

péter's picks & pans



Péter Jacsó
University of Hawaii

In spite of the obvious subjectivity, this very-well-implemented system with its free and very reasonably priced offerings is clearly a pick.

ITI InfoCentral, Clusty, ArticleFinder 2.0

My first pick requires a disclosure: ITI InfoCentral Digital Archive derives all its content from Information Today, Inc. (ITI), which in turn owns the periodicals that have published most of my columns and feature articles for the past 15 years. In spite of the obvious subjectivity, this very-well-implemented system with its free and very reasonably priced offerings is clearly a pick. The other pick is the new Clusty system from Vivísimo, showcasing its popular clustering and metasearching technology. The pan is Infotrieve's ArticleFinder 2.0 service, which abandoned its free search service. In addition to a monthly subscription fee, it still has unreasonably high document delivery charges and often levies unrealistic royalty fees.



the picks

ITI INFOCENTRAL DIGITAL ARCHIVE

I am delighted with the launch of ITI InfoCentral [www.itiinfocentral.com or <http://pqasb.pqarchiver.com/infotoday/search.html>], which I learned about from a ProQuest press release just as this column was due. (It shows that I am really not in the loop when it comes to ITI matters.) I have been bugging ITI for some time to make its large collection of library and information technology articles digitally available for both searching and direct delivery. After all, ITI publishes the leading periodicals for practicing information professionals, such as *Information Today*, *ONLINE* (ahem), *Searcher*, *Computers in Libraries*, and *DATABASE* (renamed *EContent* in 1999 with a re-launch and focus change in 2001). Most of these are already well-covered by a variety of full-text databases produced by Gale Group, ProQuest, and EBSCO, but these are subscription-based services for libraries. These publications are also included in several indexing and abstracting services, but getting the full-text or page-image version of the typically two- to five-page articles are expensive through document delivery services that typically charge about \$20 per transaction.

My wish is now fulfilled—with more than a little help from ProQuest, which makes the full documents searchable for free and charges only \$2.95 per item for immediate download. This is somewhat below the direct print delivery charges (that ITI has not increased for more than a decade) and is way below what it costs to get articles through Infotrieve. By the time you read this, I will

have an in-depth review about ITI InfoCentral in the open access archive of my database reviews, which is hosted by Gale [www.galegroup.com/free_resources/].

ITI InfoCentral covers 10 periodicals owned by ITI and will offer almost 38,000 documents by early 2005. Not all of them are available in full text, and the archive does not go back to the first volumes of all the journals. The archive starts in late 1987 for some titles and early 1988 for others but has pretty good retrospective coverage for a field focusing on information technology. I am somewhat dismayed that the DATABASE journal name has disappeared under the successor *EContent* title, even though for 20 years it was a respected journal name. If you're looking for DATABASE, search *EContent*—the articles are there. ITI's newsletters are not part of ITI InfoCentral.

The PQ Archive software is perfect for this collection. It has the power search features needed for full-text searching, such as proximity and positional operators; limiting by date range, author, one or more journal(s), and/or headline; and sorting the results by date or relevance.

The only additional search options I miss are two check boxes: one to limit the search to the abstract (usually the lead paragraph) and another one to search only documents available in full text. The short results list indicates the necessary bibliographic elements and the formats available. The content is very good and the pricing is excellent. The software has power without intimidating the user. It is no wonder that the PQ Archive software is so popular with many of the most widely read newspapers, including *The New York Times*, *The Wall Street Journal*, *The Washington Post*, *The Christian Science Monitor*, *The Los Angeles Times*, and *The Chicago Tribune*.

CLUSTY

I would not be surprised to hear that someone who had a British nanny in his or her childhood proposed the name of this lovely service. Clusty [http://clusty.com] is a metasearch engine from Vivisimo that is endowed with a smart clustering engine. (Or maybe it's the other way around.) It broadcasts your query to pre-defined groups of sites, such as Web search engines

(Gigablast, MSN, Lycos, LookSmart, Wisenut, Open Directory, and Overture), prominent news sites (Reuters, The New York Times, Yahoo! News, CNN, USA Today, and BBC News), image collections, blogs, and shopping sites. "So what?" you might say. "Other metasearch engines also do that."

What is unique about Clusty is its on-the-fly clustering of results into major groups (clusters). It is as if your incoming e-mail would be automatically clustered into labeled groups of related messages, based on content. Clustering is almost a genetic instinct, at least for librarians. We want to organ-

Advanced Search
Previous Results

Basic
Advanced

Search for:

Sort By: Relevance Most Recent First

Date Options: Last 2 Years All dates (September 1987 to present)

Date Range

From:

To:

Author: (optional)

Headline: (optional)

All

Computers in Libraries

Customer Relationship Management: CRM

EContent

ITI InfoCentral's intuitive search template with useful options

Results 1 to 10 of 20

Page: 1 2 Next >>

1.

