

6

Search Technique

This chapter introduces many of the major concepts and basic mechanics of searching. Again, we will use the DIALOG system's commands as illustrations, but also include a discussion of current Internet-based search technique at the end. As a framework for presenting the particular commands and ideas, we outline an eight-step procedure, really a codification of common sense, which we hope will help you to formulate effective search strategies and conduct successful searches. We will also use a particular search as an example to walk you through these steps.

- 1 Read the query.**
- 1a Listen to the query.**
- 1b Understand the query.**

This is only somewhat flippant. Whatever you receive to trigger a search--a phone call or electronic mail message from a patron, a written search request form, a panicky visit to the reference desk--this is your first, best resource. No matter how much the person requesting the search knows about the actual topic, he or she is the one requesting the search, and so is the only one to ask about regarding how many documents they want, what kinds of documents, what focus to put on the search, and so on. We'll talk more about the search interview process later on, but for now you should know that any information you can get from the user could be very helpful. Examples include potential search terms, known authors or titles of good documents (but be careful--these could do more harm than good, especially if the names are wrong!), the results of any previous search attempts they've made, and so on. Many search services use forms to elicit this sort of information; here are a couple of samples:

<insert Figure 6.1 here>
<insert Figure 6.2 here>

Figure 6.1 Sample Search Request Form

Figure 6.2 Internet Public Library Reference Question Form

An important note: it is entirely possible that a query presented to you for an "online search" really doesn't belong there. It might be that it's more a traditional ready-reference type question, or that a manual or Internet search would be quicker or more appropriate or more successful. Don't get seduced into believing that online searching will answer all questions, because it won't. There are situations in which a search in a commercial online service would be faster and cheaper, but this is not always the case.

So here is the sample query we will use as a demonstration. It is a real query from a psychology doctoral student working on the literature review for her dissertation.

Search Technique
September 24, 2001 2:15 PM

<insert Figure 6.3 here>

Figure 6.3 Completed Search Request Form

A quick read of the request form gives us the sense that the user has thought about this topic quite a bit and in fact has given us two known documents with authors and several potentially good search terms. These may well be quite helpful in the search. We'll talk more about the search interview process in a later chapter; for now let's assume that the conversation reinforces what we see on the form, and that she's looking for as much as we can possibly give her. This makes sense for a doctoral student in the bibliography-building stage, so we should be looking to have larger rather than smaller sets--a high-recall search. She's not sure how much is out there, but thinks it might be as many as 100 or 150 documents, perhaps more.

2 Identify the major concepts in the query.

Many requests for information that can be searched effectively online involve more than one concept. One-concept searches can certainly be searched online, but often there is a second concept lurking in the user's mind. For example, a patron seeking information on bilingual education may actually be interested in bilingual education in elementary schools, or materials used in bilingual education, or the controversy sometimes raised about it.

It is not always easy or straightforward to identify these concepts, sometimes called facets. Different people will find different concepts and act on them differently, and there is often no one "right" analysis.

Concept Analysis

We decide that an online search would help to answer this query--it does indeed have multiple concepts and the user is looking for articles, conference papers and other materials which are included in online databases--so we begin to analyze the concepts it contains.

Concept means the abstract idea of a thing, regardless of what it may be called in a given instance. This is because very often a single concept (e.g., teacher) will have more than one recognizable name (instructor, tutor, professor, lecturer, master, coach).

The controlled vocabulary in an information retrieval system is an attempt to standardize these words to one preferred term that will always be used to represent a single concept, so that we will not find the same subject entered under different headings. A golden rule for most retrieval systems is to try to gather together under one heading all the material on one subject. - GW

For this search, we identify three concepts: **traumatic brain injury, their effects, and children.** Again, there's nothing magical about this process, and it is something you get better at with practice and experience, and you probably thought of the same words when you first looked at the query. If not, don't worry, but look at your analysis and ours and see how they differ. Are they roughly similar? Did you see two concepts, or even four? How might the way you conceptualized the search affect how you do the search? This set of concepts makes sense to us, but it's not the only one, and different ones might well produce equally good (if not better) results.

Building Blocks

Search Technique

September 24, 2001 2:15 PM

Most searchers use a technique called **building blocks** in constructing their search strategies. It might help to think of your search strategy as a structure, built up from individual pieces, each of which corresponds to a concept derived from the analysis of the query. A good metaphor is Legos: each term is an individual Lego block; you put a number of them together and they form a bigger block. You can then put those bigger blocks together and make something even more complicated. If you don't like the way a particular piece looks, or if it's in the wrong place, you can move it or even take it out. Searching is like that--finding terms you think will work, putting them together in concept blocks, and then combining the concept blocks to see what they produce, and then revise the search as necessary. The steps below will walk you through that process.

3 Identify potential terms to correspond to those concepts.

Term Selection

We've identified three concepts, so all we need to do is go into the database, look them up, put them together, and go home, right? Unfortunately, it doesn't quite work that way. Remember what we said in the last chapter: we want to look for concepts, but we are forced to search for words. There's often no obvious way to go into a database and pull out just and only the documents about a particular concept. In some cases you can, when the concept is very specific, and there's really only one way to refer to it. <probably ought to have an example here, huh?>

Most of the time, though, that's not the case, so we have to try to find multiple terms which might be used to represent each concept. You might identify several and choose to use only one for strategic reasons, but we'll get to that later.

In this case, the user appears to be a good source of terms: TRAUMATIC BRAIN INJURY, CLOSED HEAD INJURY, PSYCHOSOCIAL, PEDIATRIC, and so on. But they're really all over the place. It would help to have a way to organize the process and make it easier to keep track of all of this. Have a look at the search grid.

<insert Figure 6.4>

Figure 6.4 Search Grid

It's a bit overwhelming at first, but if you look at it for a bit, it'll start to make some sense.

Look first at the boxes marked "Concept 1", "Concept 2" and "Concept 3". In those boxes, we've written in the concepts we previously identified. Right below each of these, you see a series a lines marked "S#" and "Terms". These are spaces for recording potentially good terms, and we've taken the user's terms and phrases and entered them under the corresponding concepts. (Although we haven't quite gotten there yet, you can use the "S#" spaces to record set numbers as you go, if you need help in keeping track of what's what.)

A couple of things to point out. First of all, you'll notice that in a couple of cases, we've got terms which are exactly the same as our concept names. Nothing special about this, sometimes it happens and sometimes it doesn't, as with the second concept. Don't feel you have to do that, but don't worry about it if you do.

Second, a few terms are recorded here in a slightly different way. Look at BEHAVIOR(AL) and CHILD(REN), and ADOLESCENT(S). This is Joe's

shorthand, and is a note to remember that there may be varying forms of these terms. Documents might use the word "behavior" or "behavioral", "child" or "children", "adolescent" or "adolescents". This might be one way for you to remind yourself about plurals or other variant forms. Whatever. We just suggest you find a way of working that suits you and makes you comfortable and most important helps you get quality results.

There is also no one right way to pick terms. In a group of five searchers working on the same query, you might well find they identify five different sets of terms. There would be at least some overlap, to be sure, but there is rarely only one way to go about doing a search.

I often tell my classes that there's no *right* way to do a search, but there are *bad* ways. Some conceptualizations and terms will just work better than others in producing sets of documents which the user actually wants and will find useful. It's a very difficult process shot through with ambiguity at more than one level. Don't be intimidated--just keep at it and keep improving. - JWJ

4 Select alternative (narrower, broader, or related) terms to use if the original strategy needs help.

When doing term selection, you may come across terms that you think are possibly or marginally useful but don't immediately grab you. They may in fact be good, but if your initial instinct is ambivalent, you might want to hold them out as reserves or alternatives, and perhaps put them in the right-hand column of the grid. Although you may find this hard to believe, your initial, beautifully honed crafted search strategy might not be perfect. You may well find that you need more terms, narrower terms, broader terms, different terms, or even fewer terms. Thus, it is usually a good idea to have a few additional terms in your back pocket--just in case.

In this search, we have a couple of really interesting (and specific) terms from the user in the first concept, but they might not work or be a bit too specific. We may try a few other ideas, combining some of the same words but in different ways.

5 Determine logical (Boolean) relationships between terms.

We have a good list of terms to use in finding documents for our patron, but we can't just enter them all and have the answers come out. We have to group them by concepts and then combine those concepts in the appropriate ways. There's a specific way to do this, based on the logic which underlies the construction of many information retrieval systems. This logical understructure comes to us from set theory and is usually called **Boolean logic**.

Boolean Logic and Boolean Searching

Boolean logic is part of a set of techniques used in mathematics for manipulating sets in a rigorous, logical fashion. It is named for the English mathematician George Boole, who developed the framework on which it is based. Boolean logic provides three ways in which sets can be combined, and online systems use all three.

When a search term is entered, a set of documents that contain that term is created. Boolean search techniques allow the searcher to manipulate and combine these sets to provide the user with a set that corresponds to the logic of the initial query. We will discuss each of these three Boolean operators in turn.

