

# 7

## Searching Using Controlled Vocabulary

The search we conducted in the previous chapter on the aftereffects of traumatic brain injury in children was somewhat successful. We were able to retrieve a reasonable set of 72 documents, many of which seemed to be relevant to the user's query. However, it also seemed as though there must be good documents we weren't able to get, and many of the ones we did retrieve were not really appropriate (including a document on the use of a "rating scale of attentional behaviour"). And yet we tried many good terms from the user, including an alternative or two, and even tried pearl growing terms from good documents.

In fact, those terms seemed to help quite a bit; it is the use of terms like that which we'll discuss in this chapter. The use of **controlled vocabulary** terms--subject headings, descriptors, index terms--can often aid in database searching where they are available. We'll talk about why they can be so useful, in what kinds of situations, why they sometimes don't work so well, and see some techniques for their use in searching on DIALOG. We'll conclude, once again, with a look at how this works in the Internet environment.

### Why Controlled Vocabulary Searching?

What do you call people whose profession is to help people learn? Are they teachers? Instructors? Faculty? Educators? Tutors? Docents? Professors? Lecturers?

How do you refer to systems that help people find items in a library? Catalogs? Online catalogs? OPACs? Online public access catalogs? Card catalogs?

Who wrote *Huckleberry Finn*? Mark Twain? Samuel Clemens? Samuel Longhorn Clemens? Twain, Mark?

Where would you look in a catalog for the author who wrote *One Hundred Years of Solitude*? Under Marquez, Gabriel Garcia? Or Garcia Marquez, Gabriel?

What do you think of when you read the word "mercury"? A planet? A car? A god? A metal? A thermometer?

These are all illustrations of the need for and advantages of controlled vocabularies, and their cousin, name authority files. Both of these have been developed by librarians over the last few hundred years as ways of making it easier, more often than not, to find items in a collection.

Although we'll focus primarily in this chapter on "controlled vocabulary" in the sense of subject headings for concepts like "teacher" and "metal", the use of name authority lists for proper names is closely related, and a number of databases have both. Sometimes they are separate fields, sometimes they are combined, but they work in quite similar ways. - JWJ

There are two related problems at work here. Remember what we've said before--we want to look for concepts, but are forced to search for words. Those words, however, sometimes let us down.

First of all, there are often several words or phrases for any given concept. This is the teacher/faculty/instructor issue, and it's called **synonymy**. The inverse problem is **ambiguity** or **polysemy**--more than one concept for the same "word" or series of characters: Mercury the planet and Mercury the god.

The assembly and appropriate use of controlled vocabularies can assist searchers in both of these situations. When an indexer processes a document, she selects from among a set of possible terms for any given concept. If the preferred term is FACULTY, she assigns that to the document, and subsequent searchers use that term later to retrieve documents about "people who teach". By the same token, a decision is made that subject headings such as MERCURY (MYTHOLOGY) and MERCURY (PLANET) will be used to resolve ambiguity.

There are several advantages to the use of controlled vocabularies in both indexing and searching:

--it facilitates the **gathering of like items**: assembling a set of documents about people in medical school by using and searching on the term MEDICAL STUDENTS

--it helps with **comprehensiveness**: if the indexing is consistent, you have a greater confidence that you have all or nearly all of the documents about medical students by using that term

--it also helps with the **precision** of your results: searching on MEDICAL STUDENTS will not get documents about law students, dental students, graduate students, and so on, unless those documents are also about medical students

--it can help **broaden understanding** of the topic, either by the searcher or the user. Looking for appropriate terms can often help you (or the user, especially if they're not that familiar with the subject area) to better select the terminology, refine your strategy, and generate new ideas for terms to use.

It is by no means the case that using controlled vocabulary is a panacea, and we'll see in the following chapter where and how you might choose other techniques, but it is often a good start for most searches.

## Thesaurus Structure and Use

So where do we find these terrific terms? Many database producers compile and distribute lists of subject terms that they use for indexing the documents in their files and that we can use for searching. These lists are usually called **thesauri**. You're probably familiar with the kind of thesaurus we use in writing, like *Roget's*, which are collections of synonyms and antonyms for words in the English language. Thesauri for databases are somewhat similar, but much richer. They contain not only synonyms but also information about the relationships between terms. In addition, they aid in selecting the best terms to search for a given concept.

As an example, let's look at one such controlled vocabulary--the *Thesaurus of ERIC Descriptors*, produced by the federally-funded Educational Resources Information Center (ERIC). ERIC is the producer of the most comprehensive and most often used database in the field of education, also called *ERIC*, although

it covers a number of other areas, including information and library studies. Much of the discussion that follows is very specific to this thesaurus. Thesauri can differ greatly. For comparison's sake, we'll look at another one briefly before we leave the issue, but if you really want to know how a thesaurus (and, thus, a database) is created and used, your best bet is to read the explanatory material included with a thesaurus. Experience doesn't hurt either--use a few thesauri and you become skilled at figuring out how unfamiliar ones work!

A look through the *ERIC Thesaurus* gives some examples of how a controlled vocabulary could be useful in searching. If you're looking for information about drunken driving, for example, you find that the term used to index this concept in *ERIC* documents is DRIVING WHILE INTOXICATED. A search for documents about discontinuation of programs would use the term PROGRAM TERMINATION (as opposed to PROGRAM DISCONTINUANCE, PROGRAM ELIMINATION, PROGRAM PHASEOUT, or TERMINATION OF PROGRAMS). If you were looking for the preferred term to refer to materials that are used in programmed instruction, you would find that from 1966 to 1980, the term was PROGRAMED MATERIALS (note the spelling difference), but that it was changed to PROGRAMED INSTRUCTIONAL MATERIALS in March 1980. In each of these instances, searching on the preferred term from the controlled vocabulary will give you an increased chance of retrieving documents which are on the topic of interest.

Let's look at a specific example of a search term from *ERIC*, and examine its entry in the thesaurus. The term we choose is INFORMATION SCIENTISTS, and this is what the entry looks like:

<b>INFORMATION SCIENTISTS</b>	<i>Jul. 1971</i>
CIJE: 230    RIE: 182	GC: 710
SN	Individuals who observe, measure, and describe the behavior of information, as well as those who organize information and provide services for its use.
UF	Information Brokers Information Professionals Information Specialists
NT	Librarians
NT	Search Intermediaries
BT	Professional Personnel
RT	Information Industry
RT	Information Science Information Science Education Library Associations

Let's take this line by line and find out what it means, beginning with

**INFORMATION SCIENTISTS**

*Jul. 1971*

This first line gives the term name (INFORMATION SCIENTISTS) and the date it was added to the thesaurus (July 1971). The *ERIC* file dates back to July 1966, so that is the earliest possible date. Some terms are ambiguous because they could be used in more than one way, and so they have a parenthetical component to resolve the ambiguity. An example is INEQUALITY. The term INEQUALITY could refer to the generic notion of inequality, specific inequalities (educational, social, economic), or even the mathematical concept of inequality. The descriptor INEQUALITIES was added in 1970 but was used inconsistently and was therefore removed in 1980. Alternatives are proposed,

such as EQUAL EDUCATION, DISADVANTAGED, or a series of descriptors referring to social, ethnic, sexual or racial bias and discrimination. But what about mathematical inequality? When INEQUALITIES was removed in 1980, a new descriptor was added for just that concept: INEQUALITY (MATHEMATICS). The parenthetical part of the descriptor is used to remove the ambiguity about what kind of inequality is being referred to. The point of this seemingly elaborated discussion is this: When you want to use a term such as this in searching, you must remember that the parenthetical is an essential part of the descriptor. If you search just on