Thoughts About Federated Searching

Péter Jacsó; Information Today; Oct 2004; Vol. 21, Iss. 9; pg. 17, 2
2.

WebFeat Introduces WebFeat 3

Anonymous; Information Today; Sep 2004; Vol. 21, Iss. 8; pg. 37, 2
3.

One-Stop Searching Bridges the Digital Divide

Roland Dietz; Information Today; Jul/Aug 2004; Vol. 21, Iss. 7; pg. S24
4.

Rules of Ruling the Road

Barbara Quint; Information Today; Jun 2004; Vol. 21, Iss. 6; pg. 7, 1

Results list from ITI InfoCentral

ize related materials (books, invoices, manuscripts, shoes, CD albums, or journal articles) so we can find them more easily. Anything can—and should—be clustered and sub-clustered, particularly if there are too many of them, which is often the case with results for a Web query.

Northern Light featured a similar sorting method using open access Web resources and a large bibliographic database fused from subsets of mega-databases. Vivísimo has shown clustering on its Web site on a number of databases since 2001 (although it limited clustering to 200 hits).

Clusty has its own site with many more groups and databases that users can customize by adding/removing any of them, i.e., adding Gigablast and removing Overture (although the sponsored sites will still appear on the top of the result lists even when they have little to do with the topic). The results are clustered into topics, with labels appearing on the side of the screen to also indicate how many items are within each (sub)cluster.

The process is not perfect (for example, LIS and Library and Information Science should not be two separate clusters), but it does give a good sense of the major facets of the hits. It is like signage in bookstores or supermarkets that steer you to the right shelves in the right aisle on the right floor, but it is much better for many reasons. It's as if the store attendant would fetch, say, all the XXL-sized clothing from the racks and pile them up in major clusters, such as shirts, slacks, and sweaters, then divide them into subclusters (short-sleeve, long-sleeve, dress, casual), affixing a label for each (sub)cluster. You can drill down in the clusters, or scroll down in the reasonable-sized hit list of the clusters, and peek into sites without leaving the hit list.

There are additional options to sort the results within clusters—all these help alleviate the increasing infoglut. This clustering technology really shines with consistently structured collections rich in metadata, such as journal article archives.



the pan

ARTICLEFINDER 2.0

ArticleFinder was originally a free indexing/abstracting (I/A) database of STM articles that acted as a gateway to the money-making document delivery services of Infotrieve [www.infotrieve.com]. Later, it became a fee-based I/A database, charging a flat-rate subscription of \$25 per month. In the long run, the introduction of this fee at a time when major STM publishers were starting to make their own I/A databases freely searchable may not have been a good idea. The largest digital facilitators such as High Wire Press, MetaPress, and Ingenta/CatchWord also made I/A records available free of charge. This may make one wonder about how a \$25-per-month or \$99-per-year subscription fee for searching an A/I database could possibly make it the best value, as Infotrieve declares.

ArticleFinder's login and FAQ pages claim the number of abstracts ranges from 6 to 8.5 million. Neither is that big a number. The free Scirus database has more than 5.2 million free abstracts with bibliographic records of articles published in Elsevier journals alone. It also has close to 4 million bibliographic citations with abstracts in the NASA/ADS database. CrossRef Search, which started with free abstracts from nine scholarly publishers (and had rather spotty coverage), had 29 publishers with much better coverage and several million free abstracts by the end of October 2004. Ingenta has several million free abstracts, and only a \$5 service fee for each online document on top of the royalty. This brings me to my biggest gripes about Infotrieve—the \$12 delivery charge per document and the often unrealistic copyright fees.

I wrote about this in the May 2002 issue of *Computers in Libraries* ("Digital Copyright and Copywrong"). (You should be able to find this on ITI Info Central.) Unfortunately, the problem that I wrote about in *CIL* persists in the new system, too. Infotrieve

The screenshot shows the Clusty website interface. At the top, there are navigation tabs: Web, News, Images, Shopping, Encyclopedia, Gossip, eBay, MyCombo, and Customize. Below the search bar, the query "digital librarianship" is entered. The search results are clustered into topics on the left side of the page:

- Science, Library and Information (21)
- Course (18)
- Books (15)
- Issues (17)
- Project (14)
- LIS (15)
- Education, Digital Libraries (9)
- Resume (10)
- Computers in Libraries (4)
- Presentations (5)