Before we begin, though, you should know that the use of Boolean logic and searching is neither universal nor identical across information retrieval systems. Most commercial systems like DIALOG use them in very similar ways, but, for example, most Internet-based systems at present have only fragmentary and simple Boolean search capabilities. We'll talk about them more at the end of this chapter, but you should be aware that although full Boolean capability has many advantages, it's not the only game in town.

OR

Of course, there is another way. We could, if we wanted to, go through the database and find all the documents with, say, the word TRAUMA in them, another with the word PSYCHOSOCIAL and a third with the word CHILDREN and then compare them to see which ones they all have in common. If we were doing the search manually, using print indexes, we'd do precisely that, but it would be tedious and time consuming and we'd make lots of mistakes. Using Boolean search tactics with the inverted file for the database will make it much easier and quicker.

We've got several terms for the "effects of brain injury", and we want to use them all. Look again at the grid where we've recorded these terms, and specifically, look on the right hand side of that box. See the brace and the big OR there? This tells us that we should use the boolean operator OR to combine those terms and create a concept block, and that is precisely what we will do.

OR is used to **build up concepts**, and can be helpful in several circumstances:

-- for **synonyms** or **equivalent terms**

GARBANZOS **OR** CHICK PEAS
STUDENTS **OR** PUPILS
OCCUPATIONS **OR** JOBS **OR** CAREERS

-- for **spelling variations**

HONOR **OR** HONOUR
ORGANIZATION **OR** ORGANISATION
JUDGMENT **OR** JUDGEMENT

-- for **related terms**

CLOSED HEAD INJURY **OR** TRAUMATIC BRAIN INJURY
PSYCHOSOCIAL **OR** NEUROPSYCHOLOGICAL **OR** BEHAVIORAL
CHILDREN **OR** ADOLESCENTS **OR** PEDIATRIC

This is necessary for a variety of reasons. Authors may use different forms of these words or variant spellings. It may also be that words in titles or abstracts may be slightly different from those used as subject headings. Finally, there simply may be more than one term or word used to represent a single concept or idea. We want the **concept** to be present in the documents we retrieve--the concept of "dogness" or "children" or "Europe"--but human languages are ambiguous and permit multiple ways of saying the same thing. So, once again, we often have to use multiple words to get at a single concept.

For example, suppose a patron is looking for documents about the trade policies within Europe. To represent the concept of "Europe", we might look for EUROPE but also EC (an abbreviation for the European Community), EEC (the European Economic Community), EU (the European Union, the more recent name for the economic community) or even the

names of individual countries. In the *ABI/Inform* database, which covers periodicals in business, we find the following postings figures:

EUROPE	80371
EC	16007
EEC	2663
EU	6757
EUROPE OR EC OR EEC OR EU	86934

Notice that the final number is less than the sum of the individual sets due to overlap between them.

A tool known as a **Venn diagram**, named after John Venn, is often used to represent sets and Boolean operations. The figure below is a Venn diagram which represents our first ORed concept:

<insert Figure 6.5>

Figure 6.5 Venn Diagram for OR

OR will create a set that will retrieve any documents which contain **any** individual term or any **combination** of the terms. Since it can only add documents, it will retrieve **more** documents than any term individually would, and it makes sets **bigger**.

You should be careful, though, of individual terms that **dominate** a concept set. In the Europe example above, most of those documents are probably about Europe *in general* and not the EU specifically, because the EUROPE set is so much larger than the others. This is not necessarily a problem, but can be. If your results seem to be too general, not specifically about a particular concept, it might well be that you need to reconstruct a concept set, dropping a dominant term and using only the more specific, focused terms.

AND

Once we have constructed sets for the individual concepts, we need some way of connecting them so we can find documents which (we hope) are "about" all of them. The way we do this is by use of the Boolean operator AND. Look again at the grid, and notice the big ANDs between the concept boxes. This is a reminder that we use AND to pull these concepts together and see what they have in common.

Combining two or more sets with AND will produce a set which contains documents in **all** of those sets only. That set must be **smaller** than the concept sets, and will therefore produce **fewer** documents. We call this set the **result set**.

Here's a Venn diagram illustrating how AND works, with two or three concept sets.

<insert Figure 6.6>

Figure 6.6 Venn Diagram for AND

Another word of caution: before you do an AND, look at the sizes of the individual original concept sets. Since AND can only reduce size, if one or more of the concepts sets are small (say, fewer than 50 documents, or about the size that the user said they wanted), you might think twice about even doing the AND. Think about it: the user says they want about 60 documents, and one of your concept sets has 47. Before blindly ANDing it in with one or two other sets, you might want to stop right there and have a look at it. Maybe *that's* the result set

Search Technique

September 24, 2001 2:15 PM

you want, or maybe you need more terms or broader terms in there. Because, if you AND it in with other concept blocks, you're likely to get a very small result set and have very little to work with. It might even come up with no documents at all. So just take a second and look at those numbers before you go on.

NOT

The third boolean operator is also the least often used, at least in the same way as AND and OR, for a couple of reasons. It's not even on the search grid, or at least is only tucked down in the corner under "other features". Many beginning searchers do not fully appreciate NOT and its uses, and it is a very powerful tool. Too powerful, in many situations. NOT is used to **exclude** items from a set, but it's a very blunt instrument and can have unintended and quite nasty consequences, especially if used in haste or panic.

It's difficult to come up with a good example of its fruitful use in working with concepts, since it does have such power. (We'll see it use in another way shortly.) In the next chapter, we'll do a search about distance education in library schools in the United States. It turns out that there's no good search term for "United States" in the *ERIC* database, since so many of the documents in there cover it. But there is a term FOREIGN COUNTRIES. So, to eliminate documents which are otherwise good but are from abroad, we could NOT out the ones with FOREIGN COUNTRIES. Even here, we could lose good stuff which is about both US and foreign programs, because **any** document which has that term will be eliminated, but in this case it is appropriate and we won't lose too many documents.

So, don't let NOT be a first thought. If you find yourself wanting to use NOT to eliminate a concept, think first whether there might be another way. If you've got a term that's giving you trouble, maybe you should reconstruct that concept set without it. Go from

OCCUPATIONS OR JOBS OR CAREERS

to

OCCUPATIONS OR CAREERS

if JOBS is producing junk, like documents about Steven Jobs.

Or perhaps you need to AND something new in, a new concept or focusing mechanism like date or language. There are places to use NOT, but it should almost always be a second choice.

Here's a Venn diagram for NOT:

<insert Figure 6.7>

Figure 6.7 Venn Diagram for NOT

Precedence of Boolean Operators, Nesting, Use of Parentheses

You also should take care in the order in which you enter search terms. You can, if you like, enter terms from more than once concept in a single statement. We don't recommend it, especially for beginning searchers, since it can be confusing, but you should know how to do it correctly. Most (but not all) retrieval systems will perform NOTs first, followed by ANDs, and finally ORs, and will allow the use of parentheses to override this order.

So, a statement such as

Search Technique

September 24, 2001 2:15 PM

(ORATORIO OR OPERA) AND HANDEL

would be different from

ORATORIO OR OPERA AND HANDEL

The first (correct) statement would be interpreted this way: things in parentheses are done first, so the system would find all documents with either the word ORATORIO or OPERA or both. Then, those would be ANDed with the documents which have HANDEL. This produces a set that, we hope, contains documents about vocal works of Handel. Here's the Venn diagram for that:

<insert Figure 6.8>

Figure 6.8 Handel's Vocal Work

The second statement, though, would be interpreted quite differently. Because there are no parentheses, the AND goes first, so the system would find all documents that satisfy OPERA AND HANDEL. Then those would be ORed with all the documents with ORATORIO. The resulting set would consist of all documents about Handel's operas mixed in with everything about oratorios--not what we had in mind. See how different this Venn diagram is:

<insert Figure 6.9>

Figure 6.9 Oratorios, with Handel's Operas

There's nothing really wrong with doing something like the first statement, and most experienced searchers would be quite comfortable doing it, especially for something with only three terms and a simple relationship between them. Just be sure you've got it right!

6 Begin the search.

Here is where you find out if your well-planned strategy will work or not. Even the most experienced and skillful searchers have times when it just doesn't work, when they just can't find anything of use or the search comes up dry. Sometimes, there just isn't anything to be found--there aren't any documents in that database in the area the user wants. And sometimes, you can't find the right search terms or combinations to pull up the good stuff.

Actually, there are times when the user probably doesn't *want* anything to come up. There are circumstances, for example in patent searching, when finding something is a bad sign. Someone else has patented your device, or written a book on your dissertation topic, so your work has gone for naught. Another reason why it's good to know what the user wants, and *why*. - JWJ

You will probably find it most effective to enter all the terms (ORed together) for each concept block together, see how many hits you get for each concept block, and then AND the remaining sets together to produce a result set.

In some cases, that result set will be a good one and will contain useful documents for the patron. Often, though, it could be improved (more about that in step 8). Pay attention at these early stages so you have a sense of what's going on and what you might do at later steps.

