



4-1-1984

To Err is Human: Seven Common Searching Mistakes

Carol Tenopir
University of Tennessee - Knoxville

Follow this and additional works at: https://trace.tennessee.edu/utk_infosciepubs



Part of the [Library and Information Science Commons](#)

Recommended Citation

Tenopir, Carol, "To Err is Human: Seven Common Searching Mistakes" (1984). *School of Information Sciences -- Faculty Publications and Other Works*.
https://trace.tennessee.edu/utk_infosciepubs/263

This Article is brought to you for free and open access by the School of Information Sciences at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in School of Information Sciences -- Faculty Publications and Other Works by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

TO ERR IS HUMAN: Seven Common Searching Mistakes

By Carol Tenopir

LEARNING to do online searching is in most cases not particularly difficult, but it involves memorizing a query language and sometimes learning to think in new patterns. Most of the commercial bibliographic systems are designed for knowledgeable searchers and offer little or no online help for the novice. In my semester-long classes and shorter continuing education classes for working librarians, I have noticed that beginning searchers seem to make the same errors and have the same problems year after year. Many of these problems are not terribly serious and all are mastered with practice, but I have collected them through the years in the hope that we can learn something about how search systems could be improved and, less seriously, to let beginning searchers know they are not alone! Because the majority of my classes spend the most time on DIALOG, many of my examples come from that system. Other systems would provide an equal number but slightly different array of problems. I apologize to DIALOG in advance.

Logical errors

Boolean logic is not intuitive to most people. Even if they learned Venn Diagramming and set theory in grade school, most people don't translate it to the real world until they must do it for online searching. Beginning searchers struggle with ANDs and ORs as they learn to ignore the English-language logical flaw that says when a patron

asks for information on raising goats AND sheep, in Boolean terms he really means raising goats OR sheep. No matter how good some students get at doing Venn diagrams on paper, they always seem at least once to use AND when they should use OR.

The capability of using parentheses for nested logic both helps and hurts the situation. When parentheses should be included and are left off, the system still will yield a result, erroneous as it may be. It takes a while for most students to identify why they seem to have so many false drops when they enter "Education OR training AND online reference services" in the ERIC database.

Some new searchers go overboard and decide they had better insert parentheses whenever there may be a chance they would be needed. Computers AND libraries thus gets searched as (Computers) AND (Libraries), just for safety sake. Luckily, this doesn't adversely affect search results in most systems, although it causes a few extra key strokes.

To space or not to space

Probably because our widely-used online search systems began with limited features and have added on capabilities as time goes by, there is much inconsistency in how to enter certain commands. When a space is needed and when it is not seems to be a source of frustration for nearly every beginning searcher.

In DIALOG, it is possible to enter Type 1/5/1-2 with a space after the type command or Type1/5/1-2 without the space, but SelectStepsgoatsorsheep with no spaces is disastrous. Goats OR sheep has to be entered with spaces before and after the Boolean OR in any of the major systems, but if you are working with set numbers, 10R2 with no

spaces will work just fine, sometimes. In DIALOG, Combine10R2 without spaces works (combine is one of those old-fashioned commands from an earlier era), but it must be SS S1 OR S2 with spaces to achieve the same results with a different command.

There is some consolation to students. In most cases, it works to say "when in doubt put a space"; it is much worse to not include one that is required.

A one or a set number?

In the example above from DIALOG, an S must be used to designate a previous set number when using the Select or Select Steps commands (SS S1 AND S2). Without the S, the system will free-text search for the literal number 1 anywhere in the basic index. When using the combine command, however, since it can only be used for set numbers and not for text searching, no S is needed (C 1 AND 2). Students often get puzzling results such as citations on 1,2-dichlorobenzene when they really wanted Set 1 and Set 2.

In BRS, on the other hand, a set number is assumed whenever a number is entered. If you want to search for a literal number, it must be enclosed in quotes. This causes the opposite problem of combining previous sets 1 and 2 when you really wanted citations on 1,2-dichlorobenzene.

Keyboarding problems

Speaking of numbers, 1980 and 1980 may look alike to a typist, but not to a computer. There are times when it is a disadvantage to be a good typist, especially one who has trained on a typewriter without a one or zero key. (The IBM "library" typing element leaves these off as I recall.) Even experienced searchers who had this habit in-

Carol Tenopir is Assistant Professor at the Graduate School of Library Studies, University of Hawaii, Manoa



grained early in their schooling find themselves automatically going for the "el" key at times.

The opposite problem—no previous typing skills—is much more common these days. How it hurts library school educators to tell students that they really need to know how to type to be efficient online searchers! Painfully slow input, spelling errors due to pressing incorrect keys, or even simplifying their search strategy for fear of having to type too many terms, plague some new searchers.