The main content area shows the top 171 results for the query "digital librarianship". The first result is "HP Digital Music" with a description: "Introducing the Apple iPod from HP. Ultraportable. Easy to use. Designed to give you a rich digital music experience. Welcome to HP Digital Music. www.hp.com". The second result is "Digital Items on eBay" with a description: "Visit eBay Electronics for 'digital' items. Discover great deals on new and used, name-brand merchandise. www.ebay.com". The third result is "Digital Librarianship: TheoWire" with a description: "Digital Librarianship resources for digital librarianship TheoWire May 2003 Sun Mon Tue Wed Thu Fri Sat 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 Wednesday... blogs.salon.com/0002485/categories/digitalLibrarianship - Lycos, MSN, Vindex, GigaBlast". The fourth result is "Digital Librarianship: Computers in Libraries" with a description: "DIGITAL LIBRARIANSHIP" COLUMNS... 5) May, 2000, p. 56-58. What is Digital Librarianship? 20(1... 2002, p. 52-54... XML and the Digital Librarians. 22(8) September... www2.hawaii.edu/~jacco/04liblib.htm - GigaBlast, Lycos, MSN".

A "take-a-peek" window is open over the fourth result, showing a preview of the article text: "Are such mistakes in page numbers pervasive? It's hard to say what percentage million records in the Infotrieve database display excess page numbers, but it ca records for four of my five 1998 CIL columns had wrong page numbers. In the cas for which Infotrieve charges a base fee plus a per-page fee, that is an 80-percent hits you or your client in the pocket. Interestingly, I couldn't find any article where would have been less than the article really is. In all the cases the extra pages c the column, such as full-spread ads and news pages."

Clusty's clustered result list with take-a-peek windowpane

charges a \$10 copyright fee for my most current column (as of this writing), along with a \$12 service charge. For many journals, it calculates the copyright fee on the basis of the number of pages (50 cents) and a base fee of \$4.50. The Infotrieve record indicates the page range of my column to be pages 17–27, i.e., 11 pages. Actually, it is split between two pages, page 17 and 20. It is not possible to tell how many articles have inaccurate page numbers, but my tests have shown that there are many of them, which makes the royalty fees significantly higher than they should be. The tests also make Infotrieve's claim, that "researchers have found ArticleFinder to be a more robust product to use for discovery and citation verification than competitors," very questionable.

Infotrieve may keep signing new publisher partners for document delivery every day, but more and more publishers offer pay-as-you-go services

Service: US\$12.00	Copyright: US\$10.00
DocTitle: Internet Insights - Thoughts About Federated Searching - Peter Jacsó discus	
Author/s: Jacsó, Peter	
Volume: 21 Issue: 9 Page: 17-27 Year: 2004 ISSN: 87556286	
Publication Name: INFORMATION TODAY	
Remove Item	

Example of Infotrieve's inflated service charge and copyright fee

for downloading articles. Customers will realize that direct prices from publishers are much better, as finally proven by the just-released ITI InfoCentral Digital Archive. With its \$2.95 charge per article, ITI InfoCentral allows users to buy at least five times as many articles as they could get through Infotrieve.

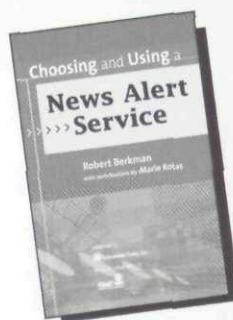
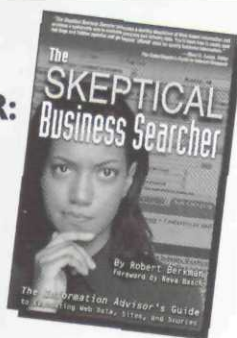
Péter Jacsó [jacsó@hawaii.edu] is professor of library & information science at the University of Hawaii's Department of Information and Computer Sciences.

Comments? E-mail letters to the editor to marydee@xmission.com.

BUSINESS INFORMATION AT ITS BEST

THE SKEPTICAL BUSINESS SEARCHER:

The Information Advisor's Guide to Evaluating Web Data, Sites, and Sources, by Robert Berkman
Foreword by Reva Basch
300 pages/ISBN: 0-910965-66-8
\$29.95

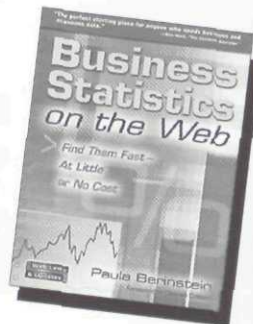


CHOOSING AND USING A NEWS ALERT SYSTEM

by Robert Berkman
with contributions by Marie Kotas;
127 pages/ISBN: 1-57387-224-5
\$79.95

BUSINESS STATISTICS ON THE WEB:

Find Them Fast – At Little or No Cost
by Paula Bernstein
240 pages/ISBN: 0-910965-65-X
\$24.95



Visit your local bookstore or order direct from the publisher.

www.infotoday.com

For more information, call 800-300-9868; outside the U.S. call 609-654-6266
Write to: Information Today, Inc., 143 Old Marlton Pike, Medford, NJ 08055
Order online at www.infotoday.com or e-mail: custserv@infotoday.com