



PROLA, SPIRES-HEP, AIP Scitation

All three databases I'm reviewing in this issue are from the field of physics (and in the last issue, one of my picks was the Astrophysics Data System). Is this some neophyte enthusiasm of mine? No, physics has never been my strong point. I hardly understand more than the prepositions and the articles (not journal articles, but the definite and indefinite articles) of physics papers. However, librarians and information professionals excel at searching and finding, even when the topic or the entire discipline is unfamiliar.

I chose PROLA and SPIRES-HEP as my picks because both show excellent examples of the smartest use of the citation data to rank results (in addition to traditional other sorting criteria) that leads even nonsubscribers to the most cited papers on a topic in physics. The Physical Review On-Line Archive (PROLA) of the American Physical Society (APS) does this for its own journals, and SPIRES-HEP, the High-Energy Physics Literature Database of the Stanford Linear Accelerator Center, does this primarily with its traditional bibliographic indexing database. I chose as my pan the Scitation database from the American Institute of Physics (which was my pick several years ago under its old name, OJPS) because it has neglected its primary role of making searchable the full text of its worthy journals and conference proceedings. Instead, it spends more energy on being an aggregator and digital facilitator of publishers of physics literature. The latter is a useful role, but others could do that, while only AIP can help users fully discover the content of all the AIP publications.



I chose PROLA and SPIRES-HEP as my picks because both show excellent examples of the smartest use of the citation data to rank results. ...



the picks

PROLA

The original *Physical Review* journal was launched at Cornell University, and until recently, I used Cornell's mirror site of the digital archive of the dozen publications of the American Physical Society. Well, the publisher revamped, and its continuously enhanced PROLA site (<http://prola.aps.org>) has become a better option. With the novel features introduced in the past few months, it sets a model for how to bring out the best of a state-of-the-art digital archive. APS is a delight to look at and work with. It has nearly half a million papers that are full-text searchable even by nonsubscribers, who also can see the bibliographic records and the abstracts for free.



True, these free options are becoming more and more common among scholarly journal publishers, but PROLA shows a model for the future.

True, these free options are becoming more and more common among scholarly journal publishers, but PROLA shows a model for the future. It uses Autonomy's Verity software, but that alone would not matter, as evidenced by Kluwer Academic's miserable search results before its content was acquired by and integrated into Springer's digital collection. It features a good interface and includes information about the citedness of papers (mashing in CrossRef data). PROLA does this, and more and better.

PROLA can sort results chronologically and by relevance, as most good systems can. But it also adds the option to sort by citedness, which is a rarity, even though it often works better than the questionable relevance sort options of most software that can come up with very different orders for the very same result set when sorting by a nondisclosed relevance ranking algorithm. This becomes obvious when relevance ranking is applied to identical result lists in various implementations of the same databases, such as MEDLINE, ERIC, and PsycINFO, on different platforms. Ranking by citedness is transparent and often makes it clear which are the most cited, "must-read" items on a topic. This could only be better if the relative citedness count (the absolute number of citations received divided by the age of publication) could be also used as a sort criterion.



Result list sorted by citedness with on-the-side cluster showing distribution by year of publications, journal title, and type of documents for one-click refinement

PROLA automatically searches not only for singular and plural versions of search terms but also for other morphological variants. For example, searching the cosmological term `eternal inflation` also brings up the phrase `eternally inflating universe`, but this does not extend to irregular plurals—`vortex` does not retrieve `vortices`. APS has its own efficient, predictable system of linking to its articles, and it also uses the DOI linking. Compare this to the huge number of unpredictable, web-hostile DOIs of Wiley with unsafe characters in the extended ASCII code. PROLA's keyword-in-context display format shows the context of the matching parts of the full text. Displaying the results list with or without the abstract is just a flip-flop click, and there is really no need for various output formats. Exporting of bibliographic records with abstracts is possible by displaying and adding records on an item-by-item level in BibTeX and EndNote formats, but marking items in the result lists and exporting all the marked ones in one fell swoop would be much better. This smart and swift software brings the best out of the much respected content from all the issues of all the ASP journals.

SPIRES-HEP

The Stanford Physics Information Retrieval System (SPIRES) was one of the pioneers of the online information systems decades ago. Then, at the end of 1991, SPIRES-HEP (www.slac.stanford.edu/spires/hep) became the first publicly accessible web server in the U.S. Beyond the physicists at the Stanford Linear Acceleration Center, the (since retired) librarian, Louise Addis, was the key person to launch this excellent eprints service, which remains one of the core resources for physicists in general and high energy physicists in particular. I mention Louise Addis by name because her role in my eyes is of the same importance as Henriette Avram's development of the MARC record format was in fostering library automation.

SPIRES-HEP is a huge bibliographic database, but it is not your grandfather's Oldsmobile. That's what makes it a long-time survivor of the traditional bibliographic databases, many of which seem to be on the endangered species list. SPIRES-HEP is such an important tool because beyond including the traditional content elements, it includes links to the full-text eprint (earlier referred to as preprint) versions of papers submitted for publications in journals and/or conference proceedings deposited in arXiv, euclides, and other preprint servers. This means usually at least half a year of lead time for researchers who want to know not just what was published on a specific topic but what is likely to be published.

