7 USING CONTROLLED VOCABULARY

The search we conducted in the previous chapter on the after effects of traumatic brain injury in children was somewhat successful. We were able to retrieve a reasonable set of 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 were not 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 will discuss in this chapter. The use of *controlled vocabulary* terms—subject headings, descriptors, index terms—can often aid in database searching when they are available. We will talk about why they can be so useful and in what kinds of situations, why they sometimes do not work so well, and some techniques for their use in searching on DIALOG. We will conclude, once again, with a look at how controlled vocabulary searching 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 hundred years or so as ways of making it easier, more often than not, to find items in a collection.

> *Although we will 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 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 is called *synonymy*. The inverse problem is *ambiguity* or *polysemy*—more than one concept for the same "word" or series of characters (e.g., 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, there is greater confidence that all or nearly all of the documents about medical students have been retrieved by using that term.
- It also helps with the *precision* of 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 searchers (or users, especially if they are not that familiar with the subject area) to select better search terms, refine 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 will see in the following chapter where and how other techniques might be chosen, 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*. Most people are probably familiar with the kind of thesauri 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 specific to this thesaurus. Thesauri can differ greatly. For comparison's sake, we will look at another one briefly before we leave the issue, but if the reader really want to know how a thesaurus (and, thus, a database) is created and used, the best bet is to read the explanatory material included with a thesaurus. Experience doesn't hurt either—use a few thesauri and 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 information about drunken driving is being sought, for example, 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 the preferred term to refer to materials that are used in programmed instruction is required, one 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 an increased chance of retrieving documents that 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:

INFORMATION SCIENTISTS *Jul. 1971*

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 document line by line and find out what it means, beginning with

```
INFORMATION SCIENTISTS          Jul. 1971
```

This first line gives the preferred term (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 (e.g., 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 were 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 the kind of inequality. The point of this seemingly elaborated discussion is this: When one wants to use a term such as this in searching, one must remember that the parenthetical is an essential part of the descriptor. If one searches just on

```
?s inequality
    S8       972     INEQUALITY
```

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

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

which will retrieve only documents indexed with that term. Be sure to include a space before the first parenthesis—the system is very picky. If the space is omitted, the result is

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

We will go back to the next line on our entry for INFORMATION SCIENTISTS:

```
CIJE: 230    RIE: 182          GC: 710
```

This line lists the number of postings and group code information. *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, and guides. (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 how many documents in each of these two collections had been indexed under 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, its use may need to be rethought. 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 the 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. Look at the next line.

```
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 the subject field is unfamiliar or when trying to choose between two terms that 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) 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 one was to look in the thesaurus under any of these, one would see

```
Information Brokers
use    INFORMATION SCIENTISTS
```

This is a reciprocal reference, rather like a *see* reference in a library catalog. Having looked up the nonpreferred term, one is 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 reindexed, so if one wants to search on a concept with old and new descriptors and get documents going back to the beginning of the database, both terms may need to be ORed together. When databases are reloaded, though, they are usually reindexed, so this is only a temporary problem.

