The Right Stuff: Research Strategies for the Internet Age

By Nancy O'Hanlon
Ohio State University Libraries
Version Date: May 23, 2005

Tom Wolfe describes the unique qualities that enabled the first U.S. astronauts to meet unknown challenges as "the right stuff." For these explorers, "the right stuff" was not simply courage, but a unique combination of daring, skill, experience and a persistent determination to succeed:

> . . . the ability to go up in a hurtling piece of machinery and put his hide on the line and then have the moxie, the reflexes, the experience, the coolness, to pull it back in the last yawning moment -- and then to go up again the next day, and the next day, and every next day . . . A career in flying was like climbing one of those ancient Babylonian pyramids made up of a dizzy progression of steps and ledges, a ziggurat, a pyramid extraordinarily high and steep; and the idea was to prove at every foot of the way up that pyramid that you were one of the elected and anointed ones who had the right stuff and could move higher and higher . . . (Wolfe 24)

Research today challenges us to navigate effectively through Internet space. At one time, knowledge of the library and skill at using printed resources was sufficient preparation for finding useful information. Today, the Internet, particularly the World Wide Web, is the vehicle for most research. The sheer volume of information available through this medium is staggering, like the Babylonian pyramid in Wolfe's vision. To be a successful researcher also requires "the right stuff" -- imagination and flexibility, well-developed searching skills, persistence and the ability to evaluate information coming from diverse sources.

Why is it important to learn research skills? During college, students are expected to find information in order to answer academic research questions, understand important societal issues and analyze complex problems. Information supplied by teachers and textbooks is no longer the sole basis for learning. College students are exposed to a wider range of information and must make judgements about the relevance and credibility of sources. In *The Craft of Research*, Booth, Colomb and Williams state that, "Those who can neither do reliable research nor reliably report the research of others will find themselves on the sidelines of a world that

increasingly lives on information." (Booth 6) They also contend that "We are inundated with information, most of it packaged to suit someone else's commercial or political self-interest. More than ever, society needs people with critical minds, people who can look at research, ask their own questions, and find their own answers." (Booth 3)

Too Much Information?

How much information is available to researchers in the Internet age? Consider the following estimates provided by a 2003 study published online by the University of California, Berkeley:

- Print, film, magnetic, and optical storage media produced about 5 exabytes of new information in 2002. Ninety-two percent of the new information was stored on magnetic media, mostly in hard disks. The amount of new information has about doubled in the last three years.

- Digital information comprises the largest amount of this total. The World Wide Web contains about 170 terabytes of information on its surface; in volume this is seventeen times the size of the Library of Congress print collections.

- Another group of Web pages is dynamically generated and stored in Web-accessible databases. This "deep Web" is estimated to be 400 to 450 times larger than the static "surface" Web. (Lyman, Executive Summary)

Much of this Web-accessible information is not available in any other format. For example, NameBase: A Cumulative Index of Books and Clippings [http://www.namebase.org] is a database containing information about people, groups and corporations who have been influential in politics, the military, intelligence, crime and the media since World War II. These names have been drawn from various books, articles and government documents recovered using the Freedom of Information Act. Because of the way it is constructed, this database allows users to find other names that appear on the same pages and thus uncover potential relationships or connections between individuals and groups. *NameBase* is a unique part of the "deep Web" that could be useful both to investigative journalists and to students.

How can you locate useful research tools like *NameBase* among billions of Web sites? Some rules of the road for effective searching of the Web information space are described in the next section of this essay.

Smart Search Techniques

In the years since 1994, when the World Wide Web was invented, a number of powerful search tools have become available to help Web users locate information hiding within this mass of billions of documents. These tools fall into two basic categories: Web directories and Web indexes (or search engines). Success hinges on choosing the right tool as well as using it effectively.

Web directories are lists of sites, along with brief descriptions of their content, selected and compiled by knowledgeable editors. Directories can be browsed by topic and are good starting points for most Internet research because they offer quick access to the best known and most useful sites.