?s inequality  
S8 972 INEQUALITY

you will get each occurrence of the word "inequality" in all four Basic Index fields. But if you wish to use the much more specific descriptor, you must search with

?s inequality (mathematics)  
S9 11 INEQUALITY (MATHEMATICS) (MATHEMATICAL EXPRESSION OR PROPOSITION C

which will get you only documents indexed with that term. You must also be sure to include a space before the first parenthesis--the system is very picky. If you do not, you'll get

?s inequality(mathematics)  
S10 0 INEQUALITY (MATHEMATICS)

Let's go back to the next line on our entry for INFORMATION SCIENTISTS:

CIJE: 230 RIE: 182 GC: 710

This line gives you the number of postings and group code information. As we mentioned earlier, ERIC is really two databases: a collection of citations to journal articles in the educational area (called the *Current Index to Journals in Education* in the printed version and identified by EJ accession numbers in the database), and a collection of citations to other kinds of documents, such as doctoral dissertations, technical reports, test banks, conference papers, bibliographies, guides, and so on. (The printed version of the second database is called *Resources in Education*, and the records are identified with ED accession numbers in the database). This line in the thesaurus tells you how many documents in each of these two collections had been indexed using the given term at the time of publication of this edition of the thesaurus. This can be of help in deciding whether or not to use a term--if it has significantly more or fewer postings, you may want to rethink using it. For INFORMATION SCIENTISTS, we see that as of 1995 (for the 13th edition of the thesaurus), it had been used for 170 journal articles and 162 "other" documents.

The "group code" gives you the very broadest category to which that term belongs. INFORMATION SCIENTISTS is in Group 710, INFORMATION/COMMUNICATION SYSTEMS. This piece of information is not particularly helpful in searching; don't worry about it.

SN Individuals who observe, measure, and describe the behavior of information, as well as those who organize information and provide services for its use.

SN stands for **scope note** and gives a brief description of the term as it is used in ERIC. Not all terms have scope notes, but such notes can be very useful, especially if you are not familiar with the subject field or if you are trying to choose between two terms which appear to be very similar. This

is perhaps the major fault with the *ERIC* thesaurus. There is frequently a multitude of terms with overlapping or similar connotations. Without scope notes it is difficult for the beginning searcher (and often the experienced searcher too!) to choose the most appropriate term to search. Personal experience suggests that *ERIC* is particularly frustrating in this respect. The scope notes are really aimed at the indexers, but as searchers we can use them, too. They often include notes about the interpretation of terms, warnings against the use of terms in certain ways, and recommendations regarding other potential terms.

UF Information Brokers  
Information Professionals  
Information Specialists

UF stands for **Use For**. This indicates that INFORMATION SCIENTISTS is the preferred term for this concept, and that these other three are not to be used. In fact, if you were to look in the thesaurus under any of these, you would see

**Information Brokers**  
use INFORMATION SCIENTISTS

This is a reciprocal reference, rather like a "see" reference in a library catalog. You have looked up the non-preferred term and are referred to the correct form of entry. In some cases, these are old terms that have been replaced by newer ones. For example,:

**College Teachers (1967 1980)**  
use COLLEGE FACULTY

This means that from 1967 to 1980, the descriptor was COLLEGE TEACHERS; in 1980 it was changed to COLLEGE FACULTY. However, old documents will not immediately be re-indexed, so if you want to search on a concept with old and new descriptors and get documents going back to the beginning of the database, you may need to use both terms ORed together. When databases are reloaded, though, they often are re-indexed, so this is only a temporary problem.

NT Librarians  
Search Intermediaries

Here's where we get to the interesting stuff. NT stands for **Narrower Term**. Terms in the *ERIC* Thesaurus (and many others) are organized in hierarchies of specificity. Just as documents vary in how much detail they give on a particular topic, so do descriptors vary. In this instance, we are told that LIBRARIANS and SEARCH INTERMEDIARIES are narrower terms than INFORMATION SCIENTISTS. If a document is strictly about "librarians," it will be indexed using that term. If it is broader, though, and talks about "information professionals," it will probably be indexed with INFORMATION SCIENTISTS. In searching, you should be aware of how narrowly your client's search is focused and what kinds of terms best reflect that level of specificity. Also, you may find in perusing the thesaurus that you enter a hierarchy at too high or too low a level; the listing of hierarchies will give you a better idea of where you should be. It may be that the best search strategy encompasses many different levels of the hierarchy, as in:

*?s librarians or information scientists*

BT Professional Personnel

BT stands for **Broader Term** and is the opposite of NT. In this particular case, I doubt PROFESSIONAL PERSONNEL is going to be much help for a search on information scientists. In some circumstances, though, you might find that a broader term would indeed be helpful.

RT Information Industry  
Information Science  
Information Science Education  
Library Associations

The final part of the display is RT, the **Related Terms**. These are terms which are not part of the hierarchy--neither narrower or broader--for this particular descriptor, but which are related (at least in somebody's opinion), and which may also be of use in searching. When constructing a search strategy, you may find that some of the terms in the RT grouping look useful and you may then decide to use them. Notice that BTs, NTs and RTs are all types of the familiar "see also" reference from the library catalog. They are suggestions of other terms that may be useful for your search. If you decide that one of them is a better term than your original choice, turn to the entry for it and start all over.

A caveat: After a while, all terms start to look good to you. Don't spend more than a few minutes looking through the thesaurus for descriptors, or you will wind up with too many terms, including some real losers, and your search won't be as effective. Find two or three, or maybe only one, that look good, and see if there are maybe a couple of others that look possible and hold them in reserve. But the longer you look, the more you will find, and that is typically counter-productive. Don't be cavalier about term selection, but too much of a good thing is undesirable here, especially if some of the terms you pick have lots of postings--just a word to the wise.

Using these print thesauri can be a big help in planning searches. There may be another, additional aid, though. In DIALOG, many databases have online versions of their controlled vocabularies that you can consult while conducting searches. We'll see how to do this shortly.

As we have said, all databases and all thesauri are different. Let's look at a couple of extracts from the *Thesaurus of Psychological Index Terms*, the controlled vocabulary for *Psychological Abstracts* and its online counterpart, *PsycINFO*. First, how would we go about searching for information scientists? There's no listing under that term, but there is this:

**Information Specialists** 88  
**PN** 4 **SC** 25338  
**B** Professional Personnel/ 78  
**N** Librarians 88  
**R** Information 67

This looks familiar. The first line shows the descriptor name and the year in which it was added to the thesaurus (1988). The second line gives postings information (PN for postings notes) showing that four documents had been indexed with this term by June 1994, for the seventh edition. The second line also shows the unique code number (SC for subject code) assigned to this descriptor, which may also be used as a search term. LIBRARIANS is a narrower term, INFORMATION is a related term and PROFESSIONAL PERSONNEL is a broader term. The slash after this last term indicates that it is an "array term,"

which "represents conceptually broad areas" and is "used in indexing and searching when a more specific term is not available."

