8 SEARCHING USING FREE TEXT

In chapter 7, we saw that searching using controlled vocabulary techniques can be effective and efficient. In many cases, it is preferable to search using controlled vocabulary terms because we often get higher-quality, more specific results. Sometimes, though, controlled vocabulary searching is either not good enough or not even possible. In these situations, we must rely on another set of techniques described in this chapter: searching free text.

Free-text searching is sometimes called keyword searching and is used in many online library catalogs and other computerized information retrieval systems. Essentially, searching with free-text techniques involves using terms from everyday or specialized language rather than controlled vocabulary terms. Take a peek at the search example at the end of this chapter now, just to get a feel for what we are describing and see how different it is from what we have done so far. It will look familiar because it is another version of the brain injury search, but now taking advantage of all the new techniques we will learn here.

In this chapter, we will explore online searching techniques used with free text, the situations in which this searching works best, where free-text terms come from, some problems with free-text searching, and how to refine searches to broaden or narrow results. A search example and a discussion of Internet free-text searching conclude the chapter.

Proximity Operators

Recall once again how the inverted file is constructed. Each word (excluding stopwords) that occurs in a document is marked with its position, and then an alphabetical list of all such words is created. Index terms that are intact phrases, such as descriptors and identifiers, are often included both as individual words and as full phrases (we said these fields are both word-indexed and phrase-indexed). We do not know yet how to search for phrases in other word-indexed fields, such as titles or abstracts.

We can, at this point, search for such phrases, but in a very crude way, using single words joined by the AND operator. If we were looking for documents about the Graduate Record Examinations, for example, we could simply search GRADUATE AND RECORD AND EXAM? and see what comes up. But there would be no guarantee that documents we retrieved would have anything to do with the Graduate Record Examinations. We could get documents that talk about examinations of record-buying habits of graduate students. Or we might retrieve a paper about record keeping of graduate schools for foreign-language examinations. AND does not allow us to specify relationships between concepts; we can only say that terms occur in the same record.

There are techniques, however, that allow us to be more specific. These are called proximity operators, and they are used to specify how close two or more words should be in the documents retrieved. There are several proximity operators, but they all work in essentially the same way.

The simplest of these allows one to retrieve documents that have two or more words in direct proximity, that is, right next to each other. In this way one can search for a phrase in word-indexed fields. For example, to search for the phrase "information industry" in titles (a rather specific tactic, by the way), use the (W) proximity operator, as in the following example from ERIC:

```
      Set    Items    Description
      ---    -----    -----------
?s information(w)industry/ti

           13082    INFORMATION/TI
            2152    INDUSTRY/TI (PRODUCTIVE ENTERPRISES, ESPECIALLY MANUFACTU...)
      S1      12    INFORMATION(W)INDUSTRY/TI
```

This command tells DIALOG to retrieve all documents that have the word "information" directly followed by the word "industry" in the title field.

More than two words can be chained together, as in the following example:

```
?s management(w)information(w)systems

           41955    MANAGEMENT
          125433    INFORMATION
           53499    SYSTEMS
      S2    2173    MANAGEMENT(W)INFORMATION(W)SYSTEMS
```

This expression will retrieve documents with these three words in this order, in any Basic Index field.

The general form of the command is

```
s <term>(W)<term>
s <term>(W)<term>(W)<term>
... etc.
```

Compare the following three expressions, again in *ERIC*:

```
?s day care

      S3    4735    DAY CARE (CARE OF CHILDREN BY PERSONS OTHER THAN
THEIR...)

?s day(w)care/de

            7651    DAY/DE
           11423    CARE/DE
      S4    6841    DAY(W)CARE/DE

?s day()care

           22753    DAY
           27177    CARE
      S5    7907    DAY()CARE
```

The first of these searches uses the bound descriptor DAY CARE. The second searches for the word DAY followed by the word CARE in the descriptor field, and retrieves more than 2,000 more documents. Why? Because several descriptors incorporate those two words in that order, including ADULT DAY CARE, DAY CARE CENTERS, FAMILY DAY CARE, SCHOOL AGE DAY CARE, and so on. S4 includes all of these. Finally, S5 includes all documents that have these two words in this order in any field and retrieves more than 1,000 more records than S4. This example illustrates the power and some of the potential problems of free-text searching.

Also note that there is no W between the parentheses in the search expression for S5. That is not a mistake—in this situation *only*, the W can be left out. (W) and () work in exactly the same way, and the W can be either upper or lower case.

Suppose one wanted to search for documents about the University of Michigan. Using this same technique, one could search on

```
?s university(w)of(w)michigan
```

but would get no hits. Not because none exist, but because OF is a stopword in the DIALOG system, as we discussed in the chapter on database construction. Recall the process: When a document containing the phrase "University of Michigan" is processed to go into a DIALOG database, each of those three words is numbered with its position in the field. Then the stopwords are eliminated, but their positions are preserved. So we have to reconstruct those phrases by allowing for the presence of stopwords (or any other words, for that matter). We do this by extending the (w) operator, and allowing for space between the words we search on. The command is

s ≤term≥(nW)≤term≥

where n is any number one or greater.[1] To search for documents containing the phrase "University of Michigan" in the Basic Index, we would search the following:

```
?s university(1w)michigan
        66777     UNIVERSITY
         6678     MICHIGAN
  S6       890     UNIVERSITY(1W)MICHIGAN
```

which retrieves all documents with the word UNIVERSITY followed by the word MICHIGAN, with at most one word in between.

This tactic can also be used when two words close together but not necessarily next to each other are wanted, as in the following example:

```
?s (online or information)(2w)retrieval
         5646     ONLINE
       125433     INFORMATION
         7013     RETRIEVAL
  S7      5555     (ONLINE OR INFORMATION)(2W)RETRIEVAL
```

This set contains all documents with either ONLINE or INFORMATION followed by the word RETRIEVAL with zero, one, or two words in between. Documents that have the phrases "information retrieval," "online retrieval," "online bibliographic retrieval," "online systems for retrieval of information," and so on might be of interest, so the search is broadened a bit.

> In practice, a number higher than (3W) or (4W) tends to be counter-
> productive, as the farther apart terms get, the more we revert to the simple
> Boolean AND. – GW

Notice that we have a Boolean expression on the left side of the (2W). There can also be set numbers here, truncations, or anything legal, as in the following:

```
?s s7(w)system?

            5555     S7
          110345     SYSTEM?
     S8      662     S7(W)SYSTEM?
?t 8/5/1

 8/5/1
EJ546260                    IR535161
   Shape Measures for Content Based Image Retrieval: A Comparison.
   Mehtre, Babu M.; And Others
   Information Processing & Management; v33 n3 p319-37 May 1997
   ISSN: 0306-4573
   Language: English
   Document Type: JOURNAL ARTICLE (080); PROJECT DESCRIPTION (141)
   Journal Announcement: CIJNOV97
   Explores the evaluation of image and multimedia information-retrieval
systems, particularly the effectiveness of several shape measures for
content-based retrieval of similar images. Shape feature measures, or
vectors, are computed automatically and can either be used for retrieval or
stored in the database for future queries. (57 references) (Author/LRW)
   Descriptors: Comparative Analysis; Content Analysis; Databases;
*Evaluation Methods; Futures (of Society); *Information Retrieval;
*Information Systems; Mathematical Formulas; *Measurement Techniques;
*Multimedia Materials
   Identifiers: *Digital Imagery; Query Processing; *Shapes; Similarity
(Concept); Vector Methods
```

Can you see why this document was retrieved? English grammar being what it is, the words wanted may not always be in the same order. An author writing a document about relativity theory might use that phrase, but she might also use "theory of relativity," "theory of general relativity," "relativity theories," or "theory of special relativity." If one was look-ing for these documents, one might be tempted to use a command like the following:

```
?s relativity(w)theory

            384      RELATIVITY
          38845      THEORY
     S9      13      RELATIVITY(W)THEORY
?t 9/5/1
 9/5/1
EJ456369            RC509090
   Relativity, Relatedness and Reality.
   Deloria, Vine, Jr.
   Winds of Change; v7 n4 p34-40 Fall 1992
   ISSN: 0888-8612
   Language: English
   Document Type: JOURNAL ARTICLE (080); POSITION PAPER (120)
   Journal Announcement: CIJMAY93
```

Anticipated the modern physics relativity theory, American Indians gained information about the natural world through careful observation based on the principle that all things are related. American Indian students could radically transform scientific knowledge by grounding themselves in traditional knowledge about the world and working this understanding into the Western scientific format. (SV)

Descriptors: *American Indian Culture; American Indian Education; American Indians; Epistemology; Higher Education; *Holistic Approach; Science Education; Scientific Attitudes; *Scientific Methodology

Identifiers: *Knowledge Acquisition; World Views

One might also want documents that contain the other phrases, so perhaps one could try the following:

```
?s relativity(2n)theor?

          384      RELATIVITY
        63161      THEOR?
   S10     49      RELATIVITY(2N)THEOR?
```

Notice the postings have gone from 13 to 49.

```
?t 10/5/1
 10/5/1
EJ531451            SE556552
  The Utilization of Fiction When Teaching the Theory of Relativity.
  Hellstrand, Ake; Ott, Aadu
  Physics Education; v30 n5 p284-86 Sep 1995
  ISSN: 0031-9120
  Language: English
  Document Type: TEACHING GUIDE (052); JOURNAL ARTICLE (080)
  Journal Announcement: CIJFEB97
  Describes a way of teaching the theory of relativity with the help of a
novel. Aims to contribute to the formation of didactic theories by means of
an evaluation of alternative methods. (AIM)
  Descriptors: *Fiction; Physics; *Relativity; Science Instruction;
Secondary Education; Teaching Methods
  Identifiers: Theory of Relativity
```

This command uses the (N) proximity operator, which retrieves documents that have the two terms near each other (hence the "N") with the possibility of intervening words, in this case as many as two. The general form

$$\text{s} \leq \text{term} \geq (n\text{N}) \leq \text{term} \geq$$

is similar to that of (W), and the same guidelines apply about using set numbers, Boolean expressions, truncation, and so on. In many cases, (nN) is more useful than (nW) because of this tendency in English to invert phrases and insert words.

