EDRS

DOCUMENT RESUME

ED 400 809 IR 018 163

AUTHOR Flanagan, Lynn; Parente, Sharon Campbell
TITLE Constructing Effective Search Strategies for Electronic Searching.
PUB DATE 96
NOTE 14p.; In: Proceedings of the Mid-South Instructional Technology Conference (1st, Murfreesboro, Tennessee, March 31-April 2, 1996); see IR 018 144.
PUB TYPE Guides - Non-Classroom Use (055) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Access to Information; *Bibliographic Databases; Computer Uses in Education; Efficiency; Electronic Text; Information Needs; *Information Retrieval; Information Services; Library Automation; Library Catalogs; Online Catalogs; *Online Searching; Optical Data Disks; Problem Solving; Relevance (Information Retrieval); *Search Strategies; User Needs (Information); Users (Information)
IDENTIFIERS *Online Search Skills; *User Training

ABSTRACT

Electronic databases have grown tremendously in both number and popularity since their development during the 1960s. Access to electronic databases in academic libraries was originally offered primarily through mediated search services by trained librarians; however, the advent of CD-ROM and end-user interfaces for online databases has shifted the emphasis from mediated to end-user searching. Unfortunately, research studies have indicated that many end-users do not understand basic search concepts and, consequently, do not employ effective search strategies when using these databases. Learning basic database design and effective search strategies allows end-users to take the often neglected first steps to successful electronic searching. The paper is a guide to effectively searching electronic databases, and includes: a history, definition, and discussion of types of electronic databases; selecting an appropriate database in which to conduct a search; planning the search strategy, including selecting search terms and combining terms; refining the search; expanding the results; problem solving techniques; and evaluating search results to determine the effectiveness of the search. (Contains 12 references.) (Author/SWC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

EDRS

CONSTRUCTING EFFECTIVE SEARCH STRATEGIES FOR ELECTRONIC SEARCHING

ED 400 809

Lynn Flanagan
User Services Librarian
Middle Tennessee State University, Murfreesboro, Tennessee

Sharon Campbell Parente
User Services Librarian
Middle Tennessee State University, Murfreesboro, Tennessee

Electronic databases have grown tremendously in both numbers and popularity since their development during the 1960s. Access to electronic databases in academic libraries was originally offered primarily through mediated search services by trained librarians. However, the advent of CD-ROM and end-user interfaces for online databases has shifted the emphasis from mediated to end-user searching. Unfortunately, research studies have indicated that many end-users do not understand basic search concepts and, consequently, do not employ effective search strategies when using these databases. Learning basic database design and effective search strategies allows end-users to take the often neglected first steps to successful electronic searching.

U.S. DEPARTMENT OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Lucinda T. Lea
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

IR018163

BEST COPY AVAILABLE

EDRS

CONSTRUCTING EFFECTIVE SEARCH STRATEGIES FOR ELECTRONIC SEARCHING

Lynn Flanagan
Sharon Campbell Parente

Electronic databases have grown tremendously in both numbers and popularity since their development during the 1960s. The January 1995 issue of the *Gale Directory of Databases* cites 8,776 individual databases as compared to just 301 in 1975 when this data was first recorded.[1] Access to electronic databases in academic libraries was originally offered primarily through mediated search services by trained librarians. However, the advent of CD-ROM and end-user interfaces for online databases has shifted the emphasis from mediated to end-user searching. Unfortunately, research studies indicate that many end-users do not understand basic search concepts and consequently, do not employ effective search strategies when using these databases. Learning basic database design and effective search strategies allows end-users to take the often neglected first steps to successful electronic searching.