Because this database covers a different, although often related, subject area from *ERIC*, the terms it uses and the level of detail explored are different. Take, as an example, the term SCHIZOPHRENIA. This is a descriptor in both databases, but the entries are quite different. First, in *ERIC*:

<b>SCHIZOPHRENIA</b>	<i>Jul. 1966</i>
CIJE: 460    RIE: 127	GC: 230
UF    Dementia Praecox	
BT    Psychosis	
RT    Autism	
Echolalia	
Emotional Disturbances	
Paranoid Behavior	

Compare this with the *PsycINFO* entry:

**Schizophrenia** <sup>67</sup>  
**PN** 16934                    **SC** 45440  
**UF** Chronic Schizophrenia  
    Dementia Praecox  
    Process Schizophrenia  
    Pseudopsychopathic Schizophrenia  
    Reactive Schizophrenia  
    Schizophrenia (Residual Type)  
    Simple Schizophrenia  
**B** Psychosis <sup>67</sup>  
**N** Acute Schizophrenia <sup>73</sup>  
    Catatonic Schizophrenia <sup>73</sup>  
    Childhood Schizophrenia <sup>67</sup>  
    Hebephrenic Schizophrenia <sup>73</sup>  
    Paranoid Schizophrenia <sup>67</sup>  
    Schizophreniform Disorder <sup>94</sup>  
    Undifferentiated Schizophrenia <sup>73</sup>  
**R** Anhedonia <sup>85</sup>  
    Catalepsy <sup>73</sup>  
    Expressed Emotion <sup>91</sup>  
    Fragmentation (Schizophrenia) <sup>73</sup>  
    Schizoaffective Disorder <sup>94</sup>  
    Schizoid Personality <sup>73</sup>  
    Schizotypal Personality <sup>91</sup>

This display shows more postings, many more detailed terms and possible alternative terms. It could be quite helpful in refining the search, especially through specifying exactly what type or form of schizophrenic disorder is desired. Of course, a search on schizophrenia *per se* would be much more productive in *PsycINFO* than in *ERIC*, but a search on the impact of schizophrenia on the learning process might yield equally good results in

either database. This gives you an idea of the challenges involved in database selection, which we will return to in chapter 10.

You should also be aware at this point of the following:

1. Not all databases have controlled vocabularies. Some database producers do not have the resources or inclination to produce thesauri, and thus none exist. Other types of databases (numeric, financial, reference) have no controlled vocabulary because it would make no sense.
2. Not all controlled vocabularies are useful. Some, such as the *Thesaurus of ERIC Descriptors* and the *Thesaurus of Psychological Index Terms*, are quite thorough and helpful in searching. Others are barely more than word lists (e.g., the thesaurus for *The Philosopher's Index*). Quite a number of databases use the *Library of Congress List of Subject Headings (LCSH)* in place of a thesaurus (e.g., *Magazine Index* or *Books in Print*). This suggests the fact that both indexing languages and indexing standards vary greatly among different databases.
3. Not all indexing is done perfectly. In your experience in searching, you will undoubtedly find index terms that will confuse, amuse, or infuriate you.

## Mechanics of Controlled Vocabulary Searching

The DIALOG commands we saw in the previous chapter for creating and manipulating sets and truncating and displaying records are basic techniques, but there are more commands and ways of searching, some of which are particularly applicable for controlled vocabulary searching.

We will present several ways of searching using controlled vocabulary: searching bound descriptors, searching for an individual word in the descriptor field, and searching for a single-word descriptor, searching for major descriptors, and the **explode** feature of DIALOG.

### Searching Bound Descriptors

If, after rummaging through the thesaurus, you have found one or more terms you wish to search, you simply enter them as a search statement. For example:

?s *choral music or rock music or vocal music*

```
      117 CHORAL MUSIC (MUSIC INTENDED FOR GROUP SINGING)
      73  ROCK MUSIC
      276 VOCAL MUSIC (MUSICAL COMPOSITIONS WRITTEN FOR VOICES,
      EIT...
S11   430 CHORAL MUSIC OR ROCK MUSIC OR VOCAL MUSIC
```

Thus, for **bound descriptors** (as these intact multiple-word descriptors are often called), we merely search on the phrase as given in the thesaurus. Recall that this will only work in the descriptor field, or in some other field which is phrase-indexed. These phrases were entered into the inverted file as phrases as well as individual words, so a search on the word ROCK would get you the 73 documents above, but also any other use of the word ROCK in any field of the database.

### Searching Individual Words in the Descriptor Field



In the search in the previous chapter, we had difficulty finding documents about the behavioral aspects of head injury in children. Finding things about neuropsychological and psychosocial aspects was easier, but we never quite got our hands on these behavioral aspects. And yet, you'd think that a database focusing on psychology would have a great many documents about behavior.

In fact, it does, and that was part of the problem we had with that aspect of the search, though it wasn't obvious at the time. We searched on the word BEHAVIORAL correctly, but quite broadly--in titles and abstracts as well as in the descriptor field. We got over 60,000 hits on that term, and if you look back at the result sets of which it was a part, they were all pretty broad and unfocused. (And that was using BEHAVIORAL; had we searched broadly on the word BEHAVIOR, we'd get about 212,000 hits, and truncating on BEHAVIOR? gets you over 250,000!) We need some way of making that search a bit more specific so it will get more useful results.

There are a great many descriptors about "behavior" in the *PsycINFO* database, and we could pick our way through them all, finding a handful which appear to be most helpful. Or, we could take an easier path. Wouldn't it be nice to be able to find all the places that the word BEHAVIOR appears in the descriptor field? That means that the indexers of those documents thought that there was some "behavioral" component present in each, which is at least more specificity than we have with the really broad BEHAVIORAL search.

There is a way to do precisely that, using a DIALOG technique called **qualifiers**:

### Qualifiers (Suffix Searching)

The search statement

```
?s alcohol
S1      3228  ALCOHOL
```

searches for ALCOHOL in all the Basic Index fields. However, we may wish to search for the word only in the title field. We could use the statement

```
?s alcohol/ti
S2      1139  ALCOHOL/TI
```

to increase specificity. We **qualify** a search statement with a suffix, using the slash and the field codes of the field or fields we wish to search in, such as:

### S TERM/field code(s)

The most often used codes for bibliographic databases are AB (abstract), DE (descriptor), ID (identifier), and TI (title), but there are often many others, especially in nonbibliographic files. Other databases may have other fields and codes in their Basic Indexes, so you will need to check the bluesheets carefully. You could also use the statement

```
?s computer?/ab
S3      32375  COMPUTER?/AB
```

This statement searches for the word stem COMPUTER in the abstract field only, and will retrieve COMPUTER, COMPUTERS, COMPUTERIZATION, and so on, in that field only.