Most Specific First

Search Technique

September 24, 2001 2:15 PM

It might seem that it doesn't really matter which of your concept blocks you enter first, since you're just going to AND them all together shortly anyway. Although that's true, there's something more to it than just that. An important and widely used convention is *most specific first*, which can be very helpful. Imagine you were doing a search on the reliability of children as witnesses in child abuse trials in the *Criminal Justice Information System (CJIS)* database. Here you have three concept blocks: children, witnesses, and child abuse. (You might also think of this as two: child-witnesses and child abuse, and if there's a subject heading for child-witnesses, that might well work.)

In a database which covers criminal justice, there are likely to be many more documents on child abuse and witnesses than on children per se. If we search for "child abuse" first, we'll get this large set and really not learn very much. However, if we search for "children" first, it will be a smaller set, and will give us some indication of how many documents we're likely to get when all is said and done. If the database supported controlled vocabulary searching, a term like CHILD WITNESSES would be ideal.

If that first and most specific set is really small (say, 10 documents or less), we might decide to stop right there and either try new terms, reevaluate the search overall, or just inspect that first set--it might be easier than continuing the search. If, though, we get several hundred documents, we proceed, but in the knowledge that our overall result set is likely to be smallish. Experienced searchers can use that kind of information to help guide how they will search from that point on.

Note, though, that if we were doing this same search in, say, *Child Abuse & Neglect & Family Violence*, we'd probably enter the "witnesses" concept first, since it is probably the most specific. This is by no means an exact science, and if you search for the broadest concept set first, it probably won't hurt you, but most specific first can be very helpful in many circumstances, especially in small databases or with narrow topics.

Getting In, Choosing Files: BEGIN

The major online search services, such as DIALOG, LEXIS/NEXIS, Dow Jones, and so on, are accessed via telecommunications networks or the Internet, as described in Chapter 3. When you connect to the service, you must log on in order to start the search. To do this, you must have established an account with the system; they will provide you with an account number and password which you use to authenticate to the system. This prevents unauthorized use of the service (i.e., freeloading).

We begin our sample search with this login procedure. In search transcripts, we'll show what you as the user type in *italics*, and the system's responses like this.

```
dialog
Trying 192.132.3.254 ...
Connected to dialog.com.
Escape character is '^]'.
Trying 3106...Open

DIALOG INFORMATION SERVICES
PLEASE LOGON:
*****
ENTER PASSWORD:
*****
Welcome to DIALOG
```

Search Technique
September 24, 2001 2:15 PM

Dialog level 42.12.05B

Last logoff: 21jan97 11:52:28
Logon file001 22jan97 14:53:03

File 1:ERIC 1966-1996/Dec
(c) format only 1997 Knight-Ridder Info

Set	Items	Description
---	-----	-----
?		

You'll notice that in the above, only the system's responses are shown. The searcher has typed in her password, but the system doesn't display it for security reasons.

The greeting tells us that the last use of the system from this user number was on January 21, 1997, and that we are in the *ERIC* database, which is file 1 in DIALOG. This is the **default database** selected by the user, meaning that when she logs in, this is the file she'll be in, perhaps the file she searches most often.

Typically, at this stage, system news will come up: new databases which are available, files not working at present, revisions or reloads of older files, and so on. After this system news (here omitted) comes the **file header**. This tells us what we are currently searching in (*ERIC*), with the dates of coverage of that file (1966 to December of 1996). Then we see the headings Set, Items, and Description--the results which appear after these will tell us the numbers of the sets we create, how many documents are in them, and what they represent. Finally, we get a question mark, which is DIALOG's system **prompt**. This tells us that the system is ready and waiting for us to give it a command.

That first command should always tell the system in which file you want to search. It is quite possible that *ERIC*, which covers education, will have documents of interest to our user (and indeed, it does), but our intention was to search in *PsycINFO*, as it is likely that documents in psychology will be more helpful. To move to *PsycINFO*, file 11 in DIALOG, we use the BEGIN command, used to change files. The form of the command is

BEGIN file-number

where "file-number" is the number of the file to be entered. If you're in the middle of a search and give the BEGIN the command, it will clear out all your previously created sets and move you to the file you requested. It's also good to use if you make some horrible mistakes and just want to start over; give the BEGIN command for the same file and you have a blank slate.

Many of the most common DIALOG commands can be abbreviated, so we could also just say

B file-number

which is what we do here:

?b 11

```

22jan97 14:53:06 User007659 Session B1273.1
      $0.00      0.000 Hrs File1
$0.00 Estimated cost File1
$0.00 Estimated cost this search
$0.00 Estimated total session cost      0.000 Hrs.

```

File 11:PsycINFO(R) 1967-1997/Feb
 (c) 1997 Amer. Psychological Asso.
 *File 11: "KWIC format pricing will change effective 1/1/97.
 See HELP RATES 011 to see new prices."

Set	Items	Description
---	-----	-----
?		

Now we're in *PsycINFO*, we can proceed with our search by starting to create sets based on terms we selected.

Choosing Search Terms: **SELECT**

The command use to search for a given term is SELECT. Its form is

SELECT what-to-search-for

and is often abbreviated as

S what-to-search-for

When the SELECT command is issued, the system searches through the inverted file (DIALOG calls this the **Basic Index**) for all the documents that contain that term. You can use SELECT to search for individual words

?s children

or phrases

?s traumatic brain injury

but those phrases will only work in phrase-indexed fields. Recall our discussion about word- and phrase-indexing in the section on inverted files. You could find a descriptor TRAUMATIC BRAIN INJURY this way, but if the phrase appears in a title or abstract, we won't get it like this. We can get it, but we have not learned how to do that quite yet.

There is another form of the SELECT command that is useful if you are searching for several terms at once. It is known as SELECT STEPS and abbreviated SS. The format is

SELECT STEPS what-to-search-for

and its abbreviated form is

SS what-to-search-for

When SELECT STEPS is used, it performs the same operations as SELECT. The terms requested are searched for in the inverted file, and a set is produced. The difference between this and SELECT is that when SELECT STEPS is used, sets are created for the individual terms in the statement as well as for the overall statement. These intermediate sets can be useful if, as the search continues, the searcher decides to use these terms in other combinations. The sets have already been created, so you can save typing and time (and money!) by using them.

SS can be used anywhere S can be used, but it not necessary when searching for single terms or phrases. You'll also notice that you create a lot more sets this way, and many beginning searchers find that a bit difficult to contend with, especially when it comes to deciding which sets to use later in a search. Some people like it, others don't, and some people wind up using it after they gain some experience. Try

Search Technique

September 24, 2001 2:15 PM

it a couple of times, so you know how to use it, and then add it to your repertoire when you feel comfortable with it.

You can use AND, OR, and NOT inside SELECT statements in several ways. As we've seen, two or more terms can be joined by a Boolean operator, as in

?s children or pediatric or adolescents

Set numbers can be used in place of terms, as in

?s s3 or altruism
?s s9 and s12

Also, more than two components may be searched for, as in

?s s3 and s8 and s17

A few cautionary words about SELECT: Always be sure to put a space after the word SELECT or S or SS. The system expects it, and if you omit it, you may be unexpected results, such as

?ssystem

DIALOG sees the SS and searches for the remainder of the statement (YSTEM) and gets zero hits.

You should also be careful about spelling and typing errors. Computers are very literal beasts, and a mistake such as this will, more than likely, get you no hits:

?s infomation

Beginning searchers often either fail to notice such an error or find it difficult to recover from it. If you do this (and we all do it), just do it again and proceed as you had planned.

One of the golden rules of online searching is stated thus: Always be suspicious when you get a set with zero postings. Have you spelled the term or terms correctly? Have you entered the command correctly? Notice the computer responds by repeating your requested term or terms so see that it received what you intended. When using set numbers, check that the postings look consistent with what you had previously seen. -GW

Another common error among new searchers occurs during searches involving previously created sets. Sometimes, instead of the statement:

?s s5 and france

the searcher enters

?s 5 and france

or even (more commonly)

?s5 and france

In this instance, it is not set S5 that is ANDed with FRANCE, it is the numeral 5, and only documents containing that number somewhere in the indexed fields will be retrieved. Again, if this should happen, simply reenter the statement. These errors are especially pernicious and difficult to catch, because they create sets that may look right. However, if an error like this is carried through an entire search, it will probably result in an almost useless set of citations. You might

be able to spot this later if you get a result set which seems to be fine but missing one concept altogether; it's usually far easier, though, to catch this as it happens and correct it then.

Truncation: ?

One last piece of DIALOG mechanics before we get into the actual search. Let's say we were doing a search on the effects of technology on libraries, librarians, and the profession of librarianship. We could, of course, search for this second concept by doing

?s libraries or librarians or librarianship

or something like that. But look at those three words--they're very similar and in fact differ only in the ways they end. Wouldn't it be nice if there were a shortcut way to search for each of those (and perhaps other variants) without having to think of them all and ORing them together? Conveniently enough, we can, using **truncation**, and the truncation operator, which in DIALOG is the question mark:

?s librar?