DATABASES

History

Electronic databases were first developed in the mid to late 1960s largely through U.S. government sponsorship. The early commercial services included Lockheed's DIALOG Information Services, Inc. (1972), System Development Corporation's ORBIT (1972), and Bibliographic Retrieval Services, Inc. (1976). These services still exist today under new ownership and/or names. The early commercially offered databases were largely scientific and technical in content. Gradually, online vendors started offering end-user interfaces that allowed searching of a subset of their databases during evenings and weekends. The Knowledge Index (DIALOG) and BRS/After Dark (Bibligraphic Retrieval Services, Inc.) both began operation in 1983. However, the real turning point in end-user access occurred with the explosion of CD-ROM databases around 1985-1986. As libraries offered the most heavily requested databases of their mediated search services on CD-ROM and other subscription based electronic formats the demand for mediated searching began to decrease significantly. At the Todd Library of Middle Tennessee State University the number of mediated searches reached a peak of 696 sessions during the 1988-89 fiscal year before the installation of our first CD-ROM products. Fiscal year 1994-1995 recorded just 48 mediated searches. There is little doubt that this trend will continue. The options for end user searching of online databases has also continued to grow. OCLC offered its FirstSearch service which currently provides inexpensive interactive access to over 50 databases in 1991. Many government and private databases are now also accessible and searchable over the internet.

EDRS

Definitions

A database or file is a collection of records or data in machine readable form. The content of most databases represents either a specific subject or discipline or a type of document.[2] The information contained in a database is created by a database producer. Database producers include private or public companies, profit and nonprofit organizations or associations, or government agencies. For example, the American Psychological Association is the producer or publisher of the *PsycInfo* (online) and the *PsycLIT* (CD-ROM) databases. It also publishes *Psychological Abstracts* in print form. The database producer provides information in machine readable form to a database vendor who processes and distributes the enhanced product for a fee. Many databases are offered by more than one vendor. Databases from different producers made accessible from the same vendor have the advantage of being searchable using the same interface or search commands. Many database producers have also adopted the role of database vendor by making their databases available directly.

Types

Databases are usually classified according to type. The *Gale Directory of Databases* currently divides databases into four basic classes; word-oriented, number-oriented, image-oriented (video and still), and sound oriented (audio).[3] The largest number of databases fall into the word-oriented class which includes bibliographic, patent/trademark, full-text, directory, and dictionary subclasses.[4] The most common word-oriented databases are full-text and bibliographic.[5] Bibliographic databases contain references to the original source material. A complete bibliographic citation including the author, title, source, subject headings and often an abstract is retrieved. The ERIC database is an example of a bibliographical database. A full-text database contains the complete text of an article or document.

Structure

Databases consist of records. There is one record for each unique item contained in the database. Additionally, each record or individual item will contain multiple fields or lines of specific information.

DATABASE SELECTION

Before selecting an appropriate database the research topic must first be clearly defined. Once the topic has been decided it is also important to consider the currency, comprehensiveness, types, location, and completeness of materials needed.

EDRS

Currency

Currency is usually more important for scientific, technical, and business searches. For these searches it will be important to know how frequently the electronic database is updated. Libraries faced with financial constraints often opt for less expensive quarterly updated subscriptions making them less current than online and often print equivalents.

Comprehensiveness

Comprehensiveness will be more significant to a student researching a master's thesis than to an undergraduate assigned a five page research paper. The graduate student will likely need to search several subject specific and related databases. In contrast, the undergraduate will usually be pleased with the results from a general academic periodical database that indexes a small number of top titles in multiple subject areas. It is important for end-users to be aware that most electronic databases do not provide coverage before the mid 1960s. Databases in the humanities and social sciences frequently do not go back further than the early 1970s. However, coverage provided by the academic library's online catalog is usually an exception. Most provide electronic access to all library material records regardless of publication date.

Types

End-users must determine what types of material are included in the database they select. Do they want to search for books, journal articles, videotapes, government documents, financial information, or a mix of material formats? The type of material desired can be a factor not only in choosing the specific database, but also in selecting specific files within the database. For example, the *PsycLit* database provides access on separate disks or files to current and retrospective journal articles and to chapters in books related to psychology.

Location

Although not unique to electronic resources, location of materials may also be a factor. A search of the academic library's online catalog will pinpoint materials owned by and located at the university. In contrast, a search of the *PsycLit* database will retrieve citations to many specialized domestic and international sources that the patron may need to request on interlibrary loan. This service may not be open to undergraduates at all institutions.

Completeness: Bibliographic or Full-text?

