

Introducing Database Advisor

A new service that will make your research easier

by Christy Hightower, Jennifer Reiswig, and Susan S. Berteaux

Xena walks up to a library workstation at 9 p.m. with a new research topic in hand: waste-water reclamation projects. She's familiar with Current Contents, but wonders if there are other databases that might be better for this topic. The reference desk is closed, but there's a link from the library's Web site to the Database Advisor (DBA). She sees that this will tell her which of the more than 25 science databases will be best for her topic, and decides to try it.

She sees an input box for search terms and a Submit button. She's used to searching the Web, so she types in: +*waste water* +*reclaim**. But then she remembers that not all search engines work the same way, so she scrolls down the Search page to see what options or help are available.

First, there is an option to include "backfiles" for the databases, she's mainly interested in current articles, so she doesn't

select backfiles. There is an option to choose how long to wait for results, she's not sure how long most searches take and leaves the default of one minute. There is a list of subject categories (e.g., biology, chemistry, mathematics, etc.). She follows a link to see which databases are included in each category and de-

cidest to leave the default setting of all categories.

Finally, she comes to help information. A table presents tips on constructing queries for DBA. She sees that truncation and Boolean operators are not supported, that synonyms should be

avoided, and that queries should be kept simple. Noting this, she moves back to the input box and changes her query to *water reclamation*, and submits the search.

The screen changes to "Results" with instructions to wait while results come in. The entries begin to appear, and the list changes as each new database is added so that the

About the authors

Christy Hightower is web coordinator at the Science & Engineering Library, University of California, San Diego, e-mail: chightow@ucsd.edu; Jennifer Reiswig is electronic services librarian at the Biomedical Library, University of California, San Diego, e-mail: jreiswig@ucsd.edu; Susan S. Berteaux is instruction coordinator and assistant head of the Scripps Institution of Oceanography Library, University of California, San Diego, e-mail: sberteaux@ucsd.edu

database with the most “hits” is always at the top. After a minute, a “Results” banner appears at the top of the page, indicating the results are complete.

Xena scans the page, and clicks on a link to get help interpreting the results. She learns that the “hits” represent the number of articles in that database that would be retrieved by doing a keyword search on her terms. There are small icons indicating whether each database contains abstracts, full text, and/or full images for articles. There is also a one-line description for each database and a link to a more detailed profile on each database. For Xena’s search, GeoRef came up first, followed by MOFR (Marine, Oceanographic & Freshwater Resources). All 25-plus databases are listed in descending order by the number of hits.

At the end of the list of databases are a number of entries with “Time Out” where the results would be. One of these, Greenwire (covering governmental activities on the environment), catches her eye and she realizes that if she had chosen a longer time to wait back on the search screen, she might have seen the results from this database. At the very bottom of the Results page she finds a form to Refine the search, featuring a table with her search filled in. She leaves the keywords as they are, decides to include the backfiles, and sets the waiting time to its maximum: five minutes.

This time Greenwire shows results, only 11 articles. Interestingly, the “backfiles” of some databases have more hits than the current files, prompting her to consider using some older material from years when this was a hotter topic. Scanning down the list, she notices some databases she hadn’t heard of that might have interesting angles on her topic: BIOSIS has over 200 hits, and the ABI/Inform database covering business and management topics has over 150.

Based on the list she has, she clicks on GeoRef to start a search of that database.

Development process: How did we get here?

In 1996, the University of California, San Diego (UCSD) Library was providing online access to over 25 science and engineering databases via Web, Z39.50, and telnet. It was not always obvious to patrons or library staff