This statement will retrieve any document which contains any word which begins with the letters LIBRAR; the words listed above, but also LIBRARIES, and so on. There will be only one set, and the system won't tell you what the precise words were you searched on. You must take care, then, that you don't truncate too far to the left or inadvertently include some word with an enormous number of postings. Go as far to the left as you need, but the farther you go, the more stuff you'll get, and the greater the potential for irrelevant terms.

I vividly recall a search I was doing in a large full-text database of newspaper articles, looking for things about fast-food companies trying to reduce the fat content of their foods. Everything was fine until I was trying to do the FAT part, and I blithely searched on

?s fat?

and as soon as I'd done it, I knew I had blown it. I got FATE and FATHER and who knows what all else. Took forever, and cost a fortune. Learn from my terrible example. -JWJ

That particular error could probably have been avoided by using a variation of the truncation operator. The general one, as illustrated above, will get all words which begin with the specified characters, regardless of how much comes after. You can control that by using, for example

?s statistic? ?

which will retrieve STATISTIC itself and any word which begins with STATISTIC and has one additional character. Thus, it will retrieve STATISTIC and STATISTICS, but not STATISTICAL, STATISTICALLY or STATISTICIAN.

If more than one extra character is desired but still a limited number, use as many question marks as characters. Thus,

?s retriev??

will get RETRIEVE, RETRIEVAL, RETRIEVED, and RETRIEVER, each of which has two or fewer characters after the stem, but not RETRIEVING, which has three.

The question mark may also be useful *inside* a word, to retrieve variant spellings. Perhaps the most common example of this use of truncation is

?s wom?n

which will retrieve WOMAN and WOMEN as well as the less frequent WOMYN. Such a use of truncation would not work for most British or Canadian spelling variations, seen in such words as COLOUR and HONOUR; they must be searched using OR, since embedded truncation only allows for a single letter:

?s behavior or behaviour?

So now let's begin our search and see what we get. Our most specific concept is probably the head injury terms, and we had two good ones (or at least they looked good) from the user, so let's see what they produce:

?s closed head injury or traumatic brain injury

```

                0  CLOSED HEAD INJURY
                8  TRAUMATIC BRAIN INJURY
S1             8  CLOSED HEAD INJURY OR TRAUMATIC BRAIN INJURY

```

OK, so maybe they weren't that good. It would appear, since everything is spelled right, that CLOSED HEAD INJURY isn't a subject heading in this database, and that TRAUMATIC BRAIN INJURY either is very infrequently used or is a new term in this database and thus hasn't had many documents indexed with it. In either event, this 8 document set certainly is far too narrow, so we'll have to fall back on alternative terms right away. Let's try this:

?s (brain or head) and (trauma? or injur?)

```

                48712  BRAIN
                7682  HEAD
                8637  TRAUMA?
                9632  INJUR?
S2             4750  (BRAIN OR HEAD) AND (TRAUMA? OR INJUR?)

```

Well, that's better, but the documents we get will be far less specific on "head injuries" than we'd hoped. What we'll get is any document with the word BRAIN or HEAD which also has some form of TRAUMA or INJURY, but these words don't necessarily have to be together or even connected in the document. Still, it's better than S2. Our other terms work somewhat better:

?s psychosocial or behavioral

```

                19903  PSYCHOSOCIAL
                61970  BEHAVIORAL
S3             80360  PSYCHOSOCIAL OR BEHAVIORAL

```

?s children or pediatric or adolescents

```

                147452  CHILDREN
                2272  PEDIATRIC
                45425  ADOLESCENTS
S4             176555  CHILDREN OR PEDIATRIC OR ADOLESCENTS

```

Note that in S4, the number of documents with the word CHILDREN is dramatically more than the others. Does CHILDREN dominate that set? Perhaps, but the terms are otherwise good, and the user is interested in both. It probably means that the result set will have more in it about children than adolescents, but that may just be the way it is.

So now let's connect them with an AND and see how big our result set will be:

```
?s s2 and s3 and s4

      4750 S2
      80360 S3
      176555 S4
S5      119 S2 AND S3 AND S4
```

7 Have a look at a few documents.

Well, we know that this original strategy (somewhat modified) produced 119 documents--documents which have one or more of the terms from each concept set. But now we really need to find out whether those documents are any good or not, and we can do that by having a look at a few.

Seeing What You've Got: TYPE, Output Formats

The DIALOG command used to display records from a set is TYPE, which can be abbreviated as T. The format of the TYPE command is

TYPE set-number/format/records-to-see

T set-number/format/records-to-see

So this command

```
?t 5/8/1-9
```

will show us the first through the ninth documents of set 5, in format 8. In DIALOG, you'll usually find that documents come out in reverse chronological order (actually, reverse of the order into which they went into the database), so asking for the first few will get you the most recent, newest ones. Here's an example of format 8:

```
5/8/1
DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

01062683      84-04014
Naturalistic assessment leading to effective treatment of self-injury in
a young boy with multiple disabilities.
Major Descriptors: *HEAD BANGING; *MULTIPLY DISABLED; *OBSERVATION
METHODS; *SELF INFLICTED WOUNDS; *TREATMENT
Minor Descriptors: CASE REPORT; CHILDHOOD; SCHOOL AGE CHILDREN
Descriptor Codes: 22350; 32509; 34830; 46290; 54190; 07790; 08750
; 45540
Identifiers: naturalistic assessment & treatment of self injurious head
banging, 11 yr old male with multiple disabilities, case report
Section Headings: 3200 -PSYCHOLOGICAL & PHYSICAL DISORDERS; 3300 -
HEALTH & MENTAL HEALTH TREATMENT & PREVENTION
```

This is sometimes called the "searcher's format", since it gives you quite a bit of information which can be useful as the search goes on. It includes not only the title but also the subject headings (here descriptors and identifiers). This will help you to know whether or not the documents are likely to be of value to the user but also can be sources of good potential terms to use as the search develops.

There are many other potential formats to use in TYPEing out documents, and you can find them all in the DIALOG Bluesheets, either in print or on the Internet at www.dialog.com. A few, though, are worth mentioning here. Format 6 is often useful for getting a quick sense of what's in a set, since it gives only titles of documents. Format 2 is usually bibliographic citation, which may be what the user most wants (although it's always good to ask). Formats 5 and 9 are called "full format"; they will display all the information in the record. You should be a

Search Technique

September 24, 2001 2:15 PM

bit careful with these, since the records could be quite long and using these formats usually carries an additional charge. More databases are including full text records, so being able to get the full record in this way is a great time-saver, but be sure you want to do it!

At this stage you're looking both to evaluate the quality of the set and think of ways to modify the search to improve it. Look at the rest of this set and do both: see what you think **and** try to find some new good terms to try in the search.

5/8/2

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

01060804 84-02086
 Vocational outcome of aphasic patients following severe traumatic brain injury.
 Major Descriptors: *APHASIA; *TRAUMATIC BRAIN INJURY; *VOCATIONAL REHABILITATION
 Minor Descriptors: ADOLESCENCE; ADULTHOOD; CHILDHOOD; PRESCHOOL AGE CHILDREN; SCHOOL AGE CHILDREN
 Descriptor Codes: 03410; 54115; 56210; 00920; 01150; 08750; 40160; 45540
 Identifiers: incidence & rehabilitation course of aphasia syndromes, vocational outcome, severe traumatic brain injured 3-63 yr olds
 Section Headings: 3384 -OCCUPATIONAL & VOCATIONAL REHABILITATION

5/8/3

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

01060419 84-01575
 Local cerebral glucose metabolism in patients with long-term behavioral and cognitive deficits following mild traumatic brain injury.
 Major Descriptors: *BEHAVIOR PROBLEMS; *BRAIN DAMAGED; *COGNITIVE ABILITY; *GLUCOSE METABOLISM
 Minor Descriptors: ADOLESCENCE; ADULTHOOD; CHILDHOOD; FOLLOWUP STUDIES; SCHOOL AGE CHILDREN
 Descriptor Codes: 05650; 06790; 10050; 21165; 00920; 01150; 08750; 20040; 45540
 Identifiers: abnormal local cerebral glucose metabolic rates, traumatically brain injured 12-59 yr olds with long term behavioral & cognitive deficits, 1-5 yr followup
 Section Headings: 3297 -NEUROLOGICAL DISORDERS & BRAIN DAMAGE

5/8/4

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

01051825 83-31228
 Appraising and managing knowledge: Metacognitive skills after childhood head injury.
 Major Descriptors: *AGE DIFFERENCES; *HEAD INJURIES; *METACOGNITION; *NEUROPSYCHOLOGICAL ASSESSMENT
 Minor Descriptors: ADOLESCENCE; ADULTHOOD; CHILDHOOD
 Descriptor Codes: 01360; 22360; 31040; 33835; 00920; 01150; 08750
 Identifiers: age, metacognitive knowledge appraisal & management, 6-19 yr olds with closed head injury
 Section Headings: 3297 -NEUROLOGICAL DISORDERS & BRAIN DAMAGE

5/8/5

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

01039257 83-18123
 Diagnosis and treatment strategies for the latent sequelae of head trauma in children.
 Major Descriptors: *HEAD INJURIES; *MOTOR TRAFFIC ACCIDENTS; *DIAGNOSIS; *DRUG THERAPY; *ANTICONVULSIVE DRUGS
 Minor Descriptors: CASE REPORT; CHILDHOOD; ADOLESCENCE; FRONTAL LOBE; SCHOOL AGE CHILDREN; NEURAL DEVELOPMENT; CONVULSIONS; COGNITIVE

ABILITY

Descriptor Codes: 22360; 32340; 13990; 15380; 03110; 07790; 08750
 ; 00920; 20440; 45540; 33605; 11750; 10050
 Identifiers: diagnosis & anticonvulsant treatment of latent head trauma
 sequelae & executive dysfunction, 9 yr old male struck by automobile,
 case report, implications for frontal lobe maturation
 Section Headings: 3363 -MEDICAL TREATMENT OF PHYSICAL ILLNESS

5/8/6

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

01037116

83-15284

Behavioural training during acute brain trauma rehabilitation: An
 empirical case study.

