Commun.*) where citations in *CA* are algorithmically corrected. In contrast with *CA*, *WoS* also indexes articles from journals published in non-Roman alphabets which, by definition, are excluded from the CAplus file.

NASA Astrophysics Data System Abstract Service (http://adsdoc.harvard.edu)

The *NASA Astrophysics Data System Abstract Service* (NASA ADS) includes four bibliographic databases, which can be searched individually or in combination.

1. *Astronomy and Astrophysics* (1,064,360 records), including 103,358 abstracts from Planetary Sciences and Solar Physics journals.
2. *Instrumentation* (751,060 records), including 201,835 abstracts from SPIE conference proceedings.
3. *Physics and Geophysics* (1,907,441 records), including 371,480 abstracts from APS journals.
4. *ArXiv Preprints* (305,990 records), consisting of all the papers published in the ArXiv e-print archive.

Each database contains abstracts from articles and monographs covering all the major journals, many minor journals, conference proceedings, several observatory reports and newsletters, NASA reports, and Ph D theses. The *NASA ADS* allows searching by author, astronomical object name, and words in the title/abstract text. The results list is ranked by how closely the paper matches the query and provides a link to the full record. From the *NASA ADS* home page [http://adswww.harvard.edu/], click on 'Search References' and then click on '*Astronomy and Astrophysics*' and then click/unclick the boxes for 'Databases to query'.

The brief results list is annotated with single letter abbreviations for various links. These are defined at: [http://adsdoc.harvard.edu/abs_doc/help_pages/results.html#List_of_Links]. 'C' is the link to 'Citations to the Article' which is limited to 'citing articles' in the *NASA ADS* database.

A comprehensive discussion of the limitations of citation searching in the *NASA ADS* Database is given at: [http://adsdoc.harvard.edu/abs_doc/help_pages/citations.html]. For example, a recent (2000) article entitled '9,9′-Bianthryl and its van der Waals complexes' … had zero 'cited by' references in *NASA ADS*, while having 8 'cited by' references (from the same chemistry journals) in both *WoS* and *Chemical Abstracts SciFinder Scholar*.

Amazon.com's 'Search Inside this Book' program (http://www.amazon.com/)

Books in Amazon's 'Search Inside the Book' program have been scanned and OCR'ed to provide searchable full text. Search results, in addition to the normal Amazon display, also provide both a page number and short excerpt for the first appearance of the search term. Clicking on 'See more references to …' will display all the page numbers and excerpt references to the search term. Registering (free) with Amazon allows display of the full page (for each occurrence of the search term) and for browsing two pages forward/back. A navigation bar above the page allows searching for other terms in the book and browsing other pages.

'This book cites' references are displayed, providing the cited book's title and page references. 'Books that cite this book' information is also provided, effectively creating a Book Citation Index. For example, 'Structure Impact' by N. Jones (1997) cites 9 books, while 'Impact Strength of Materials' by W. Johnson (1972) was cited in 6 books.

Scopus (http://www.scopus.com/scopus/home.url)

Scopus is generally considered as a potential competitor to the *Science Citation Index*, since it delivers search results that include abstracts, cited references and links to citing references. *Scopus* is a 'work in progress' that is based on a variety of sources which result in some interesting retrievals. A 'Basic' search for Marcus, R. A. as an author retrieves 56 items including an article authored by Marcus, A. R. Several articles are listed twice, suggesting incomplete editing of overlapping sources for the article references. The results list displays the article title, authors, source and the number of 'cited by' references in the *Scopus* database. Each results listing is preceded by a 'Refine Results' box that allows 'limiting to' or 'excluding' by: source title, co-author name, publication year and document type. The results list is followed by 'Search within results' box that allows limiting. Entering 91125, the zip code for R. A. Marcus at Caltech, reduces the original [Marcus, R. A.] author search results list to 34 items. This, however, obviously deletes articles without full or correct addresses, and all pre-1985 articles.

Scopus has recently added 13 million patent records, including those from the US Patent and Trademark Office (USPTO), the European Patent Office (EPO), the Japanese Patent Office (JPO), and the Patent Cooperation Treaty (PCT) of the World Intellectual Property Organization (WIPO).

Scitation/Spin Web (http://scitation.aip.org/)

Scitation is the re-launch of the American Institute of Physics' Online Journal Publishing Service (OJPS), a jour-

nal hosting platform, which debuted in 1996. *Scitation* currently hosts 140 publications from 19 society publishers. *Scitation* has also expanded coverage from journals to include: conference proceedings, standards, etc., and contains over 600,000 articles, adding more than 6000 articles per month. For each journal, coverage is limited to 'full text' articles.

