

# Péter's picks & pans



Péter Jacsó  
University of Hawaii

ACM Digital Library

IEEE Digital Library

*The results of a search in ACM Digital Library on digital libraries displays items of great relevance, from a variety of types of sources, even though the relevancy rates determined by the program are low.*

The digital libraries of ACM and IEEE Computer Society, two of the most respected associations in the field of computer and information science, offer free access for anyone to bibliographic information (often with substantial abstracts) about conference papers, as well as articles published in journals of high-impact factor and worldwide prestige. The full versions of the documents are available at very reasonable prices, and are free to subscribers of the Digital Library of ACM and IEEE, respectively.

The two archives represent the perfect model for fully or partially substituting expensive searches in traditional online databases. The reliable, consistent, and predictable coverage of sources by these digital libraries is also a big asset. These two powerhouses are goldmines for information science-related research.



## the picks

### ACM DIGITAL LIBRARY

Its name, the Association for Computing Machinery, has always seemed extremely archaic to me, but many of its publications have been avant-garde—as

is its Digital Library (<http://www.acm.org>). This offers free bibliographic citations to more than 60,000 articles, conference papers, and reviews that appeared in more than 20 ACM publications. About 82% of the items have exhaustive subject index terms, and 30% have abstracts. For more than 90%, there is a PDF version of the document. Most of these journals are among the top 15 titles in terms of impact factor, in their respective sub-categories, according to the *ISI Journal Citation Reports*. The proceedings of ACM conferences about Information

No.	Article	Score
1)	35 in digital libraries: Nabil R. Adam, Vijayalakshmi Atluri and Igg Adiwijaya; <i>Commun. ACM</i> 43, 6 (Jun. 2000), Pages 64 - 72 [ <a href="#">Find Related Articles</a> ]	11
2)	Personal computing: from P-books to E-books; Larry Press; <i>Commun. ACM</i> 43, 5 (May, 2000), Page 17 [ <a href="#">Find Related Articles</a> ]	10
3)	CoRR: a computing research repository; Joseph Y. Halpern; <i>ACM J. Comput. Doc.</i> 24, 2 (May, 2000), Pages 41 - 48 [ <a href="#">Find Related Articles</a> ]	10
4)	Issues of online research repositories from the perspective of the biomedical sciences; David L. Ambruster; <i>ACM J. Comput. Doc.</i> 24, 2 (May, 2000), Pages 49 - 53 [ <a href="#">Find Related Articles</a> ]	10
5)	A response to the commentaries on CoRR; Joseph Y. Halpern; <i>ACM J. Comput. Doc.</i> 24, 2 (May, 2000), Pages 72 - 77 [ <a href="#">Find Related Articles</a> ]	10
6)	A computing research repository: why not solve the problems first? A. J. Van Loon; <i>ACM J. Comput. Doc.</i> 24, 2 (May, 2000), Pages 54 - 71 [ <a href="#">Find Related Articles</a> ]	10
7)	The dilemma of erudition vs. speed; James Frekeages; <i>ACM J. Comput. Doc.</i> 24, 2 (May, 2000), Pages 60 - 63 [ <a href="#">Find Related Articles</a> ]	10

ACM Digital Library  
 International Conference on Digital Libraries  
 Proceedings of the fifth ACM conference on ACM 2000 digital libraries  
 June 2 - 7, 2000, San Antonio, TX USA

Knowledge-based metadata extraction from PostScript files  
 Pages 77-84  
 Giovanni Giffirda, Eddie C. Shek and Jihoon Yang

Full text: PDF 386 KB  
 [ Discuss this Article ] [ Find Related Articles ] [ Add to Binder ]

**ABSTRACT**

The automatic document metadata extraction process is an important task in a world where thousands of documents are just one "click" away. Thus, powerful indices are necessary to support effective retrieval. The upcoming XML standard represents an important step in this direction as its *semistructured* representation conveys document metadata together with the text of the document. For example, retrieval of scientific papers by authors or affiliations would be a straightforward task if papers were stored in XML. Unfortunately, today, the largest majority of documents on the web are available in forms that do not carry additional semantics. Converting existing documents to a semistructured representation is time-consuming and no automatic process can be easily applied. In this paper we discuss a system, based on a novel

Abstracts for journal articles and conference proceedings in ACM Digital Library are very substantial and informative.

the superior INSPEC database, which is the best fee-based source for such conference papers, and—to a lesser extent—Compendex. However, INSPEC and Compendex databases have a price tag of around \$6.50

Knowledge Management, Web Information, and Data Management are prestigious in the information science field. Many of the computer science publications from ACM have articles and papers that are highly relevant for information scientists and information technology specialists.