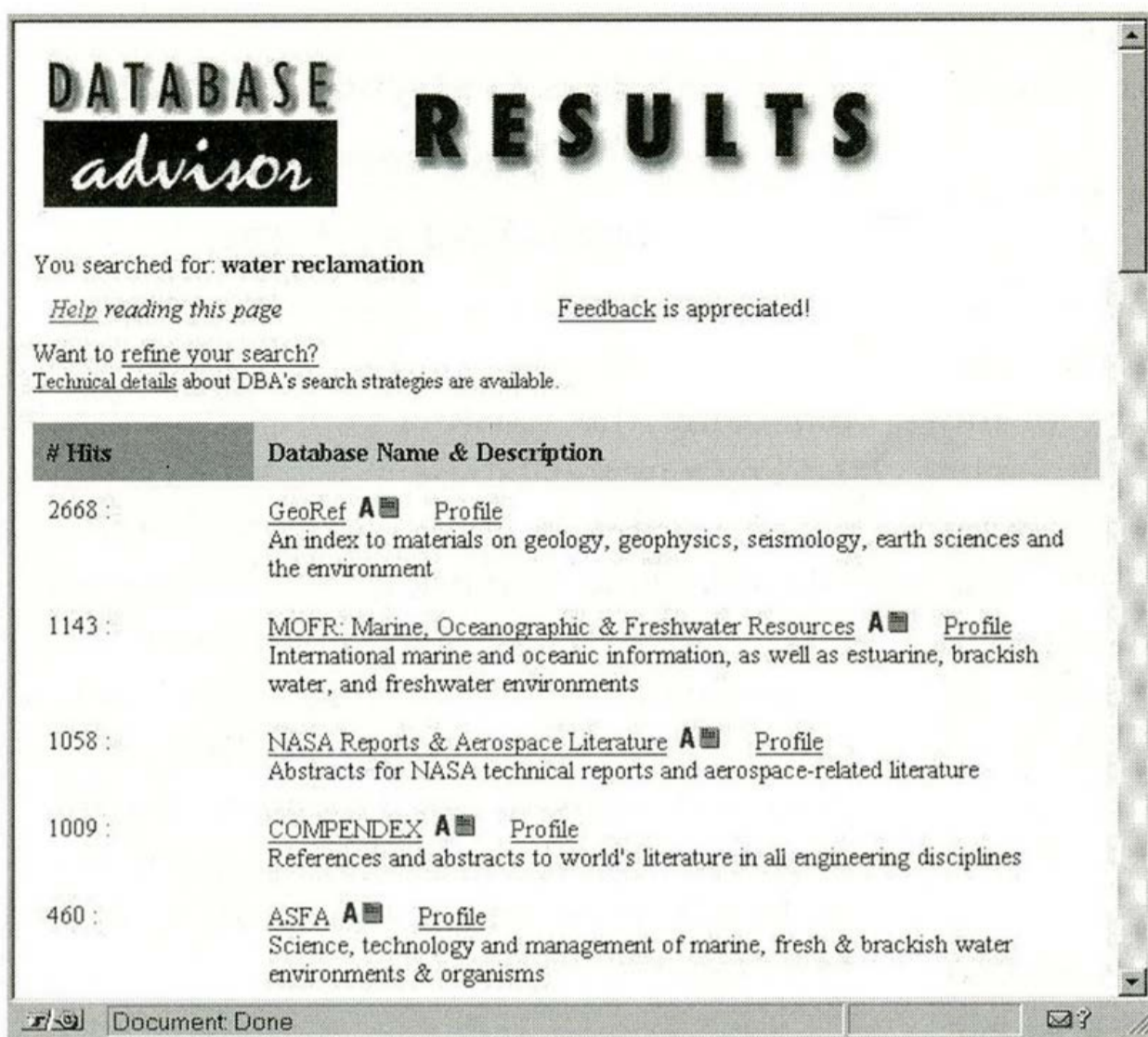
which of these bibliographic and full-text databases should be used to research a topic. Many databases also appeared to be underutilized. The libraries needed a simple, Web-based tool to quickly advise users on the best science database for their topic. The team was familiar with DIALOG’s DialIndex®, which ranks databases based on number of hits on terms supplied by the searcher.

Science librarians proposed the concept of a Web-based frontend to all the science and engineering bibliographic and full-text databases to which UCSD has remote access. A technical team consisting of two student programmers and a librarian/technical coordinator wrote the source code and scripts used to query databases. Initially, 50 hours of student programmer time was budgeted for the project. Through October 1997, the programmers logged approximately 250 hours on the DBA project.

An interface team of three science librarians (the authors) was established in February 1997 to develop the Web-based user interface, graphics, search strategies, and database profiles. Originally, basic and advanced search options were envisioned. Three models with varying degrees of complexity were presented to the UCSD science librarians. Because of an overwhelming preference for the most basic search tool, the advanced search options were not pursued in the final product.

Each librarian on the team provided profiles for one-third of the databases. The team also researched and developed search strategies comparable to MELVYL® System keyword searches for all databases covered by DBA. Although most Web databases provide basic and advanced search help, in some cases the vendor/provider had to be contacted to provide the precise search strategy that would achieve consistent, comparable results.