A couple of further examples illustrate the use of (nN): If a patron requests documents on hypothesis testing (a technique from statistics), the concept might be referred to in documents as "hypothesis testing," but "testing the hypothesis" or "a test of two null hypotheses" might also appear. So instead of HYPOTHESIS(W)TESTING, HYPOTHES?S(3N)TEST??? might be preferred to get many variant forms of the phrase.

Also, (nN) can be used to save typing. If seeking documents about public universities in Michigan, one could try a strategy such as UNIVERSIT?(1N)MICHIGAN, which would retrieve documents with "University of Michigan" and also "Eastern Michigan University," "Michigan Technological University," and even "Michigan State University." Of course, some things would be missed (notably Wayne State University), but these could be ORed in, and quite a bit of typing would still be saved.

> *The use of proximity operators is available on most of the major online systems, though their exact formats vary. On some systems, for example, the (W) operator is replaced by ADJ (adjacent to). So we may have a command in search mode, such as DAY ADJ CARE, that will be equivalent to DAY()CARE in DIALOG. Similar features are also available on the Internet, though formats vary in different search engines. – GW*

There are other, broader proximity operators in DIALOG. The (F) operator will seek documents that have two words in the same field (e.g., both in the title, both in the abstract). The field to be searched in can be specified or left unqualified. See the following examples:

$$\text{S } \leq\text{term}\geq\text{(F)}\leq\text{term}\geq$$
$$\text{S } \leq\text{term}\geq\text{(F)}\leq\text{term}\geq/\leq\text{field code}\geq$$

A search for documents about the use of CD-ROMs in school libraries might go something like the following:

```
?s cdrom? or cd()rom?
            14      CDROM?
          2029      CD
          4426      ROM?
          1431      CD(W)ROM?
    S11   1436      CDROM? OR CD()ROM?
?s school(w)(librar? or media)
        291966      SCHOOL
         41884      LIBRAR?
         34797      MEDIA
    S12   4874      SCHOOL(W)(LIBRAR? OR MEDIA)
?s s11(F)s12
          1436      S11
          4874      S12
    S13     77      S11(F)S12
?t 13/5/1-3

 13/5/1
DIALOG(R)File      1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ544898           PS526458
  Using a CD-ROM Encyclopedia: Interaction of Teachers, Middle School
Students, Library Media Specialists, and the Technology.
  Albaugh, Patti R.; And Others
  Research in Middle Level Education Quarterly; v20 n3 p43-55 Spr 1997
  ISSN: 1082-5541
  Language: English
```

Document Type: RESEARCH REPORT (143); JOURNAL ARTICLE (080)
Journal Announcement: CIJOCT97
  Observed sixth-grade students and their ways of gathering information for
a science report from Encarta 94, a CD-ROM encyclopedia. Developed
recommendations for collaboration between the classroom teacher and the
school library media specialist during the implementation of CD-ROM
technology for information gathering, as well as ways to manage a
CD-ROM-based project. (AA)
  Descriptors: Access to Information; Case Studies; Classroom Techniques;
*Computer Attitudes; Computer Uses in Education; Educational Media;
Educational Strategies; Educational Technology; Elementary School Students;
*Encyclopedias; Grade 6; Information Seeking; Intermediate Grades;
Librarian Teacher Cooperation; *Multimedia Materials; Naturalistic
Observation; *Optical Data Disks; *Student Attitudes; Teacher Attitudes

 13/5/2
DIALOG(R)File   1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ544781   IR535023
  Technology Use in North Carolina Public Schools: The School Library Media
Specialist Plays a Major Role.
  Truett, Carol
  North Carolina Libraries; v55 n1 p32-37 Spr 1997
  For related earlier study, see EJ 488 280. Journal availability: State
Library of North Carolina, 109 East Jones Street, Raleigh, NC 27601-1023.
  ISSN: 0029-2540
  Language: English
  Document Type: RESEARCH REPORT (143); JOURNAL ARTICLE (080)
  Journal Announcement: CIJOCT97
  This report on teachers and technology in North Carolina schools
continues an earlier report based on a survey of North Carolina schools
that focused on media specialists. Highlights include how teachers
incorporate CD-ROM and videodisk technologies, school library media
specialists as technology instructors, and teacher expectations of media
specialists. (LRW)
  Descriptors: *Computer Uses in Education; Curriculum Development;
*Educational Technology; Elementary Secondary Education; Learning Resources
Centers; Librarian Teacher Cooperation; Library Instruction; *Library Role;
*Media Specialists; Optical Data Disks; Public Schools; *School Libraries;
School Surveys; Staff Development; Teacher Attitudes; *Teacher Role; Use
Studies; Videodisks
  Identifiers: North Carolina; *Technology Utilization

 13/5/3
DIALOG(R)File   1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ515154   IR532273
  Government Publications: A Forgotten Treasure.
  Ekhaml, Leticia
  School Library Media Activities Monthly; v12 n4 p28-31 Dec 1995
  ISSN: 0889-9371
  Language: English
  Document Type: PROJECT DESCRIPTION (141); JOURNAL ARTICLE (080)
  Journal Announcement: CIJAPR96
  Presents information on government publications for school library media
specialists, notes problems, and identifies selection aids. Topics include:
the "Monthly Catalog" on CD-ROM, the Superintendent of Documents

```
Classification System, GPO bookstores and depository libraries, how to
purchase and promote the use of government publications, and a list of
current documents useful to teachers and students. (AEF)
   Descriptors: Classification; Depository Libraries; *Government
Publications; *Information Sources; *Library Collection Development;
*Library Material Selection; Media Specialists; Optical Data Disks; *School
Libraries
   Identifiers: Bookstores; Government Printing Office; Monthly Catalog of
US Government Publications; Superintendent of Documents Classification
```

The first two documents in S13 look promising, but the third is off the track because the strategy is so broad. However, we notice the good descriptor SCHOOL LIBRARIES, so we use the following instead:

```
?s school libraries
    S14    3479     SCHOOL LIBRARIES
?s s11 and s14
           1436     S11
           3479     S14
    S5      103     S11 AND S14
?t 15/6/1-4

 15/6/1
EJ544781            IR535023
  Technology Use in North Carolina Public Schools: The School Library Media
Specialist Plays a Major Role.

 15/6/2
EJ539702            IR534431
  Mediagraphy: Print and Nonprint Resources.

 15/6/3
EJ529708            IR533359
  Full-Text Magazine Indexes & More on CD-ROMs.

 15/6/4
EJ523175            IR532797
  What CD-ROMs Are Other Schools Using?
```

With this strategy, we get improved results. There are other proximity operators; we will see how to use some in later chapters.

Why Use Free-Text Searching?

As we have seen, controlled vocabulary searching is often an excellent way to go. But that is not always the case. The following examples illustrate several situations in which searching using terms selected from a thesaurus or controlled vocabulary is not the best method.

- *There is no thesaurus.* Obviously, if there is no controlled vocabulary, it cannot provide search items. Some databases have no thesaurus at all. Others use one, but it is unavailable because it is not published or one may not have access to it. If there is no indexing at all (often the case with newspaper databases, for example), all that are left are free-text techniques. If documents in the database have been indexed in some way, but in an unknown fashion, one can begin with free-text searching and

then use pearl-growing techniques to weave in controlled vocabulary terms. This is an often-used and successful approach.

- *There are no good controlled vocabulary terms.* In some cases, the vocabulary may simply not cover the subject area of the query very well. There may be several related, marginal terms, but no single good term.

- *The term is new.* It may also be the case that the subject area or terminology of the query is new, so no term or terms have yet been accepted into the thesaurus. In general, it will take some time, probably years rather than months, for a new term to become widely used and incorporated into the printed controlled vocabulary. Some disciplines and databases move faster than others, but in swiftly developing areas or for very new terms, controlled vocabulary may not work. Further, when the term is included in the thesaurus, older documents may not be reindexed, so they will be accessible only using free-text searching.

- *There is only one good term, and it is not an index term.* This is particularly the case when a good term is outside of a subject area. For example, if one was searching for documents about how to develop good search strategies in Web-based search engines, one might decide to search in ERIC, but ERIC, as of this writing, has no good descriptor or term for the concept "search engine." There are other possibilities—INFORMATION RETRIEVAL, ONLINE SEARCH, and ACCESS TO INFORMATION—but in this case, it might be better to search for this phrase as a free-text expression like the following:

```
?s search()engine? ?
            13234    SEARCH
             1023    ENGINE? ?
    S10        77    SEARCH()ENGINE? ?
?s s10 and search strategies
               77    S10
             2350    SEARCH STRATEGIES (COMPREHENSIVE PLANS FOR FINDING
                     INFORMATION ...)
    S11        20    S10 AND SEARCH STRATEGIES
?t 11/5/1
 11/5/1
EJ546200                 IR535073
   "Just the Answers, Please": Choosing a Web Search Service.
   Feldman, Susan
   Searcher; v5 n5 p44-50,52-57 May 1997
   Language: English
   Document Type: NON-CLASSROOM MATERIAL (055); JOURNAL ARTICLE (080);
PROJECT DESCRIPTION (141)
   Journal Announcement: CIJNOV97
   Presents guidelines for selecting World Wide Web search
   engines. Real-life questions were used to test six search engines.
   Queries sought company information, product reviews, medical
   information, foreign information, technical reports, and current
   events. Compares performance and features of AltaVista, Excite,
   HotBot, Infoseek, Lycos, and Open Text. (AEF)
   Descriptors: *Comparative Analysis; *Computer Software
   Evaluation; Computer Software Selection; Guidelines;
   *Information Retrieval; Information Services; *Online Searching;
   Search Strategies; *World Wide Web
   Identifiers: *Query Processing; *Search Engines; Web Sites
```

- *There are not many hits, either in the database or for controlled vocabulary.* Free-text strategies are inherently broader than controlled vocabulary and thus will generally retrieve more documents. If an initial controlled vocabulary search pulls up very few or no good hits, or if it is known that there will be little good material in the database, a free-text strategy might be used to widen the net and pull in more records. There is no guarantee that these will be relevant, but at least they will be a starting point. Always keep in mind that the file may not contain the right material, no matter how good the strategy.

