

péter's picks & pans



Péter Jacsó
University of Hawaii

The two picks are the digital archive of New York Times, a cultural heritage in its own, and Scirus, which turned from a frog into a prince, although still with blemishes. The pan is the Web service offered by The Haworth Press that takes away on one hand what it gives with the other and shows that it still can't genuinely embrace the Web, let alone give away something in a really useful way.



the picks

NEW YORK TIMES DIGITAL ARCHIVE, 1851-1999

Ever since the Digital Vault Initiative (DVI) was announced by ProQuest in the summer of 1998 I have been eagerly awaiting the implementation of the Historical Newspaper project, even more than I anticipated the Early English Book Collection [www.proquest.com/hp/News/Reviews/DVISBpage4.html]. Now I wait no more, because what I perceive as the most important single title in the DVI, the digital archive of *The New York Times* was released in mid-July 2002.

To put it mildly, it is awesome to have nearly 150 years of the *Times* in full-text searchable format. The paper of record is a cultural heritage that ProQuest now brings to many libraries. With the exception of LexisNexis, none of

New York Times
Digital Archive

Scirus

The Haworth Press

Hawaii Trust Wields Unusual Power

By WALLACE TITMUS

It is not often that a trust, which is a legal arrangement that allows a person to transfer property to another person for the benefit of a third party, is the subject of a lawsuit. But in the case of the Hawaii Trust, a lawsuit filed in the U.S. District Court in Honolulu last week has set off a legal battle over the trust's unusual power to control the state's natural resources.



William A. Richardson, former chief justice of Hawaii, was appointed trustee of the Hawaii Trust.

The trust, which was created in 1961, is the largest trust in the world, with assets worth more than \$1 billion. It was created by the late Governor John A. Burns, who was the first governor to be elected to the office of Governor of Hawaii.

The trust's unusual power is its ability to control the state's natural resources. The trust has the right to veto any proposed development that would affect the state's natural resources.

Richardson, who is still Chairman of the Hawaii Trust, says that the trust's unusual power is a result of the trust's unique structure. The trust is a trust for the benefit of the state of Hawaii.

The trust's unusual power is its ability to control the state's natural resources. The trust has the right to veto any proposed development that would affect the state's natural resources.

The trust's unusual power is its ability to control the state's natural resources. The trust has the right to veto any proposed development that would affect the state's natural resources.



A view of the Bishop Trust area, showing a sign that says 'Welcome to the Bishop Trust'.

The trust's unusual power is its ability to control the state's natural resources. The trust has the right to veto any proposed development that would affect the state's natural resources.

The trust's unusual power is its ability to control the state's natural resources. The trust has the right to veto any proposed development that would affect the state's natural resources.

The trust's unusual power is its ability to control the state's natural resources. The trust has the right to veto any proposed development that would affect the state's natural resources.

The trust's unusual power is its ability to control the state's natural resources. The trust has the right to veto any proposed development that would affect the state's natural resources.

Reprinted with permission of the copyright owner. Further reproduction prohibited without permission.

The PDF version of the New York Times story on Hawaii's Bishop Trust shows the high quality of the page image.

The screenshot shows the ProQuest search interface. At the top, it says "ProQuest" and "Text Version". Below that, there are tabs for "Collections" and "Search Methods". The search bar contains "Search for: bishop estate AND hawaii". To the right of the search bar, there are dropdown menus for "Article Text", "Article Text", "Article Text", "Publication type", and "Article type". Below the search bar, there are several "AND" buttons and a "Date range" dropdown menu. The "Date range" dropdown is open, showing options like "display ad", "editorial article", "editorial cartoons", "fire loss", "front page", "legal notice", "letter", "lottery numbers", "obituary/birth/marriage", "photo standalone", and "real estate transaction". There are also "Search" and "Clear" buttons. At the bottom left, there is a "Tips" section with the text "For the best results, use words or phrase."

The software offers the usual power features that are essential for full-text searching, such as proximity and positional operators, article type restrictors, and field prefixes to focus your search.

the large aggregators have substantial full-text coverage of the *Times*, let alone in the format offered by ProQuest.

With the individual articles online as very good quality PDF files, as well as entire pages in a Page Map layout, you can browse the *Times* page by page and get the context. It is an excellent idea, given the unwieldy tabloid format (even if you may be irritated, seeing from the corner of your eye an ad for a 6-month CD with the then prevailing, much higher, rate than you can get today).

Nowadays I have much less respect and enthusiasm for the *Times* than before, not only because of the stunning errors I found in the course of a few days preparing for my 2002 *Information Today* April Fool's Day column [<http://hypatia.slis.hawaii.edu/~jacso/pdf/aprilsfool.pdf>], but also because of the out-of-touch-with-the-real-world attitude so poignantly demonstrated in the reaction to the Supreme Court decision in the *Tasini* case. The intellect and managerial competence of Adolph Ochs, then his son-in-law Arthur Hays Sulzberger, and finally Alfred Ochs Sulzberger defined the prime century of the paper. During the declining past few years, Alfred Sulzberger Jr. has been at the helm without his father's guidance.

