

Secrets of the Super Searchers: A Personal and Practical Perspective
Presented at OSS '94, Washington, DC, 10 November 1994

Reva Basch, President
Aubergine Information Services
Berkeley, CA

I have found, whenever I address a group of people whose primary job function is not online searching, that I must start by making two disclaimers:

1. This presentation is not about "just pulling the information out of our computers"
2. This presentation is not about "surfing the Internet."

What I do intend to do is convey a sense of what online searching in the commercial database services has to offer, and what it takes to do it effectively. First, however, let me address these two negatives.

Surfing the Net versus Searching the Databases

Commercial database searching predates the Internet, at least in its current "information superhighway" incarnation, by several decades. The original set of dialup, publicly-searchable databases emerged roughly contiguously with the growth of ARPANET, in the 1960s and early 1970s. Not surprisingly, the earliest files mounted on systems like Dialog were government-funded -- NTIS, DOD, the National Library of Medicine and later, DOE. There was an abundance of engineering, sci-tech and medical data online. It wasn't until the early 1980s that the selection started broadening to include trade publications, market studies, wire services and newspapers, both U.S. and international.

It is important to distinguish between the sort of information that is contained in commercial databases like these and the sort available on the Internet. The Net may be "free" but, as the saying goes, you get what you pay for. Netsurfing can be enthralling, even exhilarating, as well as an enormous time-sink. I am not convinced, however, that the Internet is, or should be considered, at this stage, a primary research tool. I use it to communicate through email and some listservs to which I subscribe, to identify experts in particular fields, and to locate files and resources that I already know exist.

One *can* troll the Net, assuming sufficient time and patience, for background information. The Internet is a repository for some unique information not found in the commercial online databases -- including ftp sites, public-access library catalogs, newsgroups, listservs and online-only scholarly journals that have no print equivalent. In addition, some sources on the Net (e.g. FedWorld) contain information comparable to, but much cheaper than, that found in commercial databases. The downside of accessing it on the Net is that the search capabilities tend to be limited. You may find yourself paging through menus or long indexes rather than doing a slick Boolean search, and you may have no choice but to ftp an entire document, sight unseen, rather than downloading just the portion you need.

The commercial databases offer **control**, in the form of indexing by subject, author, source, date, etc. This makes it much more efficient to home in on exactly what one is looking for. They also offer some degree of **quality assurance**. Most of the material loaded into commercial databases is secondary in nature; i.e., it represents a published source, something that has been subjected to a certain amount of peer review and editorial control. It has been fact-checked, or at least determined by somebody other than the author to be worthy of publication. It may not be state-of-the-art by the time it gets into the database, but it usually provides a fairly reliable source of background information. Finally, the commercial databases offer **integrity**, assurance that the data has not been tampered with, that it is complete, current, and the only version in circulation. The firewalls around commercial databases have been in place for years; the Internet in general is still struggling to establish an equivalent level of security.

To put it simply, what distinguishes commercial databases from the Internet in general is the series of quality filters that material must pass through. This refinement costs money, of course -- anywhere from \$25 to \$200 an hour, typically, compared with the few dollars at most that most of

us pay for access to the Internet. In addition, the commercial providers charge for every search performed, and/or every document accessed. This economic factor is definitely an incentive to develop some searching expertise. We will address this next.

"Just Pull It From Your Computer" -- Not

There is no such thing as "just pulling it from your computer." There have been some significant advances in the last few years, e.g. search engine improvements like natural language searching and relevance ranking, and front-end enhancements like menus and point-and-click interfaces. This sort of thing has made it easier for end users, non-research professionals, to do a credible job of online searching without having to rely on intermediaries. There have been some pricing changes as well, which have tended to shift costs toward the output end rather than penalizing searchers for the amount of time they stay connected to the databases.

But online searching is still a costly and rather complex operation. It may not be rocket science, but it does require a significant investment in training and practice. Even if you contract the work out rather than do it yourself, it helps to understand what's involved, the capabilities and the limitations of the process.

I have been a database searcher since the mid-1970s; I am *still* learning. I interviewed two dozen virtuoso searchers for my book and learned something significant from every one of them. In the remainder of this presentation, I will present an overview of the most important lessons, for this audience, that have come out of our collective experience.

Super Searcher Wisdom:

- **The reference interview is critical.** The reference interview is the process whereby the requestor articulates what it is that he or she *really* needs. This is very different from coming up with a list of key words. A good search intermediary will ask not just what you want, but what you intend to do with the information once you get it. Rather than a ploy to obtain state secrets, this is a way of establishing context for the search query. A policy analyst who is trying to get some background on an entirely new area dictates a different approach from one who simply wants to track down the last bit of detail on an issue with which he or she is already quite familiar. An executive briefing demands a different strategy than an exhaustive, take-no-prisoners review of the literature. A nuclear physicist doing top-level research on super-string theory requires a more rigorous technical approach than a foreign-trade specialist with a layperson's interest in the same subject.

Database searches, unlike Gopher and Veronica searches on the Internet, can be very precise. Don't phrase your question too broadly, assuming that a general question is easier to answer than a specific one. Conversely, don't expect the databases, or the person searching them for you, to perform miracles in terms of isolating highly-specific bits of information (e.g., "What sort of laundry detergent did families of Northern European extraction in the Southeastern U.S. purchase during the first three months of 1994?").