```
NT   Librarians
     Search Intermediaries
```

Here is 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, be aware of how narrowly the client's search is focused and what kinds of terms best reflect that level of specificity. Also, in perusing the thesaurus, one may find that one has entered a hierarchy at too high or too low a level; the listing of hierarchies will give a better idea of where one 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, PROFESSIONAL PERSONNEL is unlikely to be much help for a search on information scientists. In some circumstances, though, a broader term might 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 that are not part of the hierarchy—neither narrower or broader—for this particular descriptor, but that are related (at least in somebody's opinion), and that may also be of use in searching. When constructing a search strategy, one may find that some of the terms in the RT grouping look useful. 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 a search. If one of them turns out to be a better term than the original choice, turn to the entry for it and start all over.

A caveat: After a while, all terms start to look good. Do not spend more than a few minutes looking through the thesaurus for descriptors, or you will select too many terms, including some real losers, and the search will not 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 one looks, the more one will find, and that is typically counter-productive. Do not be cavalier about term selection, but too much of a good thing is undesirable here, especially if some of the terms have lots of postings.

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 to consult while conducting searches. We will see how to do this shortly.

As we have said, all databases and all thesauri are different. Here is a 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 is 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 7th 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:

SCHIZOPHRENIA *Jul. 1966*
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 [67]
PN 16934 **SC** 45440
UF Chronic Schizophrenia
Dementia Praecox
Process Schizophrenia
Pseudopsychopathic Schizophrenia
Reactive Schizophrenia
Schizophrenia (Residual Type)
Simple Schizophrenia
B Psychosis [67]
N Acute Schizophrenia [73]
Catatonic Schizophrenia [73]
Childhood Schizophrenia [67]
Hebephrenic Schizophrenia [73]
Paranoid Schizophrenia [67]
Schizophreniform Disorder [94]
Undifferentiated Schizophrenia [73]
R Anhedonia [85]
Catalepsy [73]
Expressed Emotion [91]
Fragmentation (Schizophrenia) [73]
Schizoaffective Disorder [94]
Schizoid Personality [73]
Schizotypal Personality [91]

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

At this point, be aware 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 (e.g., 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 that both indexing languages and indexing standards vary greatly among different databases.

3. Not all indexing is done perfectly. In one's experience in searching, one will undoubtedly find index terms that will confuse, amuse, or infuriate.

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, one or more terms to search have been found, 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 that 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 retrieve the 73 documents above, but also any other use of the word ROCK in any field of the database.

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 their behavioral aspects. And yet, you would 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 was not 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 more than 60,000 hits on that term, and the result sets of which it was a part were all broad and unfocused. (And that was using BEHAVIORAL; had we searched broadly on the word BEHAVIOR, we would get about 212,000 hits, and truncating on BEHAVIOR? gets more than 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 that 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.*

Using 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 many others, especially in nonbibliographic files. Other databases may have other fields and codes in their Basic Indexes, so check the bluesheets carefully. One 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, or publication year, 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 the 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 one 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, etc.) and also any indexed with MUSIC ACTIVITIES, APPLIED MUSIC, MUSIC EDUCATION, or even just MUSIC.

Why would anyone want to do this? Sometimes, as with "behavior," one 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 there was interest in documents that discussed the use of music in foreign-language education, a search on the music concept might need to be very broad. 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 document would be entered in the Basic Index under both ORIENTAL and MUSIC. This is called *double posting.*) One may decide, then, to 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, and SLEEP WAKE CYCLE, that incorporate the word SLEEP. Furthermore, SLEEP 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, one 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: 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. (There are many instances of this in the search examples.) We called these *major descriptors* and said that the indexer had decided that these terms best described the document and were the only index terms used in the printed index. 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
```

to retrieve only those documents that 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 type

```
?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 with qualifiers we were restricting to a particular field (title or abstract), and now with descriptors 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 these writers why anyone would want to do that, but it can be done if desired.

> *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 a lot of typing and improve the breadth and recall of a search. If one explodes on a controlled vocabulary term, one will search on it and all its narrower terms, all ORed together. In some files (e.g., ERIC), only the terms that are directly narrower (one level down) in the hierarchy are returned. In other files (e.g., PsycINFO), narrower terms of narrower terms, all the way down, are returned. Check documentation and thesauri to see what a particular file does. This feature is available only in files that 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, one will get a term like JAZZ, which is a narrower term to MUSIC in the thesaurus but does not have the word "music" in it, so it would not appear by searching on MUSIC/DE.

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

We have mentioned this "online thesaurus" several times—we will now see how it works and how it might be used. To access it, we use a DIALOG command that has several other uses: EXPAND.

Viewing the Basic Index and Online Thesaurus: EXPAND

The EXPAND command's primary function is to allow one 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 one is not sure of the spelling of a word, or when 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 that 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 as an indication that this is in fact the term 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 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.

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 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 can be seen, more than one term at a time can be selected, 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
```

The EXPAND display goes up to e50. If further EXPAND statements are used, they are, in effect, overwriting in the same computer space. This means that one can only SELECT from that most recent display. The others have been lost and must be recreated to be seen or used.

