



# Searching for Images on the Web

Search software solutions now exist to help locate image and audio files

Searching for "information" usually implies searching for *textual* information. Generally speaking, for example, a textbook on information retrieval will probably discuss the principles and methods of retrieving text. A course on database searching is likely to discuss text files and ways to access them. Textual is typically the implied qualifier for information, searching, and databases. This certainly has to do with the fact that cataloging, organization, indexing, and abstracting of pictorial (let alone audio) information lags light-years behind the same processes for textual information.

No one would dispute that the Internet—especially its World Wide Web—has revolutionized the publishing of and access to textual, audio, and graphic information alike, though to a different extent for each. Information in itself still implies text, even though the most characteristic attribute of the Web is its graphic, visually appealing (sometimes appalling) form. While there are literally hundreds of search engines and meta engines that offer sophisticated ways to search the Web for textual information, such tools for searching image and audio files are much less common.

Likewise, articles about Web search engines are a dime a dozen, but they rarely discuss searching for image and audio files. And comprehensive treatises about the topic do not exist, with the remarkable exception of Paula Bernstein's book published recently by Pemberton Press (see sidebar).

In the upcoming installments of Multimedia Medley, I will be looking at various search software solutions that facilitate access to image and audio files on the Web and on CD-ROM. As for online images, I refer you to Bernstein's book, as I'm going to concentrate primarily on image search tools that appeared after her book went to press or that did not fit her book, i.e., CD-ROM search engines.

## Images on the Web

What types of images are there on the Web? Every type that you can imagine! Photos and paintings, blueprints and architectural renderings, logos, trademarks and patent drawings, clip art and cartoons, maps, and flags. The problem is that they are scattered across hundreds of thousands of Web pages, some with tens of thousands of images, others with just one or a few. But you may need just that one reproduction of a famous painting, or that shot of a particular landmark.

How do you find it? There are more ways than one.

## Web-Wide Browsing and Searching

It is fairly easy to get to some of the well-known image collections using Web directories. The classified Web directories

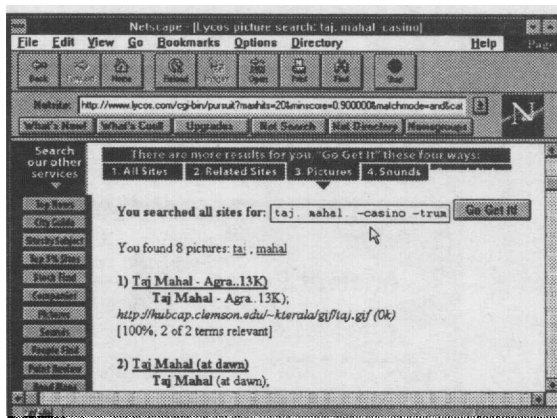


Figure 1: Lycos

may guide you to such image warehouses as Art Today, with its 500,000 images—if you realize that you have to follow the directory path "Business and Economy/Companies/Computers/Software/Graphics/Clip Art." Similarly, you can get to the gigantic stock collection of PhotoDisc through the path "Business and Economy/Companies/Photography/Stock Photography." However, with the exception of Yahoo!, these directories don't have sophisticated cross-referencing tools, and you may get lost clicking your way through the maze. Images sites can be classified under hundreds of different categories.

## Textual Query with Image Qualifier

You may also do a quick-and-dirty search for Web sites and Web pages using the query "image collection" or "stock photographs," but you will have to wade through a lot of irrelevant entries in the results list. You can easily miss important Web sites unless you formulate a really comprehensive query that includes the appropriate synonyms (pictures, photos, gallery, portfolio, etc.). This, on the other hand, may significantly decrease the precision of your search unless you use advanced search operators to specify mandatory terms and their location (such as the title of the Web site).

Considering the graphic emphasis of the World Wide Web, it is ironic that for quite a time you could limit your query terms in many of the Web-wide directories to the title field, the URL address, the first paragraph of a Web page, or to its summary, but—with one exception—there was no option in these finder tools until recently to limit your search to caption fields or to Web pages with images. The exception was AltaVista, which allows users to specify a search for Web pages where the keyword is part of the file name in a hyperlink, such as "image:pyramid." As only a single keyword can be used, it is far from perfect. Searching for an image about the White House, Big Ben, or the Temple of Dawn is, for example, quite hopeless.