The ProQuest archive includes more than 3 million pages and about 25 million articles, spanning nearly 150 years. To put it on a more human scale, it means that you will find, for example, 6,053 pieces by James Reston and 1,207 rich-in-thought columns from Anna Quindlen. You can find the front page story of the slaying of Martin Luther King Jr. and the article that warned about the dangerous power of the trustees of the Bishop Estate in Hawaii—more than 15 years before they were removed by court from their position or resigned after “earning” the utmost distrust of the public. It is equally easy to find comics, the weather report, or the crossword puzzle for a given day, if that's what you fancy.

Almost everything is included from the print editions, except for articles and photos by those freelancers who declined to sign the contract waiving their rights for the sale of their works in databases. Still it is almost “all the news that's fit to print”—and search online.

I think it was an understatement in the news release that this database “will allow researchers to find information about every aspect of American history from 1851 forward.” I prefer to say world history, with the possible qualifier—from the perspective of the United States.

The software offers the usual power features that are essential for full-text searching, such as proximity and positional operators, article type restrictors, and field prefixes to focus your search. The list of article types has some unusual entries, such as fire loss and lottery numbers. The former returned documents, but I could not see any notice about fire losses on the page; the latter worked fine. The most useful article type may be the one called front page that yields key stories about something or someone. While there are typos resulting from the OCR process, misspellings are negligible considering the enormous difficulty of applying OCR technology to newspapers, especially so many old ones.

The field prefixes of title and abstract are mostly useful, but there is no need for publication type (newspaper or periodicals), as it is a newspaper database, or publication name (except if the *Times* will be offered later in a package with the *Wall Street Journal* and the *Washington Post*). I would like to see, however, the typical newspaper restrictor, lead paragraph.

The query forms of the interface, however, need revamping. It confuses users to see options for choosing the thesaurus and using classification codes. These do not work in the *Times* database, as the records are not assigned descriptors or classification codes. Neither are some of the field-specific indexes, like DUNS or SIC/NAICS, assigned, even though they are mentioned in the help file. The stop words that are used in other ProQuest databases should not be displayed in this database, as no stop word is applied in indexing the *Times*. The text version of the interface has additional problems, such as many document types that simply do not exist in this database.

There is really no need for Basic and Advanced search, as the only difference is that the latter mistakenly offers the above options. The date-limiting options are another problem: Five of the eight (the ones which offer limiting the search to last week, last 30 days, last month, last quarter, last year) will not produce results, as the coverage ends in 1999.

In spite of the current interface problems that can be easily corrected,

the *Times* archive is a milestone product, which will be intensively used not only by researchers, but ordinary patrons in large public libraries and the employees in government agencies. Hopefully subscribers are interested not only in the daily news but also in some of the historical issues.

SCIRUS

I criticized Elsevier's Scirus [www.scirus.com] in the June 2001 issue of *Information Today* for including hundreds of thousands of inane, vulgar, and useless sites, while claiming on every page that Scirus is "for scientific information only" [www2.hawaii.edu/~jacso/extra/infotoday/scirus/scirus.html].

As of early August 2002 it still has at least 100,000 trash sites filled with vulgarities by pathetic college students (just repeat the search shown at the Web site above) and other sites that have no relevance for anyone except, possibly, the author. But now Scirus offers free abstracts for 1.6 million articles from over 1,200 Elsevier journals, 250,000 articles from Academic Press journals (Academic Press is now owned by Elsevier), almost 13 million MEDLINE records, a million patent records, and free full-text articles from BioMed Central. (You can also buy articles on the spur of the moment, but the price does not encourage impulse buying.) True, there are many duplicates and triplicates, both from ScienceDirect and BioMedNet MEDLINE, but it is not as big a problem as in fee-based databases.

In addition, Elsevier Science and search engine company Fast Search & Transfer (FAST) have significantly improved the software. It offers excellent options, in the Advanced Search module, for setting your preferences for information types (abstracts, articles, company home pages, conferences, patents, preprints, and scientist home pages), journal sources (Beilstein on ChemWeb, BioMed Central, IDEAL, MEDLINE on BioMedNet, and ScienceDirect), Web sources (Chemistry Preprint Server, CogPrints, E-Print ArXiv, Mathematics Preprint Server, NASA, Neuroscion, U.S. Patent Office, and other), and for scientific and social sciences subject areas.

Five of the eight date-limitation options will not produce results.

The search and display process is smooth and simple. The recommended terms most often co-occurring with your search term are listed on the side of the result list to refine the query, and they are excellent. You can choose to view journal results, Web results, or both. Scirus also ele-

gantly handles the links to free documents cited in MEDLINE, such as the articles from *BMJ*, the *British Medical Journal*. You need to sign up for the free BioMedNet, but then the transfers are almost transparent to and from non-Elsevier sites.

Keep an eye on Scirus. It is likely

Scirus now offers excellent options, in the Advanced Search module from search engine FAST, for setting your preferences.