- *A known item is being sought.* Known-item searches are a special case. If the document's title, or a portion of it, is known (to provide a bibliographic verification, for example), there is no need to do an elaborate controlled vocabulary search. Just do a reasonable free-text attempt. Be careful, though, not to over-specify—the user's memory may not be perfect. In looking for Marcia Bates's well-known article on "the perfect thirty-item search," one might search in Library & Information Science Abstracts for the following:

```
?s perfect()thirty()item()search/ti
            90      PERFECT
            62      THIRTY
           463      ITEM
          4890      SEARCH/TI
    S1       0      PERFECT()THIRTY()ITEM()SEARCH/TI
```

Nothing is retrieved, but then try a broader strategy like the following:

```
?s thirty()item
            62      THIRTY
           463      ITEM
    S2       1      THIRTY()ITEM
```

This approach retrieves one item, which is correct:[2]

```
?t 2/6/1

 2/6/1
168472      85-2684
  TITLE: The fallacy of the perfect thirty-item online search
?t 2/3/1

 2/3/1
168472      85-2684    Library and Information Science Abstracts (LISA)
  TITLE: The fallacy of the perfect thirty-item online search
  AUTHOR(S): Bates, Marcia J.
  JOURNAL: RQ
  SOURCE: 24 (1) Fall 84, 43-50. 11 refs
```

- *You do not want to deal with a new controlled vocabulary.* As we have seen, thesauri and vocabulary differ, sometimes widely, from database to database. If confronted with a search in an unfamiliar database, one might decide that it is not really worth learning an entirely new vocabulary for one search. If time is pressing, an initial free-text strategy might be employed to see what happens, perhaps pearl growing

from good documents as they come up. This can work, but clearly it is not a preferred method. Some people, though, are good at this sort of thing; if their style works in this setting, more power to them.

It should be pointed out early and often that neither technique—controlled vocabulary nor free-text searching—is superior to the other. In some cases, one will be preferable, but often they work in tandem: One begins with a free-text search and pearl grows using index terms from good documents; or one begins with a good index term or two, discovers a useful free-text expression, and uses it. The blending of these two techniques to produce high-quality searches is part of the real art of searching.

Choosing Free-Text Terms

Once the decision is made to use free-text techniques, how are terms generated? There may be no controlled vocabulary from which to draw. Or, a controlled vocabulary term or two may be employed, but in free-text fashion. This is not done frequently, though, so where do these terms come from?

The first and potentially best source of free-text terms is the user. He or she may have quite a good idea about the vocabulary of the subject area to be searched and will be able to provide helpful clues for further searching. This is particularly true with university faculty, researchers, and other specialists looking for new or new-to-them documents in their fields. In this situation, use the information they provide on search request forms and through interviews.

However, many users do not have that much background or experience in the areas of their topics and may not be reliable sources of terms. They are still certainly worth discussing, and if they know of any good documents, titles, or authors, these are often sources of good terms.

Further, one can also try pearl growing with free-text terms from good documents, in addition to index terms, if any. Look for additional good terms in abstracts and especially in titles. If a word or phrase is used in the title, it is often an indication that the document is really about that concept.

A balance must be struck between generality and specificity. If a very general, single-word free-text term is used, one may retrieve thousands of documents, only a few of which are of interest to the user. On the other hand, if the expression is too specific, one may retrieve very little, miss good things, or perhaps get nothing at all. This balancing act can be very tricky, but it gets easier with experience.

Problems with Free-Text Searching:
False Drops

Free-text searching is certainly not a panacea, although it can be a helpful complement to controlled vocabulary techniques. However, there are situations in which free-text searching can be problematic. The major difficulty we encounter in searching free-text is *false drops*: retrieved documents that are not germane to the topic. Because there is no control over the vocabulary in title and abstract fields when words are searched for in those fields, the author or abstracter may or may not be using them in the same way as the searcher.

The following are a few examples of common sources of false drops and a few pieces of advice to help avoid them:[3]

Problem: Reverse concepts. If one were doing a search on school libraries and used the expression LIBRAR? AND SCHOOL?, one would get not only "school library" material but also "library school" material, which is not what is wanted.

> Solution: Use proximity operators to more closely tie concepts together. A search on SCHOOL(W)LIBRAR? would not retrieve "library schools," but neither would it retrieve "libraries in schools" nor "libraries in elementary schools." One might be tempted to try SCHOOL?(2N)LIBRAR?, but that would bring the search back right where it started. In many cases, one can avoid reverse concepts by using (W), but in some instances one might have to go to controlled vocabulary.

Problem: Homographs/conflation. These are two terms for the same problem: two or more concepts that use the same word or words. Examples include words like CRACK (cocaine or seismic fault?), FIELD (part of a bibliographic record or a meadow?), and SDI (Strategic Defense Initiative or selective dissemination of information?). We say that these terms are conflated.

> Solution: Qualify or focus the expression. If crack cocaine is the preferred topic, try something like CRACK AND (DRUG? OR COCAINE). This will focus the results and eliminate seismic or other extraneous material. Alternatively, try CRACK/TI, qualifying the term down to the title field. Some nonrelevant records will still come up, but marginal or off-hand mentions of the word in the abstract field will be eliminated.

Problem: Excessive truncation. Truncation is a wonderful thing, but too much of any good thing is too much. It would not be a good idea, for example, to search on BOOK? in ABI/Inform (which will also get BOOKKEEPING), or to search on INTERN? in ERIC, which would get INTERN and INTERNS but also INTERNAL and INTERNATIONAL.

> Solution: Do not truncate too far to the left. Try to imagine all the variant forms that the word can take (e.g., COMPUTER, COMPUTING, COMPUTERS, COMPUTE, etc.) and either truncate further to the right (COMPUT? - still pretty bad) or restrict the length of your truncation (COMPUT??? will only get up to three more characters; COMPUTER? ? will only get up to one more) or do not truncate at all (COMPUTER OR COMPUTING OR COMPUTERS OR COMPUTE), depending on the database and the vocabulary.

Problem: Acronyms. Many acronyms (CBS, USA, NASA) are not a problem because they have essentially entered the language as words in their own right. Some, however, will conflate with other common words. Acronyms such as ADD (attention deficit disorder, from psychology), SAD (seasonal affective disorder), AIDS (acquired immune deficiency syndrome), and so on will retrieve many more documents than intended because they will also retrieve based on the words "add," "sad," and "aids."

> Solution: There are several things to try. Focus or qualify the set (ADD AND ATTENTION). Use the full expression (SEASONAL()AFFECTIVE()DISORDER?). Search in the descriptor field and see if the expression or acronym is an index term (ACQUIRED()IMMUN?/DE).

Problem: Negation. Suppose one is searching for documents about programs to teach older people how to use computers. A strategy like COMPUTER? AND (OLDER OR ELDERLY) might seem like a good start, but it retrieves documents that contain sentences like, "We had hoped to include elderly people in our program, but our funding wouldn't allow us to." Although many databases instruct their indexers and abstracters not to incorporate negative phrases, it is not a universal instruction, and the instructions are not necessarily followed.

Solution: Shy of completely reworking abstracts and the way they are written, there is not a lot that can be done about this. It is not a major problem, but it is particularly frustrating when it happens.

Good Places to
Use Free-Text Searching

There are several situations in which free-text techniques are especially useful—situations in which controlled vocabulary simply will not work or cannot be used. The following are a few examples:[4]

- *Geography.* Some databases have geographical descriptors (ABI/Inform, PAIS) or identifiers (ERIC), but many do not. If searching for documents that make mention of a particular geographic name, free-text techniques may be the only option: ANN(W)ARBOR, NEW(W)YORK. Keep in mind, though, that several geographic names may refer to the same area. For a search on New England, the following might be wise to do:

  (NEW()ENGLAND OR MAINE OR VERMONT OR NEW()HAMPSHIRE OR CONNECTICUT
  OR RHODE()ISLAND) OR MASSACHUSETTS)

  and even then you might miss records which refer to Boston or Providence or the White Mountains or the Berkshires.

- Other proper names. Again, some databases have personal name fields or include names as descriptors or identifiers, but if these are not available or if one is looking for other proper names, use free-text: GROUCHO(W)MARX, HOUSE(1W) REPRESENTATIVES, MICROSOFT()WORD.