The name '*Scitation*' was chosen because of the science and engineering emphasis and its new ability to provide both forward and backward reference linking. *Scitation*'s citation-linking feature is based on author-supplied article references (beginning in 1998), which are first edited and then checked against a database of over 12 million reference citations. This lookup/vetting process is analogous to ISI's KeySave process, which 'corrects' citations (if necessary) based on their related source records.

Cited references, when available in *Scitation* records, may be linked to ISI's *WoS*, *PubMed*, *Chemport/Chemical Abstracts*, *SPIN*, *INSPEC*, *EDP Sciences*, *X-ArkiV*, *SLAC-SPIRES*, publisher web sites, and other *Scitation* journals. Both the cited references and citing articles are listed on the full record screen.

The *Scitation* home page provides browsing of individual publications by title, publisher or subject category and 'quick' author/subject searching by keywords or 'advanced' full text searching that also allows limiting by publication date or volume/issue ranges. The 'advanced search' also provides for 'cited author' searches. Search results display both references and citing articles, that are hyperlinked, on the full record screen.

SPIN is a complementary database, which includes all *Scitation* records, that has been published since 1971. *SPIN* provides indexing and abstracting of more than 100 major physics, astronomy, and engineering journals and conference proceedings. It provides access to nearly 1.5 million abstracts adding about 50,000 new records each year. The *Scitation* home page [http://scitation.aip.org/] offers searching of either *Scitation* or *SPIN+Scitation* for subscribers to any *Scitation* publication.

Searching *Scitation* is freely available to non-subscribers (with personal registration). See: http://scitation.aip.org/servlet/HelpSystem?KEY=SCI&TYPE=HELP/FAQ

PROLA (Physical Review Online Archive) (http://prola.aps.org/)

PROLA is the online version of *Physical Review* (one of the few physics journals of whose older copies are still in active research use), *Physical Review Letters* and *Reviews of Modern Physics*. *PROLA* was launched online at the end of 1998, with *Physical Review* issues from 1985 to 1996. It has since grown to include the complete file of APS journals published from 1893 to the present, which comprise over 325,000 articles.

Technically, *PROLA* refers to the archival file, which currently includes 1893–2001, and is priced separately from the current file (although both are accessed from the *PROLA* interface). Each full record in the *PROLA* database will have a 'Show Articles Citing This One' link. While full-text searching and displaying citations (including abstracts) is free, subscriptions are required for display of full text and citing articles.

Citation Bridge (US Patents) (http://www.patentcitations.com/)

Following free registration and login, *Citation Bridge* searches and displays both forward and backward links from a given US Patent Number. For example, see Figure 1.

From the display screen, clicking on any hyperlinked patent number resets the display to that patent. Clicking on the patent title displays its citation/abstract data.

US Patent and Trademark Office (http://www.uspto.gov/patft/index.html)

Searching USPTO web site and displaying the record provides hyperlinks to both cited and citing patents, albeit not as convenient as *Citation Bridge*. For example, see Figure 2.

Google Scholar (http://scholar.google.com/)

Google Scholar is a scholarly literature database that includes peer-reviewed papers, theses, books, preprints, abstracts and technical reports available from academic publishers, professional societies, preprint repositories and universities, as well as those available across the web.

Search results are displayed according to relevance based on the full text of each article, the article's author, the publication source and how often it has been cited. *Google Scholar* also automatically analyses and extracts citations and presents them as separate results, even if the documents they refer to are not online. This means your search results may include citations of older works and seminal articles that appear only in books or other offline publications. For example, a search for 'zzyzx' retrieved 34 items, with each article title linking to full text, if available, at the publisher website, a 'Cited by' link to citing article records in *Google*, and 'Web Search' which provides a *Google* web search on the first author's last name.

Optics InfoBase (http://www.opticsinfobase.org)

Optics InfoBase is a database of optics research based on the content of the following journals: *Applied Optics* (1962+), *J. Opt. Soc*. (1917–1940), *J. Opt. Soc. A&B* (1984+), *Opt. Lett*. (1977+), *Opt. Express* (1997+), *J. Lightwave Technol*. (1998+), *J. Opt. Networking* (2001+), *J. Opt. Tech-*

Citation Search

Click on a patent number to make it the center of the diagram.
Click on a title to view the details of the patent.

New Search

Backward Citations		Patent	Forward Citations	
Re-Center Diagram	View Patent Details	6280986	Re-Center Diagram	View Patent Details
5482846	Ethanol production in Gram-positive microbes	Stabilization of pet operon plasmids and ethanol production in bacterial strains lacking lactate dehydrogenase and pyruvate formate lyase activities	6682912	L-glutamic acid-producing bacterium and method for producing L-glutamic L-glutamic acid-producing bacterium and method for producing L-glutamic
5424202	Ethanol production by recombinant hosts			
5000000	Ethanol production by *Escherichia coli* strains co-expressing Zymomonas PDC and ADH genes			

New Search

Figure 1.