There are two ways to enter the online thesaurus. If there is a descriptor in an alphabetical EXPAND display (such as BIAS above), one 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 there is a new display, similar to the other but slightly different. Instead of the reference numbers beginning with E, they begin with R. This means that one is 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. The full entry is not visible as it would be 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 now any terms in the display can be expanded 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 it is already known that a certain term is a descriptor, one 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 one is paying online time to work with the thesaurus. It is useful when there is no printed thesaurus available, but it is not a substitute for thorough preparation.

However, if the online controlled vocabulary is used cleverly, it can save time, money, and effort. One can SELECT from the R-display directly, which will allow one to avoid typing long descriptors and possibly making typing or spelling errors. Because the online thesaurus is sometimes updated more frequently that the print version, there may be new descriptors or relationships online to help, and certainly the postings information will be more up-to-date, so the potential size of the retrieved sets can be better gauged.

Search Example

For the sample DIALOG search, we will 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 that contain the words PSYCHOSOCIAL, NEUROPSYCHOLOGICAL (in various forms); and many with BEHAVIOR, not surprisingly. We will try searching on them in various ways, as will be seen below. Finally, the preferred terms for CHILDREN and ADOLESCENTS are children and adolescents. We begin the search with these terms.

```
File 11:PsycINFO(R) 1967-1998/May
 (c) 1998 Amer. Psychological association

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

    S1      2323    HEAD INJURIES (1973)
    S2      4677    BRAIN DAMAGE (1967)
    S3       370    TRAUMATIC BRAIN INJURY (1997)
    S4      7023    HEAD INJURIES OR BRAIN DAMAGE OR TRAUMATIC BRAIN INJURY
?ss psychosocial or neuropsych? or behavior/de

    S5     28364    PSYCHOSOCIAL
    S6     18301    NEUROPSYCH?
    S7    139787    BEHAVIOR/DE (1967)
    S8    181319    PSYCHOSOCIAL OR NEUROPSYCH? OR BEHAVIOR/DE
?ss children/df or adolescents/df

    S9     30379    CHILDREN/DF (1967)
    S10    32256    ADOLESCENTS/DF (1967)
    S11    58049    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 is 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; we will put them all together.

```
?s s4 and s8 and s11

          7023     S4
        181319     S8
         58049     S11
   S12     135     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) 1998 Amer. Psychological Assn. All rts. reserv.

01532776           1998-07266-014
Parent training.
SERIES TITLE: The LEA series in personality and clinical psychology.

DESCRIPTORS: *Behavior Modification; *Daily Activities; *Disorders; *Parent
  Child Relations; *Parent Training; Adults; Children; Developmental
  Disabilities; Head Injuries; Schizophrenia
IDENTIFIERS: planned activities parent training behavioral technique,
children & adults with normal intelligence or developmental disabilities or
head injuries or schizophrenia or other disorders
SUBJECT CODES & HEADINGS: 3200 (Psychological & Physical Disorders); 3312
  (Behavior Therapy & Behavior Modification)

 12/8/2
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01532201           1998-07011-007
Traumatic brain injury.

DESCRIPTORS: *Diagnosis; *Measurement; *Traumatic Brain Injury; *Treatment;
  Children; Epidemiology
IDENTIFIERS: description & psychological & psychiatric assessment issues &
epidemiology & medical & psychological & behavioral & pharmacological
treatments, children with traumatic brain injury
SUBJECT CODES & HEADINGS: 3297 (Neurological Disorders & Brain Damage);
  3360 (Health Psychology & Medicine)

 12/8/3
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01523141           1997-38757-001
Head injury in children.