> When searching for personal names it is worth remembering that they may be entered in some databases in inverted form (both as Groucho Marx and Marx, Groucho), so GROUCHO(N)MARX might be a more useful strategy. – GW

- *Concepts marginal to the database.* Say one is looking for documents about virtual reality systems and their potential impact on teaching. Searching in ERIC uncovers a number of descriptors on "teaching" but none on virtual reality, so one might use VIRTUAL()REALITY as a free-text expression.
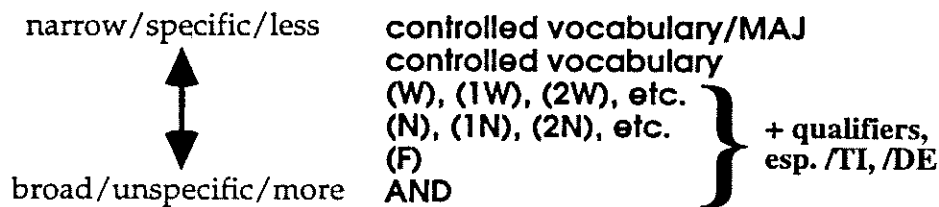
The Ladder of Specificity:
Broadening and Narrowing Searches

As we have discussed searching techniques using controlled vocabulary and free text, the idea of broadening and narrowing search statements has come up more than once. It is possible to think of these methods of searching as falling on a continuum, or ladder, of specificity.

We have already seen that controlled vocabulary searching is a very specific technique. The fact that an indexer, after evaluating a document, has assigned it a particular term, gives us a reasonably good idea that the document is about that concept. With free-text techniques, because one is dealing with parts of the document that are in natural language (titles and abstracts), one does not have that kind of confidence about the topic covered based on the simple presence of a word or phrase.
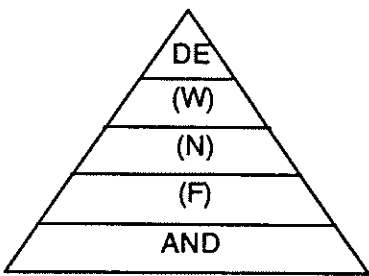
Consider the diagram in figure 8.1:

Fig. 8.1. The ladder of specificity.

narrow/specific/less   **controlled vocabulary/MAJ**

↕

broad/unspecific/more   **AND**

**controlled vocabulary
(W), (1W), (2W), etc.
(N), (1N), (2N), etc.
(F)**
} + qualifiers, esp. /TI, /DE

As one moves from the bottom of this ladder toward the top, the strategies become more specific and narrower and will retrieve fewer documents. Conversely, as one moves down, the strategies become less specific and broader and retrieve more documents. The /MAJ limit moves up the ladder, as do field qualifiers, especially /TI to qualify to titles and /DE to limit to the descriptor field.

I always think of this concept as a triangle, rather than a ladder.

This better illustrates the idea of moving up from broad to narrower or down from narrow to broader. – GW

This may be intellectually interesting, but it is also of considerable use in searching. Depending on the nature of the topic, the user, and the database, one can choose to begin a search with a strategy or set of strategies at a selected level on the ladder. After reviewing initial results, the results may be too broad, or there are too many or too few documents. One of the ways to cope with this is to move up or down on the ladder.

Similarly, if there are fewer documents than expected or even none, or if they may be a bit too narrow, move down the ladder. One can move from controlled vocabulary to free text, drop /MAJ or field qualifiers, or move to a broader free-text strategy: (W) to (1W) or (2W) or (N), (N) to (1N) or (2N) or (F), and so on.

These are not the only ways to broaden or narrow searches, though. A search can be broadened by using more terms (ORing them together), using broader controlled vocabulary terms as they are indicated in the thesaurus, or dropping a marginal concept altogether. A search can be narrowed by using fewer terms, using narrower controlled vocabulary terms, adding another concept, or NOTing out a concept.

> I am not a big fan of NOT—it is a seductively easy way of reducing the size of sets, but I think that it is often a bad idea. It is really easy to lose good documents that way. If you NOT a term out, especially using free text, any good documents that happen to have that term will also go away. If you are positive that any document that contains that term is not useful, under any circumstances, you might think about trying NOT, but I would be very careful about it. – JWJ

The following is a brief example showing this "ladder" at work in *ERIC* on the two-word phrase "test bias." We begin with the broadest strategy, by searching for the two words in the same document, using AND:

```
    Set    Items    Description
    ---    -----    -----------
?s test and bias

           71282    TEST
           14735    BIAS (AN INCLINATION, OR A LACK OF BALANCE (NOTE: ...))
    S1      3127    TEST AND BIAS
```

More than 3,000 documents in ERIC contain both the word TEST and the word BIAS. Now we move to a slightly narrower strategy, (F), and look for those words in the same field:

```
?s test(f)bias

           71282    TEST
           14735    BIAS (AN INCLINATION, OR A LACK OF BALANCE (NOTE: ...))
    S2      2767    TEST(F)BIAS
```

We get fewer documents. Let us NOT to see what we lost when we went to a narrower strategy:

```
?s s1 not s2

            3127      S1
            2767      S2
     S3      360      S1 NOT S2
?t 3/5/1
```

```
 3/5/1
DIALOG(R)File       1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.
```

EJ531824   UD519585
  Teacher Disapproval, Delinquent Peers, and Self-Reported Delinquency: A
Longitudinal Test of Labeling Theory.
  Adams, Mike S.; Evans, T. David
  Urban Review; v28 n3 p199-211 Sep 1996
  ISSN: 0042-0972
  Available from: UMI
  Language: English
  Document Type: REVIEW LITERATURE (070); RESEARCH REPORT (143); JOURNAL
ARTICLE (080)
  Journal Announcement: CIJFEB97
  Uses data from the National Youth Survey to assess the effects of
individual students" perceptions of teacher disapproval on self-reported
delinquency. Results indicate that the perceptions of teacher disapproval
are associated with subsequent delinquency. This relationship was
significant when controlling for prior delinquency, thus weakening the
argument that labeling is merely a result, not a cause of delinquency. (GR)
  Descriptors: *Delinquency; High School Students; Junior High School
Students; *Labeling (of Persons); *Peer Influence; *Secondary Education;
*Social Bias; Social Science Research; Socioeconomic Status; Stereotypes;
*Teacher Attitudes
  Identifiers: National Youth Survey

The above is one of those documents excluded by limiting to the same field (F). The words TEST and BIAS appear, but they are in different fields and are unrelated. The document has nothing to do with "test bias," and the extremely broad strategy using AND did not serve us well. We will continue to use narrower strategies like the following and view documents we exclude to demonstrate how moving up the ladder focuses searching more closely at each step.

```
?s test(10w)bias

            71282     TEST
            14735     BIAS (AN INCLINATION, OR A LACK OF BALANCE (NOTE: ...)
     S4     2103      TEST(10W)BIAS
?s s2 not s4

            2767      S2
            2103      S4
     S5      664      S2 NOT S4
?t 5/5/1-3
```

```
 5/5/1
DIALOG(R)File       1:ERIC
(c) format only 1998 The DialogCorporation. All rts. reserv.
```

EJ546738   TM520217
  Estimating the Importance of Differential Item Functioning.

Rudas, Tamas; Zwick, Rebecca
Journal of Educational and Behavioral Statistics; v22 n1 p31-45 Spr 1997
ISSN: 1076-9986
Language: English
Document Type: JOURNAL ARTICLE (080); EVALUATIVE REPORT (142)
Journal Announcement: CIJNOV97
   The mixture index of fit (T. Rudas et al, 1994) is used to estimate the
fraction of a population for which differential item functioning (DIF)
occurs, and this approach is compared to the Mantel Haenszel test of DIF.
The proposed noniterative procedure provides information about data
portions contributing to DIF. (SLD)
   Descriptors: Comparative Analysis; *Estimation (Mathematics); *Item Bias;
*Maximum Likelihood Statistics; *Test Items
   Identifiers: Item Bias Detection; *Mantel Haenszel Procedure

   5/5/2
DIALOG(R)File   1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ546729   TM520205
   A Multidimensionality-Based DIF Analysis Paradigm.
   Roussos, Louis; Stout, William
   Applied Psychological Measurement; v20 n4 p355-71 Dec 1996
   ISSN: 0146-6216
   Language: English
   Document Type: JOURNAL ARTICLE (080); EVALUATIVE REPORT (142)
   Journal Announcement: CIJNOV97
   A multidimensionality-based differential item functioning (DIF) analysis
paradigm is presented that unifies substantive and statistical DIF analysis
approaches by linking both to a theoretically sound and mathematically
rigorous multidimensional DIF conceptualization. This approach results in
the potential for DIF analysis more closely integrated with the whole test
development process. (Author/SLD)
   Descriptors: Cluster Analysis; *Estimation (Mathematics); Hypothesis
Testing; Identification; *Item Bias; *Models; Test Construction; *Test
Items
   Identifiers: Item Bias Detection; *Multidimensionality (Tests)

   5/5/3
DIALOG(R)File   1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ545074   PS526650
   Restricting a Familiar Name in Response to Learning a New One: Evidence
for the Mutual Exclusivity Bias in Young Two-Year-Olds.
   Merriman, William E.; Stevenson, Colleen M.
   Child Development; v68 n2 p211-28 Apr 1997
   ISSN: 0009-3920
   Language: English
   Document Type: RESEARCH REPORT (143); JOURNAL ARTICLE (080)
   Journal Announcement: CIJOCT97
   Used new test to determine whether 24-month olds interpret novel words in
accordance with Mutual Exclusivity Bias. Found that when asked to select
exemplars of a familiar noun, they avoided objects from previously read
story in which novel nouns were used as atypical exemplars of familiar
noun. When pronouns and proper names replaced novel nouns, toddlers did not
avoid story objects. (KB)

Descriptors: *Familiarity; Measures (Individuals); Novelty (Stimulus Dimension); *Toddlers; *Vocabulary Development
  Identifiers: *Mutual Exclusivity Bias; Naming Response; *Word Learning