Major Descriptors: *OPERANT CONDITIONING; *BRAIN DAMAGED;
 *OCCUPATIONAL THERAPY; *PHYSICAL THERAPY; *BEHAVIOR PROBLEMS
 Minor Descriptors: CHILDHOOD; CASE REPORT; POSITIVE REINFORCEMENT;
 SCHOOL AGE CHILDREN
 Descriptor Codes: 35380; 06790; 35100; 38590; 05650; 08750; 07790
 ; 39620; 45540
 Identifiers: operant conditioning during physical & occupational
 therapy, disruptive behaviors, 8 yr old female recovering from brain
 trauma, case report
 Section Headings: 3361 -BEHAVIORAL & PSYCHOLOGICAL TREATMENT OF PHYSICAL
 ILLNESS

5/8/7

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

01027149

83-02002

Pediatric traumatic brain injury: Promoting successful school reentry.

Major Descriptors: *BRAIN DAMAGED; *HEAD INJURIES; *SCHOOL ADJUSTMENT;
 *SCHOOL ENVIRONMENT; *EVALUATION
 Minor Descriptors: EDUCATIONAL MEASUREMENT; SCHOOL PSYCHOLOGY; REENTRY
 STUDENTS
 Descriptor Codes: 06790; 22360; 45510; 45610; 18260; 16140; 45740
 ; 43495
 Identifiers: physical & cognitive & behavioral characteristics & school
 reentry & adjustment & assessment & intervention, children with
 traumatic brain injury
 Section Headings: 3297 -NEUROLOGICAL DISORDERS & BRAIN DAMAGE; 3580 -
 EDUCATIONAL/VOCATIONAL COUNSELING & STUDENT SERVICES

5/8/8

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

00995549

82-23228

The Balkanization of special education: Proliferation of categories for
 "new" behavioral disorders. 17th Annual Conference of Teacher Educators for
 Children with Behavior Disorders (1993, Tempe, Arizona).

Major Descriptors: *EDUCATIONAL DIAGNOSIS; *TAXONOMIES; *SPECIAL
 EDUCATION
 Minor Descriptors: PROFESSIONAL MEETINGS AND SYMPOSIA; ADULTHOOD
 Descriptor Codes: 16075; 51550; 48930; 40740; 01150
 Identifiers: proliferation of separate eligibility categories for
 behavioral disorders & impact on special education, conference
 presentation
 Section Headings: 3570 -SPECIAL & REMEDIAL EDUCATION

5/8/9

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

00995058

82-22647

A comparison of two psychosocial interventions for parents of children
 with acquired brain injury: An exploratory study.

Major Descriptors: *SUPPORT GROUPS; *BRAIN DAMAGED; *STRESS MANAGEMENT
 ; *PARENT TRAINING
 Minor Descriptors: CHILDREN; ADULTHOOD; PARENTS

Search Technique

September 24, 2001 2:15 PM

Descriptor Codes: 50740; 06790; 50175; 36606; 08830; 01150; 36680
 Identifiers: stress management vs information & sharing support groups,
 parents of children with acquired brain injury
 Section Headings: 3361 -BEHAVIORAL & PSYCHOLOGICAL TREATMENT OF PHYSICAL
 ILLNESS

To be honest, they don't look that great. They're close--at least a few of them are in the general area of head injuries to children--but they don't really seem to focus on the areas the user mentioned in the search request. We do see the term TRAUMATIC BRAIN INJURY in document 2, which reinforces the notion that it's a brand new term which we won't find in older documents. We also see NEUROPSYCHOLOGICAL ASSESSMENT in document 4, a form of a term the user gave us, and HEAD INJURIES in document 5, which is apparently the preferred term in this database. Other than that, there's not much else here.

8 Revise and refine the search based on those initial results.

Now you get to play around a bit and try to improve your first tries.

If your first result set has very **few** documents, fewer than you expected or wanted, you probably want to get more. Think of what you know now which will produce more documents: You might want to try some of your alternative terms and see if they produce new good stuff. You may decide to truncate a bit further to the left. You might even think about dropping a concept set (going from 3 concepts to 2, for example), eliminating the *least* specific one first. Also check for errors in spelling or technique. You might use conceptually broader terms (as we did in going from S1 to S2)

If your first result set has **too many**, though, you should think about what you know which will produce fewer documents: use fewer or narrower terms, truncate further to the right, add a concept (but only if you've got a good one to add), or NOT something out (but only if you are pretty certain you won't lose good stuff this way).

If you have the **wrong** stuff, you may have made a technique error (using a digit instead of a set number), or you may just have picked poor terms. In the real world, there's no sin in doing some initial searching to see what's in there, and then logging off to re-evaluate, find some new terms, talk with the user, and get back in and try again.

We seem to be close, but not quite there. Let's try a couple of terms we spotted from the initial set. This is a tactic called **pearl growing**, and it's a very useful and efficient way of getting new search terms, especially controlled vocabulary terms which the user might not know about.

?s head injuries

S6 1817 HEAD INJURIES
 ?s neuropsycholog?

S7 9645 NEUROPSYCHOLOG?
 ?s s7 and s2 and s4

9645 S7
 4750 S2
 176555 S4
 S8 109 S7 AND S2 AND S4

We are now using our new terms for head injury and "effects" with our old children/adolescent set, and get 109 documents. We could look at this set right away, but we may well be getting a number of the same documents we've already seen. We can avoid this by using NOT in a "non-

Search Technique

September 24, 2001 2:15 PM

conceptual" way--if we NOT out the previous result set, then we'll get just new documents, which we didn't have in the previous set.

?s s8 not s5

```

          109 S8
          119 S5
    S9      86 S8 NOT S5

```

and we see that there are 86 of these. Let's have a look.

?t 9/8/1-7

9/8/1

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

```

01056943      33-75784
  Assessment of the brain injured child in the public schools: A comparison
of the Differential Ability Scales and the Wechsler Scales.
  Major Descriptors:  *NEUROPSYCHOLOGICAL ASSESSMENT;  *DIFFERENTIAL
  APTITUDE TESTS;  *WECHSLER ADULT INTELLIGENCE SCALE;  *BRAIN DAMAGE;
  *WECHSLER INTELLIGENCE SCALE CHILDREN
  Minor Descriptors:  KINDERGARTEN STUDENTS;  ADOLESCENCE;  CHILDHOOD;
  HIGH SCHOOL STUDENTS;  JUNIOR HIGH SCHOOL STUDENTS;  ELEMENTARY SCHOOL
  STUDENTS;  PUBLIC SCHOOL EDUCATION
  Descriptor Codes:  33835;  14150;  56530;  06780;  56550;  27370;  00920
  ;  08750;  22930;  27220;  16630;  42230
  Identifiers:  Differential Ability Scales vs Wechsler Scales, assessment
  of traumatic brain injury, 5-17 yr old public school students
  Section Headings:  2225 -NEUROPSYCHOLOGICAL ASSESSMENT;  3297 -
  NEUROLOGICAL DISORDERS & BRAIN DAMAGE

```

9/8/2

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

```

01050226      83-29259
  Dimensions of cognition measured by the Tower of London and other
cognitive tasks in head-injured children and adolescents.
  Major Descriptors:  *COGNITIVE PROCESSES;  *FACTOR STRUCTURE;  *HEAD
  INJURIES;  *NEUROPSYCHOLOGICAL ASSESSMENT;  *TEST RELIABILITY
  Minor Descriptors:  ADOLESCENCE;  CHILDHOOD;  MAGNETIC RESONANCE IMAGING
  Descriptor Codes:  10130;  19045;  22360;  33835;  52250;  00920;  08750
  ;  29133
  Identifiers:  reliability of Tower of London task & relation to MRI &
  various measures of executive functions, 6-16 yr olds with closed head
  injury
  Section Headings:  2225 -NEUROPSYCHOLOGICAL ASSESSMENT;  3297 -
  NEUROLOGICAL DISORDERS & BRAIN DAMAGE

