

asksam.txt

MANAGING A LOCAL AQUATIC SCIENCE DATABASE WITH ASKSAM

PETER BRUEGGEMAN  
SCRIPPS INSTITUTION OF OCEANOGRAPHY LIBRARY  
UNIVERSITY OF CALIFORNIA SAN DIEGO

In: The aquatic environment: description, management, conservation. Elizabeth Fuseler and Stephen Wiist, editors. Proceedings of the 17th annual conference of International Association of Aquatic and Marine Science Libraries and Information Centers held at Texas A&M University at Galveston, Galveston, Texas, 7-11 October 1991. Colorado State University Libraries, Fort Collins, Colorado, 1992. pp 151-159.

Maintaining a separate local database is an option for an aquatic science library wishing to maintain a comprehensive database on the local aquatic environment. Ideally information on the local aquatic environment should be integrated into one system including the database of library holdings. However administrative realities like computing equipment, software, and staffing may necessitate establishment of a separate local database in order to manage local aquatic information. As part of the University of California San Diego Library system, the Scripps Institution of Oceanography Library uses Innovative Interfaces' Innopac integrated library system for acquiring, cataloging, circulating, and accessing its holdings. The Scripps Library recognized a strong demand for access to information on the San Diego marine environment beyond that encompassed by the library's compact discs, computer catalog, and current acquisition and cataloging practices. The Library's computer catalog is acquisition-based and traditional in concept due to staffing considerations. Bibliographic coverage is restricted to items in the collection and excluded for items not in the collection. Bibliographic coverage of items in the collection is at a level that does not indicate the presence of relevant items on the San Diego marine environment ie book chapters, journal articles. Though considerable San Diego marine information is present in the computer catalog, additional coverage was needed for journal articles, technical reports, and city, county, regional, state, and federal government reports with abstracts being particularly desirable. As part of the University of California San Diego Library system, the Scripps Library uses Innovative Interfaces' Innopac integrated library system for its library holdings. Innopac could be utilized and the library holdings database enhanced to encompass San Diego marine information. However proposing to process a initially large number of references and a smaller ongoing number into Innopac would encompass an administrative review of staffing priorities. The local information would require original cataloging and such information would necessarily be deemed a lower priority compared to materials being currently acquired.

Given this necessity to establish a separate system for handling local marine information on San Diego, staffing effort remained the paramount consideration; as usual, staffing for such an effort was low priority. It was quickly recognized that the ideal San Diego marine database would involve limited effort in establishing and maintaining so that it would be viable, updated, and not subject to cancellation due to shifting priorities. To ensure that staffing effort was minimal, the database should incorporate San Diego marine information in the wide range of pre-existing electronic formats as found. Original data entry was to be avoided; the object was to acquire references on the San Diego marine environment in electronic format. While this would result in a less comprehensive database, having "something" was deemed better than having "nothing". The expedient solution was the only solution that could be implemented. Extensive reformatting of existing information into a consistent database format was to be avoided since this is time-intensive and frequently beyond in-house expertise. If the references were readable and understandable as presented in an ASCII file, then field tags, field names, and field information would be acceptable in any format including a lack of field tags/names.

In order to accept references in any format, a full-text database software capable of handling free-form information was targeted as a likely candidate for managing the San Diego marine database. Such software is also known as text management software. References could be imported from any source in any format with or without field tags or field names with no restrictions on field or record length. At a minimum, the references need only be available as a readable ASCII file as if someone had typed them up in a list. If field tags were available, then they would be left intact if they did not interfere with the comprehension of the references. Selection of a full-text database software for the San Diego marine database was complicated by the need to have a user-friendly interface. Since novice users are the target audience for the San Diego marine database, many full-text database software are undesirable with their design orientation to the experienced user. In addition the search interface presented to library users should be strictly search-only (also known as run-time) in order to avoid accidental or deliberate editing or deletion of references. A full-featured version of the database software can be used to create and update the San Diego marine database but the public should be restricted to using a search-only version of the software in order to ensure database integrity.