The above are documents that have TEST and BIAS in the same field but not within 10 words of each other in that order. A couple of these are close, referring to item bias in tests, but they are still not quite on the mark. Another attempt yields the following:

```
?s test(3n)bias

          71282   TEST
          14735   BIAS (AN INCLINATION, OR A LACK OF BALANCE (NOTE: ...)
    S6     2068   TEST(3N)BIAS
?s s4 not s6

          2103    S4
          2068    S6
    S7      69    S4 NOT S6
?t 7/5/1-3
```

7/5/1
DIALOG(R)File   1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ536974   TM519883
  The University Entrance Examinations in Turkey.
  Berberoglu, Giray
  Studies in Educational Evaluation; v22 n4 p363-73 1996
  ISSN: 0191-491X
  Language: English
  Document Type: PROJECT DESCRIPTION (141); JOURNAL ARTICLE (080)
  Journal Announcement: CIJMAY97
  Issues related to the two-stage college entrance examinations used in Turkey are explored. Focus is on the first stage, the selection phase of the examination, rather than on the second stage, the placement phase. Further study is needed of the dimensionality of the test and sources of item bias. (SLD)
  Descriptors: *College Entrance Examinations; Foreign Countries; *Higher Education; *Item Bias; *Selection; Student Placement; *Test Use; Test Validity
  Identifiers: Dimensionality (Tests); *Turkey

7/5/2
DIALOG(R)File   1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ528234   PS525393
  Anxiety and the Processing of Emotionally Threatening Stimuli: Distinctive Patterns of Selective Attention among High- and Low-Test-Anxious Children.
  Vasey, Michael W.; And Others
  Child Development; v67 n3 p1173-85 Jun 1996
  ISSN: 0009-3920
  Available from: UMI
  Language: English
  Document Type: RESEARCH REPORT (143); JOURNAL ARTICLE (080)
  Journal Announcement: CIJDEC96

Tested for bias toward shifting attention toward threatening stimuli among high-anxious children and away from such stimuli among low-anxious children, ages 11-14. Results supported the predicted attentional bias toward threat cues among high-test-anxious children. Unexpectedly, the predicted attentional bias away from threat cues among low-anxious children was found only for boys. (HTH)
   Descriptors: *Anxiety; Attention; *Early Adolescents; *Emotional Response; Personality Traits; Sex Differences; Test Anxiety

```
 7/5/3
DIALOG(R)File       1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.
```

```
EJ521000           TM519386
```
  A Blended Qualitative-Quantitative Assessment Model for Identifying and Rank-Ordering Service Needs of Indigenous Peoples.
  Loos, Gregory P.
  Evaluation and Program Planning; v18 n3 p237-44 Jul-Sep 1995
  Research supported by a consortium of human care agencies and funded by the Bishop Estate, a private trust in Hawaii.
  ISSN: 0149-7189
  Available from: UMI
  Language: English
  Document Type: EVALUATIVE REPORT (142); JOURNAL ARTICLE (080)
  Journal Announcement: CIJJUL96
  This article describes a series of qualitative and quantitative methods used to test a community-based needs assessment model that is bias free and socioculturally relevant for indigenous populations. Results of a field test involving 100 Hawaiian children are presented, and implications for policy formation are discussed. (SLD)
  Descriptors: Children; *Community Programs; Cultural Awareness; Field Tests; *Indigenous Populations; Models; *Needs Assessment; *Policy Formation; *Qualitative Research; Statistical Analysis; *Statistical Bias
    Identifiers: Hawaii

Again, the above citations are marginal at best—documents that have the two words within 10 words of each other but not within three. Narrowing still further, we get the following:

?s test(n)bias

```
           71282     TEST
           14735     BIAS (AN INCLINATION, OR A LACK OF BALANCE (NOTE: ...)
     S8     2009     TEST(N)BIAS
?s s6 not s8

           2068      S6
           2009      S8
     S9       59     S6 NOT S8
?t 9/5/1-3
```

```
 9/5/1
DIALOG(R)File       1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.
```

```
EJ519378           CG548093
```
  Perceiver Bias in Expectancies for Sexually Abused Children.
  Briggs, Kathleen; And Others

Family Relations; v44 n3 p291-98 Jul 1995
ISSN: 0197-6664
Available from: UMI
Language: English
Document Type: RESEARCH REPORT (143); JOURNAL ARTICLE (080)
Journal Announcement: CIJJUL96
Presents a study in which 134 college students judged children in vignettes varying on child gender and family history to test perceiver bias. Perceiver bias was confirmed. Perceptions of female sexual abuse victims were more biased than perceptions of male victims. (Author/SR)
Descriptors: Behavior Problems; Child Abuse; Child Behavior; Children; *Questionnaires; *Sex Bias; Sex Differences; Sex Stereotypes; *Sexual Abuse; Surveys
Identifiers: Child History Expectation Questionnaire

9/5/2
DIALOG(R)File   1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ503271   HE533895
Effect of Anonymous Test Grading on Passing Rates as Related to Gender and Race.
Dorsey, J. Kevin; Colliver, Jerry A.
Academic Medicine; v70 n4 p321-23 Apr 1995
ISSN: 1040-2446
Available from: UMI
Language: English
Document Type: RESEARCH REPORT (143); JOURNAL ARTICLE (080)
Journal Announcement: CIJSEP95
Because of concern about potential gender and racial bias in medical test grading, score patterns were examined for male and female and for white and African American medical freshmen (n=476) before and after implementation of an anonymous test grading policy. Results indicate no widespread grading bias before the policy change. (Author/MSE)
Descriptors: Black Students; Females; *Grading; Higher Education; Males; *Medical Education; Professional Education; *Racial Bias; *Sex Bias; *Testing Problems; White Students
Identifiers: African Americans

9/5/3
DIALOG(R)File   1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ497098   TM518438
Comparison of Empirical and Judgmental Procedures for Detecting Differential Item Functioning.
Hambleton, Ronald K.; Jones, Russell W.
Educational Research Quarterly; v18 n1 p21-36 Sep 1994
ISSN: 0196-5042
Available from: UMI
Language: English
Document Type: EVALUATIVE REPORT (142); JOURNAL ARTICLE (080)
Journal Announcement: CIJMAY95
A judgmental method for determining item bias was applied to test data from 2,000 Native American and 2,000 Anglo-American students for a statewide proficiency test. Results indicated some shortcomings of the

judgmental method but supported the use of cross-validation in empirically identifying potential bias. (SLD)
   Descriptors: American Indians; Anglo Americans; Comparative Analysis; *Decision Making; *Evaluation Methods; *Identification; *Item Bias; *Test Items
   Identifiers: Cross Validation; *Empirical Research

The above documents are getting closer. The words are now within three words of each other but not directly adjacent. There are probably some good retrievals in there. But we can go still further with the following:

```
?s test(w)bias

           71282    TEST
           14735    BIAS (AN INCLINATION, OR A LACK OF BALANCE (NOTE: ...)
    S10     2002    TEST(W)BIAS
?s s8 not s10

            2009    S8
            2002    S10
    S11        7    S8 NOT S10
?t 11/5/1
```

```
 11/5/1
DIALOG(R)File      1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ213055            RC503536
   In North Carolina: Overlooked Causes and Implications of School Finance
Disparities.
   Nord, Stephen; Ledford, Manfred H.
   Growth and Change; v10 n4 p16-19 Oct 1979
   Available from: Reprint: UMI
   Language: ENGLISH
   Document Type: JOURNAL ARTICLE (080); RESEARCH REPORT (143)
   Journal Announcement: CIJAPR80
   Describes model, data, results, and implications of a study attempting to
clarify two issues previously overlooked in studies of financing public
education with local property taxes: (1) that regionally aggregated data
may bias test results, and (2) that local fiscal response to grant programs
may vary with the wealth of school districts. (SB)
   Descriptors: Assessed Valuation; *Educational Finance; *Equalization Aid;
Federal Aid; *Financial Policy; *Government School Relationship; Grants;
Models; Policy Formation; *Property Taxes; Public Education; *School Taxes;
State Aid; State School District Relationship; Tax Allocation
   Identifiers: *North Carolina
```

Asking for the two words directly adjacent but in either order pulls up false drops like this. This is to be expected in this case because "bias test" is not the same as "test bias." However, a strategy such as BIAS(2N)TEST, which would look for "bias of a test," would also pull up documents such as the one above. Note that there are very few documents in this set.

As an aside, let us see what happens when we qualify one of these sets down to the title field alone:

```
?s test(w)bias/ti
        10163    TEST/TI
         1407    BIAS/TI (AN INCLINATION, OR A LACK OF BALANCE (NOTE: ...)
  S12       76    TEST(W)BIAS/TI
```

The unqualified set, S10, had more than 1,700 documents; this one has 76. Obviously, the vast majority of occurrences of "test bias" as a phrase are in the abstract, descriptor, or identifier fields. Qualification can be an important tool, but at times it may also be too specific.