We can specify more than one field to search in, as in this example:

```
?s frog/ti,de
      S4      72  FROG/TI,DE
```

This statement searches for the word FROG in either the title or the descriptor field. It is the logical equivalent of

```
?s frog/ti or frog/de
      25  FROG/TI
      54  FROG/DE
      S5      72  FROG/TI OR FROG/DE
```

This technique will only work on Basic Index fields. To search in other fields, such as author, journal name, publication year, and so on, we use prefix searching, which will be described in Chapter 8.

Qualifiers may be used also with sets that have already been created, as in the following example:

```
?s newspaper
      S6      3940  NEWSPAPER
?s s6/ti
      S7      940   S6/TI
```

The first statement produces S6, which contains all documents with the word NEWSPAPER in any Basic Index field and has 3,940 documents. Using **post-qualification** (as it is called in this instance of qualifying a previously existing set), we limit the search to the title field only and so create a new set S7, with only 940 documents.

This illustrates one of the primary uses of qualifying--to improve the quality of terms or sets by focusing them. Often a document is more likely to be about a subject if that term is in the title or descriptor fields rather than in the abstract, a longer and often less specific indicator of content. Probably your most frequently used qualifier will be DESCRIPTOR (or sometimes TITLE), as in:

```
?s cats/ti,de
```

because these two fields will provide the most relevant retrievals. Using qualifiers is a good way to narrow searches, because you can check retrieval at each step. Back to our music terms:

```
?s music/de
      S14     5537  MUSIC/DE
```

will retrieve any document with the word MUSIC anywhere in the descriptor field. That would include all the documents from the search statements above (ROCK MUSIC, CHORAL MUSIC, and so on) and also any indexed with MUSIC ACTIVITIES, APPLIED MUSIC, MUSIC EDUCATION, or even just MUSIC.

Why would you want to do this? Sometimes, as with "behavior", you may wish to search on a broad concept that has many descriptors associated with it, all of which have a certain word in common. If you were interested in documents that discussed the use of music in foreign language education, you may want to be very broad in your search on the music concept. There are many good descriptors--perhaps too many. MUSIC itself is only used for documents generically about music, but not for a document about, say, Japanese music, which would be indexed with ORIENTAL MUSIC. (Of course, this would be entered in the Basic Index under both ORIENTAL and MUSIC. This is called **double posting**.) You may decide, then, that you will provisionally accept any

document with any descriptor that includes the word MUSIC. This is a good initial strategy but will produce some false drops.

### Searching One-Word Descriptors

But what if the descriptor was just MUSIC? Not all descriptors are phrases. Many are single words. As we just said, the descriptor MUSIC is used when the document is generically about "music" but not any particular kind of music or for any particular purpose. Another example is a term such as SLEEP in *PsycINFO*. This is a descriptor unto itself, but there are also descriptors such as SLEEP DISORDERS, SLEEP APNEA, REM SLEEP, SLEEP WAKE CYCLE, and so on which incorporate the word SLEEP. Furthermore, it's a word that is relatively frequently used in abstracts but that is not as good an indicator of document content.

For example, searching on

```
?s sleep
  S1      9706  SLEEP
```

will retrieve all documents with the word SLEEP in any of the Basic Index fields.

Searching

```
?s sleep/de
  S2      6729  SLEEP/DE
```

will retrieve all documents with the word SLEEP in the descriptor field. But that is not the same thing as retrieving documents that have the one-word descriptor SLEEP. To get those, you must search

```
?s sleep/df
  S3      3945  SLEEP/DF
```

where the suffix /df ("descriptor full") qualifies the search to *one-word descriptors only*. This is an important and subtle distinction, and is not always easily grasped initially. A tip: If you want to search a one-word descriptor as a one-word descriptor, use /DF.

There are a couple of other techniques we can use to make controlled vocabulary searching more efficient or more precise: restricting to major descriptors and exploding.

### Major Descriptors

You'll recall in Chapter 6, when we discussed record structures, we saw that some descriptors were starred. (If you look at some search examples, you'll see many instances of this.) We called these "major descriptors" and said that the indexer had decided that these terms best described the document. We can use these decisions to try to improve the quality and precision of our search results.

If we wish to restrict our searching to major descriptors, we can give the command

```
?s music education/maj
```

and retrieve those documents which were assigned MUSIC EDUCATION as a major descriptor. If, as often happens, we had already gotten a set for MUSIC

EDUCATION, say S6, and now wish to improve the specificity of that set, we can do

?s s6/maj

and the set will be so restricted.

A couple of comments here: First, this command looks a lot like the suffixes we called qualifiers earlier. It is very similar, but note that there we were restricting to a particular field (title or abstract), and here we are reducing within a field. A slight difference. Technically, /MAJ is not a qualifier, but a limit, which we will discuss in Chapter 9, but it makes sense to introduce it here. Second, there is a companion limit, /MIN, which, as you might have guessed, limits to minor descriptors (the unstarred ones). It escapes me why you would want to do that, but it is there if you want it.

MAJOR descriptors are those that are also used in the printed indexes (RIE and CIJE) of ERIC. They are obviously regarded as the most important index terms, and are identified on your printout by having an asterisk beside them. They provide a useful and simple way of making your search more specific. But check your bluesheets for suffix codes, because they are not available on many files. - GW

## Explode

Exploding is a nice technique that can save you a lot of typing and improve the breadth and recall of your searching. If you explode on a controlled vocabulary term, you will search on it and all its narrower terms, all ORed together. In some files (e.g., ERIC), you will get only the terms that are directly narrower (one level down) in the hierarchy. In other files (e.g., PsycINFO), you will get narrower terms of narrower terms, all the way down. Check documentation and thesauri to see what a particular file does. This feature is available only in files which have online thesauri.

The technique is very simple:

?s music!

is the equivalent of

?s music/df or applied music or jazz/df or oriental music or rock music or vocal music

Very handy. Note that using exploding, you'll get a term like JAZZ, which is a narrower term to MUSIC in the thesaurus, but doesn't have the word "music" in it, so you wouldn't get that by searching on MUSIC/DE.

However, there are some terms in the MUSIC hierarchy you won't get by doing this. CHORAL MUSIC, for example, is a narrower term under VOCAL MUSIC, but since ERIC's explode only goes down one level, it will not be included. Again, this can save you a lot of typing and possibly retrieve documents you might otherwise overlook. Beware of overuse, though--sometimes there are undesirable narrower terms, so it pays to examine the print or online thesaurus before trying an explode.

We've mentioned this "online thesaurus" several times now--let's see how it works and how it might be used. To access it, we use a DIALOG command which has several other uses: EXPAND.