Inventors:	**Hespell; Robert B.** (late of Peoria, IL); **Wyckoff; Herbert A.** (Roscoe, IL); **Dien; Bruce S.** (Peoria, IL); **Bothast; Rodney J.** (East Peoria, IL)
Assignee:	**The United States of America as represented by the Secretary of Agriculture** ()
Appl. No.:	**201449**
Filed:	**November 30, 1998**
Current U.S. Class:	**435/161**; 435/252.3; 435/252.33; 435/320.1
Intern'l Class:	C12P 007/06; C12N 001/21; C12N 015/00
Field of Search:	435/243,248,252,252.3,252.33,69.1,190,161,320.1,471,476,477 562/479

References Cited [Referenced By]

U.S. Patent Documents

5000000	Mar., 1991	Ingram *et al.*	435/161.
5424202	Jun., 1995	Ingram *et al.*	435/161.
5482846	Jan., 1996	Ingram *et al.*	

Figure 2.

nol. (1999–2004), *Opt. Photonics News* (2002+). Planned additions in 2005 include: *J. Opt. Soc*. (1940-1984), *Opt. Spectrosc*., conference proceedings, and *Trends in Optics & Photonics Series* (TOPS).

The search interface allows title and/or keyword search terms and/or author names and/or OCIS Keywords (subject categories) and an option to limit to specific journals. 'Forward References' to other articles in Optics Infobase are listed on the 'full record' screen.

CiteSeer (http://citeseer.ist.psu.edu/)

CiteSeer is a digital library and search engine for the literature of computer and information science. *CiteSeer* was developed at the NEC Research Institute and is currently hosted at Penn State's School of Information Sciences and Technology. *CiteSeer* indexes PostScript and PDF research articles on the Web, and provides the following features.

- Autonomous Citation Indexing (ACI) computes citation statistics and related documents for all articles cited in the database, not just the indexed articles.
- *CiteSeer* allows browsing the database using citation links.
- It can show the context of citations to a given paper.
- It provides automatic notification of new citations to given papers, and new papers matching a user profile.

- It locates related documents using citation and word based measures and displays an active and continuously updated bibliography for each document.
- It shows the percentage of matching sentences between documents.
- It indexes the full-text of the entire articles and citations. Full Boolean, phrase and proximity search is supported.
- It provides the context of how query terms are used in articles.
- It analyses the graph of citations (hubs and authorities ranking).
- It allows quick and easy viewing of page images.
- It uses search engines and crawling plus document submissions to harvest papers on the Web and automatically extracts and provides metadata from all indexed articles.

Science Direct (http://www.sciencedirect.com/science/journals)

Science Direct freely provides both contents and abstracts for Elsevier titles and hosted journals from other publishers. Full text retrieval, however, is limited to subscribers. Article references are displayed from journal tables of contents and the full 'abstract' display includes a 'cited by' link to other *Science Direct* journal articles. For example:

Deformation of a chromatographic bed during steady-state liquid flow. *AIChE Journal*, Volume 44, Issue 1, January 1998, pages 2–12, Karin C. E. Östergren

was cited 6 times in *Science Direct*, 10 times in *SciFinder Scholar* and 11 times in *WoS* (including 1 in-press citation). All 6 cites in *SD* were included in *SFS* and *WoS*, both of which also indexed 4 articles from *Biotechnology Progress*, with *WoS* including an additional cite from *Numerical Heat Transfer*. The primary difficulty in depending on a commercial publisher's journal database for cited references, is that virtually none of the citations from society published journals will be found.

PsycINFO (http://www.apa.org/psycinfo/)

PsycINFO is a psychological literature database containing over 2 million references/abstracts from journal articles, books, book chapters, technical reports, and dissertations. Search by topic, author, journal title, with topic indexing based on the Thesaurus of Psychological Index Terms. *PsycINFO* records contain: Links to full-text articles with Digital Object Identifiers (DOI); links to references cited in journal articles, books, and chapters. Comprehensive coverage of cited references began in 2001, along with some records for earlier years. Links to 'Cited References' and 'Times cited in this Database' are provided for each record.

IEEE Xplore (http://ieeexplore.ieee.org/Xplore/dynhome.jsp)

IEEE Xplore provides full text access to electrical engineering, computer science, and electronics literature. *IEEE Xplore* contains over 1,000,000 full text documents from IEEE journals, transactions, magazines, letters, conference proceedings, standards, as well as from IEE (Institution of Electrical Engineers) publications. Content is generally limited to 1988+, but includes selected material from journals and conference proceedings back to the early 1950s. Each 'Abstracts Plus' record in *IEEE Xplore* includes the abstract and, if available, index terms, references and citing documents.