We can narrow still further with the following:

```
?s test bias
  S13     1951    TEST BIAS (UNFAIRNESS IN THE CONSTRUCTION, CONTENT, ADM...)
?s s10 not s13
         2002    S10
         1951    S13
  S14       51    S10 NOT S13
?t 14/5/1-3
```

```
 14/5/1
DIALOG(R)File       1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ528731          UD519339
  Black Scholars Hold a Pessimistic Outlook for African American Prospects
in Higher Education.
  Cross, Theodore, Ed.; And Others
  Journal of Blacks in Higher Education; n11 p74-77 Spr 1996
  ISSN: 1077-3711
  Language: English
  Document Type: REVIEW LITERATURE (070); RESEARCH REPORT (143); JOURNAL
ARTICLE (080)
  Journal Announcement: CIJDEC96
  Discusses survey findings that show black academics are highly
pessimistic in their view of the future of blacks in higher education.
Reasons include the following: curtailment of federal support for black
colleges over the next five years; continued built-in test bias against
blacks; no improvement in campus race relations; and persistent racial
barriers against black faculty. (GR)
  Descriptors: Affirmative Action; *Black Colleges; *Black Education; Black
Teachers; Educational Research; Educational Trends; *Financial Support;
Futures (of Society); Higher Education; Postsecondary Education; *Racial
Bias; *Racial Relations; Scores; Surveys; *Teacher Attitudes
  Identifiers: Scholastic Aptitude Test


 14/5/2
DIALOG(R)File 1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ296328          UD510634
  Beyond IQ Test Bias: The National Academy Panel"s Analysis of Minority
EMR Overrepresentation.
```

Reschly, Daniel J.
Educational Researcher; v13 n3 p15-19 Mar 1984
Language: English
Document Type: REVIEW LITERATURE (070)
Journal Announcement: CIJJUN84
Questions educational relevance of direct measures of learning such as
the Learning Potential Assessment Device, assessment of biomedical factors,
and adaptative behavior measures. Notes increased discrepancies between EMR
and average students in high school. Suggests a generic classification for
the mildly handicapped and the combining of groups for educational
purposes. (CJM)
    Descriptors: Academic Achievement; *Classification; *Educational
Diagnosis; Elementary Secondary Education; *Learning Disabilities;
Mainstreaming; *Measurement Techniques; *Mild Mental Retardation; Minority
Group Children; Racial Bias; Racial Composition; Research Needs; *Special
Education
    Identifiers: Learning Potential Assessment Device; National Research
Council

 14/5/3
DIALOG(R)File   1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ285500   EC152888
  Assessing Adaptive Behavior: Current Practices.
  Cantrell, Joan Kathryn
  Education and Training of the Mentally Retarded; v17 n2 p147-49 Apr 1982
  Available from: Reprint: UMI
  Language: English
  Document Type: JOURNAL ARTICLE (080); RESEARCH REPORT (143)
  Journal Announcement: CIJDEC83
  Twenty-nine elementary school psychologists were interviewed about
assessment of adaptive behavior. Over 95 percent reported they routinely
assess adaptive behavior skills, and 90 percent felt the assessment useful
in planning instruction. They rated methods of assessment (home observation
ranked first), cited safeguards against test bias, discussed school
policies, and recommended changes. (CL)
    Descriptors: *Adaptive Behavior (of Disabled); Attitudes; *Disabilities;
Elementary Education; Evaluation Methods; *School Psychologists; *Student
Evaluation

    Interesting. The second of these documents is about "IQ test bias" among minority
students, and so is marginal at best. The first refers explicitly to "test bias" in the abstract
but has not been indexed with the descriptor TEST BIAS. The indexers must have thought
the concept was marginal in this document. And the third one mentions safeguards *against*
test bias, yet is retrieved along with others that also contain that phrase. We have one fur-
ther step on the ladder to explore.

```
?s s13/maj

    S15     1087     S13/MAJ
?s s13 not s15

            1951     S13
            1087     S15
    S16      864     S13 NOT S15
?t 16/5/1-3
```

16/5/1
DIALOG(R)File   1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ546747   TM520226
  Flawed Items in Computerized Adaptive Testing.
  Potenza, Maria T.; Stocking, Martha L.
  Journal of Educational Measurement; v34 n1 p79-96 Spr 1997
  ISSN: 0022-0655
  Language: English
  Document Type: JOURNAL ARTICLE (080); EVALUATIVE REPORT (142)
  Journal Announcement: CIJNOV97
  Common strategies for dealing with flawed items in conventional testing,
grounded in principles of fairness to examinees, are re-examined in the
context of adaptive testing. The additional strategy of retesting from a
pool cleansed of flawed items is found, through a Monte Carlo study, to
bring about no practical improvement. (SLD)
  Descriptors: *Adaptive Testing; *Computer Assisted Testing; *Item Banks;
Monte Carlo Methods; Test Bias; *Test Items Identifiers: *Flawed Items

16/5/2
DIALOG(R)File   1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ543873   CE530793
  Career Assessment with Lesbian, Gay, and Bisexual Individuals.
  Prince, Jeffrey P.
  Journal of Career Assessment; v5 n2 p225-38 Spr 1997
  ISSN: 1069-0727
  Language: English
  Document Type: POSITION PAPER (120); JOURNAL ARTICLE (080)
  Journal Announcement: CIJOCT97
  Sexual identity development and environmental factors are central to the
career assessment of lesbian, homosexual, and bisexual clients. Counselor
biases and biases in career assessment tools must be recognized and dealt
with. (SK)
  Descriptors: *Career Counseling; Environmental Influences; *Homosexuality;
*Lesbianism; Measures (Individuals); *Sexual Identity; Test Bias
Identifiers: *Bisexuality; *Career Assessment

16/5/3
DIALOG(R)File   1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ541216   HE536582
  Educational Histories and Academic Potential: Can Tests Deliver?
  Yeld, Nan; Haeck, Wim
  Assessment & Evaluation in Higher Education; v22 n1 p5-16 Mar 1997
  ISSN: 0260-2938
  Language: English
  Document Type: EVALUATIVE REPORT (142); JOURNAL ARTICLE (080)
  Journal Announcement: CIJAUG97
  A new South African testing approach designed to assess potential
university students" ability to cope with English-medium academic education
is also designed to take into account the effects of educational
disadvantagement and minimize reliance on students" content-based secondary

school experiences. The approach incorporates principles from
constructivist learning theories. Preliminary results of the testing
approach are discussed. (Author/MSE)
  Descriptors: *Academic Achievement; Constructivism (Learning);
*Educationally Disadvantaged; English; *English for Academic Purposes;
Foreign Countries; Higher Education; *Language of Instruction; *Learning
Processes; Learning Theories; Program Effectiveness; Secondary Education;
Test Bias; *Testing Programs; Test Use
  Identifiers: *South Africa

The documents above have the descriptor TEST BIAS but not as a major descriptor, and they make up fewer than half of the total number of documents that have been indexed with this term. Clearly, these documents are about test bias, but it appears that this is not a central concept in any of them. If we look at our final /MAJ set, though, we see the following:

?t 15/5/1-3

 15/5/1
DIALOG(R)File   1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ546768   UD519986
  Equity and High Stakes Testing: Implications for Computerized Testing.
  Sutton, Rosemary E.
  Equity & Excellence in Education; v30 n1 p5-15 Apr 1997
  ISSN: 1066-5684
  Language: English
  Document Type: JOURNAL ARTICLE (080); POSITION PAPER (120); PROJECT
DESCRIPTION (141)
  Journal Announcement: CIJNOV97
  Considers equity issues of highstakes tests conducted by computer,
including whether this new form of assessment actually helps level the
playing field for students or represents a new cycle of assessment
inequality. Two computer tests are assessed: Praxis I: Academic Skills
Assessment; and the computerized version of the Graduate Record
Examination. (GR)
  Descriptors: Adaptive Testing; *Computer Assisted Testing; Educational
Assessment; Educational Testing; Secondary Education; Student Evaluation;
*Test Bias
  Identifiers: Graduate Record Examinations; *High Stakes Tests; Praxis
Series

 15/5/2
DIALOG(R)File   1:ERIC
(c) format only 1998 The Dialog Corporation. All rts. reserv.

EJ546752   UD519970
  The Overrepresentation of African American Children in Special Education:
The Resegregation of Educational Programming?
  Russo, Charles J.; Talbert-Johnson, Carolyn
  Education and Urban Society; v29 n2 p136-48 Feb 1997
  ISSN: 0013-1245
  Language: English
  Document Type: JOURNAL ARTICLE (080); PROJECT DESCRIPTION (141)
  Journal Announcement: CIJNOV97

Reviews the historical background of special education as a major factor in the placement of many children with disabilities, and examines data that reveal a disproportionately large number of students in these programs are children of color. Suggestions are offered to help lead to a more equitable placement of all children in appropriate educational settings. (GR)
  Descriptors: *Blacks; Educational Change; Instructional Improvement; Minority Groups; School Community Relationship; *Special Education; *Special Needs Students; *Student Placement; Teacher Education; Teacher Recruitment; *Test Bias
  Identifiers: Individuals with Disabilities Education Act

  15/5/3
  DIALOG(R)File   1:ERIC
  (c) format only 1998 The Dialog Corporation. All rts. reserv.

  EJ543826   TM520091
    Test Fairness: Internal and External Investigations of Gender Bias in Mathematics Testing.
    Langenfeld, Thomas E.
    Educational Measurement: Issues and Practice; v16 n1 p20-26 Spr 1997
    ISSN: 0731-1745
    Language: English
    Document Type: EVALUATIVE REPORT (142); JOURNAL ARTICLE (080)
    Journal Announcement: CIJSEP97
    Presents two approaches to evaluating gender measurement bias in mathematics testing, and discusses how to assess these approaches. The two approaches are the study of internal test structure and external test relationships. Ensuring a gender-fair test requires attention to both areas. (SLD)
    Descriptors: *Mathematics Tests; *Measurement Techniques; Psychometrics; *Sex Bias; Sex Differences; *Test Bias
    Identifiers: Internal Structure Analysis

We see immediately that this is a very focused set. For a real search, then, we may decide to begin with a controlled vocabulary term or even limiting that to a major descriptor and then broaden out if necessary. Of course, in practice, one does not have the time or opportunity to know this kind of detail about terms, so make an educated guess about where to begin and then move up or down the ladder as necessary as the search progresses.