## Viewing the Basic Index and Online Thesaurus: EXPAND

The EXPAND command's primary function is to allow you to view an alphabetical display of a portion of the Basic Index. (We will use it with prefix searching in chapter 9 and citation searching in chapter 11.) It can be very useful

when you're not sure of the spelling of a word, or when you think there may be variant spellings or misspellings in the database. The format of the command is

**EXPAND term**

which may be abbreviated as

**E term**

The result is a display like the following:

?e bias

Ref	Items	RT	Index-term
E1	1		BIARD
E2	1		BIARTS
E3	9929	8	*BIAS (AN INCLINATION, OR A LACK OF BALANCE (NOTE: ...))
E4	2		BIAS (LEONARD)
E5	1		BIAS ELIMINATION PROCEDURES
E6	1		BIAS IN ATTITUDE SURVEY
E7	2		BIAS IN ATTITUDES SURVEY
E8	1		BIASE
E9	986		BIASED
E10	2		BIASEDNESS
E11	821		BIASES
E12	99		BIASING

Enter P or E for more

The display has four columns. The first gives reference numbers which we can use later to select terms from the display. The second gives the number of postings, which is the number of documents that contain each term. The third, RT, shows the number of related terms in the thesaurus; an entry in this column indicates that this term is a descriptor. Finally, we see the alphabetical list of terms themselves. Notice that E3 is BIAS, the term we began with, and it has an asterisk in front of it. This is your indication that this is in fact the term you expanded. Also notice that there is a parenthetical expression after BIAS in E3. This is not part of the descriptor; it is the very beginning of the scope note from the thesaurus. It should be reasonably clear to you when using an EXPAND display which of these parentheticals are scope notes and which are parts of descriptors. Another point of interest is E4, BIAS (LEONARD). This is an identifier (proper nouns cannot be descriptors in ERIC, but this is not true of other files), and refers to Leonard Bias, the college basketball star who died of a drug overdose in 1986. E5, E6, and E7 are also identifiers; we know this because they are multiple-word phrases but have no related terms. E8, BIASE, is probably a spelling or typing error in one of the documents.

If you wish to see the next "page" of the display (the next 12 entries), just type

**P (for PAGE)**

or

**E (for more EXPAND)**

?p

Ref	Items	Index-term
E13	1	BIASING EFFECTS
E14	1	BIASNESS
E15	3	BIATHLON
E16	1	BIAZHUN
E17	20	BIB
E18	10	BIBB
E19	1	BIBB COUNTY INSTRUCTIONAL MATERIALS CENTER GA
E20	1	BIBBIDIBOBBIDIBOO
E21	1	BIBBINS
E22	1	BIBBITS
E23	1	BIBCON
E24	2	BIBDATA

Here we see more of the same. Because none of these are descriptors, there is no RT column (E19 is an identifier).

The reference numbers of these terms are now available for us to use in searching. We have as yet created no sets by using the EXPAND command, but we can do this by selecting terms from the display using the E numbers:

?s e3, e8-e14

This command will create a set containing documents with the terms specified. As you can see, you can select more than one term at a time, separating them with commas or using a hyphen. The selected terms are then ORed together. The command above is equivalent to

?s e3 or e8 or e9 or e10 or e11 or e12 or e13 or e14

If you use several EXPAND statements, they are, in effect, overwritten in the same computer space. This means that you may only SELECT from that most recent display. The others have been lost to you.

There are two ways to enter the online thesaurus. If you have a descriptor in an alphabetical EXPAND display (such as BIAS above), you can look at its online thesaurus entry to display the related terms by expanding on its E number, as in the following example:

?e e3

Ref	Items	Type	RT	Index-term
R1	9929		8	*BIAS (AN INCLINATION, OR A LACK OF BALANCE (NOTE: ...))
R2	1069	U	1	PREJUDICE
R3	1449	N	20	SOCIAL BIAS
R4	376	N	10	STATISTICAL BIAS
R5	1557	N	19	TEST BIAS
R6	653	N	8	TEXTBOOK BIAS
R7	82045	R	52	ATTITUDES
R8	316	R	14	EGOCENTRISM
R9	60	R	3	MENTAL RIGIDITY

Now you see a new display, similar to the other but slightly different. Instead of the reference numbers beginning with E, they begin with R. This tells you that you are in an online thesaurus display rather than an alphabetical one. The Items column is the same, but now we see an additional column, Type, which indicates whether the listed term is a Use For (U), Narrower (N), Broader (B), or Related (R) Term in the thesaurus. The final two columns are the same as before. You are not able to see the full entry as you would in the print version of the thesaurus, including scope notes, but

the relationships and hierarchies are preserved and available for online consultation.

This process can continue indefinitely, as you can now expand on any terms in the display by EXPANDING on R numbers, as in this example:

?e r5

Ref	Items	Type	RT	Index-term
R1	1557		19	*TEST BIAS (UNFAIRNESS IN THE CONSTRUCTION, CONTENT, ADM...)
R2	9929	B	8	BIAS
R3	420	R	10	CULTURE FAIR TESTS
R4	1379	R	9	ERROR PATTERNS
R5	581	R	15	OBJECTIVE TESTS
R6	1449	R	20	SOCIAL BIAS
R7	1277	R	23	SOCIAL DISCRIMINATION
R8	376	R	10	STATISTICAL BIAS
R9	302	R	11	TEST COACHING
R10	6052	R	29	TEST CONSTRUCTION
R11	2786	R	24	TEST INTERPRETATION
R12	2100	R	23	TEST ITEMS

Enter P or E for more

The other way to enter the online thesaurus does not depend on having a previous E-display in hand. If you already know that a certain term is a descriptor, you may just EXPAND directly on it by using parentheses, as shown in this example:

?e (graduate students)

Ref	Items	Type	RT	Index-term
R1	3052		12	*GRADUATE STUDENTS
R2	140	N	5	DENTAL STUDENTS
R3	169	N	5	LAW STUDENTS
R4	1680	N	8	MEDICAL STUDENTS
R5	22586	B	26	COLLEGE STUDENTS
R6	1936	R	13	COLLEGE GRADUATES
R7	1154	R	12	DOCTORAL PROGRAMS
R8	5313	R	22	GRADUATE STUDY
R9	9385	R	7	GRADUATES
R10	126309	R	29	HIGHER EDUCATION
R11	442	R	9	MASTERS PROGRAMS
R12	30	R	14	RESEARCH ASSISTANTS

Enter P or E for more

These R numbers can now be SELECTed in groups in exactly the same way as before. (Remember, we do not create sets using EXPAND; we have to SELECT from the R-display.) For example:

?s r1-r4

	3052	GRADUATE STUDENTS
	140	DENTAL STUDENTS
	169	LAW STUDENTS
	1680	MEDICAL STUDENTS
S2	4985	R1-R4

There are advantages and disadvantages to using the online version of the thesaurus. Its major disadvantage is that you are paying online time to work with the thesaurus. It is useful when you have no printed thesaurus available, but it is not a substitute for thorough preparation.

However, if you use the online controlled vocabulary cleverly, it can save you time, money, and effort. You can SELECT from the R-display directly, which will allow you to avoid typing long descriptors and possibly making typing or spelling errors. Because the online thesaurus is sometimes updated more frequently than the print version, there may be new descriptors or relationships online to help you, and certainly the postings information will be more up to date, so you can gauge the potential size of your retrieval.