Finally, it is important to ascertain if the information that will be accessed in the database is self-contained such as a directory entry or a full-text article or if it is a citation to the source material that will require another step to obtain the material.

EDRS

PLANNING THE STRATEGY

Search Term Selection

After selecting the most suitable database the research statement must be broken down into its component parts by identifying the main keywords or concepts. The search topic how **self-esteem** relates to the **academic achievement** of **third graders** can be separated into three distinct concepts. The next step involves selecting the appropriate terminology for searching these concepts.

Controlled Vocabularies: Many specialized databases have their own set of subject headings which reflect the subject matter of the discipline. For example, the MESH headings are used in many medical journal databases, special topics relating to education are listed in the *Thesaurus of ERIC Descriptors*, the Library of Congress subject headings are used in many online catalogs of books, etc. Databases with controlled subject vocabularies searchable through a thesaurus help take the guesswork out of selecting the appropriate subject headings. Most online thesaurus displays provide information about when the term was first introduced, provide a definition, and list related, broader, and narrower terms that may also be selected. It can likewise be helpful to check the database index which typically lists all terms included in the database with their frequency count. A search of the online *ERIC* thesaurus reveals that academic achievement, grade 3 and self-esteem are all valid descriptors which match the three main concepts of our search topic. However, end-users should note that if descriptors are selected through the thesaurus the search will be limited to the descriptor field of the record rather than searched throughout. Unfortunately, research studies conducted by Barbuto and Cevallos (1991) and Charles and Clark (1990 have indicated that end-users experience much difficulty with search term selection.[6]

Free-text Search: If a concept is not listed in a database's thesaurus the user should perform a free-text search where the keyword is entered and searched in all parts of the record.

Combining Terms

The Next step involves planning how the terms should be combined to retrieve relevant results. By separating our concepts into three separate search sets the boolean operators **and, or,** and **not** can be used to obtain tailored combinations. To obtain records containing all three search concepts the terms should be combined using the boolean operator **and.**

The Boolean operator **or** is used to create sets of related concepts. **Not** is used to eliminate a concept from consideration.

Common search strategies include beginning with the most specific subject headings and

expanding using the boolean operator **or**; beginning with the broadest subject headings and narrowing down as long as possible using the boolean operator **and**; and beginning with a known relevant citation or author and searching for others like it after examining the listed subject headings. A sample *ERIC* search on CD-ROM illustrating use of the boolean operators **and** and **or** to narrow and expand a search is illustrated below:

SAMPLE ERIC on CD-ROM SEARCH
ERIC 1992-12/95

No.	Records	Request
#1	5326	ACADEMIC-ACHIEVEMENT in DE
#2	463	GRADE-3 in DE
#3	1426	SELF-ESTEEM in DE
#5	6	#1 and #2 and #3
#6	2458	PRIMARY-EDUCATION in DE
#7	2673	#2 or #6
#8	14	#1 and #3 and #7

REFINING YOUR SEARCH

If your initial search produces irrelevant references or more references than you need for the purpose of your research, you may wish to refine or fine-tune the results. However, it may be better to retrieve some false drops than to attempt to construct a very elaborate strategy to eliminate them and risk also eliminating useful information. The number of records may be reduced or expanded by using several techniques.

Narrowing Results

Subject searching: By limiting your search to the controlled vocabulary of subject headings you may retrieve more relevant records.

Beginning with a free-text search on a specific term(s) and then filtering the results through a general subject category to insure that the references deal with your broad topic may be the most productive search strategy. An example might be a key word search on "blimps" followed by a subject term search on World War I.

EDRS

A technique for fine-tuning results suggested by Snow is to limit important concept keywords to major emphasis if this is available in the database.[7] This is usually applicable only in databases with a controlled vocabulary of descriptors, as in ERIC. Similar results may be achieved in more natural language indexing by requiring that concept keywords appear in the title, descriptors, or identifiers. Although this is a somewhat arbitrary approach, it may tip the scales toward precision at the expense of recall.

Searching Author and Title Fields: If you know an author or a title or part of a title, you may limit the search to the author or title field for quickly retrieving specific references.

