INFORMATION TECHNOLOGY

Plucking Images From the Digital Floods

Consider this apparently simple request: Find a picture of yellow flowers on the Internet. The simplicity is deceiving, for at present this can't be done—unless you're rescued by serendipity. Existing software can only sift through words, blind to the deluge of images, sounds, and video that increasingly swamps computer networks.

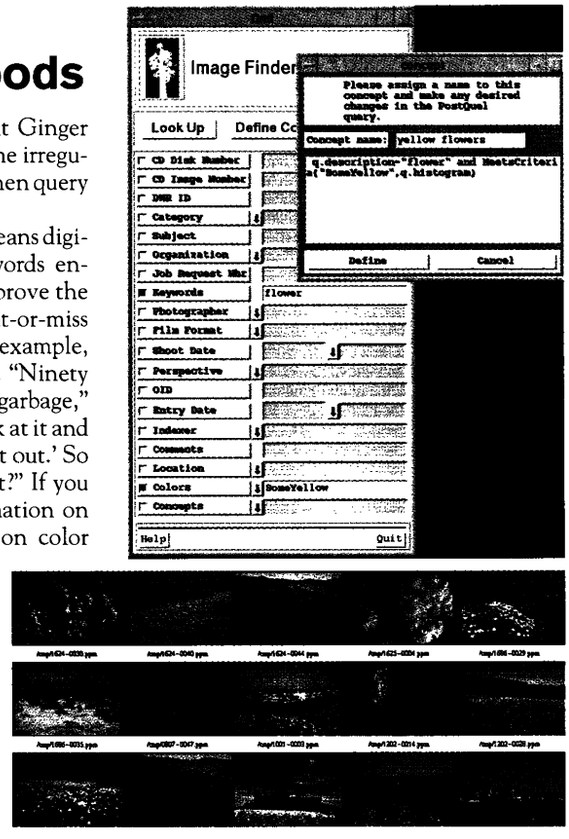
Building tools to filter drops of visual information from the digital flood is one goal of the ambitious National Science Foundation "digital library" program (*Science*, 7 October 1994, p. 20). Six university-led groups are trying to learn how best to sort, file, and retrieve information, creating pockets of digital order amid informational chaos. Two weeks ago in Atlanta, at the annual meeting of the American Association for the Advancement of Science (the publisher of *Science*), one of these groups showed that, when it comes to image searches, a picture can be worth a thousand descriptive key words.

Typically, image collections in libraries use such key words as tags for searches. But what a botanist sees as "yellow" might be "golden" to a cataloger writing the tags. As a result, the flower-gathering botanist will come up empty-handed. So when scientists at the University of California, Berkeley, began collecting images as part of their library of California environmental information, they decided to develop software that searched the images themselves.

By computing and storing the distribution of colors in an image, the prototype system allows a user to combine a key word search for "flower" with a further condition that the picture "contains some yellow." The computer delves through the archives for pictures of flowers, then examines each for splotches of yellow. Small versions of images matching these criteria pop onto the screen, and the user can click on them to retrieve full-size pictures. In contrast, says Robert Wilensky, principal investigator of the Berkeley group, a search for the key words "yellow" and "flower" with a text search program missed many of these images entirely. The original cataloger had not even entered the key word "yellow" for many of the images.

While not the first to design a system that searches through colors, "the Berkeley group is the first to apply these techniques to such a large, unified corpus of material and to hook it into a database system," says James Frew, one of the principal investigators on another digital library project, at the University of California, Santa Barbara. It's even possible for a user to scan in a picture of yellow flowers and then tell the computer to find "pictures that look like that." In the future the Berkeley system, written by grad student Ginger Ogle, may know that a "blob" is some irregularly shaped object; the user could then query for pictures with "blobs of yellow."

None of this experimentation means digital librarians have given up on words entirely. In fact, they also want to improve the way software uses text. In the hit-or-miss search through ordinary text, for example, hits can be as frustrating as misses. "Ninety percent of what you get back is garbage," Wilensky complains. "And you look at it and say, 'Hmm, this is garbage. Throw it out.' So why couldn't the computer do that?" If you use the word "film" to find information on movies, you'll also get back hits on color print film, thin-film chromatography, and maybe even soap film. What's more, the 10% that isn't garbage doesn't contain everything you need. For example, searching for the word "education" will miss most of the documents about education, as "the word 'education' doesn't occur in most of the articles about education," Wilensky says. "What occurs is 'school' and 'teacher.' "



GINGER OGLE/UC BERKELEY

Picking flowers. Digital library software, called Chabot, searches for yellow pixels in images rather than the word "yellow," and pictures of yellow flowers pop onto the screen.

The potential solution to both problems lies in semantics. A smarter system would understand relationships between words, recognizing that "education" includes "school" and "teacher." The search for "film" would include the notion of movies. The Berkeley system currently works with 4000 semantic concepts, and they hope to increase that number by an order of magnitude. There are unsolved problems—how, for instance, would a user sift through 40,000 concepts to find the desired one? But the hope is that one day picking yellow flowers on the Internet will be as easy as going outside on a summer day and picking them by hand.

–Kenneth Chang

Kenneth Chang is a science writer based in Santa Cruz, California.

PLANETARY SCIENCE

NASA Picks a Budget Lunar Trip

In the National Aeronautics and Space Administration's (NASA's) push for smaller, faster, and cheaper space missions, cheaper turns out to be a big winner. Last week NASA picked the least costly of 28 candidates for the next launch in Discovery, the agency's low-cost planetary exploration program. The $59 million spacecraft, known as the Lunar Prospector, is an orbiter for Earth's moon, and came in well below the mission budget cap of about $330 million.

"We got some dynamite proposals—very innovative and clever," says NASA Associate Administrator for Space Science Wesley Huntress. But NASA was most impressed with the return on the dollar promised by Lunar Prospector. Proposed by Alan Binder of the Lockheed Electronics and Science Corp., Prospector's price tag includes development, launch, and operations. The spacecraft will survey the composition of the moon's crust and its magnetic and gravitational fields to elucidate the crust's complex history. The craft will also search for surface ice and gas releases; ice deposits might be useful water sources during crewed exploration.

The probe, scheduled for launch in June 1997, will need to be impressive to match the cost-effectiveness of last year's Clementine lunar mission. That Department of Defense (DOD) project focused on testing Star Wars instrumentation, but provided a nearly free bonus of new scientific data from its lunar orbit. Congress will be watching to see if NASA can approach DOD's economy.

–Richard A. Kerr