The 'citing documents' is a feature in Release 2.0 and is limited to documents indexed in *IEEE Xplore*.

SPIRES HEP (http://www.slac.stanford.edu/spires/)

The *SPIRES HEP* (High Energy Physics) literature database was the first website in the United States, and the first database to be accessible over the web. *SPIRES HEP* is a joint project of SLAC, DESY and FNAL as well as the worldwide HEP community. While *WoS* will generally have additional citing references, SPIRES will also uniquely include citing conference papers.

Each *Spires HEP* bibliographic record provides links to references, keywords, cited by references, CERN Library record, publisher's full text, SLAC document server, etc.

Institute of Physics (IOP) (http://www.ioppublishing.com/)

In September 2003, the IOP announced that papers in their electronic journals had been enhanced by links to citing articles from The American Physical Society and NASA's Astrophysical Data System. This service has been further enhanced by implementation of 'cited-by' links using CrossRef's Forward Linking service for papers published in the last 10 years. The IOP will be providing this feature for their journal archive, back to 1874.

CrossRef (http://www.crossref.org)

CrossRef is a multi-member publisher trade association that is currently offering full text searching of journal articles, conference proceedings, etc., from nine of its members. Rather than holding full text content, *CrossRef* provides linkages through Digital Object Identifiers (DOI) tagged to publisher supplied article metadata.

CrossRef Search uses *Google* to supply the search technologies, while *CrossRef* supplies links to publisher

websites. *CrossRef* search is only available from individual publisher websites, e.g., Institute of Physics [http://www.iop.org/EJ/search_crossref] but accesses content from all participating publishers in this pilot project.

CrossRef also provides a Forward Linking service that allows member publishers to provide 'articles citing this article' at the item level. The IOP is providing this data for the most recent 10 years of their online content.

Eugene Garfield's recognition that the concept of citation searching, from the legal literature, should be applied to the sciences has been amply rewarded. Citation searching has surely seen a dramatic progression from the cumbersome print volumes of the *Science Citation Index*. Recent developments in hypertext linking and web browsers have led to the *WoS*, which results in the relationship between citing and cited documents being clearly evident to users.

Recent developments of 'competitors' to the *WoS*, while interesting and useful for quick links to some citing references, are clearly not a substitute for a comprehensive citation search. *WoS* currently indexes ~ 8000 journals from the sciences, engineering, social sciences and the humanities, and clearly remains the primary resource for citation searching.

1. Garfield, E., Citation indexes for science – New dimension in documentation through association of ideas. *Science*, 1955, **122**, 108–111.
2. Garfield, E., Citation indexes for science – reply. *Science*, 1956, **123**, 62.
3. Schoenbach, U. H., Citation indexes for science. *Science*, 1956, **123**, 61–62.
4. Atkins, H., The ISI web of science – Links and electronic journals; How links work today in the Web of Science, and the challenges posed by electronic journals. *D-LIB Mag.*, 1999, **5**, <http://www.dlib.org/dlib/september99/atkins/09atkins.html>

Quotations by Eugene Garfield

'The ability to search back and forth in time from the past literature to the literature, to identify cross-disciplinary developments, to eliminate the search restrictions and complexity imposed by semantic problems, and to provide an in depth index to the literature within a practical time and cost framework have all proved to be as significant in practice as they appear to be in theory.'

— Garfield
In *Citation Indexing – Its Theory and Application in Science, Technology and Humanities*
John Wiley, New York, 1979

'It is true, of course, that citation counts will not identify significance that is unrecognized by the scientific community. They are, after all, nothing more, nor less, than a reflection of that community's work and interest. To go beyond that is to begin questioning the validity of the community's perception of things, which is another area that calls for peer judgements.'

— Garfield
In *Citation Indexing – Its Theory and Application in Science, Technology and Humanities*
John Wiley, New York, 1979

'What you have to learn to deal with in American business is how to deal with mediocrity. You don't hire geniuses and keep them all very long. Even if you give them equity, which they had. I probably made the mistake of not giving other people equity in the company in the early days. But they had it. People want to remain comfortable very quickly in the game. They don't want to go on and take the risks. I guess there are times when people think that I am too much of a crapshooter. Sometimes even I agree. [laughter]'.

The Beckman Center for the History of Chemistry Eugene Garfield Oral History
Transcript of an Interview conducted by Arnold Thackray and Jeffrey L. Sturchio
at ISI, on 16 November 1987