Limiting using specialized fields: Many databases have specialized fields which are relevant to the subject covered. You should become familiar with the particular fields provided in the database(s) you use. For example, an index to business information may have a field for SIC codes, for company names, or for geographic areas. Online library catalogs may have a field for the format of the item; that is, book, serial, videorecording, etc. *ERIC* assigns a grade level(s) such as primary-education and a document type such as research reports, guides, and conference proceedings to each item. Often searching in a specific field can narrow the number of hits significantly and may prove valuable in fine-tuning results to your needs. Consider carefully what you are eliminating, however, for you may miss items containing excellent information. By choosing only book format you will eliminate all microtext, or by choosing only research reports you eliminate all conference proceedings, etc.

Limiting by date: Most databases allow for limiting results by publication date. This is often the quickest way to eliminate a large number of hits so it is very tempting to many users. Conkling and Osif point out that literature does not become obsolete as quickly as may have been assumed, and there are dangers in not studying older literature, even in scientific databases. [8]

Using the Boolean operator and: The relevance of search results can often be increased by using the operator **and** to connect two key ideas in your search. This insures that BOTH ideas are included in the reference. For example, a search on "Parkinson's Disease" with the term "treatment" connected with the operator "and" added.

Using the Boolean operator **not**: If your search results contains many hits which are not relevant and which contain a common term, use of the **not** operator will eliminate items with the term in them. This strategy should be used very carefully, however, for you may eliminate items which would be useful.

Example: A search of an online library catalog for the subject terms "science periodicals" will retrieve not only the scientific periodicals but also those dealing with social science, information science, etc. Therefore, a search in the subject headings field for "science periodicals not social information" may obtain more precise results.

EDRS

Use Proximity operators: Many databases provide a means for stipulating that words in a search query have a certain positional relationship to each other. The exact method of entering the criteria may differ from database to database, but generally the following requirements can be accomplished.

The words must be in the same field.
The words must be in the same sentence.
One word must precede the second by n number of words.
One word must be within n number of words of the second.

Relevance-ranking: Snow explains the relevance-ranking technique which is becoming available on more and more databases.[9] Three factors are applied which contribute to the "weight" of each item. The factors are: (a)Quorum - the number of query terms found in the record; (b)Proximity - Closeness of query terms to each other and other occurrences of themselves; and (c)Frequency - statistical weighing of query term frequency in the record versus the database. Relevancy-ranking allows the computer to decide which items will more nearly fit the search criteria and return the most likely ones at the top of the list. A long list of hits can be shortened by choosing to download or print only the top portion of the list.

Expanding Results

If the results of your initial search does not yield enough appropriate references or if the purpose of your research requires that you do an exhaustive literature search, then you may need to expand your search results. There are several possible ways of expanding results.

Use truncation: Some databases will automatically search for a simple plural form of a word and some will do automatic truncation in certain fields but not others. The symbol for truncation and the number of characters which are replaced will vary from database to database. You have to know the way the specific database you are using operates. Often the plural form of nouns is used in controlled vocabulary subject headings. Example: entering "librar*" retrieves "librarian," "librarians," "librarianship," "libraries," and "library.

Use the Boolean operator **or**: The Boolean operator **or** is used to link synonymous ideas. These may be expressed as keywords, phrases, codes, or the results of a previous search. Use of the "or" will serve the dual purpose of collecting additional relevant items and eliminating duplicate hits. If you receive few or no hits on a topic when you expected to find several, then try as many synonyms as possible. Lancaster et.al. in a 1994 study on end-used searching of CD-ROM databases reports that the greatest problem faced by library patrons is being able to identify and use all of the terms needed to perform a complete search.[10] Many terms may need to be used to obtain complete coverage of an idea.

EDRS

Do a free text search. Although a free text search may retrieve irrelevant items, it will serve the purpose of retrieving any possible reference to your term(s). In addition, it allows for the use of more natural language than the controlled vocabulary of a subject search. Often a free text search can be used to retrieve some items and then an appropriate subject heading can be identified which may be used to retrieve more relevant results.