After reviewing microcomputing literature, askSam software was selected. AskSam operates in the DOS environment which was mandatory at the time; an equivalent Macintosh software would be acceptable now. AskSam requires 384K of RAM and DOS version 2.0 or greater. AskSam

#### asksam.txt

can import information in any format into a database subsequently searchable by one interface. AskSam is designed to work with or without fields and field tags. An askSam database can be searched with any of three search-only interfaces available at extra cost. AskSam version 5 retails for US\$395, can be found through mail-order for as little as \$210, and is available at educational discount for US\$100. AskSam version 4.2 retails for US\$295 and at educational discount for US\$30 (US\$10 shipping/handling in US and US\$40 for Canadian orders; available from askSam, PO Box 1428, Perry, Florida 32347. Sales:800-800-1997, Technical Support:904-584-6590, Fax:904-584-7481). AskSam version 4.2, the older version, will create and maintain a local database as detailed in this paper so it is the preferential choice for the cost-conscious. The lower-cost version 4.2 can be considered a trial purchase and, if askSam becomes indispensable, upgrade to the latest version can be sought in the future. Though technical support is unavailable for version 4.2, it is unlikely it will be needed because maintaining a local database as outlined in this paper is trivial in its utilization of askSam's features. In addition to askSam itself, a search-only version should be purchased for the public usage in searching the local database. Maintaining the integrity of the database requires that the public use a search-only version that cannot alter the database. Three search-only or run-time versions are available (HyperSift, HyPeruse, and InfoSift for US\$75, US\$50, and US\$50 respectively; substantial discounts available with multiple orders) and their differences are detailed below. AskSam and/or its search-only equivalents are speedy in retrieval on a 8086/8088 microcomputer and thrilling on an 80386 or higher microcomputer. Considering that full-text searches are being executed on a seven megabyte database at Scripps Library, askSam performs well beyond expectation. In creating a database, askSam compresses the incoming information down in size and the seven megabyte database at Scripps Library represents an even larger amount of information when viewed in the original ASCII format.

The San Diego marine database at Scripps Library merges references from many separate databases under one common search interface. This is the primary attraction from a management perspective; information in differing formats is made searchable through one interface. In general, references cover the ocean environment of San Diego and Southern California including pollution and sewage in San Diego. Coverage is extended to freshwater as it relates to the marine environment in San Diego. References come from the sources listed below; as relevant sources are identified in the future, they can be readily added.

- 1: ECOLOGY OF SOUTHERN CALIFORNIA BIGHT and CALIFORNIA FISHERIES ECOLOGY databases produced by US Minerals Management Service's Pacific Outer Continental Shelf Office. These two databases were

asksam.txt

received on disk as ASCII files in SciMate database format. One database's SciMate references were readable as received with all of the source information being present in one field ie journal name, volume, issue, page numbers, and date. This database was imported into the San Diego marine database without change. The other SciMate database had its source information in separate fields ie journal name field, volume number field, issue number field, pagination field, date field. AskSam's high-level programming language was used to manipulate and rearrange the SciMate fields into a more-readable format. After reformat, the references were imported into the San Diego marine database.

- 2: COASTAL PROCESSES LITERATURE INVENTORY database produced by US Army Corps of Engineers' Coast of California Storm & Tidal Waves Study Project. This database arrived on floppy disk as a search-only (run-time) database written in Dbase software. The Inventory's references and their geographic and subject index terms were organized as a group of relational databases in Dbase format. Dbase references are unreadable in native format and, due to the relational design of the Inventory, there was no central database comprising the Inventory references in their entirety. To circumvent this problem, a global search was executed on the search-only Dbase software in order to output all of the Inventory references. Before executing a print command to print the entire database, the printer output was redirected to a disk file using a commonly-available public-domain utility software (PRN2FILE). In this way, very-readable references were redirected from the printer and were written onto an ASCII file on the hard disk. These references were then imported into the San Diego marine database using askSam.
- 3: SAN DIEGO COASTAL POLLUTION BIBLIOGRAPHY produced by Hubbs-Sea World Research Institute and University of San Diego Marine Studies Program. This database existed on a Hubbs staffperson's microcomputer as a DataPerfect database. At the request of Scripps Library, the staffperson output the DataPerfect references in a field-tagged format. The references were then rearranged using askSam's high-level programming language into a readable format for import into the San Diego marine database.
- 4: Eighteen biology, environment, energy, engineering, and policy databases were searched on an online databank or on CD-ROM for San Diego references. CD-ROM databases are particularly desirable for the initial search involving a large number of references with the equivalent online database suitable for updating. Contacting colleagues at other institutions with CD-ROM databases of interest can prove advantageous. The date of these searches were noted so that these databases can be searched

asksam.txt

in the future in order to update the San Diego marine database.

In general, askSam is full-featured and powerful but daunting at first appearance. AskSam receives good reviews but is an acquired taste. AskSam has an archaic programmers-look and does not have a contemporary visual design with drop-down menus and opening windows. The askSam search-only interfaces are far easier for the novice user but still not as intuitive as one would wish. AskSam's documentation is extensive and frequently confusing; explanation can be insufficient and page-flipping may be in order to find needed information in another section. For simple creation of a local database and subsequent import of references into it, askSam is easy to use and details are given below. AskSam's high-level programming language can successfully rearrange or reformat references with field tags. The askSam command language takes patience and trial-and-error to achieve success; documentation could be better. The author views askSam as an evolutionary step in managing the San Diego marine database and is always on the lookout for something better; a user-friendly search interface is a critical necessity.