> *Your initial choice of where on the ladder to start your search depends on your expectations of search outcome. How much material do you expect to find? Postings figures in the thesaurus, when available, can help you make this decision, as can personal experience and perhaps information from the user.*
> *— GW*

Search Example

For the search example, we will use (for the last time, I promise) the traumatic brain injury search. Now that we have all of these technical possibilities at our disposal, we can use them to create a high-quality search.

A couple of things should be noted at this point:

This is not a perfect search. There are no perfect searches. It is pretty good and is the result of doing it and refining it many times over many years. Do not despair if such a search seems too involved or complicated right now. (If it seems simple, you've found your career!)

The search uses several different techniques, both controlled vocabulary and free text, and a few other things such as exploding. Not every search uses such a variety of methods; this search is an example of how they can all contribute.

Having said all that, let us proceed.

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

    Set    Items    Description
    ---    -----    -----------
?ss brain damage and (injuries/df or trauma)

    S1     4677     BRAIN DAMAGE (1967)
    S2     1521     INJURIES/DF (1973)
    S3     8838     TRAUMA
    S4      219     BRAIN DAMAGE AND (INJURIES/DF OR TRAUMA)
```

This small set is a focused subset of documents with BRAIN DAMAGE as a descriptor; that term is used not only for traumatic brain damage but also (more commonly) for congenital brain damage. The parenthetical component restricts the set only to documents that include the one-word descriptor INJURIES or the word TRAUMA anywhere.

```
?ss head injuries or closed()head()injur?

    S5     2323     HEAD INJURIES (1973)
    S6     5001     CLOSED
    S7    11676     HEAD
    S8    15539     INJUR?
    S9      784     CLOSED(W)HEAD(W)INJUR?
    S10    2437     HEAD INJURIES OR CLOSED()HEAD()INJUR?
?s s4 or s10

           219     S4
          2437     S10
    S11   2589     S4 OR S10
?s children! or adolescents/df

        120906     CHILDREN! (1967)
         32256     ADOLESCENTS/DF (1967)
    S12 141918CHILDREN! OR ADOLESCENTS/DF
```

I chose to explode CHILDREN because it has several narrower terms, including INFANTS, and many of these narrower terms also have narrower terms. It is an economical way to get all those terms. ADOLESCENTS, on the other hand, has only two narrower terms, ADOLESCENT MOTHERS and ADOLESCENT FATHERS, both of which have very few postings. Not worth considering.

```
?s s11 and s12

              2589     S11
            141918     S12
     S13      247      S11 AND S12
?ss psychsocial or behavior/maj or neuropsycholog?

     S14        4      PSYCHSOCIAL
     S15     4565      BEHAVIOR/MAJ (1967)
     S16    14928      NEUROPSYCHOLOG?
     S17    19414      PSYCHSOCIAL OR BEHAVIOR/MAJ OR NEUROPSYCHOLOG?
?s psychosocial or s14

     S18    28364      PSYCHOSOCIAL
                4      s14
     S19    28366      PSYCHOSOCIAL OR S14
?s s19 or s17

            28366      S19
            19414      S17
     S20    47282      S19 OR S17
?s s13 and s20

              247      S13
            47282      S20
     S21       86      S13 AND S20
```

When doing this search for inclusion in this chapter, I made the spelling error you see in S14. I did not catch it until too late and was about to restart the search so it would look perfect for the book when I decided to keep it to demonstrate (a) that we all make mistakes like that, no matter how experienced we get, and (b) how to recover from an error like that. Note that I recreated that concept set by ORing it with the correct spelling. Some indexer made the same mistake I did (hence the 4 hits in S14), and those documents are included in S19.

I took three different approaches with those three terms. PSYCHOSOCIAL, as we have seen, is part of a descriptor, PSYCHOSOCIAL READJUSTMENT, but it also was mentioned by itself by the user in her original request. Because it is a fairly concrete and specific term, even in a psychology database, I chose to search it as is. "Behavior" is a far broader concept, especially in a psychology database, so I searched it very narrowly by restricting it to use in major descriptors. I could have chosen specific descriptors and ORed them together, but again the user gave us just that word, so I let it go to all descriptors, yet limited it as best I could. Finally, I truncated on NEUROPSYCHOLOG?, figuring any word beginning with a stem that specific (and that long, for that matter) would be appropriate.

```
?t 21/8/1-10

 21/8/1
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)
```

21/8/2

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)

21/8/3

01507522   1997-43861-010
Concept formation and problem-solving following closed head injury in
  children.

DESCRIPTORS: *Head Injuries; *Neuropsychological Assessment; *Severity
  (Disorders); Adolescence; Adulthood; Childhood; Longitudinal Studies;
  Preschool Age Children; School Age Children
IDENTIFIERS: Twenty Questions Test & Tower of London & Wisconsin Card
Sorting Test performance, 5-18 yr olds with mild vs severe closed head
injury, 36 mo study
SUBJECT CODES & HEADINGS: 3297 (Neurological Disorders & Brain Damage)

21/8/4

01483092   1997-42970-002
A typology of psychosocial functioning in pediatric closed-head injury.

DESCRIPTORS: *Head Injuries; *Psychodiagnostic Typologies; *Severity
  (Disorders); *Psychosocial Readjustment; Adolescence; Childhood; School
  Age Children
IDENTIFIERS: typology of psychosocial functioning, 6-16 yr olds with mild
vs moderate vs severe closed head injury
SUBJECT CODES & HEADINGS: 3290 (Physical & Somatoform & Psychogenic
  Disorders)

21/8/5

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)

21/8/6
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

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

DESCRIPTORS: *Head Injuries; *Literature Review; Adolescents; Children
IDENTIFIERS: neuropsychological or academic or psychosocial outcomes,
children & adolescents with mild head injury, literature review, 1970-1995
SUBJECT CODES & HEADINGS: 3297 (Neurological Disorders & Brain Damage)

21/8/7
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

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

DESCRIPTORS: *Behavior Problems; *Parental Attitudes; *Traumatic Brain
   Injury; *Stress; Adolescents; Children; Mothers; Adulthood
IDENTIFIERS: parental stress levels & perception of children"s behavioral
problems at least 1 yr after injury, 24-50 yr old mothers of children (aged
4.5-15 yrs) who had traumatic brain injury
SUBJECT CODES & HEADINGS: 3297 (Neurological Disorders & Brain Damage)

21/8/8
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01411695   1997-85262-001
Applicazione della Batteria Neuropsicologica Luria Nebraska nell"analisi
   funzionale di soggetti con pregresso trauma cranico e coma.
TRANSLATED TITLE: Application of the Luria-Nebraska Neuropsychological
Battery in the functional analysis of subjects with head injury and
subsequent coma.

DESCRIPTORS: *Coma; *Foreign Language Translation; *Neuropsychological
   Assessment; *Test Reliability; *Head Injuries; Adolescence; Adulthood;
   School Age Children; Childhood
IDENTIFIERS: reliability of Luria-Nebraska Neuropsychological Battery, male
10-22 yr olds with head injury & subsequent coma, Italy
SUBJECT CODES & HEADINGS: 2225 (Neuropsychological Assessment)

21/8/9
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01411183   1997-06130-003
Age at injury as a predictor of outcome following pediatric head injury:
   A longitudinal perspective.

DESCRIPTORS: *Age Differences; *Head Injuries; *Neuropsychological
   Assessment; *Recovery (Disorders); Childhood; Longitudinal Studies;
   School Age Children
IDENTIFIERS: age at injury & recovery from pediatric head injury, children
injured before vs after 7 yrs of age, 4 mo to 2 yr longitudinal study
SUBJECT CODES & HEADINGS: 3297 (Neurological Disorders & Brain Damage);
   2820 (Cognitive & Perceptual Development)

21/8/10
DIALOG(R)File 11:(c) 1998 Amer. Psychological Assn. All rts. reserv.

01409782   1996-97000-009
Attention deficits in the long term after childhood head injury.

DESCRIPTORS: *Attention; *Cognitive Ability; *Head Injuries; Adolescence;
  Brain Damage; Neuropsychology; School Age Children; School Learning;
  Preschool Age Children; Memory; Early Experience; Childhood; Adulthood
IDENTIFIERS: long term neuropsychological outcomes of memory & attention in
academic work & daily functions, 5.2-18 yr olds with head injuries
SUBJECT CODES & HEADINGS: 3297 (Neurological Disorders & Brain Damage)

These results are quite encouraging. Not perfect by any means, but several look very good, especially the sixth one (a literature review). One is in Italian, so I limited the set down to English:

?s s21/eng

    S22   82   S21/ENG

and only eliminated 4 documents. Here are the titles of the first 25 documents:

?t 22/6/1-25

 22/6/1
01523141   1997-38757-001
Head injury in children.

 22/6/2
01512432   1997-36680-006
Frontal lobe dysfunction following closed head injury in children: Findings
  from neuropsychology and brain imaging.

 22/6/3
01507522   1997-43861-010
Concept formation and problem-solving following closed head injury in
  children.

 22/6/4
01483092   1997-42970-002
A typology of psychosocial functioning in pediatric closed-head injury.

 22/6/5
01482579   1997-30097-007
Pediatric neuropsychology.

 22/6/6
01471818   1997-05606-001
Mild head injury in children and adolescents: A review of studies
  (1970-1995).

22/6/7
01427791   1996-04856-003
Behavioural adjustment and parental stress associated with closed head
  injury in children.

22/6/8
01411183   1997-06130-003
Age at injury as a predictor of outcome following pediatric head injury:
  A longitudinal perspective.

22/6/9
01409782   1996-97000-009
Attention deficits in the long term after childhood head injury.

22/6/10
01409781   1996-97000-008
Cognitive, behavioral, and motoric sequelae of mild head injury in a
  national birth cohort.

22/6/11
01409779   1996-97000-006
Discourse as an outcome measure in pediatric head-injured populations.

22/6/12
01409778   1996-97000-005
Neurobehavioral outcome of pediatric closed head injury.

22/6/13
01409774   1996-97000-001
Variability in outcomes after traumatic brain injury in children:
  A developmental perspective.

22/6/14
01409211   1996-93110-001
Differential performances on the WRAML in children and adolescents
  diagnosed with epilepsy, head injury, and substance abuse.

22/6/15
01396784   1996-17735-001
Neuropsychological deficit and academic performance in children and
  adolescents following traumatic brain injury.

22/6/16
01385762   1996-02578-040
Clinical neurological indicators are only moderately correlated with
  quantitative neuropsychological test scores in patients who display
  mild-moderate brain impairment following closed-head injuries.

22/6/17
01381178   1995-98826-019
Neurobehavioral effects of brain injury on children: Hydrocephalus,
  traumatic brain injury, and cerebral palsy.