```

9/8/3

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

```

01049041      83-28290
  Learning potential and other predictors of cognitive rehabilitation.
  Major Descriptors:  *AGE DIFFERENCES;  *NEUROPSYCHOLOGICAL ASSESSMENT;
  *LEARNING ABILITY;  *TREATMENT OUTCOMES;  *COGNITIVE REHABILITATION
  Minor Descriptors:  CHILDHOOD;  ADOLESCENCE;  ADULTHOOD;  SCHOOL AGE
  CHILDREN;  AGED;  BRAIN DAMAGED;  INCIDENTAL LEARNING
  Descriptor Codes:  01360;  33835;  27960;  54185;  10136;  08750;  00920
  ;  01150;  45540;  01370;  06790;  24700
  Identifiers:  age & neuropsychological tests of new & incidental learning
  ability, prediction of success of cognitive rehabilitation program,
  7-71 yr olds with brain injury
  Section Headings:  3380 -REHABILITATION

```

9/8/4

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

01039023 83-17768
 Editorial: Progress in pediatric neuropsychology.
 Major Descriptors: *PEDIATRICS; *CHILD PSYCHOLOGY; *NEUROPSYCHOLOGY;
 *METHODOLOGY; *SCIENTIFIC COMMUNICATION
 Descriptor Codes: 37120; 08720; 33840; 31140; 45830
 Identifiers: interface between pediatric psychology & neuropsychology &
 methodological challenges facing pediatric neuropsychologists
 Section Headings: 3297 -NEUROLOGICAL DISORDERS & BRAIN DAMAGE

9/8/5
 DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

01038990 83-17735
 Neuropsychological deficit and academic performance in children and
 adolescents following traumatic brain injury.
 Major Descriptors: *HEAD INJURIES; *NEUROPSYCHOLOGICAL ASSESSMENT;
 *ACADEMIC ACHIEVEMENT PREDICTION; *COGNITIVE ABILITY
 Minor Descriptors: CHILDHOOD; ADOLESCENCE; FOLLOWUP STUDIES; SCHOOL
 AGE CHILDREN
 Descriptor Codes: 22360; 33835; 00210; 10050; 08750; 00920; 20040
 ; 45540
 Identifiers: neuropsychological testing, prediction of academic outcome,
 9-15 yr olds with moderate to severe traumatic brain injury, 1 yr
 followup, Australia
 Section Headings: 3297 -NEUROLOGICAL DISORDERS & BRAIN DAMAGE

9/8/6
 DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

01037791 83-16170
 Validity of the Kaufman Brief Intelligence Test (K-BIT) in children with
 traumatic brain injury.
 Major Descriptors: *INTELLIGENCE MEASURES; *BRAIN DAMAGED; *CONCURRENT
 VALIDITY; *CONSTRUCT VALIDITY
 Minor Descriptors: SCHOOL AGE CHILDREN; CHILDHOOD
 Descriptor Codes: 25910; 06790; 11058; 11445; 45540; 08750
 Identifiers: concurrent & construct validity of Kaufman Brief
 Intelligence Test, 12 yr olds
 Section Headings: 2225 -NEUROPSYCHOLOGICAL ASSESSMENT; 3297 -
 NEUROLOGICAL DISORDERS & BRAIN DAMAGE

9/8/7
 DIALOG(R)File 11:(c) 1997 Amer. Psychological Asso. All rts. reserv.

01031473 83-07630
 Perils and pitfalls on the path to normal potential: The role of impaired
 attention: Homage to Herbert G. Birch.
 Major Descriptors: *POVERTY; *DEVELOPMENT; *ATTENTION
 Descriptor Codes: 39820; 13830; 04410
 Identifiers: poverty as pitfall for normal development, application of
 attention hypothesis
 Section Headings: 2800 -DEVELOPMENTAL PSYCHOLOGY

Still not all that hot, really. The fifth one looks ok but not great,
 and the rest are just all over the place. Let's see a few more, in
 format 6 (titles only):

?t 9/6/8-15

9/6/8
 01030979 83-06998
 Differential performances on the WRAML in children and adolescents
 diagnosed with epilepsy, head injury, and substance abuse.

9/6/9

01030265 83-06175
 Clinical neuropsychology and brain injury rehabilitation in Israel: A twenty-year perspective. International Neuropsychological Society Symposium: Neuropsychology in the 1990s (1993, Madeira, Portugal).

9/6/10
 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.

9/6/11
 00992772 82-19843
 Neuropsychological assessment and malingering: A critical review of past and present strategies.

9/6/12
 00981057 82-08053
 Revised scoring, reliability, and validity investigations of Piaget's Bicycle Drawing Test.

9/6/13
 00972937 32-85183
 Afasia talamica. Descrizione di un caso e revisione della letteratura. / Thalamic aphasia: Case report and review.

9/6/14
 00963345 81-38159
 Motor, visual-spatial, and somatosensory skills after closed head injury in children and adolescents: A study of change.

9/6/15
 00951381 81-27162
 Managing costs and outcome of patients with traumatic brain injury in an HMO setting.

It doesn't get much better. We're still really close, and you've got to think there's good stuff in there, but we just haven't found the right way to get at it yet. This can be pretty frustrating, but we're not done yet. First, let's review and see what we've done so far.

Seeing What You've Done So Far: DISPLAY SETS

Boy, is this helpful. After you start creating sets, it gets increasingly difficult to remember what exactly you've done, and especially difficult to remember what particular sets you've got. Just type **DS** (for **DISPLAY SETS**), and you'll get a listing of all the sets you've created so far.

?ds

Set	Items	Description
S1	8	CLOSED HEAD INJURY OR TRAUMATIC BRAIN INJURY
S2	4750	(BRAIN OR HEAD) AND (TRAUMA? OR INJUR?)
S3	80360	PSYCHOSOCIAL OR BEHAVIORAL
S4	176555	CHILDREN OR PEDIATRIC OR ADOLESCENTS
S5	119	S2 AND S3 AND S4
S6	1817	HEAD INJURIES
S7	9645	NEUROPSYCHOLOG?
S8	109	S7 AND S2 AND S4
S9	86	S8 NOT S5

Search Technique
 September 24, 2001 2:15 PM

It pays to try to determine what set numbers you expect before you get online, and then check the numbers as they come up on the screen. I find this very helpful. --GW

So what do we do now? We've used all the terms we got from the user, and there don't seem to be any new ones to pearl grow with. If we think a bit about what we've been getting, and what we haven't, we might get an idea. The documents we've seen so far have been fine in many respects: they're mostly about children and mostly about head injuries, but it's the other piece, the effects of those injuries, which doesn't seem to be quite right. And if we look back at the search, there's a word, BEHAVIORAL, which is pretty big but doesn't seem to be helping much. Maybe if we get rid of it, we might get some better-quality documents. It's worth a try, so let's reconstruct the "effects" concept set with the other two good terms:

?s psychosocial or neuropsycholog?

```

          19903 PSYCHOSOCIAL
           9645 NEUROPSYCHOLOG?
S10      29370 PSYCHOSOCIAL OR NEUROPSYCHOLOG?

```

and create a new result set

?s s6 and s4 and s10

```

          1817 S6
         176555 S4
          29370 S10
S11      72 S6 AND S4 AND S10

```

and have a look

?t 11/6/1-15

```

11/6/1
01051825      83-31228
  Appraising and managing knowledge: Metacognitive skills after childhood
head injury.

```

```

11/6/2
01050226      83-29259
  Dimensions of cognition measured by the Tower of London and other
cognitive tasks in head-injured children and adolescents.

```

```

11/6/3
01038990      83-17735
  Neuropsychological deficit and academic performance in children and
adolescents following traumatic brain injury.

```

```

11/6/4
01030979      83-06998
  Differential performances on the WRAML in children and adolescents
diagnosed with epilepsy, head injury, and substance abuse.

```

```

11/6/5
01027149      83-02002
  Pediatric traumatic brain injury: Promoting successful school reentry.

```

```

11/6/6
00998391      82-25937
  Clinical neurological trauma parameters as predictors for

```

Search Technique

September 24, 2001 2:15 PM

neuropsychological recovery and long-term outcome in paediatric closed head injury: A review of the literature.

11/6/7

00986554

82-14036

Cognitive and psychosocial outcome after head injury in children.

11/6/8

00972937

32-85183

Afasia talamica. Descrizione di un caso e revisione della letteratura. /
Thalamic aphasia: Case report and review.

11/6/9

00963345

81-38159

Motor, visual-spatial, and somatosensory skills after closed head injury
in children and adolescents: A study of change.

11/6/10

00946741

81-22194

Children's adaptive behavioural competence after head injury. Special
Issue: Issues in the neuropsychological rehabilitation of children with
brain dysfunction.

11/6/11

00946740

81-22193

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

11/6/12

00934739

81-10353

Predictors of outcome following severe head trauma: Follow-up data from
the Traumatic Coma Data Bank.

11/6/13

00930845

81-07205

Effects of intact versus non-intact families on adolescent head injury
rehabilitation.

11/6/14

00930205

81-06413

Multiple partial seizure-like symptoms following "minor" closed head
injury.