For example, the Internet Movie Database [http://www.imdb.com] is widely recognized as an excellent site for finding information on film casts, credits, plots and characters. It is relatively easy to identify using a Web directory like the Librarians' Index to the Internet [http://www.lii.org].

If your information need is not met by sources listed in Web directories, you can search one or several large Web indexes to locate relevant documents. Web indexes are created by software programs called "spiders" or "robots" because they work automatically, crawling from one Web page to any others that are linked to it, harvesting information. The spiders collect every word on each Web page and store them in a huge index or list. When you enter a search, this index is what your search words are matched against.

The companies that own search engines report that most people enter only one or two search words. When you consider the size of these Web indexes (millions of documents, with thousands of words in each document), this search strategy (or lack thereof) seems doomed to failure. You must use more than one or two search terms in Web indexes in order to narrow down search results and make them more relevant.

Search engines are quite literal and don't deal well with words that have multiple meanings. Using a variety of relevant search words can help with this problem as well. For example, a search for the word "spider" might retrieve this essay, since that word is used several times. If you were looking for information about the insect, rather than the Web software program called a spider, you could get better results by including other search words that also describe what you want (such as insect or arachnid).

Three other techniques will help to improve your chance of success when searching in large Web indexes. First, learn to use **search qualifiers** in order to specify the importance of your search terms and indicate any relationship between them. For example:

- To require that all of your search terms are included in each document found, you may put the plus sign (+) before each term.
  Example: +spider +arachnid +tarantula

- Use the minus sign (-) before each word that should be eliminated from your results.
  Example: +spider -Web -robot

- Use quotation marks to indicate when search words should be treated as a phrase.
  Example: "brown recluse spider"

It is also important to develop intelligent **limiting strategies**. In many search engines, it is possible to limit or screen search results by different variables, such as type of site (government, military, educational, commercial). This technique can be especially fruitful when you want to find varying perspectives. For example, you could limit a search for information on campaign finance reform to government Web sites, in order to learn the viewpoints of elected officials on this topic.

Finally, use **specialized search tools**, ones that focus on a particular topical area, whenever possible. By using specialized resources, like FindLaw [http://www.findlaw.com] for legal topics or HealthWeb [http://healthweb.org/] for health-related research, you are searching only the portion of Web space that is relevant for your research topic and are thus more likely to find good results.

Evaluating Information Quality

When we use the phrase "good results" in this context, we really mean search results that appear to be useful based on a quick look at the Web page. Determining the quality of information found on the Web actually takes more effort and some special techniques.

It is relatively easy and inexpensive to "publish" information on the Internet. On one hand, this open environment often permits you to see more sides of a topic than you might when relying only on printed sources. For example, when doing research on

sweatshop labor in developing countries, you may be able to locate first person accounts and discussions of working conditions in overseas factories, reports of watchdog groups and other kinds of information sources that are difficult or impossible to find in print.

At the same time, because there is no real filtering mechanism to check accuracy, one cannot accept at face value the information found on the Web. In *Evaluating Internet Research Sources*, Robert Harris notes that on the Internet, information exists on a continuum of reliability and quality:

> Information is everywhere on the Internet, existing in large quantities and continuously being created and revised. This information exists in a large variety of kinds (facts, opinions, stories, interpretations, statistics) and is created for many purposes (to inform, to persuade, to sell, to present a viewpoint, and to create or change an attitude or belief). For each of these various kinds of purposes, information exists on many levels of quality or reliability. It ranges from very good to very bad and includes every shade in between. (Harris, Introduction)

Harris further notes that when evaluating information sources, there is no single perfect indicator of quality. You must make inferences from a collection of clues. Here are three important variables to consider as you review and evaluate sources:

Purpose: Consider both the purpose of the source as well as the purpose of your research when determining which Web sources to use. Sites published by advocacy groups (such as the Sierra Club or the American Civil Liberties Union) may include fact sheets, position papers and findings from scientific research studies. These groups may present evidence to support their cause but ignore contrary findings. If you are attempting to understand an organization's mission or find evidence that is representative of a particular viewpoint, advocacy organization Web sites can be quite useful. Look elsewhere for a balanced treatment of controversial issues.