Ignoring database differences

Experienced searchers know that exactly the same search strategy will rarely work on two different databases. Presence or absence of abstracts, length of abstracts if present, presence or absence of controlled indexing, depth of controlled indexing if present, reliance on subject category or other codes, and differences caused by variation in the terminology of different subjects—all these can affect search strategies and results. A search strategy that relies on free-text searching in the ABI/INFORM database with its lengthy abstracts will yield quite different results in a database such as Magazine Index that has broad subject headings and, until recently, no abstracts.

New searchers often feel more comfortable sticking with one method of searching that they have found to work in one database—regardless of the other databases searched. Some learn to rely on free-text searching with liberal use of truncation and synonyms; others favor thesaurus terms exclusively. It takes a while to learn the best mixture of these techniques and requires database-specific knowledge to be truly expert. It seems illogical that databases should vary so much and many of us at first don't want to acknowledge that fact.

Too much faith in the computer

Since any results seem impressive when searchers do their first online search, perhaps many do not think to analyze how the results could have been better. Developing a critical eye and re-examining search strategy to determine if there could be more citations or should be fewer are skills that seem to come with experience. I've noticed when training end-users especially, the tendency at first is to accept the computer's word that there is no information on a subject, rather than to examine their search strategy for flaws.

Our bibliographic online systems offer little help in refining strategy since the choice of terms, use of controlled vocabularies, and deciding on the appropriate number of synonyms is en-

tirely up to the searcher. On most of the bibliographic systems, there is no automatic online help if a searcher forgets to input both plurals and singulars, forgets to truncate, or doesn't remember to account for variant spellings or variant forms of words. The online Expand, Root, or Neighbor commands or the possibility of truncation still must be initiated by the searcher in most systems.

What system am I on?

Almost anyone who knows more than one system's query language gets confused at times. Not only do you have to remember that it is "Expand" in DIALOG, "Root" in BRS, and "Neighbor" in Medline or Orbit, but different database structures, punctuation conventions, and hierarchy of logical operations have to be remembered. The system mode can also be confusing when searching multiple systems—DIALOG is always in command mode and requires a command before you enter a search term; ORBIT always defaults to search mode; and BRS is usually in search mode, except when it is in print mode.

It is easier to learn to search a second or third or fourth system because the patterns of thought are basically the same. Remembering specifics can be confusing at times, however, and is just plain annoying.

Solutions

The thing that strikes me most about all of these common mistakes or problems is that they have been recognized as problems for years. Experimental systems that solve almost all of them have been or are now in operation. It does take time for research to be translated into commercial systems, but perhaps in the not too distant future most of these problems will disappear.

Logical errors are being focused on again as end-users begin to search more. Ongoing research has led to the development of effective non-Boolean search systems or systems that automatically translate a searcher's natural language request into a Boolean expression. Other studies are examining the differences in language and terminology between subjects. Automatic thesauri or systems that automatically search for synonyms are being developed.

Some commercial systems have partially solved some of the common problems. NEXIS and LEXIS will automatically search some word variants and common acronyms without the searcher requesting it. Function keys are used to circumvent commands. BRS/After Dark developed their menu driven systems for the same

reason. Commercially available "front-end" systems such as the Institute for Scientific Information's Sci-Mate, solve the differences in query languages by translating input into the query language of choice.

Continuing research and developments may cause all of these problems to be solved in the next few years, as systems take on more of the power and responsibility for search strategy and error checking, and as front-end systems solve the problems of inconsistency in and among query languages. Online systems have changed as technology and experience change, but they have had no real plan. Major changes are coming soon to alleviate some of the burden of strategy from the searcher. My students and I are anxiously awaiting that day! Then we will have other problems to report.

I wish to thank my fall 1983 students for suggesting this column, and all of my students of the last six years for contributing to it.

Omni Online Database Directory

The latest in a growing list of database directories is the *Omni Online Database Directory*, edited by Mike Edelhart and Owen Davies (Collier: Macmillan, 1983). For approximately \$10 (\$8.95 from the Book of the Month Club), this new directory is a bargain. It is the first online database directory aimed at the general public rather than the information professional and is sponsored by *Omni Magazine*, the mass market science/science fiction magazine. This wide market potential probably is the reason for such a low price.

Front matter defines databases; tells what hardware and software is needed to do online searching with a personal computer; gives hints on choosing an online service and how to search; and provides some general information on costs.

The directory portion itself includes short descriptions of approximately 1000 databases in a subject-arrangement. Bibliographic, numeric, and full text databases are all included. Longer sections describe the home services, Source and CompuServe.

Information, please

I would appreciate hearing from librarians who are now using (or have used) the Source or CompuServe in their libraries. Please send me (for a future column) a description of how you use the system, amount of use, and any problems or suggestions for other libraries that may be considering use. My address is Graduate School of Library Studies, 2550 The Mall, University of Hawaii, Honolulu, HI 96822.