

LOCATING INTERNET RESOURCES IN THE AQUATIC SCIENCES

Andrea L. Duda
Networked Information Access Coordinator
E-mail: duda@library.ucsb.edu
James W. Markham
Biology Librarian
Davidson Library
University of California
Santa Barbara, CA 93106
E-mail: markham@library.ucsb.edu

ABSTRACT: Librarians at the UCSB Davidson Library are developing home pages in their subject areas of expertise. This paper describes the process the library has used to develop these pages, discusses the use of search engines, and considers methods of organizing web pages.

INTRODUCTION

The UCSB Library opened the InfoSurf gopher to the public on March 1, 1993. The gopher included information about the library, library publications, and links to resources in a variety of subject areas.

With the increasing popularity of the world wide web, it soon became clear that the library also needed web access. Construction began on the library's web <URL:<http://www.library.ucsb.edu/>> in May 1995 and it opened to the public with the start of fall classes in October.

Where gopher maintenance lent itself to being done by an individual, the world wide web allowed for more distributed development. Each collection manager in the library was given the assignment to create a page with appropriate resources for that subject area. Currently 27 librarians are developing pages on 65 different subjects.

Collection managers knew that classes would come during the summer and had time to start saving references to interesting web sites. At the same time, the library's Networked Information Access Coordinator developed a template for each subject area to be included on the web.

The Networked Information Access Coordinator and her committee of WebWeavers taught a series of three classes on how to write HTML documents for the web. Classes were taught in the library's computerized classroom and featured hands-on instruction. Participants used the first

two classes to begin developing their own home pages; in the third class they worked on their subject pages. The first class was an introduction to the web and the basics of HTML; the second class covered various kinds of lists (ordered, unordered, dictionary); the third class gave participants time to start adding resources they had collected to their pages. While these classes did not cover all the nuances of HTML, they did cover the basics that people would use most frequently. The classes were offered three separate times over the summer to accommodate vacation schedules.

While web pages are continuously under construction, resources have now been made available for the majority of subject areas.

The Biology Librarian was responsible for two web pages. As of August 1995, UCSB has two biology departments. One is the Department of Ecology, Evolution and Marine Biology (EEMB) and the other is the Department of Molecular, Cellular and Developmental Biology (MCDB). Accordingly, the library has two biology collection guidelines, EEMB and MCDB, and pages were created for each. Most of the aquatic science links appear in the EEMB web page, so that will be used for illustration here.

LOCATION OF RESOURCES

There are probably four principal means of locating resources on the world wide web. The first method is simply to collect tips from everyone, in conversation, in print, and by e-mail. E-mail is preferred because in a Windows or Macintosh environment, the URL can be copied and pasted into a search statement without rekeying or making mistakes. It is very productive to encourage a cooperative effort so that everyone searching the Internet shares any interesting discoveries.

The second method is to search the web pages and sites that others have already created and mine them for desirable items for the web page being built. For the most useful web pages, of course a link will be provided to the web page itself.

The third source is mailing lists, such as IAMSLIC, and newsgroups such as Net Happenings (comp.internet.net-happenings).

The fourth and most productive, as well as time consuming, method is to use search tools.

These are the search tools currently available on InfoSurf's "Tools for Locating Resources" page <URL:<http://www.library.ucsb.edu/tools.html>>.

Veronica

[gopher://liberty.uc.wlu.edu:70/11/gophers/veronica/](http://liberty.uc.wlu.edu:70/11/gophers/veronica/)

Veronica is a tool for searching gopherspace. It searches titles of gopher menu items and is dependent on titles provided by gopher developers. Veronica searches may include Boolean operators and searches can be limited by type of item (individual files, directories, binary files, etc.)

Feature	Availability/Comments
Boolean operators	Yes
Truncation	Yes, use *
Update frequency	Monthly
Fields searched	Menu titles
Results	List of gopher menu items

Alta Vista

<http://www.altavista.digital.com/>

Alta Vista includes 30 million pages found on 225,000 servers. The search system is very flexible. A user can enter one or more search terms and find them anywhere in the document; surrounding the words with quotation marks allows the phrase to be searched. Words that must be included in the search are indicated with a +; words to be excluded are indicated with a -. Searches can be limited to document titles, URLs, and other parts of the record.

Feature	Availability/Comments
Boolean operators	Yes; the advanced search feature allows "and," "or," "near," and "not." The simple search feature allows the use of + for terms that must appear and - for terms that must not appear
Truncation	Yes, use *
Update frequency	Continuous
Fields searched	All words in a document
Results	Title, first several lines of text on page, URL, size, date

InfoMine

<http://lib-www.ucr.edu/>

InfoMine is a virtual library housed at UC Riverside. Librarians from throughout the UC system select resources to add to InfoMine. Resources are divided into several libraries -- Biological, Agricultural, Medical Resources; Government Information; Physical Sciences, Engineering,

Computer and Math Resources; Social Sciences, Humanities, and Arts; Internet Enabling Resources; Internet Instructional Resources. Searches default to keywords, subjects, and titles.