## SAMPLE SEARCH

For the sample Dialog search, let's rerun the search from the last chapter on the various effects of closed head injuries in children. Using controlled vocabulary techniques, we are likely to get higher-quality results more quickly and easily.

First of all, a trip through the *Thesaurus of Psychological Index Terms* yields some interesting potential terms: HEAD INJURIES, BRAIN DAMAGE, and TRAUMATIC BRAIN INJURY for that concept block (probably the most specific); several which contain the words PSYCHOSOCIAL, NEUROPSYCHOLOGICAL (in various forms) and lots with BEHAVIOR, not surprisingly. We'll try searching on them in various ways, as you see below. Finally, the preferred terms for children and adolescents are, well, CHILDREN and ADOLESCENTS. We begin the search with these terms.

File 11:PsycINFO(R) 1967-1997/Dec  
(c) 1997 Amer. Psychological Asso.

```
Set  Items  Description
---  -
?ss head injuries or brain damage or traumatic brain injury

S1    1919  HEAD INJURIES
S2    3940  BRAIN DAMAGE
S3     118  TRAUMATIC BRAIN INJURY
S4    5756  HEAD INJURIES OR BRAIN DAMAGE OR TRAUMATIC BRAIN INJURY
?ss psychosocial or neuropsych? or behavior/de

S5    21405  PSYCHOSOCIAL
S6    12614  NEUROPSYCH?
S7   122616  BEHAVIOR/DE
S8   153137  PSYCHOSOCIAL OR NEUROPSYCH? OR BEHAVIOR/DE
?ss children/df or adolescents/df

S9    21595  CHILDREN/DF
S10   27954  ADOLESCENTS/DF
S11   46894  CHILDREN/DF OR ADOLESCENTS/DF
```

The "head injury" concept is by far the narrowest, as we suspected. There are obviously a lot of documents with BEHAVIOR in the descriptor field, but we can have more confidence in those documents than searching just for BEHAVIOR anywhere precisely because the word is in the descriptor field. It's a large component of that set, but not necessarily dangerously so. If all we get in the results are behavior-oriented documents, to the exclusion of psychosocial/neuropsychological ones, we may try those terms separately to see how they are contributing. The age set seems fine; let's put them all together.

```
?s s4 and s8 and s11
```



5756 S4  
153137 S8  
46894 S11  
S12 77 S4 AND S8 AND S11

This set seems reasonable; let's have a look:

?t 12/8/1-8

12/8/1  
DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

01110231 84-42833  
Mild head injury in children and adolescents: A review of studies (1970-1995).  
Major Descriptors: \*HEAD INJURIES; \*LITERATURE REVIEW  
Minor Descriptors: ADOLESCENTS; CHILDREN  
Descriptor Codes: 22360; 28580; 00950; 08830  
Identifiers: neuropsychological or academic or psychosocial outcomes, children & adolescents with mild head injury, literature review, 1970-1995  
Section Headings: 3297 -NEUROLOGICAL DISORDERS & BRAIN DAMAGE

12/8/2  
DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

01105746 34-81162  
Factors contributing to successful return to school for students with acquired brain injury: Parent perspectives.  
Major Descriptors: \*BRAIN DAMAGE; \*PARENTAL ATTITUDES; \*PSYCHOSOCIAL FACTORS; \*SCHOOL ATTENDANCE  
Minor Descriptors: ADULTHOOD; CHILDREN; PARENTS  
Descriptor Codes: 06780; 36620; 41925; 45560; 01150; 08830; 36680  
Identifiers: perspectives on medical & psychosocial & other factors contributing to successful return to school, parents of children (aged 5-18 yrs) with acquired brain injury, Australia  
Section Headings: 3297 -NEUROLOGICAL DISORDERS & BRAIN DAMAGE

12/8/3  
DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

01103462 34-78878  
Clinical neurological indicators are only moderately correlated with quantitative neuropsychological test scores in patients who display mild-moderate brain impairment following closed-head injuries.  
Major Descriptors: \*BRAIN DAMAGE; \*NEUROLOGY; \*NEUROPSYCHOLOGICAL ASSESSMENT; \*ORGANIC BRAIN SYNDROMES  
Minor Descriptors: ADOLESCENTS; ADULTS; CLINICAL JUDGMENT (NOT DIAGNOSIS)  
Descriptor Codes: 06780; 33740; 33835; 35670; 00950; 01160; 09620  
Identifiers: reliability of qualitative vs quantitative clinical neurological indicators & neuropsychological measures, 17-52 yr old patients with brain damage due to closed head injuries  
Section Headings: 3297 -NEUROLOGICAL DISORDERS & BRAIN DAMAGE

12/8/4

DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

01098489 34-76427  
Behavioural adjustment and parental stress associated with closed head injury in children.  
Major Descriptors: \*BEHAVIOR PROBLEMS; \*PARENTAL ATTITUDES; \*STRESS; \*TRAUMATIC BRAIN INJURY  
Minor Descriptors: ADOLESCENTS; ADULTHOOD; CHILDREN; MOTHERS  
Descriptor Codes: 05650; 36620; 50170; 54115; 00950; 01150; 08830; 32140  
Identifiers: parental stress levels & perception of children's behavioral problems at least 1 yr after injury, 24-50 yr old mothers of children (aged 4.5-15 yrs) who had traumatic brain injury  
Section Headings: 3297 -NEUROLOGICAL DISORDERS & BRAIN DAMAGE

12/8/5

DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

01091018 84-31853  
A taxonomy of neurobehavioral functions applied to neuropsychological assessment after head injury.  
Major Descriptors: \*NEUROPSYCHOLOGICAL ASSESSMENT; \*TAXONOMIES; \*TRAUMATIC BRAIN INJURY  
Minor Descriptors: ADOLESCENTS; ADULTS; CHILDREN  
Descriptor Codes: 33835; 51550; 54115; 00950; 01160; 08830  
Identifiers: application of taxonomy to evaluation of symptoms, individuals with traumatic brain injury  
Section Headings: 3297 -NEUROLOGICAL DISORDERS & BRAIN DAMAGE

12/8/6

DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

01079341 84-20299  
Assessing children with traumatic brain injury during rehabilitation: Promoting school and community reentry.  
Major Descriptors: \*INTERDISCIPLINARY TREATMENT APPROACH; \*PSYCHOSOCIAL REHABILITATION; \*REENTRY STUDENTS; \*REHABILITATION; \*TRAUMATIC BRAIN INJURY  
Minor Descriptors: CHILDREN; PSYCHOLOGICAL ASSESSMENT  
Descriptor Codes: 26030; 41950; 43495; 43580; 54115; 08830; 41706  
Identifiers: need for comprehensive & multidisciplinary assessment during rehabilitation to promote recovery & facilitate school & community reentry, children with traumatic brain injuries  
Section Headings: 3380 -REHABILITATION