Lycos recently introduced a feature that lets you restrict your search to images (or sound). As the search expression can

use multiple terms and even exclude terms, as in "Taj Mahal -casino -hotel -Trump," the searches can be efficient, if not perfect. The above query still would retrieve images about Taj Masjid, or Hawa Mahal, or even Mahalia Jackson. You can help the implicit truncation by putting a dot after each term (Figure 1), but you cannot specify in Lycos that both Taj and Mahal must appear. You might add Agra and/or India to the query to refine it, but extra terms may not always help.

Among the meta search engines that send your query to several search engines, SavvySearch has an option to mark in a checkbox the sources and/or types of information (e.g., images) that you are looking for. But the options for formulat-

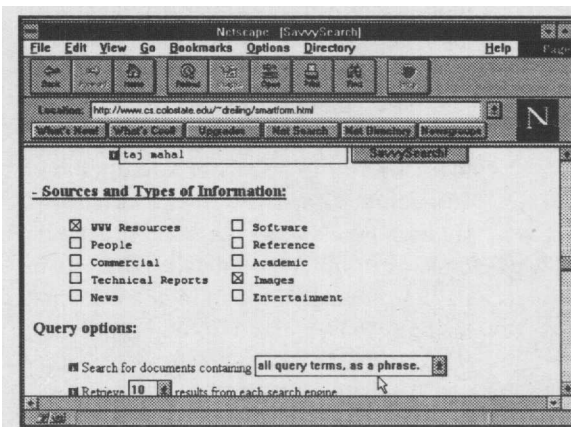


Figure 2: SavvySearch

ing the query are not comprehensive enough (Figure 2). With SavvySearch, the problem is the opposite of the limitation in Lycos. You may specify that all query terms must appear, or that the query be interpreted as a phrase, but you cannot exclude terms. My hits included a lot of irrelevant items about Taj Mahal the musician and Taj Mahal restaurants, hotels, and casinos.

HotBot, another meta search engine, offers a far more powerful set of options. You not only have full-blown Boolean operators for the textual query, but you can also specify the extensions of the image files, such as gif, jpg, and TIFF (Figure 3). Of course, the meta search engines cannot deliver better results than the underlying search engines, so be prepared for false hits as well as misses.

## Searching by Image Attributes

The problem with all of the above is that there is no "guarantee." You can't be sure that you'll find a picture of the object you specified on the Web pages you retrieve. There may be a photo of the creator of the Web page who may have men-

tioned the Taj Mahal, but that's not what you were looking for. These anomalies are avoided by two interesting search tools that are based on a thumbnail collection of images. These are retrieved, classified, and indexed from Web pages rather like site identifiers, and lead paragraphs are retrieved by traditional Web search tools such as Lycos, InfoSeek, and Excite.

Yahoo!'s Image Surfer (licensed from Interpix Software) allows you to browse thumbnail images in predefined categories and subcategories (Arts/Photography, People/Actors and Actresses) and then either jump to the original site (if it is still up), or look for similar thumbnail images. Similarity is defined on the basis of similar colors and color ratios, not the shape and form of the object, let alone the subject of the photo.

It's quite an interesting technology. Images can also be searched directly without selecting categories, using the similarity matching steps. However, the topical search options are not satisfactory. Although the features of the main Yahoo! search engine were significantly enhanced in late December, the improve-

ments do not apply to the Image Surfer, which uses a simple term-weighting algorithm. I tried to find pictures of the fabulous sunsets in Manila Bay for a presentation, but all I got was a set of 40 photos of sunsets anywhere. I had to wade through nine pages of thumbnails before I realized that none of the ones I was looking for were among them. Still, this is a good start for direct image searching.

A brand new site tool, WebSEEK at Columbia University, is based on similar concepts and seems to have a far larger thumbnail collection. The site does not have the usual publicity blurb, so I'm just guessing from the URL that it may be sponsored by Disney. I estimate the collection at 250,000 to 300,000 thumbnails from all categories and subcategories of

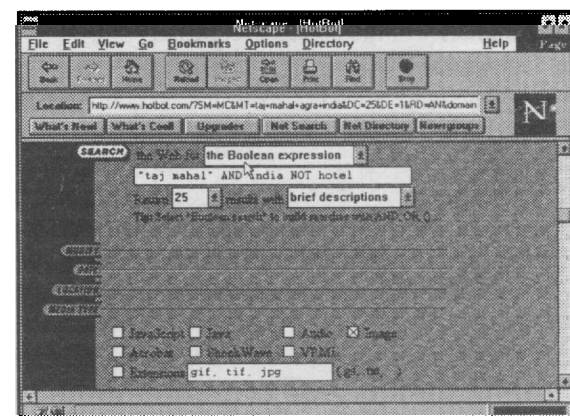


Figure 3: HotBot

topics. It is a content-based image and video search and catalog tool for the Web. It also combines a subject directory and search engine like Yahoo!'s Image Surfer. You may browse through categories and subcategories or search by keyword. Fifteen thumbnails are displayed on the screen at once (Figure 4). You may select