11/6/15

00924261

81-00115

The use of a rating scale of attentional behaviour.

Much better indeed. Many of these seem to be really close to what the
user wanted (documents 1, 3, 6, 7, 9, 10, 11, 12, and so on), and the
whole set just seems better overall. We have lost the whole
"behavioral" aspect, but obviously we just didn't have the right term or
combination of terms for that, so we might try other ideas later. But
we do have a good solid set of 72 documents which the user can evaluate
to see what she thinks.

Here's a review of the whole search:

?ds

Set	Items	Description
S1	8	CLOSED HEAD INJURY OR TRAUMATIC BRAIN INJURY

Search Technique

September 24, 2001 2:15 PM

```

S2      4750    (BRAIN OR HEAD) AND (TRAUMA? OR INJUR?)
S3      80360   PSYCHOSOCIAL OR BEHAVIORAL
S4      176555  CHILDREN OR PEDIATRIC OR ADOLESCENTS
S5      119     S2 AND S3 AND S4
S6      1817   HEAD INJURIES
S7      9645   NEUROPSYCHOLOG?
S8      109    S7 AND S2 AND S4
S9      86     S8 NOT S5
S10     29370  PSYCHOSOCIAL OR NEUROPSYCHOLOG?
S11     72     S6 AND S4 AND S10

```

Getting Out: LOGOFF

We conclude the search by logging out of the system. (In a real search, of course, we'd type out the whole result set for the user, in format 2 or perhaps 5). The command to get offline is **LOGOFF**, although lots of other words will also work (BYE, QUIT, EXIT, OFF, and so on).

?logoff

```

22jan97 15:08:27 User007659 Session B1273.2
      $3.99      0.266 Hrs File11
      $0.00 43 Type(s) in Format 6
      $0.00 16 Type(s) in Format 8
      $0.00 59 Types
$3.99 Estimated cost File11
$0.80 INTERNET
$4.79 Estimated cost this search
$4.79 Estimated total session cost 0.267 Hrs.
Logoff: level 42.12.05 B 15:08:27

```

When you get this message, you know that you're off the system and the search is completed. The sets you created are gone, and unless you saved the search (which we'll talk about later), you can't get them back again. There is a version of this command, though, called **LOGOFF HOLD**, which will allow you to get off and think about the search for a bit. It will save your sets for about a half hour; if you log back in with the same account number and password during that time, your sets should still be there.

The Internet

So far, we've talked only about how to do searching using large-scale, well-established commercial information retrieval systems such as DIALOG. There are other such systems (LEXIS, Dow Jones, and so on), and while they are all different, they bear substantial similarity to each other. So, if you're going to be using some other system, the commands will be somewhat different from those we've discussed here, but the concepts will be very much the same.

The Internet, though, is a rather different matter. In this section, we'll talk about searching using the Internet, focusing on the World Wide Web, and emphasizing similarities and differences to what we've talked about already in DIALOG. There is much more to the Internet than the Web: it supports electronic mail, discussion groups (listservs and Usenet), as well as other means of moving information around, but in this context, it makes the most sense to concentrate on the Web. To learn more about the Internet in general, there are a great many books and Web sites available for you to consult. Any bookstore will have plenty to choose from; find one that makes sense to you.

The **World Wide Web** (sometimes abbreviated as the WWW; we'll call it the Web) has been around since the early 1990s, and was developed by Tim Berners-Lee at CERN, a nuclear research lab in Switzerland. He thought

Search Technique

September 24, 2001 2:15 PM

it would be helpful to have a relatively easy way for people to make documents and information available in the distributed, networked environment of the Internet. Until that point, it was very difficult to "publish" on the Net. You could create an archive of files which could be accessed using the FTP file transfer protocol, but that was difficult and non-intuitive. You could build a menu-driven, text-only system called a gopher, but that was also limiting. Allowing people to create documents which could include images, text, and links to other documents (we call this **hypertext**) was the real breakthrough, and the Web has grown to global proportions in a few short years.

What he developed is what network people call a **protocol**, really a set of standards that define what is needed to make a document available on the Web so that it can be retrieved and displayed by other, remote machines. He called this protocol **HTTP**, for **HyperText Transport Protocol**, and it is this set of standards that forms the backbone which makes the Web work. Recall the HTML document we saw in the previous chapter. If your computer is connected to the Internet, and you write an HTML document and decide to make it available (this is called **servicing** the document, and a computer which does that is a **server**), then anybody else in the world who is similarly connected and has the right software can find it and display it (this computer is called a **client**, and so you'll hear people call this a "**client-server architecture**").

There are a few other important things to know about this environment. First of all, it is often referred to as **distributed** or **decentralized**-- this means that there isn't a single "Internet"; it is really the connections between thousands of individual networks in schools and universities and businesses all over the world. So there's no center, no central authority (other than the protocols we all agree to), and nobody to really "run" it. Thus, you can't stop anybody from making anything available (including potentially offensive material), and you can't force anybody to, for example, include indexing or subject headings.

Secondly, since the Web was developed in the 1990s, when lots of people had access to computers, it was taken for granted that any kinds of search mechanisms which would work in this environment would have to be easy to use and take little if any time to learn about. When DIALOG was started in the 1970s, computing was dominated by large, expensive mainframe computers, so not that many people would or could use them. Thus, DIALOG is a large, centralized system, and its command structure, while very sophisticated and permitting powerful searching, is intricate and difficult to learn.

Further, the HTML structure was never really intended to be a help in organizing and searching information, as was the structure of a bibliographic record. So, while we can take advantage of that structure in searching, as we saw in the previous chapter, it will be in different ways than in DIALOG: typically we'll be able to search based on what something is (an image, a link, an address) rather than what it means (an abstract, a subject heading, an author).

Also, DIALOG contracts with producers of commercial databases to make them available. This information is professionally produced, edited, organized, and indexed, and users can have a great deal of confidence in what they find there. The Internet isn't like that. There is quite a bit of very interesting and worthwhile material freely available on the Net, and the amount and quality of good stuff is increasingly rising. But there is an enormous amount of what might politely be called trivia, some things that are downright wrong, and it all sits there together. There is no "collection development" or "selection" on the Net--it just happens. Users and searchers, therefore, have to be much more vigilant

in reading and evaluating the results of searches to decide whether or not what they get is worth anything.

So the picture that emerges is one of a world where there a very large collection of "documents" (but certainly not as large as the entirety of DIALOG-accessible databases available on computers all over the world, searchable in full text but with somewhat cruder technique, and using systems which require no training and make few expectations on the users.

To further illustrate these points, let's go back through the eight steps we outlined above, commenting on how they might best work in the Web, and discussing a few specific search engines. Keep in mind that the network environment is very volatile: not only can documents and entire resources change without a moment's notice, but search systems can change and add new features and new ones can arise very quickly as well. It's entirely possible that much of the specific discussion of features and technique that follows will be radically different from what is available when you read this.

What is important, though, is the *concepts* we're looking at. If you understand how to think about searching and take advantage of the environment you're searching in, be it DIALOG or the Web or whatever comes next (and something *will* come next), you'll be in fine shape.

- 1 Read the query.**
- 1a Listen to the query.**
- 1b Understand the query.**
- 2 Identify the major concepts in the query.**

Well, there's certainly nothing wrong with these. Understanding exactly what you're looking for doesn't get any less important in the Internet environment. What might get *more* important is understanding exactly what you might be able to find there. Although the information available there is getting better, it's probably never going to be the same quality or comprehensiveness (at least not while things are still free) as you'll find in a commercial search system. The best way to know what's out there, as with any retrieval system or collection, is experience, so an investment in time in seeing what it's like will certainly pay off.

3 Identify potential terms to correspond to those concepts.

Term selection is still part of the game, but since there is very little of what we'd think of as indexing or subject description, and no consistency whatsoever, it'll be harder for you to find "standard" or "preferred" terms. (We'll talk more about the uses and benefits of controlled vocabulary in the next chapter.)

It is also worth mentioning that you have to take into account the style of writing involved in the documents you'll be searching. We'll say this again when we discuss full-text searching in DIALOG later-- newspaper files, for example, will be searched differently from more academic ones. The same thing applies here. You'll find scholarly papers and children's stories and everything in between all mixed together, along with many things in languages other than English. Again, you'll gain a better appreciation for this as you gain experience with the environment, but keep an eye out for writing style and especially word choice.

4 Select alternative (narrower, broader, or related) terms to use if the original strategy needs help.

Still a good idea as well, but you won't be able to use too many of these in most circumstances. This is because the environment in which you'll be searching here is significantly constrained, compared to what we've seen in DIALOG.

5 Determine logical (Boolean) relationships between terms.

Boolean searching is possible in most Web search engines, and it will work in much the same way as we've seen in DIALOG. You can use OR to search for one or more related terms, AND to require all terms to be present, and NOT to exclude terms, and most systems will allow you to use parentheses to affect in the order in which these are interpreted, as we also saw in DIALOG.