In August 1997, interviews were conducted with students and public services staff to obtain suggestions for improving the functionality of the interface. Interviewees were asked about their level of Web and database experience and their expectations for a Web tool like DBA. Responses provided guidance for refining help directions, individual database profiles, navigation, and



minute. (One of our cool features is that the user can control the timeout.) Those queries that take a long time (like Web) have separate processes; those that are quick (like Z39.50) are strung together. We use a small amount of JavaScript (just to select and deselect subject categories, and to warn off-campus users that they will not be able to connect to license-restricted databases from the results page).

We kept the graphic load to a minimum; the few graphics we used were created in PaintShop Pro. DBA uses the server push feature to reorder

basic screen elements. The interviews, combined with input from the technical team, reinforced the logical concepts and confirmed the usefulness of the product.

After some minor improvements, the Database Advisor was rolled out to the UCSD community on September 15, 1997. Early user feedback indicates users are indeed discovering the existence of databases that were previously unknown to them.

Technical information

The "kernel" of DBA (the script that parses user input, forks off the queries, then displays the results) is simple and small, thus portable. To run DBA, a campus needs a UNIX computer with Internet connectivity, Perl version 5 (the language DBA was written in), and preferably the Apache Web server (which is common in academic settings).

There are three types of scripts used to query databases: Web scripts, telnet scripts, and a Z39.50 script. The Z39.50 interface was made using the client written by Harold Finkbeiner at Stanford University and is our preferred method of accessing databases. All the MELVYL databases are accessed via this Z39.50 script, which is very fast. Each of the databases or database groups (telnet, Z39.50) have separate modules. These modules can be added, modified, and removed at will. This makes DBA easy to customize.

DBA is quite fast, with the majority of databases returning results within one

the search results so that the databases are listed in descending order of hits. DBA works with all graphical Web browsers, although we did encounter a problem with the way Microsoft's Internet Explorer version 3 implemented server push, which required a work-around. Internet Explorer interpreted push to mean append the pages, while Netscape's browsers interpreted push to mean replace the pages. The lesson here is not a new one: always test your programs on all browsers and all operating systems.

Another lesson is to always include a budget for maintenance; it's more time-consuming than you expect. In the first month of operation, both SilverPlatter and Cambridge Scientific significantly changed their search interfaces, another database changed its underlying search engine, another significantly changed its backfiles, and we subscribed to three new databases that required two new scripts. We also began to include databases with limited numbers of simultaneous users that required a new error message ("Too Many Users") and increased attention to the way DBA logs out so as not to tie up any ports.

Current status and future plans

When the sciences version of DBA went live in September, it was advertised via e-mail announcements, more than 50 links from the three science library Web sites (a link to DBA was placed next to each link for a da-

tabase included in DBA), bulletin board displays, and by DBA "business cards" handed out at the reference desk.

Currently DBA averages 75 searches a week. We expect that usage to increase, since we recently installed microcomputers with a Web interface in the library. We also collect information about the nature of the searches performed: the keywords and subjects used, whether the search is refined or not, what Web browser was used, etc. This will allow us to perform more detailed analysis of DBA usage patterns in the future. In addition, we link to a feedback form from every page, but after six months of operation, we have only received a single response (it was positive) this way!

Work on DBA will proceed on several fronts. Twenty-five out of thirty social science and humanities databases have already been added to a new version of DBA. We are pursuing Z39.50 access for the remaining five databases and hope to be able to release the new version in summer 1998. At that time DBA will become relevant to every student and researcher at UCSD.

Second, in response to user input, the concept of "instant gratification" is being pursued, which would take the user directly to the results of their search when they decide on a database to use.

Instant gratification was not initially offered because DBA does very generic searches in each database, and it was felt that the user would be able to get better results by redoing the search using the unique features of each database. If a user obtained 10,000 results from DBA, would those 10,000 items really be useful? However, in our limited user testing, this was one feature that was repeatedly asked for. This function presents some technical challenges and may take some time to implement.

Interested in trying this yourself?

UCSD plans to make the DBA source code available to anyone who wants it under the terms of the GNU General Public License agreement. For details, contact the authors. At other sites, the UCSD search scripts would have to be modified to send your user names and passwords, and scripts for any databases new to the system would have to be writ-

**. . . always test your programs
on all browsers and all
operating systems.**

ten using the existing scripts as models. Adding a new database in a family for which UCSD already has scripts (such as SilverPlatter or Cambridge Scientific) should require very little new programming.

Remember that anyone with Web access can use the demo version of Database Advisor, but many of the databases that show up on the Results page are restricted to UC or UCSD users and may not allow access from outside IP addresses.