Save Checked Results Email Checked Results

1. An explanation of the effectiveness of latent semantic indexing by means of a bayesian...
Story, R.E., *Information Processing & Management*, May 1996
 Latent **Semantic Indexing** (LSI) is an effective automated method for determining if a document is relevant to a reader based on a few words or an abstract describing the reader's needs. A particular feature of LSI is its ability to deal...
Full text article available from **SCIENCE @ DIRECT**
[similar results](#)

2. Latent Semantic Indexing: A Probabilistic Analysis
Christos H. Papadimitriou / Prabhakar Raghavan / Hisao Tamaki / Santosh Vempala, *Journal of Computer and System Sciences*, Oct 2000
 Latent **semantic indexing** (LSI) is an information retrieval technique based on the spectral analysis of the term-document matrix, whose empirical success had heretofore been without rigorous prediction and explanation. We prove that, under certain...
Full text article available from **IDEAL**
[similar results](#)

3. An evaluation of concept based latent semantic indexing for clinical information retr...
C G Chute / Y Yang, *Proc Annu Symp Comput Appl Med Care*, Feb 1992
 Latent **Semantic Indexing** (LSI) of surgical case report text using ICD-9-CM procedure codes and index terms was evaluated. The precision-recall performance of this two-step matrix retrieval process was compared with the SMART Document retrieval system, ...
MEDLINE Citation on **BioMedNet MEDLINE**
[similar results](#)

4. Automated semantic indexing of imaging reports to support retrieval of medical images ...
H J Lowe / I Antipov / W Hersh / C A Smith / M Mallhot, *Methods Inf Med*, Dec 1999
 This paper describes preliminary work evaluating automated **semantic indexing** of radiology imaging reports to represent images stored in the Image Engine multimedia medical record system at the University of Pittsburgh Medical Center. The authors used...

Refine your search using the following terms:

- document collections
- document retrieval
- formulate
- generalization
- information filtering
- information retrieval
- intelligent information retrieval
- latent
- patient record
- relevance feedback
- retrieval system
- retrieval systems
- singular value decomposition
- sparse matrices
- squares problems
- vector space

Refine using:

All of the words

The recommended terms most often co-occurring with your search term are listed on the side of the result list to refine the query.

to become increasingly useful as some abstracting and indexing services provide less and less current or adequate coverage of their professed fields of specialization.



the pan

THE HAWORTH PRESS

The Haworth Press [www.haworthpressinc.com] is as disappointing with its free online table of contents and abstract offerings as it has been with paid subscription services. It may be the fourth or fifth largest publisher in library and information science, but it has been on the bottom of my list of publishers for a long time. I don't know any other publisher that is so systematically late in publishing most of its journals. It frequently collapses two or more issues into a single combined one—saving a lot in mailing expenses. You can tell from a mile when you see a run of Haworth journals on the shelf by the excessive amount of combined issue numbers of the spines. Most of its journals are quarterly; still the spine labels show 1/2, 3/4, and even 1/2/3/4. From the incessant delays, and the inferior print quality—which reminds me of

Eastern European journals from the early 1960s—I would not be surprised to learn that printing and binding are outsourced to Mongolia.

Librarians voiced their dissatisfaction on the SERIALST list server last summer. You can find the main thread of their complaints about the company at the list archives [http://list.uvm.edu/cgi-bin/wa?A2=ind0107&L=serialst&D=0&P=5793]. It adds insult to injury that later issues are delivered before earlier issues, as you can see from the 2002 issues of *The Serials Librarian*. The dispatch lists for other journals show a very similar pattern.

With this background it is no surprise that Haworth is terribly behind

the curve with free Web offerings. While most other publishers recognize that it is in their own best interest to offer the bibliographic citations and abstracts of their articles, Haworth does not seem to get it. Until recently Haworth had hardly anything to offer on the Web except subscription information for its journals and purchase price for the books.

When I learned about its free table of contents and abstract service, I became a little hopeful, especially as the excellent free BUBL abstracting/indexing service (which covered among others Haworth Press journals) stopped providing the service because practically all publishers launched their own free services. It was the more disappointing then to see that the publisher chose to e-mail the tables of contents and abstracts rather than post them on its Web site at least in a browsable format. The e-mailed format is OK as an alert service, but not as a finding tool. I hope someone will write a small utility to automatically extract from e-mails the records with abstracts, aggregate them into a small database, and post the utility on the SERIALST discussion list. Even better, Haworth Press could wake up and do something useful for free in the form of a searchable database containing all the abstracts of all of its journals.

Péter Jacsó [jacso@hawaii.edu] is Associate 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.

JOURNAL DISPATCH SCHEDULE RESULTS

Serials Librarian, The

[Make another request](#) | [More journal details](#)

Volume	Issue	Anticipated Pub. Date	Cover Date	Issue Online?	Date Published	Date Mailed
41	3/4	published	2002	Forthcoming	6/10/2002	
42	1/2/3/4	published	2002	Forthcoming	6/25/2002	
43	1	9/12/2002	2002	Forthcoming		
43	2	9/22/2002	2002	Forthcoming		

The Haworth Press journal dispatch lists show that later issues are sometimes delivered before earlier issues.