Use the pearl-gathering technique: Pearl-gathering is a term used to refer to the technique of studying a relevant citation and identifying additional terms to be used in your search query. Additional words, subject headings, or specific fields may identified. A particular author may appear as an expert on your topic, suggesting a search by author.

Substitute a general term(s) for specific ones. Use a more general term(s) with the expectation that an item dealing with the general topic may include at least some information on the more specific topic you need. This is an especially useful strategy when a search for a very specific topic has returned few or no hits, or when searching for books in an online catalog where individual chapters are not indexed.

Remove some of the limiters or operators. If the boolean operators "and" and "not" have been used in your search or if a limitation by date, publication type, etc. have been applied, you may have eliminated too many hits. By removing these limiters, your results will be expanded.

Problem Solving Techniques

In an article in Online Ojala humorously reminds us that Murphy's Law is alive and well and lurking in a database search near you.[11] She admits that when she sits down to enter a search, it often seems that she becomes dyslexic, transposing letters or numbers, entering the wrong set number, or making typing errors. Any of these things could dramatically change your results.

If you are not getting the results you expect in your search, there are several techniques which may be helpful in solving the problem.

Be sure you are using an appropriate database: The first thing you need to check is to determine if you are using an appropriate database for your topic and purposes. You will not be able to locate periodical articles in a library's online catalog, and you will not be able to locate book authors/titles in a periodical index. In addition, you will not find technical scientific information in a business index, nor business information in a nursing database. Although you may retrieve a few hits while using an inappropriate database, most likely they will not be the BEST references and certainly they will not provide a complete listing of references on your topic.

EDRS

Check also the time period covered and the update frequency of the database in relation to your topic. If an event happened one month ago and the database was updated three months ago, you need another source. Or if a scientific discovery was made twenty-five years ago, and the database covers only the last ten years, you will miss the most important coverage of the topic.

Check spelling: Although this may seem like an obvious error, it is often an overlooked one. Typing errors are easy to make. British spellings are often different from American spellings. Spaces, apostrophes, hyphens, slash marks, diacritical markings, etc. may make a vast difference in the way a computer searches. A computer matches your entry against what it has stored in its file on a character-by-character basis, space for space, symbol for symbol. If you enter a journal title as "Art News" when it should be "ARTnews" or "Data Base" when it should be "Database," the computer does not guess what is wanted. Some systems will ignore case of letters and punctuation, but you should know the system to be certain. And do not expect every database to use the same form of a word. Both "Park nson's Disease" and "Parkinson Disease" are used in subject headings (controlled vocabulary) in various databases.

Terminology may change over time. Some databases have the terms African Americans, Afro-Americans, Blacks, and Negroes all still on some records, depending on what was considered correct at the time the record was entered. Educators no longer refer to "mainstreaming" but now use "full inclusion." Use of a thesaurus with scope notes can help a searcher find all the terms applied to a concept.

Be sure you browse the index or other listing of entries to check the name of a person or company. If the index uses the full first name of a person and you use the initials, you may not retrieve all the needed records. Be especially wary of names containing acronyms. For example, American Telephone and Telegraph may be spelled out, but it may be listed under AT&T, A.T. & T., A.T. and T., or other variants. The official names of companies may contain the word Inc., Co., or spell out the word "company", etc.

Names or phrases with stop or noise words within them are also tricky. How does one search for "Bank of America" or "stream of consciousness"? This may be easy if the database allows bound phrase searching. Otherwise you may have to use proximity operators (bank w/1 america). You must know how the particular database you are using handles such stop words to be successful.

Another problem related to spelling is entering an incorrect code number. If the digits in a SIC code are reversed, for example, your results will not be at all what you expect.

Reconsider any truncation: Truncation of a term may allow for inclusion of some words you did not consider. For example, an entry for polic# (in search of policy or policies) will also retrieve police.

EDRS

Reconsider search strategy: Look carefully at the search you entered. If it is a complex strategy using several logical operators or reusing search results sets, it is easy to make a mistake in entering it. If a searcher wants to combine concept A and (B or C or D), but types in A and B or C or D, the system will most likely perform the search as (A and B) or C or D. A wrong set number may be used; a "not" operator may be eliminating useful items; incorrect use of nested parentheses may be changing the results.