One useful tactic for certain types of inquiries is to pose the question: "What would the ideal article on this subject be called?" This provides a reality check for an intermediary searcher, who can use it to benchmark actual results. It is also a effective way of judging whether the subject has been adequately defined. If it cannot be expressed in a few succinct words, this may indicate that the concept is too amorphous to be reduced to Boolean logic. It may not be appropriate to an online search at all.

Some of my colleagues refer to the reference interview as the "search negotiation." It truly *is* a negotiation session, an iterative process whereby you, the subject specialist, describe what you ideally hope to find, and the searcher, the research specialist, explains what the system is capable of performing.

Another useful touchstone is to ask: "How much information do you *expect* to find?" Has a great deal of material been published, so much that we should think about qualifying the request by date, by language, or in some conceptual way? Conversely, is this a very new field, where one would expect to find very little, or nothing at all? Are you, in fact, *hoping* to find nothing -- if you are searching for a dissertation topic, or checking into the patentability of a device, "nothing" may be exactly the answer you are looking for.

- **Know your sources.** If your search does yield zero results, this does not necessarily mean that no information is available on the subject. You may have made a typo or misspelled a search term. You may be using an inappropriate database: Is the chemical literature really the best place to research a medical problem? Although the effect of electric power lines on nearby human populations might well be covered in EPRI's Electric Power Database, there is probably some unique and pertinent information in the dozen or more other environmental and health-related files available on Dialog. Newspaper databases, too, often report problems from a local perspective.

All of us who have been online searchers for some time fall into formulaic patterns. We have favorite databases that we tend to use repeatedly. Mary Ellen Bates pointed out an example in which two very well-known searchers wrote about biotechnology in the same issue of a professional publication. Each of them selected completely different sets of databases as examples of "where to check first." The moral is that it helps to examine your preconceptions.

- **Offline preparation saves online dollars, and produces better results.** There are several sets of decisions that should be made before going online, e.g. What database(s) are you going to search? Which online service will you use? What will your initial search strategy be? Are there key words or codes you should be using? This is not the time and place to go into detail about cost-effective search techniques. My book, *Secrets of the Super Searchers* [1], is replete with this kind of information, and I would be happy to share a bibliography of my published articles on the subject with anyone who is interested.

- **Take advantage of powerful system features,** e.g.: Natural language searching and relevance ranking in full-text databases; multifile searching and de-duplication; email delivery of search results.

- **Maintain a sense of scale.** Not all projects require 100% retrieval. Recognize when you have reached the point of diminishing returns, when you start seeing the same references, or documents based on the same information, over and over. Know when to stop.

- **Similarly, know when *not* to go online.** Do not automatically assume that every question is a candidate for a database search, or that every project can best be addressed that way. Sometimes the easiest and most cost-effective approach is to pick up the phone, or a reference book. Discover what you can find in basic reference works like the *World Almanac*, *Statistical Abstracts* or *Encyclopedia of Associations*.

- **Know when to call in an expert.** Proficient online searchers are as skilled in their fields as you are in yours. Consider whether your mission might not be better served by hiring in, or contracting out, that expertise, rather than trying to master it yourself.

Conclusion

The real power of database searching, from a strategic viewpoint, is the enormous range of material that one can cover in a very short time and for relatively little money. Where appropriate, the process is inherently more cost-effective than paying a researcher to hunt for the equivalent information in the library or on the telephone. An online search makes it possible to find the informational equivalent of a needle in a haystack.

A huge variety of data is available in electronic form, from the complete text of today's *New York Times*, *Washington Post* and *Wall Street Journal* -- as well as smaller city papers and international ones -- to trade publications covering computers, telecommunications, aviation and avionics, construction, and so on; to academic journals in medicine, agriculture, foreign policy and most other disciplines. There are market research studies on consumer electronics, pharmaceuticals, petrochemicals, etc., and brokerage house analyst reports on publicly-traded companies, both domestic and foreign.

Not all online information is text-based: You can do everything from chemical structure modeling to checking real-time stock quotes, to pulling statistical data, modeling it and loading it into a spreadsheet. A large, supermarket-style databank like Dialog represents one-stop shopping for research on virtually any topic. There are boutique databanks as well, that specialize in certain kinds of information, e.g. broadcast transcripts, real estate transactions or advertising rates. I recommend browsing Gale's *Directory of Databases* [2] for a sense of the variety of sources, probably in the tens-of-thousands, that are available.

References

1. Basch, Reva, *Secrets of the Super Searchers*. Wilton, CT. Online, Inc./Pemberton Press, 1993.
2. Marcaccio, Kathleen Young, ed. *Gale Directory of Databases*. Detroit, MI, Gale Research, Inc., semi-annual.

THIRD INTERNATIONAL SYMPOSIUM on NATIONAL SECURITY & NATIONAL COMPETITIVENESS: OPEN SOURCE SOLUTIONS Proceedings, 1994 Volume II - Link Page

[Previous](#) [OSS '94 Cover, Coat of Arms, Contents,](#)

[Next](#) [OSS '94 Douglas Bernhardt, 'I want it fast, factual, actionable'--Tailoring Competitive Intelligence to Executive Needs,](#)

[Return to Electronic Index Page](#)