

dbcost.txt
CUTTING THE COST OF DATABASE SEARCHES

Peter Brueggeman
Scripps Institution of Oceanography Library
UCSD
La Jolla CA 92093

In: Oceans from a global perspective: international cooperation in marine science information transfer. Carolyn P Winn, ed. Proceedings of the 13th annual conference of International Association of Marine Science Libraries and Information Centers held at Halifax, Nova Scotia, Canada, 5-9 October 1987. Office of the Research Librarian, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, 1990. pp 157- 162.

ABSTRACT: The databases of interest to the marine sciences typically have high connect-time charges and high citation charges. High search costs can have a chilling effect on the database search service offered by a marine science library. Software and stratagems which can reduce database search costs are the focus of this paper.

A marine science library's database search service may access a multidisciplinary range of databases in the course of serving the information needs of its clientele. Geology, zoology, fisheries, physics, meteorology, engineering, geophysics, and other disciplines are represented in marine science and thus a wide range of databases may be accessed for the literature of these disciplines. Typically most science databases have high connect-time charges and high citation charges. These high costs can have a chilling effect on a library's search service. If the search service is free, then search costs need to be controlled in order to keep within budget. If the search service is recharge-based, then high search costs hinder frequent usage of the service, particularly by graduate students who typically have limited access to funds for database searches. If high-cost searches are consistently delivered, then service may become limited to those with the ability to pay substantial search costs; the search service may only be serving its well-funded clientele. Cost reduction can be achieved through effective searching techniques and efficient equipment. Cost reduction efforts can target both connect-time and citation charges to receive incremental savings.

MINIMIZING CONNECT-TIME

Connect-time charges can be reduced two ways: minimize online time and/or use less expensive connect-time. Online connect-time can be minimized by searching at baud rates higher than 1200 and also by using microcomputer software, particularly search aid software. DIALOGLINK (1,2) and PROSEARCH (3,4) typify search aid software. Search aid software is a hybrid of telecommunications and wordprocessing software and is designed expressly for database searching. Search aid software is used to keyboard search terms while offline, to logon to the databank with one keystroke, to upload search terms to the database in a line-by-line mode, and, to download and edit search results. Search aid software integrates the most useful telecommunications and wordprocessing features for database searching into one user-friendly software package. This paper focuses on the use of search aid software; however some telecommunications and wordprocessing software can be used in tandem to achieve similar functionality.

Search aid software can be used to keyboard search terms while offline; after logging into the database, search aid software line-by-line uploads those terms. This saves on the connect-time spent keyboarding, especially time lost due to keyboarding error. Search aid software automatically downloads the entire search session into the microcomputer's temporary RAM memory; disposition of the search results can be addressed after exiting the databank and stopping the connect-time clock. Since the entire search session is captured in RAM memory, the search results can be printed or downloaded-to-disc while offline. The searcher thus avoids using connect-time to execute downloading and/or printing protocols or to fiddle with the printer's paper.

USING LESS EXPENSIVE CONNECT-TIME

Less expensive connect-time can be found by databank shopping; one databank can offer substantially lower connect-time cost on a particular database compared to another databank. For example, BRS currently offers substantially less expensive connect-time than Dialog on several databases of interest to marine science. The cost of connect-time may be the main factor in a particular search and not the charge per citation. If a preliminary search on the databank's database that indexes its databases (like Dialog's DIALINDEX

dbcost.txt

or BRS' CROS) indicates a low number of citations for a particular database, then switching to another databank in order to use its less expensive connect-time may be worth considering.

Inexpensive connect-time can be used for printing citations as well as for searching; Pam Mofjeld's paper (presented at this meeting) describes printing citations online using Dialog DIALMAIL's substantially less expensive connect-time instead of using the more expensive connect-time of the database associated with the citations.