(continued on next page)

## NEED A NEW IMAGE? THIS BOOK'S FOR YOU

**Berinstein, Paula. *Finding Images Online***  
 Wilton, CT: Pemberton Press, 1996. \$29.95

This is an exceptionally well-written, well-structured, and well-illustrated book that fills a niche left empty by the hundreds of books about searching online information sources. Beyond offering a crash course on digital image formats, visual literacy, and copyright, the book presents an in-depth analysis and comparison of the alternative sources of digital images and the tools and options of searching for images. In an impressively consistent structure and with equally impressive competence, Berinstein discusses the pros and cons of using the consumer online services (CompuServe, AOL, Microsoft Network), the professional online services (DIALOG, LEXIS-NEXIS, STN, Questel-Orbit), the various ftp, gopher, Usenet, and World Wide Web sites in general, and the most important image warehouses accessible on the Internet, such as Kodak's Picture Exchange, the Eastman Exchange, Publishers Depot, and Muse.

For each, the author discusses the coverage, characteristics, charging schemes, and browsing and/or search options, and provides tips and informative case studies.

The author's enthusiasm for the subject is well complemented by her realism and street-smart advice. The 400-page book has a substantial and current subject index to online image resources that is as well done as the index to the book itself. By the time you read this, there will be an online update to the book on the publisher's Web site (<http://www.onlineinc.com/pempress>). It is meant to provide information about new and enhanced sources for those who want to search for images. Those who read the book will not only search for images online but will find them—just as the title of the book promises. The price of the book is sure to be offset by the amount of time you save after just a few searches.

—P. J.

## Multimedia Medley

(continued from previous page)

any of them and initiate another search in the whole WebSEEK catalog or in the category to find similar images.

Again, similarity is based on color

and thumbnails only for those sites where the images are stored in Web pages. If the images are stored in a database, they are not accessible to the Web-wide search tools. This is similar to the limitations of the traditional, text-oriented tools. Just as records within SilverPlatter's or Ovid's databases need their own information retrieval software, so do the image databases on the Web sites of the largest museums, stock photo agencies, and some other art-related sites. Some of these have stunningly powerful features that I will discuss next month.

*Péter Jacsó is associate professor of library and information studies at the*

*University of Hawaii. He writes for this and other professional magazines, speaks at professional conferences, and regularly offers his online/CD-ROM workshop series. His e-mail address is [jacsop@hawaii.edu](mailto:jacsop@hawaii.edu).*

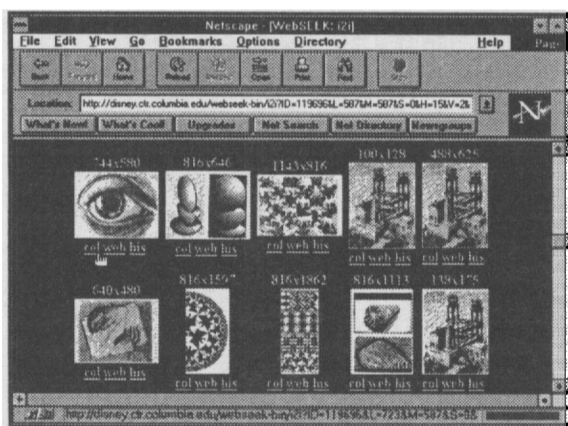


Figure 4: WebSEEK

composition just as in Image Surfer, not on the similarity of the object. WebSEEK may find Leonardo's torso of a man more similar to Escher's famous "Hand Drawing a Hand" than to Leonardo's anatomical sketch about a woman's hand (which may have inspired Escher, by the way). I would be more interested in a shape-recognition algorithm to find related works of arts, but color matching seems to work very well.

In WebSEEK, you can even tweak the color histogram of the sample picture to refine the search criteria. It is not unlike assigning different weight factors to search terms in a text query. WebSEEK could be a good tool for text-based image searches even without the image-matching capability, but the single keyword search capability is not enough for this. While you may limit your search to video, color photos, gray images, and graphics, you cannot choose two or more categories at the same time. Despite the single keyword limitation, this is an impressive search tool.