12/8/7

DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

01076735 34-86477  
Neuropsychologie traumatickeho poskozeni mozku u deti. (Neuropsychology of traumatic brain damage in children).  
Major Descriptors: \*TRAUMATIC BRAIN INJURY; \*NEUROPSYCHOLOGY  
Minor Descriptors: CHILDREN  
Descriptor Codes: 54115; 33840; 08830  
Identifiers: neuropsychology & traumatic brain damage, children  
Section Headings: 3297 -NEUROLOGICAL DISORDERS & BRAIN DAMAGE

12/8/8

DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

00998391                    82-25937  
Clinical neurological trauma parameters as predictors for  
neuropsychological recovery and long-term outcome in paediatric closed head  
injury: A review of the literature.  
Major Descriptors: \*LITERATURE REVIEW; \*HEAD INJURIES; \*PROGNOSIS;  
\*METHODOLOGY; \*NEUROPSYCHOLOGY  
Minor Descriptors: CHILDREN; SEVERITY (DISORDERS); RECOVERY  
(DISORDERS)  
Descriptor Codes: 28580; 22360; 40830; 31140; 33840; 08830; 46824  
; 43390  
Identifiers: use of clinical neurological trauma parameters, prediction  
of neuropsychological recovery & long term outcome, children with  
closed head injury, literature review  
Section Headings: 3297 -NEUROLOGICAL DISORDERS & BRAIN DAMAGE

Pretty good--a reasonably-sized set, with many potentially interesting  
documents (including the one in Czech). We might try to improve it a bit, by  
focusing the BEHAVIOR term down to documents in which it is part of a major  
descriptor. We first reconstruct that set, with the narrower version of  
BEHAVIOR, then recombine the concept sets to form a new result set:

?s s5 or s6 or behavior/maj

	21405	S5
	12614	S6
S13	67076	BEHAVIOR/MAJ
S14	98704	S5 OR S6 OR BEHAVIOR/MAJ

?s s4 and s14 and s11

	5756	S4
	98704	S14
	46894	S11
S15	71	S4 AND S14 AND S11

and it appears to make very little difference. We only lose 6 documents by  
doing this. You probably wouldn't do this in a real search, but let's take a  
moment and see what those 6 are. We can use the Boolean operator NOT to do  
this:

?s s12 not s15

	77	S12
	71	S15
S16	6	S12 NOT S15

?t 16/8/1-6

16/8/1  
DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

00791508                    78-13090  
The head injured adolescent: A case report and review of the issues.  
Major Descriptors: \*HEAD INJURIES  
Minor Descriptors: ADOLESCENTS; COGNITIVE ABILITY; SOCIAL BEHAVIOR;  
ADULTHOOD  
Descriptor Codes: 22360; 00950; 10050; 48080; 01150  
Identifiers: pathophysiology & psychiatric & cognitive & social sequelae

during adolescence, 18 yr old male with closed head injury  
Section Headings: 3290 -PHYSICAL & PSYCHOSOMATIC DISORDERS

16/8/2

DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

00432607

70-01587

Hemichorea-hemiballismus as conversion reaction following head trauma.

Minor Descriptors: HEAD INJURIES; ADOLESCENTS; CASE REPORT;  
CONVERSION NEUROSIS; BEHAVIOR MODIFICATION

Descriptor Codes: 22360; 00950; 07790; 11730; 05640

Identifiers: behavior modification, hemichorea-hemiballismus conversion  
reaction following head injury, 13 yr old female

Section Headings: 3330 -BEHAVIOR THERAPY & BEHAVIOR MODIFICATION

16/8/3

DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

00278788

61-06234

Organismic characteristics as predictors of problem behavior for boys in  
residential treatment.

Minor Descriptors: ADOLESCENTS; INTELLIGENCE QUOTIENT; IMPULSIVENESS;  
MOTOR COORDINATION; BRAIN DAMAGE; BEHAVIOR PROBLEMS; MALE DELINQUENTS  
; DISORDERS ; SYMPTOMS

Descriptor Codes: 00950; 25920; 24650; 32230; 06780; 05650; 29220  
; 14520; 51140

Identifiers: organismic characteristics, dimensions of behavior  
problems, latency-aged male delinquents

Section Headings: 3230 -BEHAVIOR DISORDERS & ANTISOCIAL BEHAVIOR

16/8/4

DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

00031857

42-19165

STUDIES ON THE HYPERACTIVE CHILD: IV. AN EMPIRICAL ANALYSIS OF THE  
MINIMAL BRAIN DYSFUNCTION SYNDROME.

Minor Descriptors: BRAIN DAMAGE; BRAIN DISORDERS;  
ELECTROENCEPHALOGRAPHY; NEUROLOGY; BEHAVIOR PROBLEMS; CHILDREN

Descriptor Codes: 06780; 06800; 16530; 33740; 05650; 08830

Identifiers: HYPERACTIVE CHILD WITH MINIMAL BRAIN DYSFUNCTION

Section Headings: 3200 -PHYSICAL AND PSYCHOLOGICAL DISORDERS

16/8/5

DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

00024579

42-11013

GROUP WORK WITH MILDLY BRAIN DAMAGED CHILDREN: EFFECT ON SOCIAL  
FUNCTIONING AND SPEECH.

Minor Descriptors: BRAIN DAMAGE; HANDICAPPED; CHILDREN; GROUP  
DYNAMICS; GROUP PROBLEM SOLVING; ETHNIC GROUPS; SOCIAL GROUPS;  
COLLECTIVE BEHAVIOR; ANIMAL SOCIAL BEHAVIOR; SPEECH CHARACTERISTICS;  
SPEECH DEVELOPMENT; VERBAL COMMUNICATION

Descriptor Codes: 06780; 22220; 08830; 21760; 21800; 18080; 48230  
; 10250; 02860; 49070; 49080; 55520

Identifiers: GROUP WORK, SOCIAL FUNCTIONING + SPEECH, 6-9 YR. OLDSE

Section Headings: 3200 -PHYSICAL AND PSYCHOLOGICAL DISORDERS

16/8/6

DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

00007974 41-09252  
PRESUMED MINIMAL BRAIN DYSFUNCTION IN CHILDREN: RELATIONSHIP TO  
PERFORMANCE ON SELECTED BEHAVIORAL TESTS.  
Minor Descriptors: BRAIN DAMAGE; CHILDREN; BEHAVIOR ; MEASUREMENT ;  
TESTING  
Descriptor Codes: 06780; 08830; 05670; 30200; 52330  
Identifiers: BEHAVIORAL TESTS, CHILDREN  
Section Headings: 3200 -PHYSICAL AND PSYCHOLOGICAL DISORDERS

Not so hot, but again, it's a small change. This search is much improved over the previous version, but there is still room for improvement. The techniques we pick up in the next chapter, for searching free-text, will help even more. In the meantime, here are the next 20 or so titles, to give you a better idea of what this set looks like:

00946740 81-22193  
Head injury during childhood. Special Issue: Issues in the  
neuropsychological rehabilitation of children with brain dysfunction.

00935087 81-10902  
The transition from hospital to home: Family readjustment and response to  
therapeutic intervention following childhood-acquired brain injury.

00918261 80-42034  
Cognitive effects of mild head injury in children and adolescents.