DESCRIPTORS: *Brain Damage; *Head Injuries; *Literature Review; *Traumatic
  Brain Injury; Children
IDENTIFIERS: head injury in children, literature review
SUBJECT CODES & HEADINGS: 3297 (Neurological Disorders & Brain Damage)
```

12/8/4
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01522010 1997-36671-015
Central nervous system dysfunction: Brain injury, postconcussive syndrome, and seizure disorder.
SERIES TITLE: Issues in clinical child psychology.

DESCRIPTORS: *Behavior Therapy; *Central Nervous System Disorders; *Drug Therapy; *Medical Treatment (General); *Rehabilitation; Brain Concussion; Children; Convulsions; Students; Traumatic Brain Injury
IDENTIFIERS: medical management & rehabilitation & pharmacological issues & behavioral interventions, students with pediatric traumatic brain injury or postconcussion syndrome or seizure disorders
SUBJECT CODES & HEADINGS: 3570 (Special & Remedial Education); 3300 (Health & Mental Health Treatment & Prevention)

12/8/5
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01512432 1997-36680-006
Frontal lobe dysfunction following closed head injury in children: Findings from neuropsychology and brain imaging.

DESCRIPTORS: *Brain Disorders; *Head Injuries; *Literature Review; *Neuropsychological Assessment; *Tomography; Children; Neuropsychology; Prefrontal Cortex
IDENTIFIERS: neurobehavioral sequelae of & neuroimaging techniques for & performance on neuropsychological tests & prefrontal brain dysfunctions following closed head-injuries, children, literature review
SUBJECT CODES & HEADINGS: 3297 (Neurological Disorders & Brain Damage)

12/8/6
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01494841 1996-93815-012
Behavior modification with brain-injured children: A brief note on directions for research.

DESCRIPTORS: *Behavior Modification; *Experimentation; *Traumatic Brain Injury; Children
IDENTIFIERS: directions for research on behavior modification with brain-injured children
SUBJECT CODES & HEADINGS: 3297 (Neurological Disorders & Brain Damage)

12/8/7
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01482579 1997-30097-007
Pediatric neuropsychology.

DESCRIPTORS: *Brain Disorders; *Developmental Stages; *Neuropsychological Assessment; *Neural Development; Adolescents; Preschool Age Children; School Age Children; Oral Communication; Motor Skills; Head Injuries; Cognitive Processes
IDENTIFIERS: development of nervous system & motor & speech & higher cognitive functions & application to development of neuropsychological assessment approach, 2-5 & 6-14 yr olds with neuropsychological deficits
SUBJECT CODES & HEADINGS: 3297 (Neurological Disorders & Brain Damage)

```
 12/8/8
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01482317           1997-09204-008
Children with brain injury.

DESCRIPTORS: *Biopsychosocial Approach; *Brain Damage; *Neuropsychological
  Assessment; *Systems Theory; Children
IDENTIFIERS: nature of various brain disorders & methodological &
biobehavioral systems approaches to neuropsychological assessment,
children with brain injury
SUBJECT CODES & HEADINGS: 2225 (Neuropsychological Assessment); 3297
  (Neurological Disorders & Brain Damage)
```

Pretty good—a reasonably sized set, with many potentially interesting documents. 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

          28364    S5
          18301    S6
           4565    BEHAVIOR/MAJ (1967)
    S13   50523    S5 OR S6 OR BEHAVIOR/MAJ

?s s4 and s13 and s11

           7023    S4
          50523    S13
          58049    S11
    S14     116    S4     AND S13 AND S11
```

and it appears to make very little difference. We only lose 19 documents by doing this. This would probably not be done in a real search, but we will take a moment and see what those look like. We can use the Boolean operator NOT to do this:

```
?s s12 not s14

            135    S12
            116    S14
    S15      19    S12 NOT S14
?t 15/8/1-6

 15/8/1
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01532776           1998-07266-014
Parent training.
SERIES TITLE: The LEA series in personality and clinical psychology.

DESCRIPTORS: *Behavior Modification; *Daily Activities; *Disorders;
  *Parent Child Relations; *Parent Training; Adults; Children;
  Developmental Disabilities; Head Injuries; Schizophrenia
IDENTIFIERS: planned activities parent training behavioral technique,
children & adults with normal intelligence or developmental disabilities
or head injuries or schizophrenia or other disorders
SUBJECT CODES & HEADINGS: 3200 (Psychological & Physical Disorders); 3312
  (Behavior Therapy & Behavior Modification)
```

15/8/2
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01522010 1997-36671-015
Central nervous system dysfunction: Brain injury, postconcussive syndrome, and seizure disorder.
SERIES TITLE: Issues in clinical child psychology.