### WEB SITE REFERENCES

**AltaVista**

<http://www.altavista.com>

**Lycos**

<http://www.lycos.com>

**SavvySearch**

<http://www.cs.colostate.edu/~dreiling/smartform.html>

**HotBot**

<http://www.hotbot.com/index.html>

**Image Surfer**

<http://ipix.yahoo.com>

**WebSEEK**

<http://disney.ctr.columbia.edu/WebSEEK>

## Xiphias Corp. Releases DVD-ROM Encyclopedia

Xiphias Corp. has announced the release of Encyclopedia Electronica, a DVD-ROM that features Omniview, Inc.'s new PhotoBubble imaging technology. The digital encyclopedia debuted in November 1996, and, according to Xiphias, it is the first product built from the ground up for DVD-ROM, MMX technology, and Internet exploitation.

Whether over the Internet, on CD-ROM, or on DVD-ROM, PhotoBubble technology enables users to navigate a scene via mouse or keyboard input, viewing all aspects of an image within a 360-degree panorama. "We've selected a brilliant new technology called PhotoBubble imaging, which will revolutionize how people think of pictures," said Peter Black, president of Xiphias. "The PhotoBubble technology allows 360-degree viewing of photographic images. It provides a knockout visual experience."

Prepared in collaboration with Toshiba, Intel, and WEA (Warner Electra Asylum) Manufacturing, the software package created by Xiphias exploits the full 4.7 GB of space on a DVD-ROM. It delivers 30 PhotoBubble images, more than 80 minutes of MPEG-II video, and AC3 multichannel audio integrated into a digital encyclopedia. The product will be optimized for Intel's new generation of MMX technology.

PhotoBubble images are created by taking two opposing photographs with a standard fisheye lens and joining them via computer image remapping. The Omniview Web site (<http://www.omniview.com>) provides the Photo-Bubble viewer free for downloading as well as sample PhotoBubble images.

Source: Omniview, Inc., Knoxville, TN, 800/916-6664, 423/690-5600.

## IDC/LINK Publishes Internet Telephony Report

According to "Internet Telephony: An Alternative Dial-tone?," a report from IDC/LINK (a subsidiary of International Data Corporation), Internet telephony entered the market in 1996 as a clever client software application riding the universal Web browser success. In 1997, Internet telephony is a key driver of telephone "minutes of use" (MOUs) in the long-distance and international-telecommunications market. In addition, real-time telephony over the Internet is becoming more than a stand-alone product. It is catalyzing the development of multimedia telephony applications (MTA) such as whiteboards, application and document sharing, multiuser data conferencing, and ultimately real-time video. According to the report, which was published in January 1997 by IDC/LINK's Residential Broadband Service program, this will lead to new revenue

streams and opportunities and will account for 12.5 billion MOUs by the year 2001.

"Multimedia telephony applications have appeal to business and consumer market segments," said Rona Shuchat, author of "Internet Telephony" and manager of IDC/LINK's Residential Broadband Services research program. "The MTA scenario will permit live interaction with an operator or interaction with voice-messaging systems, while viewing the products and simultaneously placing orders online. This has applicability to customer service, help desk support, electronic shopping, and distance learning."

The report is available for purchase by contacting Janis Dempsey at 800/343-4952.

Source: International Data Corp., Framingham, MA, 508/935-4764.

## Gale Releases History CD-ROMs

Gale Research has announced the release of two new educational CD-ROMs for students: *DISCovering World History* and *DISCovering U.S. History*. According to Gale, the CDs led students to a critical understanding of the past through an exploration of the interplay between key persons, places, and events.

*DISCovering World History* features thousands of in-depth essays dealing with historical periods, with special attention paid to often-neglected events in Asian and African history; social history; notable individuals; and movements or events in disciplines such as the arts, reli-

gion, philosophy, and technology. *DISCovering World History* also features illustrations, a searchable timeline, a glossary, maps, primary documents, and more.

*DISCovering U.S. History* includes many of the same features, as well as primary source documents such as the Declaration of Independence and the Emancipation Proclamation. The Windows stand-alone version of both discs is \$500, and the LAN price (for two to eight users) is \$700.

Source: Gale Research, Detroit, 800/347-4253, 313/961-2242; <http://www.gale.com>.