MINIMIZING CITATION CHARGES

With the retrieval of a relatively small amount of citations, citation charges are not an overwhelming component of the total search cost yet it would be desirable to reduce them. However, with large citation retrievals, citation charges and the connect-time necessary to print them will be the major component of search cost. Examining a large retrieval in order to figure out a way to reduce it consumes connect-time to examine results; more connect-time is consumed executing additional search steps that may or may not offer significant reduction. Thus reduction of large retrievals can become counterproductive in many cases; the searcher has to spend money on connect-time in order to save money on citation charges. In addition, the deleted citations may contain some citations of interest.

To reduce citation charges for both small and large retrieval searches, a two-step search methodology (5) can selectively retrieve the relevant citations scattered throughout a retrieval and also minimize connect-time. Citation charges and connect-time can be reduced by minimizing the number of complete citations printed and not necessarily the number of citations retrieved. A retrieval can be reduced offline rather than online; the connect-time clock is stopped while the retrieval is reduced. First, a quick search retrieves a reasonable (and possibly large) set of citations. All of the retrieved citations are displayed as partial citations consisting of titles and database accession numbers only. The databanks usually assess little or no citation charge for this output format. Partial citations can be output quickly; approximately fifty titles and accession numbers are output per minute at 1200 baud and approximately 100 per minute at 2400 baud. The requestor then scans the titles retrieved and selects interesting titles. Second, a followup search uses

dbcost.txt

the database accession numbers corresponding to selected titles as search terms in order to retrieve and print the complete citations. With this methodology, citation charges are accrued for only those citations of interest and not the entire search retrieval. No connect-time is expended in devising the means to reduce the retrieval. The requestor benefits both from cost reduction and from comprehensive retrieval. The searcher makes few compromises in delivering results and is relieved from the intellectual strain of narrowing a retrieval.

This two-step method is best accomplished on a microcomputer with search aid or with telecommunications and wordprocessing software. It can also be accomplished using a "save search results" capability like Dialog's "logoff hold" command; however Dialog only saves search results for ten minutes which may be insufficient time for the requestor to examine the titles. Using search aid software, the titles and accession numbers are downloaded. The downloaded titles are scanned by the requestor on the searcher's microcomputer or printed on paper. The requestor indicates the titles of interest and the searcher edits the downloaded results to delete everything but the database accession numbers corresponding to the titles of interest. Next, a search strategy integrating the accession numbers is composed with search aid or wordprocessing software; it is subsequently uploaded into the database in order to retrieve complete citations.

Using BIOSIS on Dialog as an example,

FIRST, the searcher downloads the retrieval as partial citations composed of titles and accession numbers (Dialog's citation format six)

6/6/1

0016040193 BIOSIS Number: 81019109

MERCURY CONCENTRATIONS OF THE AXIAL MUSCLE TISSUES OF SOME MARINE FISHES OF THE CONTINENTAL SHELF ADJACENT TO TASMANIA AUSTRALIA

6/6/2

0016040188 BIOSIS Number: 81019104

MERCURY DISTRIBUTION IN A POLLUTED MARINE AREA RATIO OF

dbcost.txt

TOTAL MERCURY METHYLMERCURY AND SELENIUM IN SEDIMENTS MUSSELS
AND FISH

6/6/3

0016040158 BIOSIS Number: 81019074
FLUX OF CADMIUM THROUGH THE MARINE NEMATODE ENOPLUS-BREVIS

6/6/4

0016040132 BIOSIS Number: 81019048
CADMIUM BINDING SITES ON CELLS OF A MARINE PSEUDOMONAD

6/6/5

0016038482 BIOSIS Number: 81017398
REDUCTION OF GROWTH RATE AND RESTING SPORE FORMATION IN A
MARINE DIATOM EXPOSED TO LOW LEVELS OF CADMIUM