```
22/6/18
01352975          1996-97159-000
Pediatric traumatic brain injury.
SERIES TITLE: Developmental clinical psychology and psychiatry, Vol. 31.


22/6/19
01338383          1995-25937-001
Clinical neurological trauma parameters as predictors for
  neuropsychological recovery and long-term outcome in paediatric
  closed head injury: A review of the literature.


22/6/20
01330143          1995-14036-001
Cognitive and psychosocial outcome after head injury in children.


22/6/21
01314586          1994-97656-014
The impact of neuropsychiatry upon forensic issues related to children
  and adolescents.


22/6/22
01305613          1994-38159-001
Motor, visual; spatial, and somatosensory skills after closed head injury
  in children and adolescents: A study of change.


22/6/23
01265744          1994-10353-001
Predictors of outcome following severe head trauma: Follow-up data from
  the Traumatic Coma Data Bank.


22/6/24
01253977          1993-97880-009
Traumatic brain injury.


22/6/25
01244067          1993-38257-001
Long-term outcome of head injuries: A 23 year follow up study of children
  with head injuries.
```

Let us look at a summary of the complete search strategy:

?ds

```
Set     Items     Description
S1       4677     BRAIN DAMAGE (1967)
S2       1521     INJURIES/DF (1973)
S3       8838     TRAUMA
S4        219     BRAIN DAMAGE AND (INJURIES/DF OR TRAUMA)
S5       2323     HEAD INJURIES (1973)
S6       5001     CLOSED
S7      11676     HEAD
S8      15539     INJUR?
S9        784     CLOSED(W)HEAD(W)INJUR?
```

```
Set     Items      Description
S10      2437      HEAD INJURIES OR CLOSED()HEAD()INJUR?
S11      2589      S4 OR S10
S12    141918      CHILDREN! OR ADOLESCENTS/DF
S13       247      S11 AND S12
S14         4      PSYCHSOCIAL
S15      4565      BEHAVIOR/MAJ (1967)
S16     14928      NEUROPSYCHOLOG?
S17     19414      PSYCHSOCIAL OR BEHAVIOR/MAJ OR NEUROPSYCHOLOG?
S18     28364      PSYCHOSOCIAL
S19     28366      PSYCHOSOCIAL OR S14
S20     47282      S19 OR S17
S21        86      S13 AND S20
S22        82      S21/ENG
```

The Internet

The most important thing to understand about searching using Internet search engines is that the available techniques are far less sophisticated and powerful than that of DIALOG or other commercial search vendors. The databases involved are huge; most search engines have indexed the full text of many tens of millions of Web documents, and as such, the overhead involved in creating the kind of inverted file we have been working with is unrealistic. So while most search engines will allow Boolean searches, there are no proximity operators, for example. AltaVista does, however, have a NEAR operator, which works like a (5N), finding target words within five words of each other.

Two other important factors: First, the underlying documents are encoded in HTML, so there is less information in that structure than in a bibliographic document record—no author, maybe a title, and certainly nothing like subject headings or abstracts, which gives the searcher less to work with and fewer opportunities to use sophisticated search techniques in refining or polishing searches. Those kinds of techniques might become more feasible if and when metadata, information about documents, becomes more widely prevalent in documents. Such metadata is now seen in META tags in HTML documents, listing keywords, authors, publishers, etc. Schemes such as the Dublin Core, which attempt to codify or standardize such metadata, might also assist searchers, but they would require wide compliance, which has not yet occurred.

Second, the type of writing found in Web documents varies greatly. There is a great deal more casual, metaphorical, ironic, and sarcastic usage, not to mention terrible spelling, new words, and new uses of old words (who would have thought "spam," now used to describe unwanted E-mail, would be so common?). This means many more false drops, especially with words or phrases with multiple uses.

On the other hand, if you have a really specific name or word or phrase, it will get better results in this environment than perhaps anywhere else. Take advantage of these circumstances rather than viewing them solely as problems. One of the most common reference questions is "What is the other word in the English language, besides "angry" and "hungry," that ends with -gry?" This is the sort of question that drives reference librarians nuts because there is no really good way to look for the answer, other than in a book of language curiosities, or just knowing it from having dealt with the question several times before.

The Internet is the ideal place to look for this type of query. Do a search on almost any search engine on the words "angry" and "hungry," and besides some junk, what will come

up are several, perhaps contradictory, answers to the question. I answered a reference question for someone looking for an old poem. He remembered a fragment of it from childhood (these are horrible questions, too—invariably the lines are half-correct, and these poems are never in books, especially if they are Ann-Landers-type poems), and it had the phrase "piddling pup" in it. I figured there couldn't be *that* many documents on the Net with a phrase like that, so I searched for it, and found exactly one, a page from Australia with the full text of the poem.

Remember that most search engines work this way: if words are entered in the search box and the button is pushed, the words will be searched together; documents will be returned, ranked in order by some proprietary (and therefore unexplained) mechanism; and the first ones on the list are allegedly the best. This is, for all intents and purposes, a big OR, getting all documents that have any of the words asked for, and then ranking them by their frequency of occurrence in the document, perhaps giving higher weight to words in the title, <h1> tag, <meta> tag, or early in the page.

Imagine a search on the Net for recipes that use coconut milk (perhaps someone has a lot of coconut milk about to go bad and needs to use it up). I would probably try a strategy something like

> + "coconut milk" recipe

which would retrieve documents with the phrase "coconut milk" and prefer documents that also include "recipe." I would really prefer to truncate on "recipe," but not all engines allow you to (most notably HotBot at the moment), while others will implicitly truncate unless specifically told not to.

These search engines continue to evolve, both in operation and interface, so it pays to check the help pages, especially if the front or search pages change, to see if the command language or operation has also changed.

There is a lot of money to be made here. As the Internet becomes more of a presence in lives and commerce, and as more people use it for information, a search engine that consistently outperforms the others will be very popular, generate more business, get more traffic and advertising, and make more money. All of a sudden, information retrieval is big business. Therefore, there is a great incentive to make these engines work as effectively as possible.

In fact, there is a fascinating Web site devoted just to this matter: Search Engine Watch (http://www.searchenginewatch.com). This table, taken from one of their introductory pages, gives an idea of the kinds of information available there. (See Fig. 8.1 on p. 174.)

Fig. 8.1. Search engine watch pages.

Section	Designed For	Description
A Webmaster's Guide To Search Engines	Webmasters	This section explains how search engines find and rank web pages, with an emphasis on what webmasters can do to improve how search engines list their web sites.
Search Engines Facts And Fun	Search Engine Users	This section provides background about search engines, tips on how to use them better, some history, and even a game to test your search engine knowledge
Search Engine Status Reports	Anyone	This section provides some insight on how search engines are performing in different areas. Check out some of the material, and make note of reports you want to keep an eye on.
Search Engine Resources	Anyone	A collection of links to search engine related resources across the web.
Search Engine Report Mailing List	Anyone	It's free, and it will keep you updated on the latest search engine news.

Research in areas such as computational linguistics, natural language processing, improved categorization or classification, intelligent agents, and so on, might well prove to make the difference, but so far, no one engine has emerged as the obvious winner . . . yet. It would be nice to believe that librarians and librarianship will be in on this, helping to design and build technological assistants based on our traditions and expertise in searching and understanding the needs of users.

Notes

1. At present, DIALOG allows this number to be up to 127 in most files; it is unlimited in files that provide the full text of documents—we'll talk more about this later.

2. Notice the hyphen in the title: "thirty-item" as it appears in the title. In DIALOG, all internal punctuation (hyphens, apostrophes, slashes, quotation marks, etc.) is removed and replaced by spaces. Thus, when this document was processed, the hyphen was removed, and the words "thirty" and "item" went into the inverted file as separate words.

3. These are based on categories given in an article by Elaine Wagner (September 1986), "False Drops—How They Arise . . . How to Avoid Them." *Online* 10(5): 93–96.

4. Markey et al. (1980), "An Analysis of Controlled Vocabulary and Free Text Search Statements in On-Line Searches." *Online Review* 4(3), 225–36.

Additional Reading

Morton, Douglas (1993), "Refresher Course: Getting Next to Proximity Operators" *Online* 17(6): 55–58.

Rowley, Jennifer (1994), "The Controlled Versus Natural Indexing Languages Debate Revisited: A Perspective on Information Retrieval Practice and Research," *Journal of Information Science* 20(2): 109–19.