Importing a file of references into an askSam database is best accomplished by ensuring that each reference is bracketed by a "record delimiter". A record delimiter is the demarcation between individual references and is used to chop up an ASCII file of references into individual references. While askSam can be used for full-text searching of text without delimiters between references, it is desirable to chop a file into individual references so that irrelevant references do not appear onscreen preceding and following the relevant reference being retrieved. AskSam can recognize line spaces as record delimiters but it is best to insert a character-based record delimiter using the global search-and-replace feature of word processing software. Line spaces can appear in unexpected places and it is preferable to maintain control over specification of individual references through use of a delimiter. AskSam can recognize a one or two character record delimiter; it is preferable to use two characters that are unlikely to appear elsewhere eg @#. Insert the two-character delimiter at the appropriate interval between references using the global search-and-replace feature of software. If each reference starts with a field name like AU (standing for "author"), then search for all occurrences of "AU" in a file of references and replace with "@# AU". If each reference is separated by several line spaces, then search for several hard-returns and replace with "@#" bracketed by hard-returns. If possible, avoid a search-and-replace based on two hard-returns since this can occur in the middle of a reference (and would appear as one line space in that reference). Always keep backup files so that new stratagems can be tested for inserting record delimiters; it may take trial-and-error.

#### asksam.txt

After record delimiting, use askSam to CREATE NEW FILE (create a new database) or FILE SELECT (open an existing database). Next, use askSam's MODIFY MODES function to specify that incoming references are a straight ASCII file (not askSam's unique format) and that references over twenty lines long (the size of an onscreen display) have each twenty-line fragment linked (DOCUMENT mode). Next, askSam's IMPORT function is selected, the incoming import filename is specified, and DOCUMENT TERMINATION is specified as "@#". IMPORT executes quickly. After successive files of references are imported into the local database, it is best to run a disk optimizing or defragmenting software so that the local database is written contiguously on the hard disk. Commercial disk defragmenting software include PC TOOLS, NORTON UTILITIES, MACE UTILITIES, and DISK OPTIMIZER. Writing a large database contiguously on the hard disk will enhance askSam search speed.

After creating a local database, it should be implemented for public access with a search-only or run-time version of askSam. AskSam has three search-only versions: HyperSift, HyPeruse, and InfoSift. InfoSift is the simplest search interface and offers multiple word searching with Boolean commands (AND, OR, NOT). HyPeruse has single word searching without Boolean commands for linking multiple words; HyPeruse does have a hypertext function. The askSam hypertext function is enabled when the searcher uses the cursor keys to move a highlighted box down through onscreen text to a desired word; press the ENTER key and that word is searched. This hypertext feature is nice but not earth-shattering; it selects a single word displayed onscreen which could have been typed faster than moving the highlight box down to it. The lack of Boolean commands in HyPeruse makes it unsuitable for searching a bibliographic database wherein linking multiple search words with AND, OR, and NOT is usually necessary.

HyperSift is currently being used at the Scripps Library for public access. HyperSift offers multiple word searching with Boolean commands, hypertext highlighting of an onscreen word for subsequent searching, and additional search commands like the VICINITY command for specifying word proximity. HyperSift's word proximity searching using askSam's VICINITY command is too advanced and cryptic for casual infrequent searchers. InfoSift's basic features are adequate for almost all bibliographic search situations: simple searches on multiple words with Boolean commands. InfoSift is the better choice for the cost-conscious since it costs US\$25 less. Purchase both InfoSift and the older askSam version 4.2 to put together an askSam-based database system for US\$80 plus shipping/handling.

The askSam software or a search-only software like InfoSift or HyperSift can be loaded from the DOS command line with a specified database. A batch file was created at the Scripps Library so that

asksam.txt

introductory screens explain database contents and basic search features followed by HyperSift loading with the San Diego marine database. At startup, HyperSift displays the first reference in the database. To search word(s), use cursor arrows to move the highlight box down to the colon on the askSam query line. At the query line, search words are typed and the first typed letter must over-type the colon on the query line (another user-unfriendly feature of HyperSift). If the colon is left onscreen and search words typed after it, then the search will not be executed. HyperSift's unintuitive necessity to move a highlight box down the screen into the query line in order to begin typing search words is the primary deficiency of the HyperSift interface. Ideally, the searcher should be placed automatically at the query line wherein search words are typed. InfoSift is being purchased by the Scripps Library to investigate its greater simplicity compared to HyperSift; a review (Duffy, 1987) indicates that the searcher is placed at the prompt to type a search request. This small change would make InfoSift much more intuitive than HyperSift for the novice user.