00899773 30-87174  
Epidemiologicky vyzkum znaku lehkych mozkovych dysfunkci. 1. etapa:  
konstrukce a overeni metody--vyskyt znaku ve zkoumanem souboru. /  
Epidemiological investigation of signs of minimal brain dysfunctions: I.  
Design and testing of method: Incidence of signs in the investigated group.

00893526 80-18394  
Cognitive and behavioral abnormalities in children whose mothers smoked  
cigarettes during pregnancy.

00893517 80-18385  
Head injured children and education: A need for greater delineation and  
understanding.

00809915 78-24808  
Epidemiologic characteristics and sequelae of closed head-injured  
children and adolescents: A review.

00791508 78-13090  
The head injured adolescent: A case report and review of the issues.

00789006 28-53228  
Family reaction and adaptation to a child's traumatic brain insult and  
neuropsychological recovery.

- 00788767                    28-52989  
The effects on children of a father's severe closed head injury.
- 00781911                    53-03500  
Immediate, short-term and residual effects of acute head injuries in children: Neuropsychological and neurological correlates.
- 00781910                    53-03499  
Disordered brain function in young children and early adolescents: Neuropsychological and electroencephalographic correlates.
- 00781908                    53-03477  
Behavioral correlates of cerebral damage in children aged 9 through 14.
- 00779834                    77-28444  
Comment on presentation of nonorganic factors in Balottin et al.
- 00775256                    78-08210  
Psychosocial consequences of head injury in children and adolescents: Implications for rehabilitation.
- 00754335                    27-58024  
Psychosocial adjustment in parents of head-injured children.
- 00750214                    27-77280  
Die Bedeutung neuropsychologischer Testbatterien für die Diagnostik von Hirnfunktionsstörungen im Kindesalter. / The relevance of neuropsychological test batteries for the diagnosis of brain dysfunctions in childhood.
- 00750213                    27-77279  
Testtheoretische Probleme der Diagnostik von Hirnschädigungen im Kindesalter am Beispiel der Validität neuropsychologischer Testverfahren. / The validity of neuropsychological tests in diagnosing brain damage in childhood: Methodological problems.
- 00700511                    76-37135  
The Trail Making Test: A review of research in children.
- 00671142                    76-12449  
Children and adolescents with traumatic brain injury: Impact on the family.
- 00667820                    76-09127  
Issues in neurobehavioral assessment of mild head injury.

## Controlled Vocabulary Searching on the Internet

This won't take long, because there really isn't anything which could sensibly

be called "controlled vocabulary searching" in the Internet environment. The search engines say they "index" Web documents, but they don't mean that in the sense that the ERIC clearinghouses or Psychological Abstracts do. They don't create abstracts, add descriptors or even have any real contact with documents. When they say "indexing", they really mean "inverted file creation", perhaps with a stoplist.

So why is this section here, and why isn't it over yet? First of all, to make you aware of this situation, but secondly to give you a few ideas of ways to search for networked resources using techniques which could only loosely be referred to as vaguely resembling the use of controlled vocabularies. In general, though, searching on the Net is full text and free text all the way, and we'll deal with those in more detail in subsequent chapters.

If the point of a controlled vocabulary is to lead the searcher to a collection of resources all on a single topic (what cataloging theorists would call the "gathering" function), then the **category structure of Yahoo!** probably comes the closest on the Net. Yahoo! is the de facto "catalog" of the Net, but you should in no way believe it shares important characteristics with library catalogs or indexed DIALOG files--for example, they make no claim of comprehensiveness or selectivity (beyond the "cool" ratings on some sites).

However, it is a good place to start in looking for a list of potentially interesting sites, especially if the words you have to describe the concept of interest is not very specific. (For example, if I were looking for sites about the US Open Tennis championships, *Babylon 5*, or Jimmy Carter, I'd probably start with Yahoo!). Their category labels serve as a very broad categorization, and in fact, when searching in Yahoo, I try to search for words I think will be in category labels rather than site names or their very brief (sometimes nonexistent) descriptions. Search engines are also starting to add category features, but they so far cover dramatically fewer sites than Yahoo!.

If Yahoo! doesn't help, you might also work with search engines, but try to identify the right, really specific word or phrase which everybody uses to describe a particular concept. (Assuming there is one, of course, which there may well not be.) Book/movie/TV/album titles, personal and geographical names are easy, but there aren't name authority files out there, either, so you may have to search variants. More generic concepts can be harder--you might actually try using database thesauri to identify "official" terms such as "attention deficit disorder" or "hebephrenic schizophrenia", but as always on the Net, you're at the mercy of the people who write the documents. "Official" terms will help you to find documents written by people who use "official" language, and miss altogether alternative or nontraditional points of view which might use different language. Welcome to the Net.

A final idea is to directly search for Web sites by guessing at domain names. There are some obvious ones (ford.com for the Ford Motor Company, nbc.com for the NBC network, and so on), and some which are slightly less obvious but can be worked out. (ala.org for the American Library Association, umich.edu for the University of Michigan). There's interesting information about the impact of eating eggs on cholesterol levels from the American Egg Board. If I worked there, I'd suggest they be at eggs.org, but they went for aeb.org, as I discovered from one of their commercials. A little guesswork, possibly supplemented by searching using Yahoo or a search engine, could pay off.

But it still remains the case that the Net is almost entirely a controlled-vocabulary free zone and probably will be for the foreseeable future. This makes the techniques of free-text searching very important there, and we'll explore those in the next chapter.





addl stuff on bound descriptors and identifiers

Why don't we use field qualifiers to ensure that these terms are only searched in the descriptor field? Let's see what happens if we do:

?s *choral music/de or rock music/de or vocal music/de*

```
      117 CHORAL MUSIC/DE (MUSIC INTENDED FOR GROUP SINGING)
        9 ROCK MUSIC/DE
      276 VOCAL MUSIC/DE (MUSICAL COMPOSITIONS WRITTEN FOR VOICES,
        EIT...
S12    368 CHORAL MUSIC/DE OR ROCK MUSIC/DE OR VOCAL MUSIC/DE
```

What happened? Recall the discussion on word versus phrase indexing from Chapter XX. In *ERIC* (most other databases are similar), the descriptor and identifier fields are phrase-indexed. That means that both individual words (CHORAL, MUSIC, ROCK, VOCAL) and complete phrases (CHORAL MUSIC, ROCK MUSIC) are incorporated in the Basic Index. The title and abstract fields, on the other hand, are word indexed. That means that only individual words are entered into the Basic Index. Thus, if we search for the pattern CHORAL MUSIC, the only place it can be found is in the descriptor or identifier fields.

For CHORAL MUSIC and VOCAL MUSIC, the results are the same, because both are terms from the thesaurus. But ROCK MUSIC retrieved 73 documents as is and only nine when qualified to the descriptor field. ROCK MUSIC used to be an identifier, but after several years it was promoted to the status of a descriptor. In the following example, we can see the remaining 64 documents were retrieved under its previous field:

?s *rock music/id*

```
S13    64 ROCK MUSIC/ID
```

This is a relatively infrequent occurrence, but it is something you need to keep in mind.

---