Maybe it is not your error: Remember that the database itself may contain errors and omissions. Many producers are now using scanners to input full text which causes some errors to occur. PAIS is spelled PALS in a recent full-text article. Data entry personnel often make typing errors or omit a field altogether. You may miss a useful record if you are searching for all companies with a specific SIC code or a specific publication date and no information is entered in the field.

Most databases have a limit to the number of words and the number of characters in each word which the computer actually checks in response to each query. The PALS OPAC system uses only the first four words of a term, title, or subject search. If you enter the title "Journal of the American Medical Association" in the catalog at Todd Library, you get over fifty hits because the search stops after the word "American" and retrieves all matches to that point. A title term search using only the four most important words in the title would be more exact. OCLC also uses only four title words and uses the search pattern 3,2,2,1. That is, it uses only three letters of the first word, two letters of the second and third words, and the first letter of the fourth word. PALS uses an 8,6,6,4 search pattern. There is also a limit to the number of letters checked in author names, etc. If you obtain strange results, this method of computer operation may be the cause.

Try several strategies: If the purpose of your research requires an exhaustive literature search, then you need to try several different strategies and even several different databases, if possible, and compare the results.

EVALUATION OF RESULTS

It is easy to be so dazzled by the technology of online searching that you accept whatever results you get on the first try as the best information available. However, you should scan the references immediately, noting the words and phrases which are relevant to your needs, and asking questions similar to the following:

Are my hits appropriate for my purposes?
Are they from popular, trade, or professional sources?
Are they from known authors in the field?
Are they what I expected?

EDRS

If you are not satisfied with the results, then do some creative thinking about your search query and try again!

CONCLUSION

The advent of end-user electronic databases is an exciting development for both patrons and librarians. There is little doubt that these resources are a popular and glamorous addition to the information arsenal. However, too often end-users are misinformed about what they are searching, how it is organized, how it should be searched. Without acquiring what Loomis and Fink refer to as a "cognitive gateway with which to approach any access source," obtaining relevant search results will remain a haphazard proposition at best.[12] End-users must realize that it is essential to obtain instruction in the design and use of these powerful products in order to reap their full benefit. At the same time librarians must encourage users to be realistic about the capabilities and limitations of these products.

1. Martha E. Williams, Industry Analysis in Gale Directory of Databases. Volume 1, Online Databases by Kathleen Lopez Nolan, ed. (Detroit, Michigan: Gale Research Inc., 1995), xix.

2. Searching DIALOG: The Complete Guide (Dialog Information Services, Inc., 1989), 3-1.

3. (Williams, 1995, xxi)

4. Ibid., xxiii.

5. Ibid.

6. See Domenica M. Barbuto and Elena E. Cevallos, "End-User Searching: Program Review and Future Prospects," RQ 31 (Winter 1991): 223; Susan K. Charles and Katharine E. Clark, "Enhancing CD-ROM searches with Online Updates: An Examination of End-User Needs, Strategies, and Problems," College and Research Libraries 51 (July 1990): 327.

7. Bonnie Snow, "TARGET for the Biomedical Searcher," Online 18 (Nov-Dec 1994): 58-63.

8. Thomas W. Conkling and Bonnie Anne Osif, "CD-ROM and Changing Research Patterns," Online 18 (May 1994): 71-74.

9. (Snow, 1994, 58-63)

10. F. W. Lancaster, Cheryl Elzy, Mary Jo Jeter, Laura Metzler, and Yuen-Man Low, "Searching Databases on CD-ROM: Comparison of the Results of End-User Searching with Results from Two Modes of Searching by Skilled Intermediaries," RQ 33 (Spring 1994): 379.

EDRS

11.MaryDee Ojala, "Troubleshooting Your Search: Whatever Can Go Wrong, Will Go Wrong," Online 19 (Nov-Dec 1995): 59-61.

12.Abigail Loomis and Deborah Fink, "Instruction: Gateway to the Virtual Library," chap. in The Virtual Library: Visions and Realities, ed. Laverna M. Saunders (Westport, Connecticut: Meckler Publishing, 1993), 49.