Authority: The most credible information is provided by writers who have education, training or life experience in a field relevant to the information. You can read the jacket blurb to find out more about the author and the scope of a printed book. On the Web, look for an "About Us" link to find brief information about site authors and their credentials.

For example, the nationalissues.com Web site [http://www.nationalissues.com] has an "About Us" page that provides the names and educational/employment

backgrounds of the principal staff as well as a statement about this site's intent to provide a balanced viewpoint on issues such as gun control, school reform, taxes. The authors appear to be individuals with substantial experience in public policy research, government and journalism.

But inquiring minds may want to check credentials or learn more than Web site authors tell us about themselves. If the author is an organization rather than an individual, what is their reputation? You have access to the tools that will help you to locate this kind of information. Search in a Web index to find other documents that the site authors may have published. Use your college library's catalog and online reference databases (of newspaper, magazine and journal articles) to locate works in print written by or about individuals and organizations.

Content: When evaluating the content provided by a Web site, there are a number of questions to consider. These are the same questions that one would ask when evaluating the content of any information source.

- First, does the information appear to be **accurate**? Factual information may be checked in other sources.

- Is there an attempt by the author to be **objective**? Opposing viewpoints should be presented in an accurate manner. Harris notes that "there is no such thing as pure objectivity, but a good writer should be able to control his or her biases." Does the author use a calm, reasoned tone or resort to inflammatory language? Does the site author have a vested interest in the topic, one where "the messenger will gain financially if you believe the message. . ."? (Harris, Objectivity)

- Finally, does the author cite **evidence to support claims** and then document (provide references to) these sources, so that they are easily available for further research? Harris recommends that you "triangulate" an important information source, finding at least two other sources that support it. References supplied by an author can help in this process, but you should look for other sources as well.

A World Beyond the Web

Does all important information live on the World Wide Web, somewhere among the exabtyes? In a word, no. Despite it's size and scope, not everything useful for research is on the Web. There are numerous holes in this fabric - many important

printed books and articles will never be available online, because their conversion to digital format is not cost-effective.

For example, when studying about focus group interviews for a marketing or sociology class, you should read *The Focused Interview*, the seminal work on this subject written by Robert K. Merton, an eminent sociologist who invented the technique. This book was first published in 1956 (and a second edition appeared in 1990). Neither work is Web-accessible.

Although it is important to learn how to search effectively on the Web and to evaluate the information found there, it is equally important to understand that there are projects for which this type of research will not be sufficient. The college library will continue to be important to successful research, despite the phenomenal growth of online information. It is the place where students can expect to find the most important or influential works on many subjects as well as personal assistance and advice about research problems from librarians. Having "the right stuff" for research in the Internet age involves the willingness to explore all spaces where information and knowledge abide, not just the most obvious ones. Booth states that by learning and practicing the craft of research, you

> . . . join the oldest and most esteemed of human conversations - the conversation conducted by Aristotle, Marie Curie, Booker T. Washington, Albert Einstein . . . all those who by contributing to human knowledge have freed us from ignorance and misunderstanding . They and countless others once stood where you now stand. Our world today is different because of their research. It is no exaggeration to say that, done well, yours will change the world tomorrow. (Booth 7)

Works Cited

Booth, Wayne C., Gregory C. Colomb and Joseph M. Williams. *The Craft of Research.* Chicago: University of Chicago Press, 1995.

Harris, Robert. *Evaluating Internet Research Sources.* Version Date: November 17, 1997. 24 May 2005 http://www.virtualsalt.com/evalu8it.htm.

Lyman, Peter and Hal R. Varian. *How Much Information? 2003.* University of California, Berkeley: School of Information Management and Systems. 24 May 2005 http://www.sims.berkeley.edu/research/projects/how-much-info-2003/.

Wolfe, Tom. *The Right Stuff.* New York: Farrar, Straus, Giroux, 1979.

This essay is also reprinted in *Foundations: A Reader for New College Students*, edited by Virginia N. Gordon and Thomas L. Minnick. 2nd ed. Wadsworth, 2001.