For example, a phrase search for digital libraries in the title, abstract, and index fields found 54 records from the *ACM Transactions on Database Systems*, *Communications of the ACM*, and *ACM Computing Surveys*. While the relevance ranking algorithm assigned low scores to the items, the result list in its entirety was closer to 100% from my perspective. When extending the search to conference papers, the result list increased to 240 items. Once again, mostly items of high and very high relevance were retrieved from dozens of conferences sponsored by ACM.

When items have abstracts, they tend to be substantial and informative. Some records even include a full-text review of the article from *ACM Computing Reviews*. Conference papers are particularly important because proceedings of information science conferences sponsored by the ACM Special Interest Groups, such as SIGMOD and SIGIR or the Information Knowledge Management conference, are not abstracted or indexed by traditional commercial databases with information science and technology coverage (LISA, H.W. Wilson's files, or ISI), or are covered sporadically and unpredictably (Information Science Abstracts). The notable exceptions are

per DialUnit, and more than \$2 for the bibliographic record with index terms and abstract. The ACM Digital Library charges only for the full page-image document, and a very reasonable fee at that—\$10. If you have a subscription for a yearly fee of \$185, then everything is free and instantaneous. Journals are typically covered from their first volume, even if it means a retrospective conversion from 1954 onwards as is the case with the *Journal of the ACM*. Only two journals are not covered from their first issue, *Communications of the ACM* (starts with volume 27) and *Intelligence* (starts with volume 10).

The software is very intuitive, and at the same time offers powerful search features, like fuzzy searching, Soundex searching, and proximity operations. You can limit the search to one or more of the several data elements that make up an item, as well as to document categories (journal articles or conference papers) and specific journals. It has two output formats for the result list. The layout of the full bibliographic record could

*IEEE Computer Science Digital Library features the powerful Autonomy software that allows both natural language and Boolean searches, including exact phrase searching.*

be made tighter, and term highlighting would be useful. The result list can be sorted by publication name, date, or score. It would help in picking the most promising items from the result list if they could be marked while browsing the list.

IE 5.0 users are warned that there is a problem with login, but by the time this review is published it may be fixed. I had no problem with the login process using IE 5.5. The whole search procedure is easy and usually very rewarding. This is a top-notch database with very good software and superb content at an excellent price, that has highly relevant results for researchers in information science.

### IEEE COMPUTER SOCIETY DIGITAL LIBRARY

The Computer Society of the Institute of Electrical and Electronics Engineers publishes 12 magazines, 10 of which are covered by the *Science Edition of the Journal Citation Reports of ISI*. Of the nine transactions that the Computer Society publishes, eight are covered by ISI. Many of these are relevant not only for computer specialists, but also for information scientists and educated practitioners. The JCR does not cover conference proceedings so I can't quote impact factors, but the numerous conference proceedings of the Computer Society are widely respected.

Bibliographic citations and abstracts for all these publications are available free of charge in the Digital Library of the Computer Society (<http://www.computer.org>). It has more than 32,000 items from all of its journals/

HOME INDEX SEARCH HELP CONTACT CART

computer.org

**Main Site and Digital Library Search**

Search for:

Search in:  Full text in the Digital Library  Abstracts, Titles, and Authors  Free content

Use natural language, BOOLEAN, or "exact phrase".