DESCRIPTORS: *Behavior Therapy; *Central Nervous System Disorders; *Drug Therapy; *Medical Treatment (General); *Rehabilitation; Brain Concussion; Children; Convulsions; Students; Traumatic Brain Injury
IDENTIFIERS: medical management & rehabilitation & pharmacological issues & behavioral interventions, students with pediatric traumatic brain injury or postconcussion syndrome or seizure disorders
SUBJECT CODES & HEADINGS: 3570 (Special & Remedial Education); 3300 (Health & Mental Health Treatment & Prevention)

15/8/3
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01494841 1996-93815-012
Behavior modification with brain-injured children: A brief note on directions for research.

DESCRIPTORS: *Behavior Modification; *Experimentation; *Traumatic Brain Injury; Children
IDENTIFIERS: directions for research on behavior modification with brain-injured children
SUBJECT CODES & HEADINGS: 3297 (Neurological Disorders & Brain Damage)

15/8/4
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01481495 1997-09033-008
A positive, communication-based approach to challenging behavior after ABI.

DESCRIPTORS: *Behavior Problems; *Classroom Behavior Modification; *Special Education; *Communication Skills Training; *Traumatic Brain Injury; Adolescents; Cognitive Rehabilitation; School Age Children
IDENTIFIERS: critical intervention themes in & communication-based approach to preventing evolution of problem behaviors, child & adolescent students with acquired brain injury
SUBJECT CODES & HEADINGS: 3570 (Special & Remedial Education)

15/8/5
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01481494 1997-09033-007
Understanding and overcoming the challenging behaviors of students with ABI.

DESCRIPTORS: *Behavior Problems; *Classroom Behavior Modification; *Special Education Students; *Traumatic Brain Injury; Adolescents; School Age Children
IDENTIFIERS: problem behaviors & behavior change strategies, students with acquired brain injury
SUBJECT CODES & HEADINGS: 3570 (Special & Remedial Education)

```
 15/8/6
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01480852           1997-08859-009
Pediatric traumatic brain injury: Challenges and interventions for
  families.

DESCRIPTORS: *Coping Behavior; *Family Therapy; *Traumatic Brain Injury;
  *Needs; Advocacy; Behavior Therapy; Parents; Roles; School Psychologists;
  Professional Consultation; Parent School Relationship; Family Members;
  Children; Client Education
IDENTIFIERS: coping experiences & needs for & psychologist's role in
education & support & advocacy & behavioral therapy & home-school
consultation, children with traumatic brain injury & their families
SUBJECT CODES & HEADINGS: 3313 (Group & Family Therapy)
```

Not so good, but again, it is 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 first 20 titles of that good set 14, to give a better idea of what this set looks like:

```
 14/6/1
01532201           1998-07011-007
Traumatic brain injury.

 14/6/2
01523141           1997-38757-001
Head injury in children.

 14/6/3
01512432           1997-36680-006
Frontal lobe dysfunction following closed head injury in children: Findings
  from neuropsychology and brain imaging.

 14/6/4
01482579           1997-30097-007
Pediatric neuropsychology.

 14/6/5
01482317           1997-09204-008
Children with brain injury.

 14/6/6
01481491           1997-09033-004
The role of neuropsychology in educating students with ABI.

 14/6/7
01481334           1997-08987-030
Traumatic brain injury in children: Neuropsychological, behavioral, and
  educational issues.

 14/6/8
01481171 1997-08958-008
Recent advances in neuropsychological assessment of children.
SERIES TITLE: Critical issues in neuropsychology.
```

14/6/9
01480853 1997-08859-010
Children and adolescents with traumatic brain injury: Reintegration challenges in educational settings.

14/6/10
01480848 1997-08859-005
Neuropsychological consequences of traumatic brain injury in children and adolescents.