Internet Search Engine Technique

There are a couple of important variations, however. Some systems require you to use AND NOT rather than simply NOT. Strictly speaking, from a set-theory point of view, this is correct, but commercial systems typically don't make you do this. Be aware of it, though, because leaving the AND out of that expression means you'll be searching on the word "not" rather than using the NOT as a command.

Some systems also allow you to use the + and - signs in searching. Putting a + in front of a word or phrase requires it to be in documents (like AND); putting a - there excludes it (like NOT).

So a search such as this is possible in AltaVista:

+noir +film -"pinot noir"

which would retrieve documents with both the words "film" and "noir" but not the phrase "pinot noir".

This illustrates another technique--the use of quotes to define a phrase. Although there are almost never subject headings or descriptors in Net documents, the search engines are able to retrieve based on phrases anywhere in those documents. (We'll discuss how to do this in DIALOG later.) Therefore, a search such as

"stupid pet tricks"

in Infoseek will retrieve documents with those three words exactly in that order.

Capitalization is an issue on the Net. You will notice we never discussed it in DIALOG; that's because all characters are treated as capitals regardless of how they appear in original documents. That's not the case here. Typically, searches are conducted in lower case, but if you wish to search on a word or phrase which contains capitals, you may do so. Thus, searching on

Turkey

in AltaVista will retrieve precisely that--the word "Turkey", capitalized. Many such documents will likely be about the country Turkey, but some will be references to other kinds of turkeys where the word is somewhere capitalized, as in the first word of a sentence. It will not, though, retrieve documents where the word "turkey" appears but

is never capitalized. This might be useful in a number of situations, including of course searching on proper nouns.

There is one very large difference between searching here and in commercial systems which probably dramatically affects the way in which you search. So far, search engines in the Net world do not allow you to create and manipulate sets.

Now, of course, since we've said this, the day the book goes to press, some system will announce this as a new feature! - JWJ

Every search in the Net is a one-shot deal. You don't get to create separate concept sets and then combine them into a result set. This isn't as dramatic as it sounds; you can certainly create a sophisticated search in a single statement and then redo it as you review results, but it is just a different way of thinking about constructing and performing searches.

Given the nature of the Net and the information you'll find there, it's often best to do shorter and more specific searches anyway. You're more likely to find searches like

something AND ("America Online" or AOL)

here than the extended strategy we used for the head injury search earlier.

6 Begin the search.

In discussing searching in commercial services, our first piece of advice was to search on the most specific concept block first to give you some idea of how many things you're likely to get. That advice is also useful here. Since you don't get to create and manipulate sets and since many other pieces of technique might not be available to you, it makes sense to search on the most specific aspects of what you're looking for. Decide on what you think are the narrowest terms, the ones that will retrieve the fewest things and yet still be of interest.

Alternatives to Search Engines

There's an exception to this. If you're looking for a group of documents on a similar or related topic, it might be more productive to use a service like Yahoo! or the Argus Clearinghouse or the Internet Public Library.

If, say, you were trying to find documents about the TV show *ER*, you might think about using Yahoo, searching through its menus on Entertainment, Television, Shows and so on, down to the category for *ER* to find all the sites it knows about. Searching on "ER" would be difficult if not impossible--some search engines won't search on anything shorter than 3 or 4 letters.

Other sites which serve to collect and organized related information resources might be helpful in similar ways. The IPL (<http://www.ipl.org>) can point you to sites about philosophy, for example, in its Ready Reference Collection, selecting ones of high quality and useful content, describing each, giving author and publisher information, and collecting them to make them easier to find and access.

The guides to subject-oriented resources in the Argus Clearinghouse (<http://www.clearinghouse.net>) are an excellent way to know more about what is available in many topics, and act in many ways as pathfinders do

Search Technique

September 24, 2001 2:15 PM

in libraries. Each guide is evaluated on a series of criteria about the guide, including the resources in the guide, its design and organizational scheme, and evaluative techniques it uses.

Both of these do some of the things which libraries and librarians usually do (and in fact both are staffed by people with library education and background): find, evaluate, describe and organize information resources so they can more easily be found and used.

Ranking of Retrieved Documents

In DIALOG, we said that documents would be retrieved in reverse order of input to the database, so you get the most recent ones first. That's not the case here. Typically, Internet search engines will use some algorithm to rank the documents according to how closely they think they match your query. This sounds great, but it's not without problems, and it doesn't always work the way you'd necessarily want it to or think it should.

AltaVista will raise the score of a document (i.e., put it towards the top of the list) if the words you searched for are in the first few words of the document, if your query words are "close to one another" in the document, and if those words appear more than once. Infoseek uses very similar criteria, but rather than looking at proximity, it will score words higher if they are relatively rare in the database of all documents--in other words, uniqueness helps.

The use of these kinds of rating schemes has led to a fascinating phenomenon: trying to influence how search engines rank pages. For a while, people could simply add hundreds of occurrences of words and phrases to inflate their scores when those words are searched. Sometimes those words were relevant to the actual content, sometimes not. The search engine people got wise, and changed their procedures, and then people came up with new ideas. Amazing stuff, really--a cottage industry devoted to trying to "fix" information retrieval. Who knew our field could be so intriguing? -JWJ

Excite says that it is able to search by concept rather than simply by words, and can look "for ideas closely linked to the words in your query". In their early-1997 description of how to use their service, they say "Our search engine can figure out that relationships exist between words and concepts--that the term "elderly people" is related to "senior citizens". It learns about related concepts from the documents themselves, and learns more from each new document it indexes." When it presents results, it also gives you the opportunity to do a new search for "more like this" for each document to find what it thinks are similar documents.

In no case do you get any further information than this on what actually goes on. Ranking of documents in relation to queries is an old idea from information retrieval research, but has only recently been implemented in commercial systems, since it requires yet more overhead and results have been less than perfect.

It is perhaps worth noting that DIALOG-type systems are binary--records either match the search strategy, or they do not. The Web, on the other hand, is a partial match system. The search retrieves anything which matches your search statement or any part of your search statement, so postings are large. And don't expect that the document you consider to be most relevant will necessarily be output first, or even towards the top! --GW

The increased investment and attention that the Internet has brought to the world of information retrieval means that ideas such as these will

Search Technique

September 24, 2001 2:15 PM

probably appear with some regularity. This could well be a major boon to the search for information, and these systems do work reasonably well at present. It will, however, more than likely take some time before these systems can do the kind of reasoning and interpretation we take for granted in people.

Truncation

There are a couple more mundane details about searching here to be discussed. The first is truncation--most systems permit it, but in different ways. AltaVista uses the * as the truncation operator, either at the end of words or in the middle, and in all cases it will match an arbitrary number of characters. So here you can search on

col*r

to get both "color" and "colour", (but also "collector" and "collider"; perhaps

colo*r

would be better) and

antiq*

will get "antique", "antiques", "antiquities", "antiquated" and so on.

Lycos, on the other hand, will automatically interpret words you give it as to be truncated, unless you tell it otherwise. So a search on

match

here will get "match", "matches", "matching" and so on. To get only "match", you must end the word with a period:

match.

to stop truncation.

Searching Using Structure

There is structure in HTML documents, as we saw in the last chapter, and some systems allow you to take advantage of it in searching. Since the kind of structure here is different than we find in bibliographic records (remember it's used here to describe the internal components of a document, not the fields which describe a document), the searching will also be different, but it can also be of great help.

AltaVista permits searching on a number of these parts of documents, including the title tag, image tags, links to other documents, and the URL address of a page. Therefore, searches like these are possible:

title:"ESPNET" and "Steffi Graf"

to get pages with ESPNET in the title and the phrase "Steffi Graf" anywhere

image:cow

will get things with "cow" in an image tag, not necessarily images of cows!

link:albany.edu

Search Technique

September 24, 2001 2:15 PM

will get pages with at least one link to any Web site at the University at Albany

echinacea and url:*.org

will get documents containing the word "echinacea" which come from not-for-profit organizations

7 Have a look at a few documents.

You typically have less control over what you see when you review the results of your searches. Rather than seeing simple counts of number of hits (although AltaVista permits this), you will see a list of the first 10 or so documents. You may be able to ask for more at a time, and may be able to indicate how much information to see about each: the document's title (from the title tag, which many documents don't use), the URL, and perhaps its size, when it was last visited by the search engine, and a line or two from the document.

In many cases, it will be possible for you to make initial decisions about which of the retrieved documents will be of interest, but you almost always will have to inspect them more closely, and of course that's easy to do by simply clicking on the link.

8 Revise and refine the search based on those initial results.

Many search engines provide, along with the list of retrievals, an active window showing the search you entered. This makes it easier to make changes to that search or simply clear out the window and try again, without having to go back to the main screen. This is rather different from the process we suggested for DIALOG searching, but is a convenience for the searcher.

As we mentioned previously, this is a fluid and rapidly changing area. The best way for you to know what kinds of searching are possible and the techniques you can use is to look at the documentation the service provides: help pages, sample searches and so on. Knowing what you want to try to do and experience in searching in all kinds of environments will assist you greatly in understanding what's available and how to use it most effectively.