Search the main Web site and all publications, or select individual areas below:

**Magazines:**

- Annals of the History of Computing
- Computing in Science & Engineering
- Computer magazine
- Computer Graphics & Applications
- Concurrency
- Design & Test of Computers
- Intelligent Systems
- Internet Computing
- IT Professional
- Micro
- MultiMedia
- Software

**Site areas:**

- Conference Proceedings
- CS Store
- Main Web site (non-publication areas)

**IEEE Transactions on:**

- Computers
- Knowledge & Data Engineering
- Parallel & Distributed Systems
- Pattern Analysis & Machine Intelligence
- Software Engineering
- Visualization & Computer Graphics

# Power Searching with the Pros

- Intensely practical...
- Top experts...
- Concentrated instruction...
- Up-to-the-minute reports...
- No marketing hype...

September 9-11, 2001  
Hyatt Regency/Reston, VA

# websearch

UNIVERSITY

For information & registration: 800/248-8466 Ext. 524

magazines, from six of the nine transactions, and from 500 of its proceedings. It features the powerful Autonomy software that allows both natural language and Boolean searches, including exact phrase searching so you can distinguish between management information systems and information management systems. The only weak point of the software is that you can't mark items on the result list as you browse it, the same limitation I mentioned for the ACM Digital Library.

You may select all of the publications for searching, just the conference proceedings, or just specific titles by checking the box next to the titles. You can limit the search to the abstract, title, and author index, or let it run against the index of the full text of the documents. The full documents are available in HTML or PDF formats (often in both). Many conference papers are available for free in HTML and/or PDF format for any users. The charge for the rest of the documents is a very reasonable \$10.

The Digital Library is a gold mine on such hot information science research topics as artificial intelligence,

expert systems, information visualization, data mining, knowledge discovery, human-computer interface, informatics, and natural language processing—to name just a few. The search for *artificial intelligence* as a phrase retrieved 374 items from such prestigious sources as the *Transactions on Knowledge & Data Engineering*, *Transactions on Visualization & Computer Graphics*, *IEEE Computer*, *Intelligent Systems*, and several conference proceedings. The search for *data mining* as a phrase just in the title and abstract field yielded 371 items, about one-fourth from journals and transactions, and the rest from conference proceedings. Free versions of the full document were available for many conference papers, even though the result list identified only the abstracts as free.

It is an added bonus that at the IEEE site you can search for additional items from other IEEE publications using the Inktomi search engine. IEEE publishes well over a hundred journals and transactions, and a large number of conference proceedings. About 80 of the journals

are covered by the *Science Edition of the Journal Citation Reports of ISI*, and many of them are top-ranked periodicals. The latest edition of JCR, for example, ranks the IEEE Transactions on Information Theory as #4 by impact factor, and proves that it is by far the most cited #1 periodical in the computer and information system category of ISI, in the company of 62 other prestigious scholarly publications.

True, you have more sophisticated search options on Dialog, if you find the really appropriate databases. Two databases claim explicitly in their BlueSheets to cover the subject of information theory: NTIS and ISA. The former, however, does not cover transactions or journals. ISA has 2,032 records from the most cited periodical in information theory, but it does not have a single record for the current year, only one for 2000, and four for 1999. This is a drastic decline from the 200-item average of the years before 1999, especially when you consider the prominent mention of the topic of information theory on the ISA bluesheet, and the clout of the IEEE Transactions on Information Theory. While you figure this out, you have spent more money at the \$6.40 DialUnit and \$1.30 display rate of this database than it would cost to receive a relevant and current full document instantly from the IEEE site. INSPEC, Compendex, and PASCAL do justice to this topic and source, but not even they can compete with the free abstracts of the Computer Society's Digital Library and don't offer the free or reasonably priced document delivery.

Information scientists and researchers can benefit greatly from the Digital Library that is enhanced by the searchable archive of many other IEEE serials. They will find very reasonable the subscription fee (\$50 if you are a member) or the \$10 for instant, pay-as-you-go document delivery for those who don't subscribe to the Digital Library.

---

*Péter Jacsó* ([jacsop@hawaii.edu](mailto:jacsop@hawaii.edu)) is associate professor of Library & Information Science at the University of Hawaii's Department of Information and Computer Sciences.

Comments? Email letters to the editor to [marydee@xmission.com](mailto:marydee@xmission.com).