Feature	Availability/Comments
Boolean operators	Multiple terms are combined with "and"
Truncation	Yes, use #
Update frequency	New records are being added continuously
Fields searched	Title, Keywords, Subjects
Results	Title and description of matching items

InfoSeek Net Search

<http://www2.infoseek.com/>

InfoSeek presents the searcher with a simple dialog box in which to enter words to be searched. InfoSeek does not permit Boolean searching; searches must be either words or phrases. Results of a search show the title of the page and the opening text on the page.

Feature	Availability/Comments
Boolean operators	No; multiple terms are treated as phrases
Truncation	No
Update frequency	Unknown
Fields searched	Text on page
Results	Title, URL, first several lines of text on the page

Lycos

<http://www.lycos.com/>

Lycos collects web addresses, the links within documents, titles, headings, subheadings, plus the 100 highest weighted words, plus the first 20 lines of text. Lycos provides a simple option to enter a search term. The results show the URL, title, keywords, and an excerpt from the page.

Feature	Availability/Comments
Boolean operators	Yes (default for multiple terms is "or")
Truncation	Defaults to a substring search
Update frequency	Unknown
Fields searched	Keyword, title, URL, excerpt
Results	URL, title, keywords, abstract

Open Text Web Index

<http://www.opentext.com:8080/>

The Open Text Web Index lets you search every word of every web page that they have indexed for a particular word, or phrase, or combination of words and phrases. The Open Text Web Index has indexed the full text of close to one million web pages.

Feature	Availability/Comments
Boolean operators	Yes; the default is to treat multiple words as a phrase; "and" and "or" are available options
Truncation	Single words are treated as substrings
Update frequency	"We're spending most of our capacity indexing new pages, and somewhat less revisiting old pages"
Fields searched	Every word on every page
Results	URL, title, first several lines of text on the page

SavvySearch

SavvySearch searches multiple search engines at once. It simplifies searching by providing a common interface, but loses some of the special features of the different search engines.

Feature	Availability/Comments
Boolean operators	Yes; options to use "any query term" or "all query terms"
Truncation	Yes, use *
Update frequency	Depends on the individual search engine
Fields searched	Depends on the individual search engine
Results	Title, first several lines of text on the page

WebCrawler

<http://www.webcrawler.com/>

The WebCrawler records web addresses (URLs), keywords from the contents of the documents, and keywords from the document titles and hotspot text (the clickable words in links). The WebCrawler Project began as a research project at the Department of Computer Science and Engineering at the University of Washington in Seattle. The project has since moved to America Online, where it continues to operate as a freely available Internet search tool.

The WebCrawler uses relevance rankings for each document that is retrieved. To compute the relevance number for a particular document, WebCrawler takes the total number of times each of

the words in your query appears in the document and divides it by the total number of words in the document. This is computed for each document that contains all of the words in your query. The document with the largest number computed in this fashion is arbitrarily assigned a score of 100, and the scores of the remaining documents are scaled to it.

The WebCrawler database is about 100MB and contains information on over 190,000 different documents that the WebCrawler has explored.

Feature	Availability/Comments
Boolean operators	Yes, "and" and "or"
Truncation	Yes; default is a substring search
Update frequency	Weekly
Fields searched	Most words in the document; does not search common words like "web" or combinations of letters and numbers
Results	Title and relevancy ranking

WWW Virtual Library

<http://www.w3.org/hypertext.DataSources/bySubject/Overview.html>

The Virtual Library is a subject catalog of Internet resources. The catalog is distributed among servers worldwide. Items added to the virtual library are selected by the individual maintaining the particular subject area and are of generally good quality. The Virtual Library includes fewer resources than the general indexing services, but they are selected references of good quality.

Feature	Availability/Comments
Boolean operators	No; this is a subject-oriented collection that does not include a search capability
Truncation	No; no search capability
Update frequency	Each library area is updated individually
Fields searched	No search capability
Results	Each subject area is different

Yahoo!

<http://www.yahoo.com/>

Yahoo! is a hierarchical subject-oriented guide for the world wide web. Yahoo! lists sites and categorizes them into appropriate subject categories. Most of the entries in Yahoo! come from user submissions. Some links are added through an automated search robot.

Feature	Availability/Comments
Boolean operators	Yes; "and" and "or"
Truncation	The default is a substring search
Update frequency	Unknown
Fields searched	Title, URL, Comments
Results	Title; may include a description

ORGANIZATION OF RESOURCES

There are many ways to organize the resources located. It can be useful to look at a variety of other web pages to gain an idea of possible organization schemes. The type of resources located will make a difference in how they should be organized. As updating continues, new categories may be located that must be incorporated and subsections may be broken out of larger categories. Better organization schemes may also be located.

One category of resources will be links to other major sites. These have been called by various webmasters Gateways, Indexes, Collections, Super Sites, Megasites, or various other names.

When designing a page for aquatic biology, it seems important to provide links to the home pages of organizations or institutions that deal with aquatic science, so a lot of the links we have provided are to various organizations, which in turn provide other useful links.

There are also links to data sets, some of them interactive. On this home page we have provided a link to the latest tides for several areas, which we organized along with a site for wave information under the general heading **Oceanography**.

Another set of sites provide access to images of various organisms, which we have organized at the moment under the broad categories **Botany** and **Zoology**.

However the materials for the home page are located and organized, it must be kept in mind that the page cannot remain static, but will always be somewhat under construction, so that the schemes of locating and organizing resources reported this year may be very different next year.