URLs

1. Demo version of Database Advisor: http://scilib.ucsd.edu/Proj/dba/dba_public.html

2. DBA Search Strategies: http://scilib.ucsd.edu/Proj/dba/search_strat.html

3. DIALOG's DialIndex ® (free via the Web): <http://www.dialogweb.com/cgi/dwframe?context=databases&href=topics/All/> ■

INFORMATION LITERACY: ST&R

Information Literacy: Search Strategies, Tools & Resources (ST&R) © 1998

A comprehensive Web-ready information literacy program.

- **ST&R Book:** 9 chapters ranging from basic information access concepts and search strategies to learning how to use and evaluate Internet sources, online catalogs, and periodical literature. Includes ST&R WebLinks with annotated Internet links to selected digital libraries.

- **ST&R Show:** Microsoft® PowerPoint 97/98 presentation programs for Windows 95 and Apple® PowerPC.

- **ST&R WebBook:** PDF version of ST&R Book for lab use with Internet browsers and search engines;

- Problem solving assignments; extensive references; glossary; index.

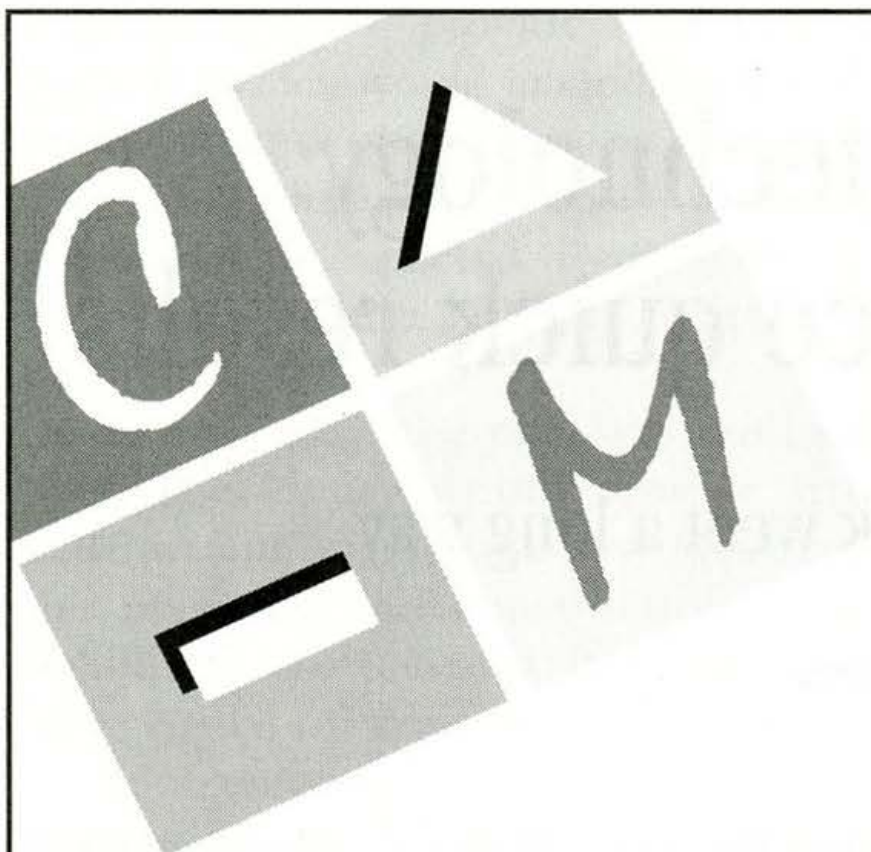
- \$59 (ST&R Book+ST&R WebLinks+ST&R Show), \$34 (ST&R Book+ST&R WebLinks), \$30 (ST&R Show), \$7 (ST&R WebLinks). For ST&R WebBook site license contact InfoEN (infoen@lainet.com).

- Developed for instructional librarians in college and public libraries. Customization available.

InfoEN Associates, Box 66276, Los Angeles, CA 90066 • 310/391-3923 Voice/FAX

- infoen@lainet.com

- <http://www.lainet.com/infoen/>



Collection Manager

Breakthrough
Collection
Development
Services

- ✔ Intuitive presentation of approval profiles
- ✔ Online review of new title announcements
- ✔ Flexible selection and ordering features
- ✔ Adaptable to local systems and workflows
- ✔ Developed in partnership with librarians

To begin using Collection Manager, the complete workflow solution, contact...

BLACKWELL'S
BOOK SERVICES

Matt Nauman
Lake Oswego, Oregon
Telephone: 1-800-547-6426
E-mail: mktg@blackwell.com
WWW: <http://www.blackwell.com>