It shows the shrewdness of SPIRES-HEP current developers that they were among the first to display the citedness of the papers, which is a very important clue when selecting the most promising items from the result list of 257 records, as happened in my case when I searched for `rotating black holes`. The results list can be

HEP :: HEPNAMES :: INSTITUTIONS :: ... EXPERIMENTS :: JOBS :: V

FIND K "ROTATING BLACK HOLES" Paper 1 to 25 of 257

Browse Author | Format: Standard (incl. cites) | Sort: Approx. Citation Count

Display again

Standard (incl. cites)
Citesummary
LaTeX(US)
LaTeX(EU)
Harvmac
BibTeX
CV format
RSS

1) Rotating black holes: synchrotron radiation. frames, energy extraction, and scalars
James M. Bardeen, William H. Press, Saul A. Teukolsky. OAP-288, 1972.
Published in *Astrophys.J.*178:347,1972.

TOPCITE = 100+ ←
LaTeX(US) | LaTeX(EU) | Harvmac | BibTeX | Cited 246 times
Journal Server
ADS Abstract Service ←
Bookmarkable link to this information

2) Rotating black holes - separable wave equations for gravitational and electromagnetic perturbations.
S.A. Teukolsky (Caltech, Kellogg Lab). OAP-291, 1972.
Published in *Phys.Rev.Lett.*29:1114-1118,1972.

TOPCITE = 100+ ←
LaTeX(US) | LaTeX(EU) | Harvmac | BibTeX | Keywords | Cited 181 times
Journal Server
ADS Abstract Service ←
Phys. Rev. Lett. Server
Bookmarkable link to this information

Innovative result list with important clues and links

FIND K "ROTATING BLACK HOLES" Browse Author | Format: Citesummary | Sort: Approx. Citation Count

Display again

Generated on 05/25/08 (mm/dd/yy)
Your result was: 257 [eligible papers = 247 (published or arXiv E-prints)]

Breakdown of search results by citation

	All papers	Published only
Renowned papers(500+ cites):	0	0
Famous papers (250-499 cites):	0	0
Very well-known papers (100-249):	7	7
Well-known papers (50-99):	14	14
Known papers (10-49):	98	93
Less known papers (1-9):	96	67
Unknown papers (0):	32	14
Total eligible papers analyzed:	247	195
Total number of citations:	4755	4595
Average citations per paper:	19	24

The citation summary provides a superb at-a-glance report card about the result set.

SAO/NASA ADS Astronomy Abstract Service

- [Find Similar Abstracts \(with default settings below\)](#)
- [Full Refereed Journal Article \(PDF/Postscript\)](#)
- [Full Refereed Scanned Article \(GIF\)](#)
- [References in the article](#)
- [Citations to the Article \(388\) \(Citation History\)](#)
- [Refereed Citations to the Article](#)
- [SIMBAD Objects \(1\)](#)
- [Also-Read Articles \(Reads History\)](#)
- [HEP/Spires Information](#)

• [Translate This Page](#)

Title: Rotating Black Holes: Locally Nonrotating
Authors: Bardeen, James M.; Press, William H.; Teukolsky, Saul A.
Publication: Astrophysical Journal, Vol. 178, pp. 347-372
Publication Date: 12/1972
Origin: ADS
Comment: A&AA ID. AAA008.066.088

Additional clues about the importance of the paper in the ADS service

sorted by citedness, and the items are labeled to cluster them into various citedness-level categories. The links take the users to the official version of the paper at the publisher site or, if the library does not have a subscription to the digital edition, to digital collections with preprint copies in a variety of formats for free.

In this example there is such a link to the ADS service (which was my pick for a good reason in the last issue),

which regales the searchers not only with links to the full-text versions but also with details about the citedness of the paper by sources covered by ADS, distinguishing between different types of citing sources.

SPIRES-HEP also offers a citation summary, which provides a perfect visual clue to see how the 257 papers on my topic are distributed by citedness, reflecting the awareness and recognition of the papers in the set. This at-a-glance “report card” further facilitates efficient cherry-picking by the searcher. There are other sorting criteria as well.