Word truncation is signified with an asterisk, eg COAST\* retrieves COAST, COASTS, COASTAL. The Boolean AND is implied between two or more words. Typing {OR} between words indicates that any word should be retrieved eg ABALONE\* {OR} HALIOTIS. Typing {NOT} between words will exclude the word after {NOT} from search results, eg PLANKTON\* {NOT} DIATOM\*. The OR and NOT Boolean commands are typed within braces so that askSam will recognize them as Boolean commands and not search them as words. After the first reference matching the search criteria is retrieved, press the ENTER key to view successive references. If a reference is longer than the twenty lines appearing onscreen, press the PGDN cursor key to view successive screens for that reference; abstracts and descriptors may be extensive and continue on successive screens. If askSam locates the search term(s) in the middle or ending screens of a multiscreen reference, press the PGUP cursor key to page up to the first screen of that reference which typically has the bibliographic citation. Any word appearing onscreen can be hypertext searched by moving the highlight box down to that word and pressing the ENTER key. To print, move the highlight box to PRINT; to quit, move the highlight box to QUIT.

This paper details the operational constraints of the Scripps Library's San Diego marine database and the selection of askSam and HyperSift to manage it. While certainly not an ideal database on the San Diego marine environment, the San Diego marine database involved minimal effort in bringing a variety of pre-existing information together under one search interface. This is its main virtue; if the effort was not minimal, the database would not exist due to other priorities. The database's deficiencies are several. It is not integrated with the library catalog which is remotely accessible and

asksam.txt

supports multiple users. The San Diego marine database currently resides on a stand-alone single-user microcomputer; future planning may make it accessible over the campus network for remote and multiple users. The San Diego database is passive in acquiring references and thus not comprehensive in coverage. The database offers access only to information available elsewhere in electronic format; no original information is keyboarded at this time though that is always a possibility. The database's coverage of local government reports could be stronger but this would require a strong staffing commitment since this information is the most difficult to acquire. Liaison with government departmental librarians is planned in order to identify electronic records for incorporation into the San Diego marine database. Duplication of references exists in the San Diego marine database because references are accepted from several sources. Indexing for subject or geography is not consistent and varies depending on the source of the references. Abstracts are not consistently available in the database. Since searches are full-text, certain words are problematic ie FISH since they may appear in non-subject fields ie journal title abbreviation. Even with these deficiencies, the San Diego marine database has been successful in serving library users. Local information is the most difficult to identify and to find and is frequently requested by library users. This irony of reference service lead to the selection of askSam and HyperSift for the San Diego marine database at Scripps Institution of Oceanography Library. It may not be an ideal database or database software for coverage of the marine environment in San Diego but it is a substantial step in that direction.

References:

Duffy, Richard. "InfoSift: it is a small, easy, fast version of askSam". PC WEEK 4(45):129, November 10, 1987.

Grunin, Lori. "A \$75 runtime version of askSam". PC MAGAZINE 8(2):54, January 31, 1989 [reviews HyperSift].

Kittle, Paul W. "askSam for your data! A look at text-based management". DATABASE 12(1):99-101, February 1989.

Marshall, Patrick. "AskSam 5.0 adds databaselike relationality". INFOWORLD 13(32):67-71, August 12, 1991.

Nielsen, Brian. "askSam: fitting a tool to a complex job". DATABASE 14(1):78-80, February 1991.

Perez, Ernest. "AskSam database features strong flexibility, speed". INFOWORLD 11(16):78-86, April 17, 1989.

asksam.txt

Pruett, Nancy Jones. "Using askSam to manage files of bibliographic references". ONLINE 11(4):46-52, July 1987.

Rubenking, Janet. "askSam adds programming oomph to strong, but cryptic app". PC MAGAZINE 10(10):49, May 18, 1991.

Van Name, Mark L and Bill Catchings. "HyperSift: the ticket to askSam". PC-COMPUTING 2(3):39-40, March 1989.

Westman, Stephen. "Database manager/text search program, askSam, version 4.2". LIBRARY SOFTWARE REVIEW 10(4):281-284, July-august 1991.

Yakal, Kathy. "Information management: askSam ver. 5.0". PC SOURCES 2(6):309, June 1991.

-----