14/6/11
01480845 1997-08859-002
Assessing children with traumatic brain injury during rehabilitation: Promoting school and community reentry.

14/6/12
01480843 1997-08859-000
Childhood traumatic brain injury: Diagnosis, assessment, and intervention.

14/6/13
01479633 1997-08457-002
Pediatric brain injury: Mechanisms and amelioration.
SERIES TITLE: Critical issues in neuropsychology.

14/6/14
01471818 1997-05606-001
Mild head injury in children and adolescents: A review of studies (1970-1995).

14/6/14
01461782 1997-97163-001
Specificity of brain-behavioural relationships revisited: From epileptic personality to behavioural phenotypes.

14/6/16
01457377 1997-03731-002
A taxonomy of neurobehavioral functions applied to neuropsychological assessment after head injury.

14/6/17
01456899 1997-03303-005
Factors contributing to successful return to school for students with acquired brain injury: Parent perspectives.

14/6/18
01451651 1996-98606-010
Cognitive rehabilitation for children with traumatic brain injury.

14/6/19
01437777 1996-06915-007
Assessing children with traumatic brain injury during rehabilitation: Promoting school and community reentry.

14/6/20
01427791 1996-04856-003
Behavioural adjustment and parental stress associated with closed head injury in children.

Controlled Vocabulary Searching on the Internet

This will not take long, because there really is nothing that could sensibly be called "controlled vocabulary searching" in the Internet environment. The search engines say they "index" Web documents, but they do not mean that in the sense that the ERIC clearinghouses or Psychological Abstracts do. They do not create abstracts, add descriptors, or even have any real contact with documents. When they say "indexing," they really mean "inverted file creation of words in the full document," perhaps with a stoplist.

So why is this section here, and why is it not over yet? First, to promote awareness of this situation, and second, to provide a few ideas for ways to search for networked resources using techniques that could only loosely be referred to as resembling the use of controlled vocabularies. In general, though, searching on the Net is full text and free text all the way, and we will deal with those types of searching in more detail in the next chapter.

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 one 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 looking for a list of potentially interesting sites, especially if the words for the concept of interest are not very specific. (For example, if I were looking for sites about the US Open Tennis championships, *Babylon 5*, or Jimmy Carter, I would probably start with Yahoo!). Yahoo!'s 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! does not help, one might also work with search engines but try to identify the right, really specific word or phrase that everybody uses to describe a particular concept. (Assuming there is one, of course, which there may well not be.) Book/movie/TV/album titles and personal and geographical names are easy, but there are no name authority files out there, either, so variants may have to be searched. More generic concepts can be harder—one might actually try using database thesauri to identify "official" terms such as "attention deficit disorder" or "hebephrenic schizophrenia," but as always on the Net, one is at the mercy of the people who write the documents. "Official" terms will help to find documents written by people who use "official" language, but miss altogether alternative or nontraditional points of view that might use different language. Welcome to the Net.

A final idea is to directly search for websites by guessing at domain names. There are some obvious ones (e.g., ford.com for the Ford Motor Company, nbc.com for the NBC network), and some that are slightly less obvious but can be worked out (e.g., ala.org for the American Library Association, umich.edu for the University of Michigan). There is interesting information about the impact of eating eggs on cholesterol levels from the American Egg Board. If we worked there, we would suggest they be at eggs.org, but they went for aeb.org, as we discovered from one of their commercials. A little guesswork, possibly supplemented by searching using Yahoo or a search engine, can 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 will explore those in the next chapter.

Additional Reading

Bates, Marcia J. (1988), "How to Use Controlled Vocabularies More Effectively in Online Searches," *Online* 12(6): 45–56.

Fidel, Raya (Winter 1992), "Who Needs Controlled Vocabulary?" *Special Libraries* 83: 1–9.

Morton, Douglas (1994), "Refresher Course: Expanding Your Outlook," *Online* 18(3): 77.

Tenopir, Carol (November 15, 1987),"Searching by Controlled Vocabulary or Full Text?" *Library Journal* 112: 58–59.