SPIRES-HEP provides a highly rewarding tool for finding the potentially most important papers on a topic. It is also perfect for checking the productivity and citedness of researchers, journals, and institutions and for determining their h-index. The cited and citing references also have their links to the official toll-access and/or the open access eprint versions of the papers. The SPIRES-HEP team is conservative in counting citations and treating documents that may not have been refereed, such as conference papers, in different ways from the ones published (and implicitly peer reviewed). That is one of the reasons that the citedness count may be less than in other repositories or journal archives. The designers of SPIRES provide objective arguments for this decision in a background file, and they warn users to take citation counts with a grain of salt. SPIRES-HEP perfectly illustrates the power of linking and citation searching—two of the dominant new features in online searching.



the pan

AIP SCITATION

Many years ago AIP Scitation (<http://scitation.aip.org>) was among my picks, so my current judgment on it may be surprising. I am disappointed because AIP never moved beyond making the metadata elements of its own journals and conference proceedings searchable. I think it shot itself in the foot by not offering this increasingly important feature—it keeps hidden many of its relevant papers on subjects that may not have the search terms in the title or the abstracts. It seems to have been too busy playing the role of aggregator and digital facilitator for the digital collections of other publishers in physics and beyond physics, thus stretching itself too thin.

It is quite telling how few hits it finds for the exact phrase query on **rotating black holes**, where 16 of its 17 journals came up empty-handed. (The AIP Conference Proceedings subset had 15 hits). The only journal that yielded hits (three of them) was the *Journal of Mathematical Physics*. Searching for the plural format (which is not done automatically in this archive) increased the number of hits to five from this journal.

Actually, there are at least 10 additional papers in that journal alone that include the query term. The mere presence of rotational black holes would not necessarily mean that the article is about that topic, but the first one that I found (“Static Bondi Energy”) through the excellent arXiv repository proves the validity of the expression presented in the paper by saying that it was successfully applied to rotating black holes. It even cites another paper by the author about gravitational energy of rotating black holes.

The argument that limiting the search to metadata would limit the retrieval of less relevant items is not convincing, and a relevance-based sorting (which is available in the

archive along with chronological sorting) could or should take care about this—if it is a good one—by ranking higher items in which the query term occurs more frequently and/or more prominently such as in the title and/or abstract. There is no reason not to index the full text these days, when elementary school students have PCs with fast processors and 250 Gigabyte hard drives in their backpacks.

I am even more concerned that this limited indexing policy also applies to other databases of other publishers hosted by Scitation. Take as an example the APS journals searchable through AIP Scitation. The exact phrase query **Euclidean instanton** finds only seven items through Scitation, six from *Physical Review D*, and one from *Physical Review Letters* when searching the full bibliographic records, the most comprehensive option.

When doing the search in PROLA directly, the same query in the same index finds 11 hits (10 from *Physical Review D* and one from *Physical Review Letters*) and 95 when searching the full-text index. This is absurd.

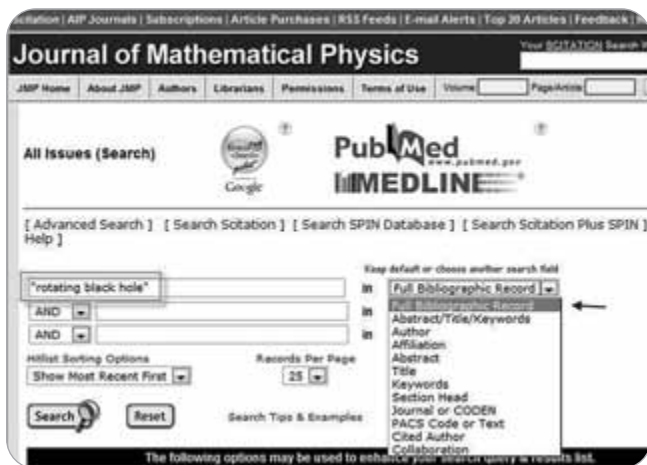
It adds insult to injury that for reasons unknown to me the current year and the previous 3 years of APS journals are available for browsing and displaying only through AIP but not APS. Luckily, they still can be searched through PROLA, but users are then forwarded to the Scitation database. In order to avoid flip-flopping between the APS and AIP archive, searchers may stay at the AIP site and not realize how much they are shortchanged in discovering pertinent papers using the limited search engine of AIP. It is not lost on me that the word citation is part of the name of the AIP service, but it offers none of the smart features that I mentioned for PROLA and SPIRES.

I hope that Scitation can make it again to the picks, but in order to do so, it must make the full text fully searchable (at least of its journals), incorporate its “mirror site” of arXiv in Scitation to make users aware of the availability of open access version of papers, and make some powerful option available for citation-based searching.

In an interesting event just as I was finishing this manuscript, Microsoft announced it will close down its Live Book and Live Academic sites. The latter was my pan earlier. When I saw its absurdly incompetent and irresponsible efforts to add citedness counts to its result list, I made that system the subject of my April review in the Péter’s Digital Reference Shelf column (www.gale.cengage.com/reference/peter), hoping that it was only a bad April Fool’s Day joke. It was not, but luckily it will get axed. It is a shame that Microsoft—with one of the best-known corporate names in the world—was incubating such an inferior service for 2 years and then came out with a primitive, reckless service. PROLA and SPIRES show good practices for academic databases.

Péter Jacsó (jacso@hawaii.edu) is professor of library & information science at the University of Hawaii’s department of information and computer sciences.

Comments? Send email to the editor (marydee@xmission.com).



There’s no option for searching the full text—the full records option means full bibliographic records.