6/6/6

0016026838 BIOSIS Number: 30017046
SILICONES IN ESTUARINE AND COASTAL MARINE SEDIMENTS

6/6/7

0016026837 BIOSIS Number: 30017045
PARTITIONING OF POLYCHLORINATED BIPHENYLS IN MARINE
SEDIMENTS

6/6/8

0016019861 BIOSIS Number: 81009319
ESTIMATION OF TOXICITY TO MARINE SPECIES WITH STRUCTURE-
ACTIVITY MODELS DEVELOPED TO ESTIMATE TOXICITY TO FRESHWATER
FISH

6/6/9

0016005239 BIOSIS Number: 30005239
PHYSIOLOGICAL EFFECTS OF MARINE POLLUTANT STRESS A
MANAGER'S PERSPECTIVE

6/6/10

0016002990 BIOSIS Number: 30002990
PREDICTION OF THE BIODEGRADABILITY IN THE AQUATIC MARINE
MEDIUM THE CHEMOSTAT AS MODEL SYSTEM

SECOND, the requestor scans all retrieved titles and
indicates the desired titles -- the titles mentioning cadmium
and mercury

dbcost.txt

THIRD, using search aid or wordprocessing software, the searcher edits out everything except the desired database accession numbers

0016040193
0016040188
0016040158
0016040132
0016038482

FOURTH, the searcher composes a search strategy utilizing the downloaded accession numbers

b5
t0016040193/3
t0016040188/3
t0016040158/3
t0016040132/3
t0016038482/3
logoff

FIFTH, a subsequent line-by-line upload into the database retrieves the desired complete citations

USING DATABASE ACCESSION NUMBERS TO MINIMIZE CITATION CHARGES

For many databases of interest to marine science, stopping with the retrieval of titles and database accession numbers only can be a cost reduction service option if the library has the corresponding printed abstract/index. The requestor receives a printout of titles and accession numbers only and then looks up the corresponding complete citations in the appropriate printed abstract/index. Most or all citation charges are thus avoided. If search results are delivered with title and accession number only, then inexpensive connect-time is of great interest since citation charges will be a negligible or nonexistent component of the total search

cost.

AQUATIC SCIENCES AND FISHERIES ABSTRACTS, OCEANIC ABSTRACTS, ZOOLOGICAL RECORD, INSPEC (Physics Abstracts), GEOREF (Bibliography and Index of Geology), WATER RESOURCES ABSTRACTS, and BIOSIS (Biological Abstracts) are some databases which provide accession number access to the corresponding printed abstract/index. Oftentimes the database's accession number is the same as the printed abstract/index's volume/year and its abstract/citation number. In other cases, the database's accession number field contains the printed abstract/index's volume/year and abstract/citation number in addition to an unrelated database accession number. Examine the databank's database documentation and citation charge structure to explore this possibility and try it before relying on it as a low-cost service option.

For example, Dialog's BIOSIS offers reliable accession number access to the printed Biological Abstracts; BRS' backfile of BIOSIS however suffers from duplicate erroneous accession numbers because BRS loaded years into some accession numbers instead of volume numbers. BRS offers free accession numbers for access to the printed ZOOLOGICAL RECORD whereas Dialog does not. Dialog's GEOREF's accession numbers can be used as access to the Bibliography and Index of Geology citations only from 1977 to present.

REFERENCES

1. AGGI RAEDER. 1986. Dialoglink: product evaluation of a new communication software from Dialog. Database End-User. 2(4): 17-21.
2. JOANNE WITIAK. 1986. Dialoglink: a review of Dialog's search assistance software. Online. 10(6):39-42.
3. MARGARET BELL HENTZ. 1986. Pro-Search: online search aid. Information Today. 3(9):15,33.
4. BARBARA QUINT. 1986. Menlo Corporation's Pro-Search: review of a software search aid. Online. 10(1):17-25.
5. PETER BRUEGGEMAN. 1986. Converting search results into search statements: using Dialoglink and Pro-Search search aid software. Database End-User. 2(9):26-28.

□